IBM InfoSphere Identity Insight

User Guide

Version 9 Release 0

IBM InfoSphere Identity Insight

User Guide

Version 9 Release 0

Note

Before using this information and the product it supports, read the information in "Notices" on page 375.

Edition notice

This edition applies to version 9 release 0 of IBM InfoSphere Identity Insight (product number 5724-L71) and to all subsequent releases and modifications until otherwise indicated in new editions.

Contents

Preface vii
Contacting IBM Software Support viii

Chapter 1. Overview of IBM InfoSphere Identity Insight 1
Product architecture 2
Acquisition programs 3
Universal Message Format (UMF) 4
Pipelines. 4
Pipeline nodes 5
Application monitor 5
Transports 6
Data sources 6
Entity database 7
User interfaces. 7
Web services 8
Core concepts 9
Entities 9
Identities 10
Attributes 10
Entity resolution. 11
Recognize 11
Resolve. 14
Relate 16
Scoring 22
Event Manager 24
Events 24
Event alerts 25
Event types 25
Event rules 25
Getting started with Event Manager 25
Configuring the Event Manager CEP module . . 27
Guidelines for configuring event rule results . . 34

Chapter 2. System requirements and planning 41
Detailed System Requirements 41
System requirements when running on IBM AIX 41
System requirements when running on Linux for System x 42
System requirements when running on Linux for Power Systems 43
System requirements when running on Linux for System z 44
System requirements when running on Microsoft Windows Server 45
Defining the system architecture 46
Pipeline deployments 46
Creating a protected user for non-Windows installations 46
User roles and responsibilities 46

Chapter 3. Setting up the databases . . 49
Setting the environment variables 49
DB2 environment variables 49
Oracle environment variables 50
Granting Oracle users CREATE VIEW privileges . . 51
Creating and configuring the databases 51
Creating the entity database 52
Configuring client authentication 52
Sizing the Oracle Statement Cache. 53

Chapter 4. Administering 55
Administering Users 55
Basic Authentication 55
Administering the Console 56
Configuration Console. 57
User roles and responsibilities 57
Optimum browser settings for using the Configuration Console. 58
Logging in to the Configuration Console . . . 59
Logging out of the Configuration Console . . . 59
User accounts for the Configuration Console . . 59
Managing access to the Configuration Console . 60
Help topics 64
Administering System Configuration settings . . . 64
Deploying patches and hot fixes 64
Modifying Database information 64
Modifying HTTP ports 65
Modifying HTTP Transport 65
Modifying the host name. 65
Running reports from the Configuration Console. . 66
Viewing statistical reports 66
Running the Configuration report 73
Exporting reports 77
Administering the Visualizer 79
Visualizer 79
User roles and responsibilities 80
Optimum browser settings for using the Visualizer 81
Logging in to the Visualizer 82
Closing the Visualizer 82
Managing access to the Visualizer 83
Configuring activity codes for the Visualizer . . 85
Administering System Configuration settings . . . 88

Chapter 5. Configuring the system for data 89
Configuring data in the system. 89
Configuring characteristic types 89
Configuring number types 92
Configuring name data 95
Configuring DQM rules 105
Configuring lookup codes 108
Configuring generic data values 111
Configuring roles 113
Configuring role alert rules. 115
Configuring entity types. 119
Overview of Degrees of Separation 122

Configuring UMF documents 125
Configuring the data source 126
Configuring event types 133
Configuring entity resolution 135
Entity resolution 136
Configuring resolution configurations 136
Configuring resolution rules 138
Customizing the candidate builder 151
Configuring confirmations and denials 154
Configuring system parameters 157
Configuring system parameters for name scoring 157
Configuring system parameters for the Name Manager 157
Configuring system parameters for the database 158
Configuring system parameters for the logs . . 159
Configuring system parameters for confirmation and denial 159
Configuring system parameters for role alerts 160
Configuring system parameters for attribute alert generators 160
Configuring the system parameters for concurrency 160
Configuring system parameters for data quality management 161
Configuring system parameters for product options 161
Configuring system parameters for Event Manager 162
Configuring system parameters for the Visualizer 162
Setting the default path for Centrifuge 162
Setting the default path for UMF files 163
Attribute and scoring customization 164
Storing large attribute data 164
Configuring source characteristics for large attribute data 167
Configuring resolution characteristics for large data 167
Configuration reports for attribute and scoring customization 168
Configuring custom scoring plug-ins 169
Developing custom scoring plug-ins for IBM InfoSphere Identity Insight 170

Chapter 6. Managing pipelines 175
Pipelines 175
Pipeline configuration check 175
Pipeline nodes 176
Starting pipelines 176
Stopping pipelines. 178
Configuring pipelines 178
Pipelines registration 179
Registering pipelines 179
Viewing registered pipeline details 180
Editing pipeline registrations 181
Deleting pipeline registrations. 181
Help topics 182
Configuring routing rules 183
Routing rules 185
Help topics 186
Deleting routing rules 188
Pipeline status and statistics 188
SNMP agents 188
Starting SNMP agents 189
Stopping SNMP agents 190
Checking pipeline status in the Configuration Console 190
Checking pipeline status using the command line. 191
Viewing application monitor events 192
Viewing UMF exceptions 193
Viewing new identities 194
Help topics 194

Chapter 7. Loading data. 201
Adding a new data source 201
Converting data to UMF 202
Acquisition programs. 202
Transferring UMF files to a queue 202
Queue utility 202
Configuration file for the queue utility 203
Command syntax of the queue utility 204
Converting UMF files to appropriate formats. . . 207
UMF formatting utility 207
Command syntax of the UMF formatting utility 207
Extending the entity model. 208
Universal Message Format (UMF) 208
Analyzing source data 208
Reviewing the default UMF specification . . . 209
Mapping UMF segments to the entity database 209
Address standardization with IBM InfoSphere QualityStage and AddressDoctor 215
QS-AVI address cleansing requirements and task overview 215
QS-AVI troubleshooting 216

Chapter 8. Analyzing data 219
Analyzing data using the Visualizer 219
Setting up the Visualizer 219
Starting the Visualizer 229
Analyzing alerts in the Visualizer. 234
Finding entities. 244
Analyzing entities 253
Adding data through the Visualizer 262
Running reports from the Visualizer. 271
Analyzing data with the Analyst toolkit 293
Reporting on data with the IBM Cognos reports 293
Analyzing data using the graphing tool . . . 301

Chapter 9. Developing 323
Web services 323
Web services software requirements 324
Starting Web services pipelines 324
Testing Web services 326
srd.wsdl file 327
wsutil.jar 328
Building queries against the entity database . . . 330
Web service pipeline searches 330
Building Web services queries to find a specific entity 331

Building Web services queries to find entities with similar attributes 338

Chapter 10. Troubleshooting and support 345

Troubleshooting overview 345
Troubleshooting IBM InfoSphere Identity Insight 347
Pipelines troubleshooting checklist 347
Analyst toolkit web applications troubleshooting checklist 348
Visualizer troubleshooting checklist 349
System Health 353
Database tables that affect system performance 354
Large Entities query 355
Total Unique Numbers by Entity query. 356
Unique Number Shared by Multiple Entities query 357
Searching knowledge bases. 358
Messages overview 359
UMF parsing errors 360
Logs 360
Pipeline log files 360
Analyst Toolkit web application log files . . . 367
Visualizer log files. 367
Event Manager log files 370
Tracing 370
Getting fixes. 370
Learning about fixes and service updates . . . 371
Service updates. 372
Contacting IBM Software Support 372

Notices 375

Index 379

Preface

IBM InfoSphere Identity Insight helps organizations solve business problems related to recognizing the true identity of someone or something ("Who is Who") and determining the potential value or danger of relationships ("Who Knows Who") among customers, employees, vendors, and other external forces. This analysis occurs in real time and in the context of existing business applications. IBM InfoSphere Identity Insight provides immediate and actionable information to help prevent threat, fraud, abuse, and collusion in all industries.

About this publication

IBM InfoSphere Identity Insight V8.1 is a scalable entity resolution and analysis platform for fighting threat and fraud. This guide provides information about how to use and apply its identity and relationship disambiguation technology to your organization's ability to recognize Who is who? Who knows who? and Who does what? By accumulating identity context over time, InfoSphere Identity Insight V8.1 uses various enterprise sources of information to determine whether persons really are who they say they are. You can apply sophisticated entity algorithms along with patented multicultural name analysis to determine whether a person has been identified before, whether a person is new to your organization, or whether there is some previous assumption that should be corrected based on new facts.

Intended audience

This guide is intended for system administrators, application developers, data analysts, and IBM Professional Services personnel to effectively use the product in your environment.

Prerequisite and related information

This user guide is a subset of the information found in the online information center (http://publib.boulder.ibm.com/infocenter/easii/v8r1m0/index.jsp). It is provided as a convenience. Other product information sources include:

- IBM InfoSphere Identity Insight Version 8 Release 1 Release Notes
- WebSphere Application Server documentation
- Your database software documentation
- IBM Cognos Business Intelligence software documentation
- IBM ILOG Visualization Software documentation
- Depending on your deployment, any of the following information:
  - Your message queuing software documentation
  - Your address correction software documentation
  - Your ETL tool software documentation

How to send your comments

Your feedback is important in helping to provide the most accurate and high-quality information. If you have any comments about this book or any other IBM InfoSphere Identity Insight documentation, use the following form to send us your comments:

http://www.ibm.com/software/data/rcf/

You can also go to the information center and use the embedded feedback forms and related feedback options.

Contacting IBM Software Support

IBM Software Support provides assistance with product defects.

Before you begin

Before contacting IBM Software Support, your company must have an active IBM software maintenance contract, and you must be authorized to submit problems to IBM. For information about the types of maintenance contracts available, see "Enhanced Support" in the *Software Support Handbook* at techsupport.services.ibm.com/guides/services.html

About this task

Complete the following steps to contact IBM Software Support with a problem:

Procedure

1. Define the problem, gather background information, and determine the severity of the problem. For help, see the "Contacting IBM" in the *Software Support Handbook* at techsupport.services.ibm.com/guides/beforecontacting.html
2. Gather diagnostic information.
3. Be prepared to provide the following information in the problem report to assist IBM Software Support:
   - Product name and version
   - Database type and version
   - Operating system name and version
4. Submit your problem to IBM Software Support in one of the following ways:
   - Online: Click **Submit and track problems** on the IBM Software Support site at http://www.ibm.com/software/support/probsub.html
   - By phone: For the phone number to call in your country, go to the Contacts page of the IBM Software Support Handbook at techsupport.services.ibm.com/guides/contacts.html

What to do next

If the problem you submit is for a software defect or for missing or inaccurate documentation, IBM Software Support creates an Authorized Program Analysis Report (APAR). The APAR describes the problem in detail. Whenever possible, IBM Software Support provides a workaround that you can implement until the APAR is resolved and a fix is delivered. IBM publishes resolved APARs on the Software Support Web site daily, so that other users who experience the same problem can benefit from the same resolution.

Chapter 1. Overview of IBM InfoSphere Identity Insight

IBM® InfoSphere Identity Insight helps organizations solve business problems related to recognizing the true identity of someone or something ("who is who") and determining the potential value or danger of relationships ("who knows who") among customers, employees, vendors, and other external forces. IBM InfoSphere Identity Insight provides immediate and actionable information to help prevent threat, fraud, abuse, and collusion in all industries.

In many organizations, the raw data that represents identities and relationships already exists. The problem with most systems is that there is no simple way to manage, analyze, and resolve the volume of data that you need to gain the maximum insight from it.

With IBM InfoSphere Identity Insight, organizations can manage, analyze, and integrate data in real time from any source, such as customer databases, vendor lists, employee databases, regulatory compliance lists, and streaming data feeds. IBM InfoSphere Identity Insight sends real-time alerts to analysts, security, or other personnel for further investigation. IBM InfoSphere Identity Insight can also help identify the network value of customers or their market segments, based on a comprehensive view of the customer.

Using IBM InfoSphere Identity Insight, organizations can construct a central, dynamic entity database that can be used as a platform for all their knowledge-based applications. IBM InfoSphere Identity Insight integrates with other enterprise systems through a wide variety of protocols and technologies.

Recognizing identities

Using the core process of entity resolution, IBM InfoSphere Identity Insight resolves inconsistent, ambiguous identity records into comprehensive entities across multiple data sets, despite deliberate attempts at misrepresentation.

During entity resolution, IBM InfoSphere Identity Insight:

- Determines when multiple records that seem to describe different entities are actually a single entity.
- For each resolved entity, integrates the disparate identity records into a composite view of the entity, while maintaining full attribution for each record. Full attribution ensures that data is never lost and is always traceable back to its original source.
- As new data is loaded into the system, IBM InfoSphere Identity Insight updates and manages the information in context for the entities in the entity database. It can completely comprehend the meaning of new or changed data as it is loaded, making the most of each transaction and enhancing the comprehensive view of each entity in the entity database.

Detecting relationships

Building on the entity relationship process, IBM InfoSphere Identity Insight detects relationships between entities in the entity database, as records from multiple data sources are loaded and processed.

During the entity resolution process, IBM InfoSphere Identity Insight:

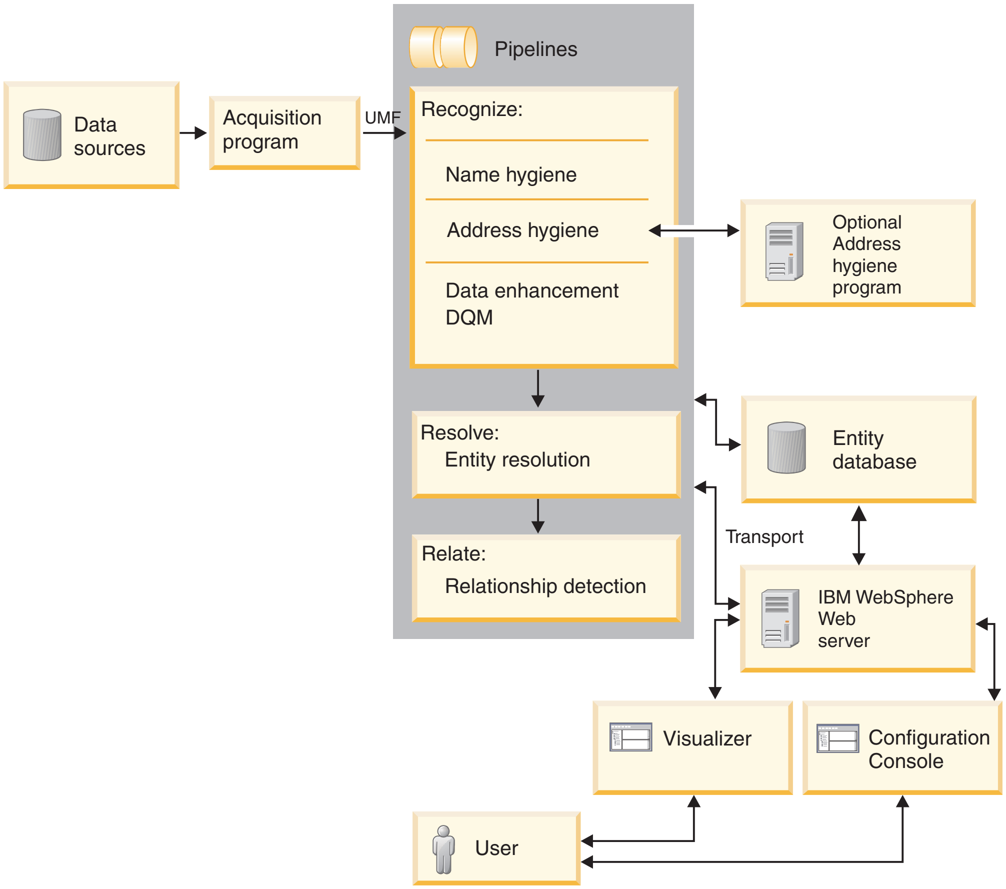
- Links entities by identity attributes, such as telephone numbers and addresses, to uncover relevant, yet non-obvious relationships.
- Assembles networks of associations and entities using individual data attributes (such as identification numbers and names), locations (such as IP addresses), facilities (such as warehouses, schools, airports, or hotels), organizations (such as cells, clubs, associations, or gangs), money (such as cash or wire transfers), and accounts (such as loyalty clubs, banks, checking, credit, or savings).
- Identifies suspect or interesting relationships, even those that are hidden or disguised, and sends real-time alerts, based on a set of user-defined rules. IBM InfoSphere Identity Insight enables analysts and investigators to perform sophisticated searches against the entity database to further explore each related entity and every entity or attribute that those entities are linked to.

IBM InfoSphere Identity Insight also supports customizable rule-based exception reporting, so organizations can specify which entities resolved or which relationships detected trigger alerts.

Product architecture

IBM InfoSphere Identity Insight is a multi-tier system, in which data from data sources is loaded into the system from acquisition programs and processed by the pipelines that are hosted by pipeline nodes. The results of the processing are written to the entity database and can be routed to other systems or other databases.

In a typical deployment, enterprise data from multiple data sources is sent to acquisition programs, where the data is transformed into Universal Messaging Format (UMF). Each acquisition program uses a transport to send the data into one or more pipelines. Many of these transports are bi-directional, and the system can be configured to provide the acquisition program with responses.



One or more pipeline processes run on pipeline nodes. Each pipeline maintains its own connection to the entity database. As the pipeline receives the UMF data from one or more acquisition programs, it processes the data record-by-record through its three core processes: recognize, resolve, and relate. As each record is processed, the pipeline stores the results of the processing in the entity database.

Users interact with the system using these interfaces:

- Configuration Console, which is used to configure and monitor the system
- Analyst Toolkit applications, which can be used to analyze and disposition alerts, explore relationships, perform searches, and generate reports
- Command line interfaces, which are used to run pipelines
- Web services, which can be used to run the pipelines or integrate the product with other enterprise systems, including customized user interfaces

IBM InfoSphere Identity Insight uses IBM WebSphere Liberty. This application server hosts the Configuration Console, Analyst Toolkit elements, and the Web services.

This robust architecture provides scalability for any deployment. Pipelines can be deployed on any number of small or large machines. Pipeline performance can scale to any desired level, given enough database capacity.

Acquisition programs

An acquisition program contains the tools and programs that acquire data, transform it into Universal Message Format (UMF), and then submit the transformed data to the pipeline for processing.

You can use the acquisition program utilities provided with the product to transform data into UMF, or you can use extract, transform, and load (ETL) tools, such as WebSphere® QualityStage, as your acquisition programs.

Universal Message Format (UMF)

Universal Message Format (UMF) is an extensible XML dialect used for structuring data source files. UMF contains standard tags that represent key pieces of identities, relationships, and activities. Before data can be processed by the pipelines, it must be converted into UMF and follow the UMF specification.

UMF consists of these hierarchical components:

UMF documents
The collection of UMF segments that structure the data and indicate the type of data source record.

UMF segments
The part of the UMF document that structures the data for the data source.

UMF elements
XML tags and values that define the data within a UMF segment of a UMF document.

The UMF specification lists the specific types of UMF documents, the UMF segments within each UMF document type, and the valid UMF elements within each UMF segment.

Pipelines

Pipelines are the components that perform name and address hygiene standardization, data quality management, and entity resolution. The pipelines also perform relationship resolution and generate alerts, based on the system configuration.

Pipelines perform three core processes:

- Recognize, which involves optimizing incoming data by performing data standardization, hygiene, enhancement, and quality checks
- Resolve, which involves resolving entities
- Relate, which involves detecting relationships and generating alerts

Pipelines are hosted by pipeline nodes.

You can configure pipelines for parallel processing, so that one pipeline command spawns multiple parallel pipeline processing threads, which enables the system to concurrently process multiple data requests. This feature can help improve system performance, reduce data processing time, and mitigate hardware memory constraints.

The parallel pipeline processing feature is configured in two places:

- The global concurrency setting is controlled by the `Pipeline default concurrency` parameter on the **System Configuration** tab in the Configuration Console. The value here determines the number of parallel processing threads started from a pipeline start command. The default value for this parameter is 1, meaning that unless this parameter is edited, only one pipeline processing thread starts.

- A local concurrency setting (by pipeline node) can be configured in the pipeline configuration file. If you specify a concurrency parameter and value in the pipeline configuration file by pipeline node, that value overrides the global system parameter. When you issue a pipeline start command on that pipeline node, you start the same number of concurrent pipeline processing threads as specified in the pipeline configuration file.

Pipeline nodes

Pipeline nodes are the physical machines that host one or more pipeline processes.

The pipeline node is where you install and start the pipeline executable that runs the pipeline processes. You configure and maintain the pipeline configuration file for all pipelines that are hosted by this machine. The system also writes the pipeline messages to the log files on the pipeline nodes.

Pipeline nodes connect pipeline processes to these components of the product architecture:

Acquisition programs

As part of the extract, transform, and load (ETL) process, acquisition programs use transports to send UMF data into pipelines for processing. You use the transport method appropriate to the type of acquisition program to connect to the pipelines. For example, if you use the UMF file utility as an acquisition program, you use the file transport.

Entity database

The entity database contains entity information. Pipelines access entity information while processing incoming records for entity and relationship resolution. The pipeline node must have the appropriate database client installed and configured, so that the pipelines can access the entity database.

Queues

If your system uses queues as transport methods to send data to the pipelines for processing, you must install and configure the appropriate message queuing software on each pipeline node.

Address hygiene servers

If your system uses address hygiene products from other companies for additional address cleansing, each pipeline node must be configured to connect to the address hygiene servers.

Web services

You must use an HTTP transport to connect the pipeline processes on the pipeline node to the Web services.

Application monitor

The Configuration Console includes an application monitor that you use to monitor pipelines (their status, statistics, and errors) and route results between pipelines and other systems or databases.

To take advantage of the application monitor, you must register the pipelines that you want to monitor or route results from in the Configuration Console.

Monitoring pipelines

The application monitor works with an SNMP agent that runs on the pipeline node hosting the pipelines that you want to monitor. The SNMP agent sends statistics on all registered pipelines on a pipeline node to the application monitor, which publishes them in the Configuration Console. The application monitor refreshes pipeline status and statistics every 60 seconds.

Routing pipeline results

The application monitor enables you to route the results of the data processed by the pipelines to other systems or databases. To route the results of pipeline processing, you use the Configuration Console to configure routing rules, which specify which pipeline to route from and where the results are routed to.

For example, rather than have analysts build report queries against the entity database (which might be cumbersome), some organizations choose to route a subset of the results to a reports database. The analysts build and run their investigative report queries against the reports database, which contains only the entity and relationship information important to the analysts.

Transports

Transports move data from one place to another – between acquisition programs and pipelines, between pipelines and the entity database, and even between pipelines and external systems.

To transport data, you must use a syntax format specific to the type of transport mode you are using, which includes an Universal Resource Identifier (URI).

IBM InfoSphere® Identity Insight supports several transport methods:

- Databases
- Files
- HTTP
- Message queues (IBM WebSphere MQ)

Data sources

Data sources contain the identities that you want to process for entity resolution and load into the entity database. Data sources contain identifying data (unique, personal identifiers for an identity) and non-identifying data (other attributes and data points for an identity). The identity records in the data source must be exported as Universal Message Format (UMF) before they can be processed by the system or loaded into the entity database. Examples of data sources include, but are not limited to, employee lists, watch lists, customer lists, and vendor lists.

Data sources contain vital information, such as the information about the original source (because the original data was transformed into UMF) or the external reference for the data source. These details make each data source unique in the system.

During entity resolution, if two entities are unresolved, the system uses the data source information to determine which information belongs with which entity.

Data source locations and source systems

You can organize incoming data sources by creating source locations and source systems and associating them with your data sources. You can use source locations and source systems to distinguish among similar types of data sources.

For example, if you are processing reservation data and human resource data from more than one location, you can use a data source location to distinguish which location is contributing the data:

- Property X Reservation data
- Property X Human Resource data
- Property Y Reservation data
- Property Y Human Resource data

Configurations by data source

To maximize the results of entity resolution and relationship detection, configure each data source using these settings:

Roles Because data sources are groupings of the same type of data, you can automatically assign the same role to every identity record in the same incoming data source. For example, by associating the Employee role to a human resources data source, all incoming records from the employee list are automatically assigned the Employee role.

Load levels
You can determine whether to load all the data in an incoming data source or only the data that resolves or relates to one or more entities.

Relationship resolution settings
You can configure the level of relationship detection by data source. For example, you can turn off relationship resolution for a data source or select the number of degrees of separation for detecting relationships within that particular data source.

Entity database

The entity database is the database that stores identities, entities, and data that is used for relationships, resolutions, and alerts.

The entity database is the persistent store of all resolved entities and their relationships. As pipelines process incoming UMF records, the new data is constantly compared against the data that is already in the entity database. Therefore, entity resolution and relationship detection occur against composite entities that contain all of the accumulated attributes of all prior records.

User interfaces

IBM InfoSphere Identity Insight offers several user interfaces to interact with the product features.

Configuration Console

The Configuration Console provides a task-oriented interface to help you more easily do some of the most essential tasks to get up and running with Identity Insight.

The Configuration Console is hosted by IBM WebSphere Liberty.

Managing system configuration

The Configuration Console is used to configure most of the system parameters and options in a set of simplified, streamlined interfaces. The console then writes the changes to the configuration database. Changes made directly to the configuration database are not supported; these changes most likely result in the product not working properly

Explorer

The Identity Insight Explorer is a thick-client Windows application based on the Identity Insight plug-in for i2 Analyst's Notebook. The Identity Insight Explorer allows users to search for and investigate entities, analyze and disposition alerts, and visualize entity networks using the thin-client graph.

Graph

The InfoSphere Identity Insight graphing tool gives users the ability to analyze web-based graphs that visualize Identity Insight alerts, entity relationships, and other entity information.

EntitySearcher thin client

The EntitySearcher thin client combines the best of find-by-attribute and find-by-resolution search capabilities in a browser-based client

Cognos Reports

The Analyst Toolkit provides a set of Cognos reports that can be used to create customized Identity Insight reports.

Identity Insight Plug-in for i2

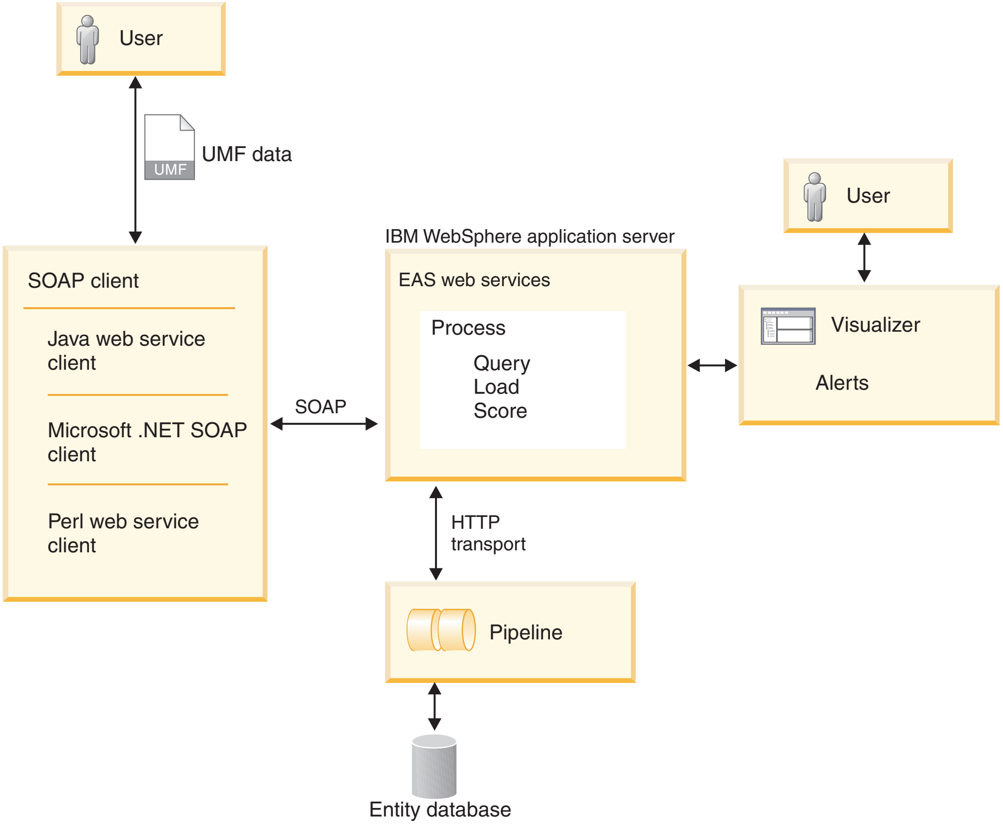
The Identity Insight plug-in for i2 combines the power of Identity Insight's resolution, relationship and alert detection with i2's visualization and analytical tools. Seamlessly integrated with IBM i2 Analyst's Notebook, the plug-in allows users to view and search for entities and alerts. Entities and alerts can be visualized directly in IBM i2 Analyst's Notebook, which allows users to visually investigate entity networks and take advantage of the analytical tools provided by Analyst's Notebook.

Command line interfaces

The product uses command line interfaces to run the pipelines. You start and stop pipelines issuing commands on a command line.

Web services

IBM InfoSphere Identity Insight provides a set of Web services that you can use to build external applications that can load Universal Message Format (UMF) data for pipeline processing or search for entities in the entity database. You use the bi-directional HTTP (hypertext transfer protocol) transport method, which is a standard feature in the pipeline.



IBM InfoSphere Identity Insight Web services use four SOAP (Simple Object Access Protocol) methods: process, search, load, and score The product supports SOAP version 1.1.

The product includes several components to help you get started using Web services.

srd.wsdl
: This file contains a Web services description language (WSDL) definition of the product Web services. You can use this file with any SOAP toolkit or technology to start the Web services. It can be found by starting WebSphere Liberty and loading the file from http://hostname:port/easws/resources/wsdl/srd.wsdl

wsutil.jar
: This file is a Web services test client provided for testing your Web services installation and configuration. This utility can be found in the `ibm-home/easws` directory.

Core concepts

To use IBM InfoSphere Identity effectively, you must understand its key concepts, such as entities, identities, and attributes.

Entities

An entity is a collection of one or more identities that represent the same person, organization, place, or item. Entities are stored in the entity database.

Although entities are often thought of as people, entities can also be things such as businesses or vehicles. In fact, you can use the system's extensible configuration to map your organization's data and create any type of entity that you want to resolve or relate.

Entities are often composed of identities that come from several different source systems. Entity resolution determines which identities are really the same entity and creates a composite entity that contains all the identities associated with that composite entity. The system maintains full attribution of the records, identifying the source associated with each identity in the composite entity.

You configure the system to resolve and relate entities in a way that meets the goals of your organization.

Identities

Identities are a collection of attributes from a data source that represent a person, organization, place, or item.

Through entity resolution, identities are resolved and composite entities are created from the individual identities, when the identities share common attributes with the composite entity.

Previously, identities might have been called accounts.

Attributes

Attributes are characteristics or traits that describe a person, organization, place, or item. Common attributes include information such as names, addresses, phone numbers, credit card numbers, tax identification numbers, and license numbers.

The system supports the following kinds of attributes:

Names
Name attributes define the name of the person, organization, place or item, as defined by the entity model and the incoming identity. Name attributes typically represent persons and businesses, but they can extend to the names of vehicles (such as cars, trucks, ships, or planes), groups, or any other type of entity that your business defines in its entity model.

Addresses
Address attributes define a location of the identity and typically contain standard address information: street name and number, unit or building number, city, state, country, and postal code.

Numbers
Number attributes are comprised of data that is usually described as a number, such as credit card numbers, phone numbers, and passport numbers. Numbers are not limited to numeric characters only though, because many numbers use alphanumeric characters.

Characteristics
Characteristic attributes define other identity traits or information that is not expressed through the other kinds of attributes. You can use characteristic attributes to customize the system to define identity characteristics that you want to use for resolving entities or detecting relationships. Common types of characteristics include dates of birth and gender.

E-mails
E-mail attributes define Internet e-mail addresses. E-mail addresses tend to be unique; some studies have suggested that people who tend to use more than one name still tend to use the same one or two e-mail addresses.

In Universal Message Format (UMF), the various kinds of attributes are expressed in UMF segments. Each kind of attribute is its own UMF segment.

Entity resolution

Entity resolution is the process that resolves entities and detects relationships. The pipelines perform entity resolution as they process incoming identity records in three phases: recognize, resolve, and relate.

Recognize

During entity resolution, pipelines must recognize the data by validating, optimizing, and enhancing the incoming identity data. During this recognize phase of the pipeline process, the pipelines cleanse and standardize the data values, as well as perform data quality checks on the data to protect the integrity of the entity database.

Data Quality Management (DQM)

Data quality management (DQM) is the pipeline process that checks the data for required values, valid data types, and valid codes. You can also configure DQM to correct the data by providing default values, formatting numbers and dates, and adding new codes.

Data quality management, along with name hygiene and standardization and address hygiene and standardization, is designed to optimize and enhance data quality. This data quality preparation is an essential step in entity resolution, because it increases the confidence in the resulting resolved entities and detected relationships.

To apply data quality management to the data loaded into the system, you configure data quality management rules (or DQM rules). DQM rules can perform a variety of repair, clean up, and standardization functions on incoming identity data values, such as properly formatting numbers, identifying and correcting clerical or transposition errors, and identifying and correcting intentional inaccuracies introduced by those intent on trying to conceal their identities.

The product comes pre-configured with several DQM rules by UMF segment that handle the most typical data quality issues for that UMF segment. But you can configure additional DQM rules, as needed. Before you do so, however, you must be familiar with the original quality of the data and the ETL (extract, transform, and load) process that was used to transform the identity data into UMF. After you know what further data enhancement is necessary, you can select the right DQM rules, functions, and values to apply to each type of identity data that needs further data quality optimization.

Example of using a DQM rule

For example, the date format for your system is DD/MM/YYYY. But in several of your data sources, the date values are formatted as MM-DD-YYYY. You can add the DQM rule 204 to the `<NUMBER>` UMF segment, configuring it to fix all incoming dates formatted as MM-DD-YYYY to the date format of DD/MM/YYYY.

Name hygiene and standardization:

During pipeline processing, names are cleansed and standardized to prepare the identity record for optimal entity resolution processing.

Pipeline processes provides the most accurate name information about entities for current, future, and historical use. As new or changed identity name data enters the system, it is compared against the product name standardization dictionary, which contains a list of root names and their known derivatives, to identify the root name. When the root name is identified, the system keeps both the root name and the original name for the incoming identity record.

For example, the following table shows two examples of possible derivatives of the same root name, including the various ways to spell the name. The names on the left are all derivatives of the root name on the right.

Table 1. Examples of some possible derivatives for the root names of Richard and Mohammad

Derivatives	Root
Dick, Dickie, Ricardo Rich, Richie, Rick Rickey, Ricki, Rickie Ricky, Rikki, Ritchie	Richard
Mohamad, Mohammad Mohamed, Mohammed	Mohammad

The name hygiene and standardization process also corrects any misspellings, if necessary, but again, the system keeps both the original spelling and any corrections as part of the record. Most other systems (including ETL and database marketing tools) do not.

Name hygiene and standardization are an important step to increase the confidence levels of entity resolution. This process is especially important because the average person uses as many as five different versions of his or her name for official and consumer purposes.

Address hygiene and standardization:

Address hygiene and standardization is the pipeline process that normalizes and standardizes address information to correct possible errors and transpositions and to prepare the identity record for optimal entity resolution processing.

As part of the address hygiene process, the pipelines parse and standardize address information. For example, `Street` to `St` or `123-A Main St`to `123 Main St Apt A`.

This pipeline process also verifies new or changed information against a global address database and standardization software provided by the IBM InfoSphere QualityStage product or by another address hygiene product, such as the Group 1 Software CODE-1 product. The chosen address hygiene product determines if the address information is correctly formatted, corrects any detected misspellings (such as misspelled street names), and corrects any missing or incorrect information (such as updating the city name to match the postal code and address).

For example, the following table shows examples of address cleansing and standardization from the original address to the corrected, standardized address.

Table 2. Examples comparing two original addresses with the resulting standardized address

Original address	Standardized address
460 Oak Street Mill Valleu, CA 94914	460 South Oak Street Mill Valley, CA 94914
4737 Simeron Drive Easton, MA 02334	4737 Cimmeron Drive Easton, MA 02334

The address hygiene and standardization pipeline process retains both the original address, as well as the corrected and enhanced address, to enhance the confidence levels of later entity resolution and relationship detection. Retaining this information also provides better historical information.

Data quality check:

As identity data comes into the system for processing, the pipeline checks the quality of the data to protect the integrity of the entity database. Each incoming identity record is tested for proper Universal Message Format (UMF) construction, required values, valid data types, and configured data source codes.

As the process checks the data quality, it attempts to correct the problems, if it is possible and if the system is configured to do so. When determining whether or not to correct data quality problems, the system uses the configured data quality management (DQM) rules. DQM rules define which data quality defects on incoming identity records are acceptable for the system to correct and which defects are acceptable to leave as-is but still process the records.

To view the data quality for a particular data source, you can view or print the Load Summary report. The Quality summary section can give you helpful insights into the overall data quality for that data source or for a particular set of identity records loaded from that data source. Using this information, you can adjust your ETL process, as necessary, for a particular data source.

The standard logging and error handling logs all data quality errors and corrections, as well as errors that the system could not or did not correct. Check the system logs frequently, so that you are aware of data quality errors that were not corrected by pipeline processing. In most cases, you will need to correct the data quality errors, and then reload the corrected identity records into a pipeline for entity resolution processing.

Data quality check examples

The system can automatically add codes that are not recognized as new codes, if it is configured to do so. The UMF_EXCEPT log shows the results of new codes added by the system or records rejected and not processed, because the system did not recognize a code and was not configured to add it as new.

The table below shows two examples of codes on incoming records that were not already configured in the system.

Table 3. Examples of two codes not configured in the system and the result of system processing

Code	Quality check	UMF_EXCEPT log
Addr_Type x	New code added	write to log
Num_Type xxx	New code rejected	write to log

- In the first example, the system is configured to automatically add the new address type code.
- In the second example, the system is not configured to automatically add the new code or allow the record to be processed for entity resolution.

In both cases, the system logs the action to the appropriate log file.

Resolve

During entity resolution, the pipelines resolve identities into entities. After the data values in the identity records have been cleansed, standardized or enhanced, the pipeline uses sophisticated search algorithms to compare the data values in the incoming identity record against existing entities in the entity database to determine if they are the same entity.

Resolving entities involves these phases:

Generating candidate lists
The system uses the information on the incoming identity record to match against entities already in the entity database to create a list of potential entity resolution candidates. Each candidate shares enough attribute values to continue evaluating the candidate for entity resolution. You can configure the criteria that the system uses to generate the candidate lists.

Performing entity resolution
After generating candidate lists, the system then applies the resolution rules to each entity on the candidate list, using a scoring method that calculates a resolution score to determine if the incoming identity and the existing entity should be resolved. You can configure resolution rules and set the thresholds for the resolution scores to determine how closely the attribute values must be for the incoming identity and the candidate entity to be resolved into one entity.

Candidate lists

Candidate lists are the lists of entities that have the potential to match the incoming identity record. The candidate list is built by retrieving those entities that share attributes with the incoming identity, based on the attributes that are specified in the candidate builder configuration.

The entity resolution process only uses the entities on the candidate list for resolving entities and resolving relationships.

Because entity resolution and relationship detection are determined based on attributes, you want to carefully consider the attributes in your data sources to determine which attributes create the strongest candidates.

After the candidate list is generated, the entity resolution process compares the incoming identity to the first candidate on the list using the configured resolution rules. The system uses the resolution rules, in order, to compute a resolution score that represents how closely the incoming identity attributes match the attributes of the candidate entity. If the incoming identity attributes meet or exceed the resolution score for that rule, the incoming identity record is resolved into the candidate entity.

If the resolution score does not meet or exceed the resolution score set for that resolution rule, the system goes to the next resolution rule until the incoming identity record has been resolved into a candidate entity or all resolution rules have been exhausted.

If the incoming identity record is not resolved into an existing entity, the system resolves the record into a new entity and stores the new entity in the entity database.

Resolution rules

Resolution rules are a set of criteria that the system uses to define how compared entities resolve (if they are or are not the same entity) and relate (if entities are not resolved to the same entity, how many attributes they share).

When defining resolution rules, you must specify thresholds that contribute to the total resolution score, which determines whether an incoming identity resolves into an existing entity:

- Candidate thresholds specify which attribute data values are compared to determine whether an identity and an entity will be resolved into one composite entity. The threshold is the minimum score at which a particular attribute value must match between the incoming identity and an existing entity to satisfy the resolution rule.
- Confirm/deny thresholds specify how much scoring weight (positive or negative) is given to matching or conflicting attribute data values when you enable the use of denials.

You can also specify how conflicting values for the same attributes affect the resolution score. These conflicting values are called denials. You can configure resolution rules that specify that the rule is not met if there are any conflicts (denials) in the attribute values. You can also adjust the thresholds for a resolution rule to create automatic denials, based on the comparison scores not meeting one or more specified threshold scores. The higher a threshold score is set, the more exact the match must be in order to satisfy the resolution rule.

Re-resolve

The re-resolve process occurs during the entity resolution process when two entities are resolved as the same entity, and a composite entity record is created. Entity resolution uses the new composite entity record to start the process all over again, to see if the new composite entity can be resolved to any of the other entities in the entity database.

Just as with a new incoming entity, the entity resolution process attempts to generate a candidate list of entities from the entity database. If a candidate list can be generated, the entity resolution process begins entity resolution, comparing each candidate on the list against the new composite entity. If a candidate list cannot be generated, the entity resolution process continues to the relationship detection process.

Unresolve

The unresolve process occurs as part of the entity resolution process when the attribute values on the incoming identity provide new information that indicates that a composite entity is actually made up of two entities and the composite entity record is split into the two entities. The system knows which records belong to which entity because of the data source associated with each record. Once the unresolve process completes, the system then begins the re-resolve process.

Unresolve example

Previously, the system resolved an incoming identity record for Will Smith at address 1234 Main Street, Anytown, USA, phone number (201) 555-2244, and e-mail address of jrsmith@internetprovider.com to a William Smith, Sr. at that same address with that same phone number.

Now a new incoming identity record is processed for Will Smith, Jr. at e-mail address jrsmith@internetprovider.com with credit card 123-54-9999.

Based on the new information of Junior and the credit card number, the system can determine that the William Smith, Sr. composite entity record should be unresolved into William Smith, Sr. and William Smith, Jr. After the one entity is split into two entities, the system begins the re-resolve process to check if any other entities in the database now resolve to William Smith, Jr, based on the new information.

Relate

During entity resolution, pipelines also complete the relationship detection process, which detects relationships between identities and entities and generates alerts for relationships of interest.

The system uses roles which are the classification of an identity that defines the focus or purpose for that identity to detect and establish relationships between entities. In the system, you define roles and then assign roles to identities by data source or as part of transforming the original data source data into Universal Message Format (UMF).

When the pipeline processes incoming identities for entity resolution and resolves the identity to an existing entity, the two records have a 0-degree relationship; that is, the incoming identity and the entity are the same. But the entity resolution process can go beyond 0-degree relationships, depending upon how the system is configured.

After the pipeline exhausts all possibilities in the entity resolution resolve phase, the relationship detection process evaluates the entities that remain on the candidate list, or those entities that did not resolve to the incoming identity, to see if a relationship exists between them. Typically, entities that are on the candidate list are linked to the incoming identity at 1-degree of separation for at least one attribute, meaning that both entities share the same attribute data values for at least one attribute, which is why the entity is on the candidate list.

After the process detects a relationship, the system compares the assigned roles between the identity and the entities to the configured role alert rules. If the system finds that the roles assigned to the identity and an entity meet the criteria for that rule, the system generates an alert, indicating it has detected a relationship of interest. The relationship can be at 0-degrees, 1-degree, or multiple degrees, depending upon how the system and the role alert rules are configured.

Relationships

Relationships are links between two or more entities. Relationships are detected at the end of entity resolution process, when two entities share several data attribute values.

Relationships can be based on links discovered by the system, disclosed by an analyst, or both. However, not every relationship is interesting enough to warrant generating an alert for further analysis or investigation. You define relationships of interest by configuring role alert rules that specify which combination of roles assigned to entities need to generate alerts.

Relationship examples

Examples of relationships that might be detected during entity resolution:

- A customer is also a vendor. Based on the policies and procedures of your organization, this might be considered a relationship of interest.
- An employee knows a customer. Unless the policies and procedures of your organization forbid such an association, or perhaps depending on the data that is shared by the employee and the customer, this might not be considered a relationship of interest.
- A customer knows another customer. If one of the customers has a high net worth to your company, knowing whom your customer knows might be a good way to use your customer network to market within that network.

Overview of Degrees of Separation:

The Degrees of Separation feature extends the relationship matching capabilities of IBM Relationship Resolution.

The default behavior of IBM InfoSphere Identity Insight identifies high-interest relationships and matches entities at one degree of separation from an inbound identity resolved to an entity. The enablement of the Degrees of Separation feature extends these capabilities to an almost limitless range of user-defined degrees of separation from an inbound identity resolved to an entity.

The Degrees of Separation feature uses separation configurations, roles, role alert rules, and relationship scores, to make real-time link analysis against very large data sets.

When an inbound identity is resolved to an entity, an entity graph is created using the one degree relationships that IBM InfoSphere Identity Insight detects. The entity graph uses the one degree relationships to build multi-degree relationship chains stemming from the entity the inbound identity was resolved to. A role alert chain can then be created by linking two multi-degree relationship chains, each stemming from the entity the inbound identity was resolved to. The role alert chain can then be used to find a relationship between the entities at the end and inclusive of each multi-degree relationship chain.

Degrees of Separation reduces work by evaluating all paths that connect two entities and using the strongest path strength in reporting relationships. Degrees of Separation can be configured to report one role alert for each configured role alert rule per entity the inbound identity was resolved to.

The Degrees of Separation configuration can be set in the Console by using the **System Configuration** tab, Degrees of Separation value.

Impersonal awareness:

Impersonal awareness is a product feature that extends the traditional relationship resolution process to find and analyze impersonal relationships. The relationship detection process finds relationships between entities based on attribute values associated with those entities. Sometimes, it is important to find relationships between entities based on activities or other impersonal identifiers. These relationships between entities based on activities or other impersonal identifiers are referred to as *impersonal* relationships, and activities or impersonal identifiers that relate people are called *relating facts*.

Impersonal relationships always exist at two or more degrees of separation, because the relating fact is, itself, an entity. So to enable impersonal awareness and find impersonal relationships, configure your data sources to use the Degrees of Separation feature, which extends entity and relationship resolution to detect relationships at more than two degrees of separation.

For example, a telephone transaction contains data about telephone numbers - both the calling number and the receiving number. Even though a person placed the telephone call to another person, from the telephone transaction alone, no common data can be attributed to the persons. Often, the relating fact (the telephone call) is known before any other information about the related entities (the two people involved in the telephone call) is known. Since these relating facts cannot be attributed to a person, they must be represented as separate entities that are not people, but relate to people. However, impersonal awareness recognizes that a relationship between two persons exists as a consequence of the phone call.

UMF includes an entity type functionality, which allows you to define relating facts as entity types. Using this functionality, relating facts become separate entities in the entity database and can be used to detect relationships between Person entities. By configuring new entity types, specifying the appropriate entity type in UMF, and creating new resolution configurations, these relating facts may be used to automatically find impersonal relationships and conflicts between entities.

Entities of differing entity types never cross-resolve, even if the resolution rules allow it, and even if the data supports the resolution. This means that an entity type of Phone call never resolves to an entity type of Person .

The Analyst Toolkit graphs and reports impersonal relationships and any associated alerts, just as it does personal relationships and associated alerts.

Impersonal awareness example

For example, if you wanted to find impersonal relationships using telephone calls, you would create a new entity type of Phone call and adjust your acquisition node to correctly tag each telephone call record with the *Phone call* entity type.

When the telephone records are ingested into the system, standard entity and relationship resolution detects a one degree relationship between the Phone call entity and the calling entity (Person). It also finds a one degree relationship between the person called and the Phone call entity. By itself, the system does not detect a relationship between the persons.

However, when Degrees of Separation is configured, it continues the analysis and detects the two degree impersonal relationship between the caller and person called. An impersonal relationship exists, based on the telephone numbers that are attributes of the Phone call entity type. Degrees of Separation then analyzes the impersonal relationship and generates an alert if a conflict is found.

Roles

A role is a classification of an identity that defines the focus or purpose of that identity. You can associate one or more roles with an identity. As identities are resolved into entities, entities inherit all associated roles.

You use roles to configure role alert rules, which define relationships of interest and generate alerts.

Every identity is assigned a role in one of two ways:

By incoming data source
When you configure a new data source, you associate a role with that data source, which will assign that role to all identities containing that data source code.

By UMF
When you transform the data source into Universal Message Format (UMF), you can directly assign roles as part of the UMF record using the `<SEP_ROLES>` UMF segment with the `<ROLE_CODE>` UMF tag. If you configure by UMF, DQM rules and a lookup table will need to be added.

Examples of useful roles might include employees, vendors, customers, or watch list.

Example of assigning roles using UMF

To assign the role of employee to an identity record using UMF, you would enter the following `<SEP_ROLES>` UMF segment and `<ROLE_CODE>` UMF tag for the identity record:

```
<SEP_ROLES>

  <ROLE_CODE>Employee</ROLE_CODE>

</SEP_ROLES>
```

Alerts

Alerts are messages or other indications that signal an event has occurred.

Alerts are generated in one of two ways:

- Attribute alerts are generated whenever entities match a specified collection of attributes.
- Role alerts are generated whenever one or more entities that are linked through a relationship share roles that the user has defined as *of interest* or *conflicting*.

It is important to define what alerts meet your organization's goals. A good place to begin is to ask which relationships between entities are of interest to your organization. Relationships are based on user-configured roles, which are assigned to incoming data records by source system. When two entities share enough attribute data values without resolving to the same entity, those entities form a relationship. Make sure that the role alert rules configured for your organization clearly define which entity roles create a relationship that your analysts want to investigate further.

Alert examples

Some examples of relationships of interest that your organization might want to generate alerts include:

- One of the people employed by your organization is also a vendor supplying goods or services for payment to your organization.
- One of your customers shares an address and a name similar to a person listed on a government watch list.
- Two of the people who filed slip-and-fall reports at your organization have similar names and addresses and share a phone number.

Attribute alerts:

Attribute alerts are alerts that are produced by attribute alert generators, which create a persistent system query looking for specific attributes or identities in the entity database. Whenever attributes for entities match the criteria of the attribute alert generator, the system creates an attribute alert.

Visualizer users create their own personal attribute alert generators. If you are looking for a specific identity or any identities or entities that match a specific set of attributes, you can create your own personal attribute alert generator that searches for matches until the specified expiration date.

Examples of possible entity attributes you might want to be notified about include:

- Name and unique number (such as a credit card number)
- Name and phone number
- Address
- Name and non-unique number

Attribute alert generators are configured and viewable using the Visualizer. The attribute alert generators that you create are only available to you.

Example of an address attribute alert

You are watching the address of 675 Hickory Street Las Vegas, NV. You can configure an attribute alert generator to create an attribute alert whenever that address is associated with an incoming identity record added to the entity database.

Role alerts:

A role alert identifies when one or two entities linked through a relationship that meets or exceeds a configured role alert rule. Role alerts are based on configured roles and role alert rules. They can indicate a warning or a problem (such as a customer knows a bad guy) or simply indicate relationships of interest (such as a customer knows an employee).

You define relationships *of interest* or as *conflicting* by configuring role alert rules that identify which roles should not exist in a single entity or cannot be linked between one or more entities. Use the Configuration Console to configure filters for role alerts, which determine if the system re-alerts when there is new information (such as a new identity or a new data source code).

During entity resolution, the pipeline evaluates relationships between the incoming identity and entities on the candidate list. After determining a relationship exists between the incoming identity and a candidate entity, the system then evaluates whether the roles assigned meet a configured role alert rule. If so, the system generates a role alert.

A role alert identifies entity data at the time the role alert was created. The Role Alert detail screen shows the entity data as it existed at the time the role alert was created. As entity data changes over time, the entity resume contains the latest entity data. If you want to see the current data for a particular entity, view the entity resume.

You can view and work with role alerts in the Analyst's Toolkit components (Cognos reports, the Identity Insight plug-in for i2, and the Identity Insight Explorer).

Role alert rules:

Role alert rules are user-defined rules that identify one or more roles that cannot exist in a single entity or cannot be linked between multiple entities. During entity resolution, if the criteria for a role alert rule is met, the system generates a role alert.

Although most role alert rules specify when roles conflict, you can also define a role alert rule where an entity that is assigned a role knows another entity that is assigned the same role. For example, you might find it interesting to know whenever your customers know one another, and define a role alert rule (*customer knows customer*) that generates a role alert whenever one customer entity relates to another customer entity in the entity database.

Because entities consist of multiple records (often from different data sources), and because roles are typically assigned by data source, one entity can be assigned multiple roles. So it is also possible to define a role alert rule that generates a role alert whenever one entity is assigned both the customer role and the bad guy role, based on incoming data.

Note: Keep in mind that when a system is configured to use a large number of roles, the number of role alert rules increases exponentially.

Although the system detects each relationship that violates a role alert rule, by default it only reports one role alert for each entity. For example, if the system detects that an entity that is assigned an employee role is related to two different vendor entities, and a role alert rule is configured to generate a role alert when employee knows vendor, both conflicts are detected and written to the database, but by default, only one role alert is reported.

When configuring role alert rules, you can also specify alert filters that govern whether the system re-alerts (generates a new alert) when new identities or new data source codes are introduced to existing entities involved in a previously generated alert.

Role alert invalidation:

As data is processed through entity and relationship resolution, entities and the relationships between them change over time. Those changes, based on the perpetual analysis of new and existing data, can cause role alerts to become invalid. The role alert invalidation feature of InfoSphere Identity Insight provides the most current context to analysts, so analysts do not spend their time researching conflicts that are no longer valid.

Role alert invalidation removes relationship-based role alerts that are still in the pending status. Typically, alerts in the pending status have not yet been viewed or processed by an analyst. If a role alert has another status, such as completed or assigned, even though the data supports invalidating that role alert, the alert is not invalidated. Only one status can be assigned to an alert so if the alert is already in an Assigned or Completed state, it is not invalidated.

Role alerts that occur at 0-degrees are also invalidated, when an identity is deleted or unresolved from the entity.

How role alert invalidation works

Relationship-based role alerts can become invalid for several reasons:

- If an entity changes its entity ID as part of the re-resolve or unresolve processes during entity resolution, the relationship can either go away or be transferred to a new entity ID.
- If a single entity becomes two separate entities based on new data, the new entities are each assigned a new entity ID. Through full attribution, all the data that belongs to the new entity is removed from the old entity and added to the new entity, including roles that create relationship-based role alerts
- When data is deleted from the entity database, either an entire entity or a key component of a relationship can be removed, causing a role alert to become invalid.
- When data is marked as generic, its ability to be used in detecting relationships is diminished or removed. If a relationship is removed, all role alerts that depend upon that relationship become invalid.

Replacement role alerts

Whenever a role alert is invalidated, the pipeline automatically re-evaluates each conflict along the relationship path, looking for data to support an alternate relationship-based conflict.

A *relationship path* is the chain of entities and attributes that link one entity to another entity. The length of the relationship path is determined by the configuration for degrees of separation. The separation configurations are set through the Configuration Console.

Scoring

During entity resolution, the system computes how closely the attributes for an incoming identity match the attributes of an existing entity. The results of this computational analysis are scores that the system uses to resolve identities into entities and detect relationships between entities.

Resolution scores

The resolution score is the value that is assigned during entity resolution as a result of the confirmation and denial processing and that defines the likelihood that the compared identities represent the same entity. This score is user-defined and is used to resolve a new identity to an existing entity.

As the pipeline processes incoming identities for entity resolution, it compares the shared attribute values for the attributes of the incoming identity and of each entity on the candidate list. Part of the comparison includes computing scores that represent how closely the attribute values match. These scores are then compared against the configured thresholds and resolution score for each resolution rule. After the entity resolution process uses a confirmation and denial process to prevent false positives, the system creates a baseline resolution score for both the incoming identity and the entity on the candidate list.

If one or more attributes is configured to be used for further confirmation denial, the process then evaluates those attributes. The results affect the baseline resolution scores of the incoming identity and the candidate entity. If the attribute values match, the resolution score can be positively affected by adding the configured number of points. If the attribute values do not match, the relationship score can be negatively affected by subtracting the configured number of points. When you configure an attribute to be used for confirmation or denial, you specify the number of points up or down to adjust the baseline resolution score.

The system then compares the resulting resolution score of the incoming identity and the candidate entity to each resolution rule. If the resolution score meets or exceeds the configured resolution confidence score for the resolution rule, the system resolves the incoming identity into the candidate entity, creating one composite entity in the entity database.

Relationship scores

The relationship score is the value that is assigned during entity resolution as a result of applying the resolution rules and that defines how closely the two compared identities are related to each other. This score is user-defined and is used to relate entities.

During entity resolution, the pipeline compares the incoming identity (which may not be resolved into an entity) against the remaining entities on the candidate list. While these candidate entities may not resolve to the incoming identity, they are still evaluated for relationships.

During the relationship detection process, the pipelines determine relationships by calculating a relationship score for each attribute data value that is shared between the incoming identity and the entities on the candidate list, starting with the first entity:

- If the relationship score satisfies the criteria configured for relationships (by degrees of separation), the system determines that the two entities are related. The relationship is written to both composite entity records. The system then checks the configured role alert rules to determine if the relationship is considered a relationship of interest. If so, the system generates an alert. If not, it moves to the next entity on the candidate list.
- If the relationship score does not satisfy the criteria configured for relationships, the process moves to the next entity on the candidate list, until all entities have been evaluated for relationships.

Event Manager

Event Manager extends the capabilities of IBM InfoSphere Identity Insight by combining near real-time event analysis and event monitoring with identity and relationship resolution. When enabled, Event Manager provides your organization with the ability to track business events and alert on suspicious events or events of interest, so that you can take the appropriate business action in a timely manner, to assist your organization in its fight against threat and fraud.

Because the threat and fraud scenarios are constantly changing, Event Manager provides you with the flexibility to define the types of events to track, and configure the business rules for processing events and generating event alerts. These rules are a set of criteria that Event Manager uses to determine how events are processed and what triggers an event alert. You configure the business rules, based on your business needs and scenarios.

You also decide what constitutes an event alert. Event alerts are not typically triggered by a single event, but by a series of complex events that all happen at different times, within different contexts. For example, you might define a business rule that aggregates money transfers by customer over a given time period, and generates an alert if the total amount exceeds the legal limit. Or you might define a business rule that alerts you when two credit card purchases using the same credit card number occur within the same hour at locations more than 200 miles apart.

How event processing works

The Event Manager feature of IBM InfoSphere Identity Insight works with the IBM Active Middleware^{tm} Technology complex event processor, which consists of two parts: the CEP engine and the Eclipse^{TM}-based Rule Author tool. You configure the business rules for events and event alerts in the Rule Author tool, then export that configuration as the `CEP.XML` file. After you enable Event Manager, whenever the pipeline detects incoming UMF data formatted using the EVENT data segment, the pipeline processes the data for identity resolution and then passes on the processed data to the CEP engine. The CEP engine processes the event data against the event business rules configured in the `CEP.XML` file, and returns the decision information back to the IBM InfoSphere Identity Insight pipeline, where the event information is stored in the entity database. If there are event alerts associated with an event or combination of events, you can configure Event Manager to display those event alerts in the Visualizer or another visualizer tool for further analysis and disposition.

You can also configure your client application so that the CEP engine can return immediate decisions to the client application, providing your organizational representatives with on-the-spot information. For example, the CEP engine could immediately alert your customer service representatives to stop a transaction, such as a wire transfer that would exceed the legal dollar limit allowed for a customer to transfer within a 24-hour time period.

Events

Events represent information about something that happened in the business domain, such as "a customer opens an account" or "a customer wires money". In Event Manager, events contain attributes, which are based on their corresponding event types.

Event alerts

An event alert occurs when one of more complex events meets set criteria over a specified life span. Event alerts are based on complex event business rules and other configurations contained in an event rules file (`cep.xml`). These alerts can indicate situations of interest, such as "Two or more purchases of more than $10,000 U.S. dollars occurred in the last hour at locations 200 miles from each other".

Event types

Event types categorize events and define the unit of measure for the value associated with events in Event Manager. Examples of event types include wire transfer, account opening, or credit card transaction.

Event types are required for event processing, because the user-defined business rules that the event processor uses call a specific event type. If the event type does not exist, the event processor cannot process the event.

Event rules

Event rules are a set of business rules that determine how incoming event records are processed by the complex event processing (CEP) engine, and what type of event response (such as an event alert) is returned to the pipeline and the client application. You configure event rules in the Eclipse-based^{tm} complex event processing Rule Author tool. Event rules are grouped under a CEP project and exported into an event rules file named `cep.xml`.

You configure event rules to return information and alerts based on items of interest to your organization or analysts. Event rules can be configured to alert on the data from a single incoming event record. But most event rules group a collection of complex event data and trigger an alert after a particular threshold or condition is met.

In the Rule Author tool, event business rules are called *situations*. For more information, refer to “CEP terminology” on page 29.

Common event rules contain summing or counting functions. For example, you can configure an event rule to generate an event alert when any entity wires more than $15,000 U.S. dollars in a 24-hour time period.

Getting started with Event Manager

Use the following steps as a checklist to configure and use Event Manager.

Procedure

1. Required: Install the Eclipse-based CEP (complex event processor) Rule Author tool. The Eclipse^{TM}-based Rule Author tool is not automatically installed with the product. (Event Manager functionality and the CEP engine are automatically installed.) The Rule Author tool is included in a ZIP file in the product download.
2. Required: Use the Rule Author tool to create a CEP project to group all the event rules and configurations for Event Manager.
3. Required: In the Rule Author tool, import the `cep.xml` event rules file into the CEP project and customize the file by creating the event rules that meet your

business event processing and alerting usage scenarios. Before altering an original starting file, back up or copy the file to another directory, as a precaution.

Important: The case used for naming the event rules file is very important, especially in the Unix environment. The file name must be in lower case only.

4. Required: Export the `cep.xml` event rules file. The CEP engine and Event Manager use this event rules XML file to process events and determine when to generate alerts. The exported XML file must be named `cep.xml`, and it must be located in the following directory: `*product_installation_home*/ibm-home/gem/`.
5. Required: Configure Event Manager system parameters in the Configuration Console.

   Remember: Before system configuration changes take effect, you must stop and restart all running pipelines. You can either stop all running pipelines before configuring Event Manager system parameters and event types or stop and restart all running pipelines after you configure Event Manager system parameters and event types.
6. Required: Configure event types in the Configuration Console.

   Remember: Before system configuration changes take effect, you must stop and restart all running pipelines. You can either stop all running pipelines before configuring Event Manager system parameters and event types or stop and restart all running pipelines after you configure Event Manager system parameters and event types.
7. To see event alerts in the Analyst Toolkit applications, do the following:
   a. Optional: Identity Insight already contains default activity codes for dealing with event alerts (Pending, Assigned, and Closed). But you can create additional activity codes for event alerts in the Configuration Console, if you want. Stop all running pipelines before creating the activity codes, and then restart the pipelines after the activity codes are created.
   b. Optional: You can review event alerts, change the status of event alerts, assign event alerts to yourself, or assign event alerts to other analyst alert groups.
   c. Optional: If you want to view the full details about a specific event alert, you can generate the Event Alert Detail report.
   d. Optional: You can view the event alert history for an entity on the entity resume.
   e. Optional: From the entity resume, you can click **Show Events** to view all the events associated with the entity, even events that did not generate an event alert. Or you can click **Report** to print an All Events report that also shows all event associated with the entity.
8. Required: Use the EVENT data segment definitions to include event processing information in the UMF data that you convert to send to the pipelines.
9. Optional: If you want to send system messages (including Event Manager messages) to your client application, make sure to use an HTTP pipeline, and make sure that your client application can receive messages from the standard SYSTEM_MESSAGE return document.
10. Optional: After Event Manager processes events, you can review the Event Manager log files and the associated Configuration Console log files.

Enabling Event Manager in the Configuration Console

Before you can process events using Event Manager, you must enable and configure Event Manager in the Configuration Console.

About this task

Procedure

1. In the Configuration Console, click the **System Configuration** tab.
2. To enable event processing, modify the **Enable event processing** value.
3. To configure the Universal Resource Indicator (URI) to CEP, modify the **Event processor URI** value. The default should be http://localhost:13510/gem
4. To increase the total event processing time setting, modify the **Event processor timeout** value. This setting indicates in seconds amount of time that the pipeline waits for a response from the external event processor (CEP) before timing out with an error.
5. To modify the number of days of event history sent to the pipeline for use in evaluating a new inbound event, modify the **Event history window** value.
6. Click **Save**.

Configuring the Event Manager CEP module

In IBM InfoSphere Identity Insight, *CEP* refers to the complex event processing tools packaged with the product. These tools are the components within Event Manager that extend identity and relationship resolution to process event transactions and generate event alerts. This section provides information on how to configure the CEP tools to work specifically with Event Manager.

Architecture

The CEP component of Event Manager consists of two tools:

Eclipse^{tm}-based Rule Author tool

The Eclipse-based complex event processing Rule Author tool is the component that you use to configure and export event rules in the `cep.xml` file. The event rules file determines how events are processed and what triggers an event alert.

When you install IBM InfoSphere Identity Insight, you also install a compressed file containing the Rule Author tool and its User's Guide. However, you must first uncompress the tool files before you can begin configuring event rules.

Complex event processing engine (CEP engine)

The complex event processing engine (CEP engine) is the component that processes incoming event data against the event rules configured in the `cep.xml` file.

When the pipeline receives data formatted using the EVENT data segment of an incoming UMF document, it sends that data to the CEP engine for event processing. After the CEP engine evaluates the event data against the configured `cep.xml` file, it sends the results back to the pipeline. If the event data meets or exceeds any configured event rule, the CEP engine also sends back a signal to the pipeline to generate an event alert. Whether or not an event alert is generated, the final event data that the pipeline receives is written to the entity database.

The CEP engine is installed by default with IBM InfoSphere Identity Insight.

These complex event processing components are part of a specific version of IBM Active Middleware^{tm} Technology that are included within Event Manager. These complex event processing components are included in your product purchase.

The cep.xml file

The `cep.xml` file contains the event rules and other settings necessary to process event data and generate event alerts. The Event Manager functionality in the pipeline and the CEP engine can only process events against the event rules file named `cep.xml`. This file is in E**x**tensible **M**arkup **L**anguage (XML) format, because data that comes into the pipeline is formatted in **U**niversal **M**essaging **F**ormat (UMF), a format based on XML.

A starting `cep.xml` file is included with your product installation that contains the many of the required configuration settings that Event Manager needs to work with the CEP engine. You can import the starting `cep.xml` into a CEP project, and then configure the event business rules..

Note: Before you import the `cep.xml`event rules file and make any changes or export that file, make a back up copy of the original and store the original file in another directory. Consider using a versioning or source control system whenever you modify the event rules file.

Additional resources for CEP

For more in-depth usage information about using the Eclipse-based Rule Author tool, refer to the User's Guide for the tool. The guide, named `AMT3.0.UserGuide.PDF`, is located at*`install_path`*`/cep/`.

Installing the Eclipse-based complex event processor Rule Author tool

Complete the following steps to install the Eclipse^{TM}-based Rule Author tool onto a workstation. Both the Event Manager and the CEP engine are installed with the product installation program. But you must install the Rule Author tool from a ZIP file that is included in the install.

Before you begin

The Rule Author tool only works on a Microsoft Windows operating system and requires Java version 1.5 or higher.

About this task

You use the Rule Author tool to configure the rules and thresholds that are used to monitor your business, and then you export that information into the event rules file (`cep.xml`). Event Manager and the complex event processor (CEP) engine use the event rules file to process events and detect event alerts. Event alerts can be associated with a single event or a combination of events. You can configure Event Manager to display those event alerts in the Analyst Toolkit or another visualization tool for further analysis.

To install the Eclipse-based Rule Author tool from the ZIP file:

Procedure

1. Browse to the product installation directory.
2. Browse to the `/cep` subdirectory.
3. Copy the `CEP_3.0.1.1.03.zip` file to a Microsoft Windows client machine.
4. Uncompress the `CEP_3.0.1.1.03 file` to *`drive letter`*`:/CEP/`

What to do next

For in-depth usage information about the Rule Author tool, refer to the User's Guide located in the `cep/AMT3.0_UserGuide.PDF` file.

Starting the Rule Author tool:

To use the Eclipse-based^{tm} Rule Author tool, you must first start the tool. The Rule Author tool is installed and started separate from IBM InfoSphere Identity Insight components.

About this task

The Rule Author tool only works on a client with a Microsoft Windows operating system and requires Java version 1.5 or higher.

Procedure

1. Open Microsoft Windows Explorer and navigate to the directory where the Eclipse-based Rule Author tool is installed.
2. Double-click the batch script named `AmitIDE.cmd`. The batch script opens the Rule Authoring tool executable.

CEP terminology

Some of the terms used in the Eclipse-based^{tm} Rule Author tool might differ slightly from terms used in IBM InfoSphere Identity Insight and its components. This glossary can help you understand the complex event processing terms and how they relate to Event Manager and other components.

cep.xml file

The `cep.xml` file contains all the event business rules and required complex event processing configuration settings required for Event Manager and the CEP engine to process incoming event records. An event rules file by this name must be exported to the *`product_installation_directory`*`\ibm-home\gem`\location.

Important: The file name must be all lower case, especially in Unix environments.

You maintain and export the event rules file using the Rule Author tool.

A starting `cep.xml` event rules file is included with your IBM InfoSphere Identity Insight product installation. This starting file already contains many of the required settings and configurations needed to work with Event Manager. You can import the starting `cep.xml` event rules file into the Rule Author tool, making a back up copy of the original file first, to add event business rules and to export the file to the required location. Consider using a versioning or source control system to store the file before and after modifying it.

CEP engine

The complex event processing engine (or CEP engine) is the mechanism that processes incoming event data from the pipeline and evaluates the data against the rules defined in a CEP project. The CEP project is defined in the `cep.xml` file, which is configured and exported by the Rule Author tool.

The CEP engine that Event Manager currently uses is part of the IBM Active Middleware^{tm} Technology product. The version of the CEP engine required by Event Manager is included and installed as part of IBM InfoSphere Identity Insight. However, you must configure Event Manager in the Configuration Console and event rules in the Rule Author tool before you can successfully process events with the CEP engine.

CEP projects

Projects are a top-level group that the complex event processor uses to contain a grouping of events, lifespans, and rules. To use Event Manager, you create **one** CEP project that contains all the event information, including business event rules, for the events that you want to monitor. Event Manager only uses one CEP project at a time, but any single project can test multiple event types and multiple rules per event type.

You create and maintain the CEP project inside the Rule Author tool.

Rule Author tool (Eclipse-based Rule Author tool)

This tool enables you to define CEP projects, events, and other configurations that are part of the `cep.xml` event rules file that the CEP engine uses to process events and generate event alerts.

Event classes

To use Event Manager, your CEP project must contain the following event classes, which come preconfigured in the starting `cep.xml` file:

- `EAS_START.event`: Used to indicate the Event Manager lifespan initiator.
- `EAS_STOP.event`: Used to indicate the Event Manager lifespan terminator.
- `EVENT.event`: Used to define the Event Manager business rules (or situations) that are used to process incoming event data and generate event alerts.

EVENT.event

This CEP event class maps the input data that is passed from the pipeline to the CEP engine for processing. The mapping corresponds directly to the `GEM_EVENT` table in the entity database. You use the Rule Author tool to ensure the attributes associated with `EAS_EVENT` match the data mappings in the `GEM_EVENT` table.

Lifespans

In CEP, lifespans are the time intervals during which situations (event rules) are relevant. A lifespan always starts with an initiator and always ends with a terminator. Lifespans are associated with an event class.

For Event Manager, the event class EVENT must contain the lifespan initiator `EAS_START` and the lifespan terminator `EAS_STOP`.

Situations

Situations in the Rule Author are equivalent to *event rules*. You use the Rule Author tool to configure situations that define the business rules determining which events or combination of events are meaningful to your organization and what triggers an event alert.

Situations are associated with a CEP project and event class and are contained in the `cep.xml` event rules file.

As UMF data comes into the pipeline, records (or UMF_ENTITY input documents) that contain an EVENT data segment definition are sent to the CEP engine. The CEP engine evaluates this incoming event data against the configured situations in the cep.xml event rules file. If an event meets or exceeds a defined situation, the CEP engine sends an event alert back to the pipeline, which can be displayed in the Analyst Toolkit applications or a visualization tool of your choice.

Threshold condition

You define threshold conditions as part of an event rule (situation). Think of threshold conditions as data filters or quick data checks. During processing, the CEP engine checks the incoming event data to see if it meets the specified threshold condition before it processes the data against the rule. If the data meets the threshold condition, the CEP engine processes the event data against the rule; if the data does not meet the threshold condition, the CEP engine moves to the next event rule.

For example, to only process events that occurred at Branch 102, build a threshold condition that specifies EVENT_LOC='102'.

UMF_LOG_ID key

The UMF_LOG_ID is a unique sequential number assigned to each record as it is processed. In a CEP project, the UMF_LOG_ID is a grouping key that is associated with all of the required Event Manager event classes and lifespan indicators. This grouping key ensures that all the incoming records with the same UMF_LOG_ID are processed together.

If you import the starting cep.xml file included with your product into a CEP project, the UMF_LOG_ID key is already configured and assigned to the Event Manager event classes and lifespan indicators.

Configuring the cep.xml event rules file

The information configured in the cep.xml event rules file determines how Event Manager and the CEP engine process incoming event data, and what responses are returned to your client application, the pipeline, the entity database, and the applications. Event rules are a big part of the information contained in the cep.xml file, but the rules are not the only required information. There are several other elements and settings that must also be included to properly process events through Event Manager.

Your product includes a starting cep.xml event rules file that contains the necessary elements and settings, already configured for you. If you import the starting cep.xml file, you do not need to configure or change these elements or settings, but you can focus on configuring the event business rules and adding the rules to the cep.xml file. Because event rules are unique to each organization, the starting cep.xml file doe not include any pre-configured event rules (or situation types).

Required elements and settings for the cep.xml file

This information is provided for your reference. If you choose to not to import the provided starting cep.xml file, but rather create your own file from scratch, use this information to assure that the file contains all the required elements and settings. If the cep.xml event rules file that you export to use with Event Manager is not complete (does not include this information), Event Manager cannot process incoming event data.

Event classes

Event classes describe the different event structures that the CEP engine needs to be aware of. To process events, the following event classes must be part of the cep.xml event rules file:

EAS_START.event
: This event class becomes the Event Manager lifespan initiator, or the signal to the CEP engine to start processing the event.

EAS_STOP.event
: This event class becomes the Event Manager lifespan terminator, or the signal to the CEP engine to stop processing the event.

EVENT.event
: This event class is the basis for every event business rule that you create. It contains the information that maps the incoming event record data to the Event Manager table (GEM_EVENT) and to the EVENT data segment.

Lifespan

In CEP, a lifespan is a time interval during which particular event rules are relevant. Because the pipeline processes near real-time data, the only real purpose of the lifespan is to signal the beginning and end of an event record.

The lifespan information required for Event Manager processing includes the following elements:

EAS_START
: This element is the required lifespan initator and signals the start of an event. You set this lifespan element in the **Event Initiators** table on the **Lifespan: Initiators** tab.

EAS_STOP
: This element is the required lifespan terminator and signals the end of an event. You select the terminator selection for **Terminate By Event** on the **Lifespan: Terminators & Keys** tab.

UMF_LOG_IDgrouping key
: A UMF_LOG_ID is a unique sequential number assigned to each record as it is processed. In a CEP project, the UMF_LOG_ID grouping key ensures that all incoming records with the same UMF_LOG_ID are processed together. This grouping key is assigned to all event classes and lifespan indicators.

EVENT.event attributes

The required attributes for this event class map directly to the EVENT data segment, which are the fields in the GEM_EVENT table in the entity database. If any of these required attributes are missing from the EVENT.event, event processing fails. You might see one or more error messages, such as errors that mention 'invalid or malformed XML' or 'missing information in the CEP configuration XML file'.

Specify these attributes on the **Situation**: **General & Event** tab of each event rule.

Creating a CEP project:

CEP projects are are a grouping of event rules, lifespans, and other event information used by Event Manager and the CEP engine. CEP projects are part of the cep.xml event rules file and are created and maintained in the Eclipse-based^{tm} CEP Rule Author tool. Before you can configure event business rules for Event Manager, you must first define a CEP project.

Before you begin

- The CEP Rule Author tool must already be installed, and its files uncompressed.
- The CEP Rule Author tool only works on a Microsoft Windows operating system and requires Java version 1.5 or higher.

Procedure

1. In the CEP Rule Author tool, select **File** > **New** > **Project**.
2. Select **Event Processing Project** and click **Next**.
3. Click **Finish**. The CEP project displays in the left navigational pane.

What to do next

Import the starting `ibm-home\gem\cep.xml` event rules file included with your product installation. This file already contains the required elements and settings to work with Event Manager. After you import these required objects into the CEP project, you can configure the event business rules, and then export the final `cep.xml` event rules file to start processing events through Event Manager.

Importing the cep.xml event rules file:

The `cep.xml` event rules file contains the information that the CEP engine and Event Manager use to process events and generate event alerts. A starting `cep.xml` file that already contains the elements and settings required to work with Event Manager is included with your product installation. So rather than starting from scratch, import the existing `cep.xml` file into a CEP project.

Before you begin

- Make a back-up copy of the original `cep.xml` event rules file, so that you can return to the original file, if needed. Consider keeping the file in a versioning or source control system.
- The Eclipse-based^{tm} Rule Author tool must be installed, and its files uncompressed.
- Keep in mind that the Rule Author tool only works on a client with a Microsoft Windows operating system and requires Java version 1.5 or higher.
- You must have already created a CEP project in the Rule Author tool.

Procedure

1. In the Rule Author tool, select **File** > **Import**.
2. Select **Event Processing Definition** and click **Next**.
3. Browse to select the `cep.xml` file. Remember to change the default file type from `DEF` to `XML`. Typically, this file is located in the *`product_installation_directory`*`\ibm-home\gem` directory.
4. Check the following items:
   - Make sure that all contents of the file are selected. (Expand the top folder to examine the contents of the file, if necessary.)
   - Make sure that the correct CEP project name displays. (Browse to select the project, if necessary.)
5. Click **Finish**. Click **OK** to override the existing file, if you receive that message. When the file is successfully imported, several plus signs display in the Rule Author tool left navigational pane.

What to do next

Add business event rules and then export the cep.xml event rules file to the *product_install_directory*\ibm-home\gem\ directory.

Exporting the cep.xml event rules file:

For Event Manager to execute the complex event processing rules, you must export the cep.xml event rules file that you configured in the Eclipse-based^{tm} Rule Author tool.

Before you begin

The Rule Author tool only works on a client with a Microsoft Windows operating system and requires Java version 1.5 or higher.

About this task

- If the CEP engine is already up and running when you export the event rules file, you must reload the file on the IBM WebSphere server for the changes in the new exported cep.xml event rules file to take effect.

Procedure

1. In the Rule Author tool, select **File** > **Export**.
2. Select **Event Processing Definition** and click **Next**.
3. Select the CEP project.
4. Set the event processing definition file to the new cep.xml event rules file. The file is typically located at *product_installation_directory*\ibm-home\gem\cep.xml.
5. Click **Finish**. If the system warns you of overwriting an existing cep.xml file, click **OK**.
6. Optional: If the IBM WebSphere server is currently running, reload the CEP rules. When the CEP engine starts on the product application server, CEP loads the current cep.xml event rules file. If the WebSphere server is running when you export the file, the changes do not take effect until you reload the new cep.xml file.
   a. Open a Web browser window and navigate to the WebSphere server. For example, http://localhost:13510/gem.
   b. Click **Reload Rules**.

      Note: The WebSphere server does not visibly acknowledge that the rules have been reloaded.

Guidelines for configuring event rule results

Event rules define how to process events and what situations generate event alerts. Event rules (called *Situation Types* in the Eclipse-based^{tm} CEP Rule Author tool) are included in the cep.xml event rules file that Event Manager and the CEP engine use to process incoming event data. The complex event rules that you define are unique to your organization.

Before you begin defining event rules, keep the following considerations in mind, so that the rule works with Event Manager:

- Remember that the event rules should center on an entity and the transactions that an entity can do. Entities are typically persons, but an entity can also represent a place or a thing. For example, an entity can be a ship.
- Event rules must be expressed either as a declarative statement (such as 'Location=Texas') or as a mathematical expression (sum, count, average) expressed over time.

Required situation attributes for each event business rule

To return event data from the complex event processor to the entity database, you must manually add the required situation attributes to each event business rule that you create. These attributes are not part of the starting `cep.xml` event rules file, so importing that starting file does not automatically create event business rules (situations) or add these attributes to any new or existing rules.

These situation attributes map event data directly to the Event Manager `GEM_EVENT` table (and match the UMF from each incoming event record). Without these required attributes, none of the data processed by the CEP engine is returned to Event Manager through the pipeline.

Table 4. Required situation attributes for complex event business rules

Attribute name	Attribute type	Attribute expression	Attribute description
`EVENT_SIT_STATUS`	`string`	`"PENDING"`	Indicates the event alert status for the event alert. In the i2 plug-in, Explorer, and Cognos Alert Summary report, the event alert status is displayed as part of the Alert Summary. All newly generated alerts typically receive the pending status, indicating that an analyst needs to analyze and disposition that alert. Keep in mind that an event alert status can be anything that makes sense for your organization and is configured as an event status in the Configuration Console. If you do not want the event to display in the Analyst Toolkit component user interfaces, use the "`CLOSED`" event alert status.

Table 4. Required situation attributes for complex event business rules (continued)

Attribute name	Attribute type	Attribute expression	Attribute description
REASON_DESC	string	"*<Description of event rule or alert>*"	Describes the event rule that triggered the event alert. Make this description as meaningful as possible for your analysts. For example, if the event rule generates an alert when an entity transacts over $1500 in a 24-hour period, you might enter "SumOver1500" as the REASON_DESC.
ALERT_GROUP	string	"*<alert group>*"	Indicates which alert group to assign event alerts generated from this event rule. Typically, this value is "DEFAULT", but you can enter any alert group configured in the Configuration Console.

Displaying the detail of event alerts

Typically, event alerts are triggered from more than one complex event. You can display event alerts in the Analyst Toolkit applications or a client application, but by default, the details of the events that made up that alert are not included.

If you want to include the details of the events that make up the event alert, you must include the following situation attribute:

Table 5. Settings needed to create the EVENTS situation attribute in an event rule

Name	Type	Expression	Dimension (Show Advanced button)
EVENTS	integer	Event.EventID	[] (to indicate that the EventID is an array] You must edit the Situation Attribute and click the Show Advanced button to see and define the setting for this column.

Best practices

If you display your event alerts in the Analyst Toolkit applications, keep the REASON_DESC situation attribute a simple string of text, rather than adding values from the event to the message. The Analyst Toolkit groups common alerts into one alert summary that includes a count of the number of alerts included in the summary. Analysts click on an alert summary to disposition all the alerts contained in that summary.

If you define values from the event in the REASON_DESC, each event alert displays as a separate alert summary with a count of 1, which means that your analysts see every event alert in both the alert summary and the alert detail areas of the Alert Summary window.

Creating an event rule to sum complex events

Create a basic SUM event rule to sum the totals of events and create an event alert if the sum of those events exceeds a set threshold. For example, you could create an event rule that sums all money transfers sent by one person within 24 hours and send an event alert if the sum of those money transfers (events) is more than $15,000.

Before you begin

You must have an existing CEP project, which groups events rules and all rules configuration.

About this task

These steps provide the basic instructions for creating a simple business rule that sums the value of your choosing. For some steps, there are multiple ways to achieve the same end result. For more options, refer to the *Situations* section of the IBM Advanced Middleware^{tm} Technology User's Guide (guide for the Eclipse-based^{tm} CEP Rule Author tool), which is included with your product.

Procedure

1. In the left navigation pane, right-click **Situation** and select **New** > **Situation**. Make sure the correct project name displays in **Event Processing Project**.
2. Enter a unique rule name in **Situation name**. The situation name is the event rule name that displays in the entity database and in the Visualizer component, if you choose to display event alerts there. Make the name meaningful to those who analyze the event alerts. For example, if you are creating a rule to sum the value of all events and then send an alert if the sum of the events crosses the $15,000 limit, you might name this rule SumOver15K.
3. In **Select source**, select **Empty of Type** and then select **atleast** from the drop-down list. The **atleast** situation can sum event values, as well as preserve the information of each event that met the event rule. For more information about situation types, refer to *Situation Properties* in the User's Guide .
4. Click **Finish**. When the main situations screen displays, you might notice several errors in the **Problems** section. These errors indicate missing values, but you can ignore these errors for now. As you complete these steps, the errors go away.
5. In the **Events** section, select EVENT as the base event for this rule. EVENT is the always the base event for every event business rule. It contains the necessary mapping to the entity database GEM_TABLE and the EVENT data segment.

6. Optional: You can build a *threshold condition* to filter events before they are evaluated against this rule so that events must meet the threshold condition specified to be considered.
7. To build the summing expression, click **Show Advanced** and then **Edit**.
8. In **Quantifier**, select `each`. This selection ensures that each incoming event record that satisfies the conditions of this event rule are included in the total sum.
9. In **Weight**, click **...** to edit the field. Use the **Expression Builder** to select the event field to sum. Make sure that the expression displays in the **Expression Builder Text** area and then click **OK**. By default, the weight of each event is equal to 1. When the event rule is evaluated, the sum of all weights is compared to the **Quantity** attribute on the **Condition & Results** tab. When the total equals at least the quantity indicated, an event alert is generated. For example, to sum the values of each event that meets the event rule, select `EVENT.EVENT_VALUE`.
10. Optional: If the field selected as the weight contains decimal digits (double type), use the Expression Builder to build an expression to do the following: .
    a. Multiply the calculation results by 100 to keep the decimal digits a convert the dollars to cents.
    b. Convert the double type to an integer. You can achieve this using functions.

    For example, if you are totalling the values of events (`EVENT.EVENT_VALUE`), you might enter `EVENT.EVENT_VALUE*100` in the **Expression Builder Text**. area. Then you might select **Functions** > **Math** > **Round** to round the result to the closest integer value. The final expression displays as `Round(EVENT.EVENT_VALUE*100)`.
11. In **Sum Expression**, click **...**to edit the field and select the event field to sum. For example, to sum the value of each event that meets or exceeds the event rule, select `EVENT_VALUE` .
12. Optional: To sum only events that meet a specific condition, enter the condition in **Threshold Condition** or use the Expression Builder to assist you. For example, to only sum the values of events that occurred at Branch 102, enter EVENT.EVENT_LOC="102". This field acts as a filter, automatically skipping events that do not meet or exceed the condition.

    Tip: To simplify your view and see **Threshold Condition** more easily, click **Hide Advanced**.
13. On the **Condition & Results** tab in **Lifespan**, select `EASLifeSpan`. Notice that until you make the selection, this field displays in red. The red color indicates that this is a required field, and it is one of the errors listed in the **Problems** section. When you make the lifespan selection, the error disappears from the **Problems** section.
14. In **Quantity**, enter the "atleast" quantity that the event rule sums up to before generating the event alert. Remember to multiply dollar amounts by 100. For example, to generate an event alert when the sum reaches at least $15,000, enter `150000`.
15. In **Detection Mode**, notice that `immediate` is selected. Keep this selection. The detection mode determines when to calculate and report the results of the events. The `immediate` selection generates an alert as soon as the sum reaches the quantity.
16. In `Situation Attributes`, enter the required situation values for the following situation attributes:
    - `EVENT_SIT_STATUS`

- REASON_DESC
- ALERT_GROUP

17. Optional: To preserve the details of all the events that make up the sum, add the EVENTS situation attribute, using the following information:
    a. In **Name**, enter EVENTS,
    b. In **Type**, enter integer.
    c. In **Expression**, enter EVENT_ID (or select it in the **Expression Builder**).
    d. Click **Show Advanced** to display the **Dimensions** column, and enter [] in the column to indicate that the type is an array of events.

    These values instruct CEP to send the internal EVENT_ID of each event included in the total sum back to the pipeline along with the event alert. The pipeline writes each EVENT_ID to the entity database and sends the information to the Visualizer or the client application used to display event alerts. The EVENT_ID is an internal sequential number (ID) created by the pipeline when it sends event data to the CEP engine.
18. Save the event rule.

Creating an event rule to count complex events

Create a basic COUNT event rule to count events and create an event alert if the total count exceeds a set threshold. For example, you could create an event rule that counts all wire transfer transactions within 24 hours and sends an event alert if the transaction count is more than 500.

Before you begin

You must have an existing CEP project, which groups events rules and all rules configuration.

About this task

These steps provide the basic instructions for creating a simple business rule that counts the value of your choosing. For some steps, there are multiple ways to achieve the same end result. For more options, refer to the *Situations* section of the IBM Advanced Middleware^{tm} Technology User's Guide (guide for the Eclipse-based^{tm} CEP Rule Author tool), which is included with your product.

Procedure

1. In the left navigation pane, right-click **Situation** and select **New** > **Situation**. Make sure the correct project name displays in **Event Processing Project**.
2. Enter a unique rule name in **Situation name**. The situation name is the event rule name that displays in the entity database and in the Visualizer component, if you choose to display event alerts there. Make the name meaningful to those who analyze the event alerts. For example, if you are creating a rule to count all the events that occurred at a particular branch location, you might name this rule CountBranch102Transactions.
3. In **Select source**, select **Empty of Type** and then select one of the following values from the drop-down list:
    - **atleast**: At least *n* or more events have arrived during the life span.
    - **atmost**: No more than *n* events have arrived by the end of the life span.

    Both situation types can count event values, as well as preserve the information of each event that met the event rule. For more information about situation types, refer to *Situation Properties* in the User's Guide .

4. Click **Finish**. When the main situations screen displays, you might notice several errors in the **Problems** section. These errors indicate missing values, but you can ignore these errors for now. As you complete these steps, the errors go away.
5. In the **Events** section, select EVENT as the base event for this rule. EVENT is the always the base event for every event business rule. It contains the necessary mapping to the entity database GEM_TABLE and the EVENT data segment.
6. Optional: You can build a *threshold condition* to filter events before they are evaluated against this rule, so that the events must meet the threshold condition specified to be considered.
7. On the **Condition & Results** tab in **Lifespan**, select EASLifeSpan. Notice that until you make the selection, this field displays in red. The red color indicates that this is a required field, and it is one of the errors listed in the **Problems** section. When you make the lifespan selection, the error disappears from the **Problems** section.
8. In **Quantity**, enter the "atleast" or "atmost" quantity that the event rule counts to before generating the event alert.
9. In **Detection Mode**, notice that immediate is selected. Keep this selection. The detection mode determines when to calculate and report the results of the events. The immediate selection generates an alert as soon as the count reaches the quantity.
10. In Situation Attributes, enter the required situation attribute names, types, and expressions:
    - EVENT_SIT_STATUS
    - REASON_DESC
    - ALERT_GROUP
11. To preserve the details of all the events that make up the count, add the EVENTS situation attribute, using the following information:
    a. In **Name**, enter EVENTS,
    b. In **Type**, enter integer.
    c. In **Expression**, enter EVENT_ID (or select it in the **Expression Builder**).
    d. Click **Show Advanced** to display the **Dimensions** column, and enter [] in the column to indicate that the type is an array of events.

    These values instruct CEP to send the internal EVENT_ID of each event included in the total sum back to the pipeline along with the event alert. The pipeline writes each EVENT_ID to the entity database and sends the information to the Visualizer or the client application used to display event alerts. The EVENT_ID is an internal sequential number (ID) created by the pipeline when it sends event data to the CEP engine.
12. Save the event rule.

Chapter 2. System requirements and planning

This reference section contains information about supported platforms, system requirements, and system architecture.

Detailed System Requirements

These requirements identify the hardware and software products that you must install and use before opening a problem report with the IBM Support team.

System requirements when running on IBM AIX

The following list identifies the products that are supported when IBM InfoSphere Identity Insight runs on the AIX® operating system.

Table 6. System requirements when running on IBM AIX

Operating Systems	• IBM AIX 7.1L
Hardware Requirements	• POWER7® (64-bit) • POWER6® • POWER5
Java™	The following is installed with the product: • IBM 64-bit Java Runtime Environment, Version 8
Databases	• IBM DB2® Database for Linux, UNIX, and Windows 11.1 • IBM DB2 Database for Linux, UNIX, and Windows 10.5 • Oracle 12c • Oracle 11g Release 2 (11.2.0.1, 11.2.0.2, or greater)
Database Clients	• DB2 client v11.1 when connecting to IBM DB2 Database for Linux, UNIX, and Windows 11.1 • DB2 client v10.5 when connecting to IBM DB2 Database for Linux, UNIX, and Windows 10.5 • Oracle 12c client when connecting to Oracle 12c. • Oracle 11g Release 2 client when connecting to Oracle 11g Release 2.

Table 6. System requirements when running on IBM AIX (continued)

Java Database Connectivity (JDBC) Clients	- DB2 client v11.1 JDBC driver when connecting to IBM DB2 Database for Linux, UNIX, and Windows 11.1. - DB2 client v10.5 JDBC driver when connecting to IBM DB2 Database for Linux, UNIX, and Windows 10.5. - Oracle 12c JDBC drivers when connecting to Oracle 12c. - Oracle 11g JDBC drivers when connecting to Oracle 11g.
Web Browsers	- Mozilla Firefox
Message Queuing Software	- IBM WebSphere MQ
Other	- IBM C++ Runtime Environment Components for AIX, For more information about this requirement, review this support information: http://www-01.ibm.com/support/docview.wss?uid=swg24025181

System requirements when running on Linux for System x

The following list identifies the products that are supported when IBM InfoSphere Identity Insight runs on the Linux for System x operating system.

Table 7. System Requirements when running on Linux for System x

Operating Systems	- Red Hat Enterprise Linux AS, Version 7.0 - Red Hat Enterprise Linux AS, Version 6.0
Hardware Requirements	- Intel x86_64
Java	The following is installed with the product: - IBM 64-bit Java Runtime Environment, Version 8
Databases	- IBM DB2 Database for Linux, UNIX, and Windows 11.1 - IBM DB2 Database for Linux, UNIX, and Windows 10.5 - Oracle 12c - Oracle 11g Release 2 (11.2.0.1, 11.2.0.2, or greater)
Database Clients	- DB2 client v11.1 when connecting to IBM DB2 Database for Linux, UNIX, and Windows 11.1 - DB2 client v10.5 when connecting to IBM DB2 Database for Linux, UNIX, and Windows 10.5 - Oracle 12c client when connecting to Oracle 12c. - Oracle 11g Release 2 client when connecting to Oracle 11g Release 2.

Table 7. System Requirements when running on Linux for System x (continued)

Java Database Connectivity (JDBC) Clients	• DB2 client v11.1 JDBC driver when connecting to IBM DB2 Database for Linux, UNIX, and Windows 11.1. • DB2 client v10.5 JDBC driver when connecting to IBM DB2 Database for Linux, UNIX, and Windows 10.5. • Oracle 12c JDBC drivers when connecting to Oracle 12c. • Oracle 11g JDBC drivers when connecting to Oracle 11g.
Web Browsers	• Mozilla Firefox
Supported Message Queuing Software	• IBM WebSphere MQ

System requirements when running on Linux for Power Systems

The following list identifies the products that are supported when IBM InfoSphere Identity Insight runs on the Linux for Power Systems operating system, Little Endian version.

Table 8. System Requirements when running on Linux for Power Systems

Operating Systems	• Red Hat Enterprise Linux AS, Version 7.0 • Ubuntu, Version 15
Hardware Requirements	• IBM Power System, POWER8, Little Endian
Java	The following is installed with the product: • IBM 64-bit Java Runtime Environment, Version 8
Databases	• IBM DB2 Database for Linux, UNIX, and Windows 11.1 • IBM DB2 Database for Linux, UNIX, and Windows 10.5 • Oracle 12c • Oracle 11g Release 2 (11.2.0.1, 11.2.0.2, or greater)
Database Clients	• DB2 client v11.1 when connecting to IBM DB2 Database for Linux, UNIX, and Windows 11.1 • DB2 client v10.5 when connecting to IBM DB2 Database for Linux, UNIX, and Windows 10.5 • Oracle 12c client when connecting to Oracle 12c. • Oracle 11g Release 2 client when connecting to Oracle 11g Release 2.

Table 8. System Requirements when running on Linux for Power Systems (continued)

Java Database Connectivity (JDBC) Clients	• DB2 client v11.1 JDBC driver when connecting to IBM DB2 Database for Linux, UNIX, and Windows 11.1. • DB2 client v10.5 JDBC driver when connecting to IBM DB2 Database for Linux, UNIX, and Windows 10.5. • Oracle 12c JDBC drivers when connecting to Oracle 12c. • Oracle 11g JDBC drivers when connecting to Oracle 11g.
Web Browsers	• Mozilla Firefox
Supported Message Queuing Software	• IBM WebSphere MQ

System requirements when running on Linux for System z

The following list identifies the products that are supported when IBM InfoSphere Identity Insight runs on the 64-bit Linux for System z® operating system.

Table 9. System requirements when running 64-bit Linux on System z

Operating Systems	• Red Hat Enterprise Linux AS, Version 7.0
Hardware Requirements	• IBM System z
Java	The following is installed with the product: • IBM 64-bit Java Runtime Environment, Version 8
Databases	• IBM DB2 Database for Linux, UNIX, and Windows 11.1 • IBM DB2 Database for Linux, UNIX, and Windows 10.5 • Oracle 12c • Oracle 11g Release 2 (11.2.0.1, 11.2.0.2, or greater)
Database Clients	• DB2 client v11.1 when connecting to IBM DB2 Database for Linux, UNIX, and Windows 11.1 • DB2 client v10.5 when connecting to IBM DB2 Database for Linux, UNIX, and Windows 10.5 • Oracale 10g Release 2 (10.2.0.2.0) client when connecting to Oracle 11g Release 1 (11.2.0.1) or 11g Release 2 (11.2.0.2)
Java Database Connectivity (JDBC) Clients	• DB2 client v11.1 JDBC driver when connecting to IBM DB2 Database for Linux, UNIX, and Windows 11.1. • DB2 client v10.5 JDBC driver when connecting to IBM DB2 Database for Linux, UNIX, and Windows 10.5. • Oracale 10g Release 2 (10.2.0.2.0) client when connecting to Oracle 11g Release 1 (11.2.0.1) or 11g Release 2 (11.2.0.2)

Table 9. System requirements when running 64-bit Linux on System z (continued)

Web browsers	• Mozilla Firefox
Supported Message Queuing Software	• IBM WebSphere MQ

System requirements when running on Microsoft Windows Server

The following list identifies the products that are supported when IBM InfoSphere Identity Insight runs on an Microsoft Windows Server 64-bit operating system.

Table 10. System requirements when running on Microsoft Windows Server

Operating Systems	• Microsoft Windows Server 2008 R2 • Microsoft Windows Server 2012 R2
Hardware Requirements	• Intel x86_64
Java	The following is installed with the product: • IBM 64-bit Java Runtime Environment, Version 8
Databases	• IBM DB2 Database for Linux, UNIX, and Windows 11.1 • IBM DB2 Database for Linux, UNIX, and Windows 10.5 • Oracle 12c • Oracle 11g Release 2 (11.2.0.1, 11.2.0.2, or greater)
Database Clients	• DB2 client v11.1 when connecting to IBM DB2 Database for Linux, UNIX, and Windows 11.1 • DB2 client v10.5 when connecting to IBM DB2 Database for Linux, UNIX, and Windows 10.5 • Oracle 12c client when connecting to Oracle 12c. • Oracle 11g Release 2 client when connecting to Oracle 11g Release 2.
Java Database Connectivity (JDBC) Clients	• DB2 client v11.1 JDBC driver when connecting to IBM DB2 Database for Linux, UNIX, and Windows 11.1. • DB2 client v10.5 JDBC driver when connecting to IBM DB2 Database for Linux, UNIX, and Windows 10.5. • Oracle 12c JDBC drivers when connecting to Oracle 12c. • Oracle 11g JDBC drivers when connecting to Oracle 11g.
Web Browsers	• Windows Internet Explorer 10 and above • Mozilla Firefox
Supported Message Queuing Software	• IBM WebSphere MQ

Defining the system architecture

You must plan out the database and server configurations of your product installation.

Pipeline deployments

Pipelines can be installed on a single server or multiple servers, depending on the system requirements and the server resources.

When deploying pipelines, consider the following performance factors:

- Pipelines can be run in single form, or configured to run concurrent parallel processing threads.
- Each CPU can handle 1.5 to 2 pipelines or parallel processing pipeline threads.
- Parallel processing pipelines can receive data from multiple data sources at once, so you do not need to split the files manually to equal the number of single pipelines.

When deploying pipelines, also consider the following factors:

- Pipelines can be executed on any supported hardware and operating system configuration.
- Although possible, do not run the pipelines on the machine where the database is located.
- Parallel processing pipelines are less work to configure than multiple pipelines.
- Multiple server configurations require more work and maintenance to administer.
- Single server configurations require expensive hardware that increases exponentially with the number of CPUs.

Creating a protected user for non-Windows installations

For all non-Windows platforms, create a protected user to run the product installation program.

About this task

Do not run the product installation program as a ROOT user.

User roles and responsibilities

User roles help categorize the typical tasks that must be completed to effectively deploy and use IBM InfoSphere Identity Insight. Many different types of users might use IBM InfoSphere Identity Insight for various purposes; that is, users take on the responsibilities of one or more roles in using the product.

You can define groups of users based on the various user roles and responsibilities.

The most common user roles include these roles:

Analyst

Analyzes the data and reviews entities, relationships, and alerts. The analyst defines what results are most valuable and makes sure that the system returns those results. The analyst works closely with the operator and application administrator.

Operator
Loads data into the system, runs the pipelines, and verifies that the system is running acceptably, providing load-quality reports as necessary. The operator also reviews the results, exceptions, and events. The operator works closely with the analyst, data source administrator, and application administrator.

Data source administrator
Prepares the data for loading it into the system, which includes converting the data to a UMF file and validating the file. The data source administrator works closely with the operators, application administrators, and database administrators.

Application administrator
Configures the application, including the configuration of the data, entity model, and rules. The application administrator works closely with the data source administrators and operators to define the entity model, and coodinates configuration changes with the database administrator, data source administrator, and operators. The application administrator also coordinates and consults with overall system administrators, if they exist.

Database administrator
Ensures that the database is configured and tuned appropriately for use with the application. The database administrator works closely with the operator, data source administrator, and application administrator.

System architect
Sizes and estimates the hardware and software requirements in planning for the deployment of the application. The system architect works closely with the installer, database administrator, data source administrator, and application administrator to ensure the deployment achieves the vision, strategies, and objectives and integrates into your business processes while delivering expected results.

Installer
Manages the installation and initial configuration of the application. The Installer sets up initial users in the system. Frequently, IBM Professional Services works with the system architect to complete these responsibilities.

Programmer
Designs and develops graphical interfaces or customizes graphical interfaces for the various functions, such that the deployment of the application integrates seamlessly into your environment. The programmer works closely with the system architect and the application administrator, often to disseminate alerts to the appropriate people in the most effective manner for your environment.

Security architect
Ensures that the project team adheres to and implements a secure system. The security architect works closely with the system architect, installer, and database administrator.

Chapter 3. Setting up the databases

Before you install the product, you must set up the required databases.

Setting the environment variables

For DB2 or Oracle databases, you must set environment variables.

DB2 environment variables

Set all of the following required environment variables for your operating system on the target machine.

AIX environment variables

Note: You must ensure these environment variable values prepend any existing entries of the same environment variables.

All environment variables must be capitalized.

Table 11. AIX environment variables for DB2 databases

Environment Variable	Value	Conditions
DB2DIR	DB2 software installation path	where *DB2DIR* is the location where the DB2 client/server software is installed.
DB2INSTANCE	DB2 database instance name	where *DB2INSTANCE* is the name of the DB2 database instance you have created.
LIBPATH	`$`*`DB2DIR`*`/lib64:`*`INSTALLDIRECTORY`*`/lib`	where *DB2DIR* is the location where the DB2 client/server software is installed, and where *INSTALLDIRECTORY* is the location where the product will be installed.

Linux environment variables

Table 12. Linux environment variables for DB2 databases

Environment Variable	Value	Conditions
DB2DIR	DB2 software installation path	where *DB2DIR* is the location where the DB2 client/server software is installed.
DB2INSTANCE	DB2 database instance name	where *DB2INSTANCE* is the name of the DB2 database instance you have created.

Table 12. Linux environment variables for DB2 databases (continued)

Environment Variable	Value	Conditions
LD_LIBRARY_PATH	`$DB2DIR/lib64:INSTALLDIRECTORY/lib`	where *DB2DIR* is the location where the DB2 client/server software is installed, and where *INSTALLDIRECTORY* is the location where the product will be installed.

Microsoft Windows environment variables

You must use the Microsoft Windows 8.3 naming convention when setting up environment variables in a Microsoft Windows environment. The environment variables must not contain any spaces.

Table 13. Microsoft Windows environment variables for DB2 databases

Environment Variable	Value	Conditions
DB2DIR	DB2 software installation path	where *DB2DIR* is the location where the DB2 instance was created. Some versions of DB2 instead set *DB2_HOME* or *DB2PATH*. The installer will look for these if *DB2DIR* is not found.
DB2INSTANCE	DB2 database instance name	where *DB2INSTANCE* is the name of the DB2 database instance you have created.
DB2CODEPAGE	Set equal to the CODEPAGE value of the DB2 database.	A mismatch can cause encoding issues for Latin-1/UTF-8 data on data-load.

Oracle environment variables

Set all of the following required environment variables for your operating system on the target machine.

Note: You must ensure these environment variable values prepend any existing entries of the same environment variables.

All environment variables must be capitalized.

AIX environment variables

Table 14. AIX environment variables for Oracle databases

Environment Variable	Value	Conditions
ORACLE_HOME	Oracle client software installation directory	where *ORACLE_HOME* is the location where the Oracle client software is installed.

Table 14. AIX environment variables for Oracle databases (continued)

Environment Variable	Value	Conditions
LIBPATH	$*ORACLE_HOME*/ lib:<*product install directory*>/lib	where *ORACLE_HOME* is the Oracle client software installation directory, and where <*product_install_directory*> is the location where the product will be installed.

Linux 64-bit environment variables

Table 15. Linux 64-bit environment variables for Oracle databases

Environment Variable	Value	Conditions
ORACLE_HOME	Oracle client software installation directory	where *ORACLE_HOME* is the location where the Oracle client software is installed.
LD_LIBRARY_PATH	$*ORACLE_HOME*/ lib:<*product install directory*>/lib	where *ORACLE_HOME* is the Oracle client software installation directory, and where <*product_install_directory*> is the location where the product will be installed.

Microsoft Windows environment variables

You must use the Microsoft Windows 8.3 naming convention when setting up environment variables in a Microsoft Windows environment. The environment variables must not contain any spaces.

Table 16. Microsoft Windows environment variables for Oracle databases

Environment Variable	Value	Conditions
ORACLE_HOME	Oracle client software installation directory	where *ORACLE_HOME* is the location where the Oracle client software is installed.

Granting Oracle users CREATE VIEW privileges

In order for the product to run successfully, Oracle database users need to be granted CREATE VIEW privileges.

About this task

The CREATE VIEW privileges must be assigned to the user directly, and not a role-based assignment.

Creating and configuring the databases

You create a single database, known as the entity database for all components of the product to use.

Creating the entity database

You must create a database for the pipeline to store identities, entities, relationships, and alerts, and to also store Configuration Console configuration information and the application monitoring information.

About this task

See your database documentation for instructions about creating new databases.

Use UPPERCASE letters for database names.

Configuring client authentication

Client authentication allows users to connect to the entity database without supplying additional user name or password credentials in the pipeline's `.ini` file.

About this task

Client authentication is also known as trusted OS database authentication. Client authentication allows the connection to be made by using the currently logged in user name. This authentication scheme trusts that the operating system already properly authenticated the user. Client authentication can be used on DB2, and Oracle database platforms. The pipelines and IBM WebSphere processes must be executed by the O/S user who can access the entity database in trusted mode. If multiple users must execute these processes, contact IBM Support for further details.

Configuring client authentication for DB2 databases

Set up DB2 to use client authentication.

Procedure

1. Set the following global database server configuration options:
   a. Set **`authentication`** to the value `client`.
   b. Set **`trust_allclnts`** to the value `yes`.
   c. Set **`trust-clntauth`** to the value `server`.
2. Catalog the product databases using the **`authentication client`** parameter of the **`db2 catalog database`** command.
3. Synchronize the operating system and DB2 database user names.
4. Ensure you have the DB2 JDBC Type-2 Driver in addition to the standard DB2 JDBC Type-4 Driver. This should be contained in the file `db2java.zip`.
5. Enable trusted authentication when installing the product.

Configuring client authentication for Oracle databases

Set up Oracle to use client authentication.

Procedure

1. Set the following global database server configuration options:
   a. Set **`os_authent_prefix`** to the value `OPS$`.
   b. Set **`remote_os_authent`** to the value `TRUE`.
2. Create Oracle database users so the user can use both external and database authentication methods. Example syntax:

```
CREATE USER OPS$<user> IDENTIFIED BY <dbpassword> DEFAULT
TABLESPACE <tablespace> TEMPORARY TABLESPACE <temp-tablespace>
QUOTA UNLIMITED ON <tablespace>;
GRANT CONNECT, RESOURCE TO OPS$<user>;
```

3. Ensure you have the Oracle JDBC Type-2 Driver in addition to the standard Oracle JDBC Type-4 Driver. For Oracle this should be contained in the file `ojdbc16.zip`
4. Enable trusted authentication when installing the product. Provide a username with the `OPS$` prefix when asked for database credentials in the product installer.

Sizing the Oracle Statement Cache

Oracle database administrators must appropriately size their statement cache.

About this task

The product can be very statement intensive, which means that the Oracle statement cache can grow very quickly and exceed the default Oracle database settings. For more information about sizing and tuning these parameters, see your Oracle documentation.

Procedure

Configure the following parameters at the server level using the **ALTER SYSTEM SET** Oracle command:

SESSION_CACHED_CURSORS
A good value for this parameter is about 20 simultaneous cursors per pipeline or parallel processing pipeline thread.

OPEN_CURSORS
A good value for this parameter is about 20 simultaneous cursors per pipeline or parallel processing pipeline thread.

CURSOR_SHARING
This parameter greatly affects performance. Configure this parameter based on the fact that the product widely uses bind variables, and the application will benefit from cursor sharing

Chapter 4. Administering

Administering tasks include configuring and maintaining system settings for the user interfaces and updating global configuration settings. Administrators use the Configuration Console to perform the administrative tasks.

Administering Users

All Identity Insight applications and plug-ins now recognize the WebSphere Liberty authentication model. By default, Identity Insight ships with Basic Authentication, but this can be extended using any of the WebSphere Liberty supported implementations for authorization and authentication.

Basic Authentication

Identity Insight's authentication framework has been standardized to take advantage of the many authentication options provided by WebSphere Liberty. By default, the simplest file-based Basic Authentication is configured.

There are many methods for storing users, groups and passwords. For more detailed descriptions of other authentication options, see the WebSphere Liberty Core documentation for Authentication.

When Identity Insight is installed, a userid and password are selected for the administrator account. By default, this user has access to all of the user interfaces. (Default user name: admin) Other users must be added by the administrator.

The file containing the user information is found at `wlp/usr/servers/iiServer/users.xml`. You can add, remove and manage users by editing this file.

User management

Modification of users for the default file-based user registry.

About this task

How to add, modify and delete users using the default file-based user registry.

Note: This is the simplest form of authentication that WebSphere Liberty provides. It is strongly advised that a much more secure system be used in production for managing user accounts.

Note: WebSphere Liberty includes a **`securityUtility`** command-line tool that is available in the `wlp/bin` directory. When you run the **`securityUtility encode`** command you can encode an input from the command line or from an input prompt. Copy this value into the password field as required.

Procedure

1. Open the `wlp/usr/servers/iiServer/users.xml` file.
2. Scroll down to the `<basicRegistry>` element.
3. Modify the `<user>` tags
   a. To add a new user: Add a new line following the format: <user name="name" password="password"/>

b. To modify a user: Find the line containing the user's name and modify the password.
c. To remove a user: Remove the line containing the user name to be deleted.

Group management

Modification of groups for the default file-based user registry.

About this task

How to modify groups using the default file-based user registry.

Procedure

1. Open the `wlp/usr/servers/iiServer/users.xml` file.
2. Scroll down to the `<basicRegistry>` element.
3. Modify the `<group>` tags. By default, there are two groups created: admins and groups. These are used by the various applications to limit access to certain functionality. More groups can be generated to create a more fine-grained control over user access.
   a. To add a group: Add a new `<group>` element to the `<basicRegistry>`.
   b. To remove a group: Remove the `<group>` element from the `basicRegistry`.
   c. To add a user to a group: First define the user and then add them as a `<member>` element.
   d. To remove a user: Remove the `<member>` element from the `<group>`

Application management

Control in an application can be granted to a specific set of users or groups

About this task

Access to various applications can be controlled by setting the groups assigned to various security roles within WebSphere Liberty.

Note: Reducing functionality of some elements may cause applications to not work properly. It is advised that you work with IBM Support when modifying the application-level security elements.

Procedure

1. Open the `wlp/usr/servers/iiServer/applications.xml` file.
2. Find the `<application>` whose authentication you would like to manage.
3. Find the `<security-role>` element. Most of the applications contain a single role, but some may have a more granular set of security groups.
4.

   Note: Do not modify the security-role names.
   Modify the `<group>` elements within the security role. The `<special-subject>` element seen in some `<security-role>` elements allows any authenticated user to use this particular application.

Administering the Console

To use the Console effectively you must configure the browsers, set up accounts for the appropriate users and manage access to the Console.

Configuration Console

The Configuration Console provides a task-oriented interface to help you more easily do some of the most essential tasks to get up and running with Identity Insight.

The Configuration Console is hosted by IBM WebSphere Liberty.

Managing system configuration

The Configuration Console is used to configure most of the system parameters and options in a set of simplified, streamlined interfaces. The console then writes the changes to the configuration database. Changes made directly to the configuration database are not supported; these changes most likely result in the product not working properly

User roles and responsibilities

User roles help categorize the typical tasks that must be completed to effectively deploy and use IBM InfoSphere Identity Insight. Many different types of users might use IBM InfoSphere Identity Insight for various purposes; that is, users take on the responsibilities of one or more roles in using the product.

You can define groups of users based on the various user roles and responsibilities.

The most common user roles include these roles:

Analyst

Analyzes the data and reviews entities, relationships, and alerts. The analyst defines what results are most valuable and makes sure that the system returns those results. The analyst works closely with the operator and application administrator.

Operator

Loads data into the system, runs the pipelines, and verifies that the system is running acceptably, providing load-quality reports as necessary. The operator also reviews the results, exceptions, and events. The operator works closely with the analyst, data source administrator, and application administrator.

Data source administrator

Prepares the data for loading it into the system, which includes converting the data to a UMF file and validating the file. The data source administrator works closely with the operators, application administrators, and database administrators.

Application administrator

Configures the application, including the configuration of the data, entity model, and rules. The application administrator works closely with the data source administrators and operators to define the entity model, and coodinates configuration changes with the database administrator, data source administrator, and operators. The application administrator also coordinates and consults with overall system administrators, if they exist.

Database administrator

Ensures that the database is configured and tuned appropriately for use with the application. The database administrator works closely with the operator, data source administrator, and application administrator.

System architect

Sizes and estimates the hardware and software requirements in planning for the deployment of the application. The system architect works closely with the installer, database administrator, data source administrator, and application administrator to ensure the deployment achieves the vision, strategies, and objectives and integrates into your business processes while delivering expected results.

Installer

Manages the installation and initial configuration of the application. The Installer sets up initial users in the system. Frequently, IBM Professional Services works with the system architect to complete these responsibilities.

Programmer

Designs and develops graphical interfaces or customizes graphical interfaces for the various functions, such that the deployment of the application integrates seamlessly into your environment. The programmer works closely with the system architect and the application administrator, often to disseminate alerts to the appropriate people in the most effective manner for your environment.

Security architect

Ensures that the project team adheres to and implements a secure system. The security architect works closely with the system architect, installer, and database administrator.

Optimum browser settings for using the Configuration Console

The Configuration Console is a Web-based application that requires specific settings for the browser that you use to access it.

Use the following browser settings to best view the Configuration Console:

Table 17. Optimum browser settings

Setting	Value	Description
Resolution	800 x 600 minimum; 1024 x 768 or greater recommended	
Text size	Medium	
JavaScript	On	
Cookies	On	At a minimum, first party session cookies must be enabled.
Security - Trusted Web site	HTTP address of the Configuration Console	Make sure that the HTTP address of the Configuration Console is included on the list of trusted Internet Web sites.
Security - Download options	Enabled	Make sure that all download options for trusted Internet Web sites are enabled.
Popup blockers	Allow pop-ups from the HTTP address of the Configuration Console	Make sure that HTTP address of the Configuration Console is on the list of Web sites that allow pop-ups.

Logging in to the Configuration Console

Logging into the Configuration Console allows you to view and change system configuration settings.

Before you begin

Your system administrator must have created a user account for you to use to log in.

Procedure

1. Open the Configuration Console:
   a. Open the browser in which you want to run the Configuration Console.
   b. Enter the URL for the Configuration Console using the following syntax: `http://<servername>/console/`.
   c. Press the **Enter** key.
2. In the **Login** window, type your user name and password.
3. Optional: If you are a system administrator and need to edit the current system configuration, select the **Edit Configuration** option. If you are editing the current system configuration, you typically need to stop all pipelines, to prevent new data from being processed until your configuration changes are complete.
4. Click the **Login** button.

What to do next

If your user name and password match those set up for the Configuration Console, the Configuration Console opens. Otherwise, an error occurs, and you must log in again after determining the appropriate user name and password.

Logging out of the Configuration Console

You can log out of a current Configuration Console session without exiting the application. If there is no activity for 60 minutes, the Configuration Console automatically logs out the current user.

Procedure

Click **Sign off** in the upper-right corner of any Configuration Console window.

What to do next

You are now logged out of the Configuration Console session, and you must log in again to continue using the Configuration Console.

User accounts for the Configuration Console

To log in to the Configuration Console, your system administrator creates and gives you a user account. User accounts include a user name and a password which you can change.

You cannot log in multiple times with the same user account. If you share a user account with other people, you cannot log in to the Configuration Console at the same time. If you attempt to log in using a user account that someone else is currently using, their session will be terminated and your session will begin.

The system administrator can create additional user accounts at any time. Additionally, the system administrator can restart the Configuration Console to force a timeout.

Managing access to the Configuration Console

Each user of the Configuration Console must be given access to it and use a user name and password to log in to it. You can manage the user names and passwords using the application-specific file provided by the Configuration Console. Or if your users have RDBMS user accounts that allow them to access the entity database, you can use those user accounts and the database administration tools to manage users' access to the Configuration Console. These user names and passwords are separate from those configured to access the Visualizer and are not necessarily the same as Visualizer user names and passwords.

Managing access to the Configuration Console using database login information

You can manage access to the configuration Console using the same user ID and password as the entity database.

Before you begin

Ensure that no one is logged in to the Configuration Console, to prevent configuration conflicts

Procedure

1. Launch the configuration utility by going to the `<install location>/installer/util/` directory and typing one of the following commands:
   a. For Windows, type `eacfg.bat –i –l ../logs/`.
   b. For Unix, type `eacfg –i –l ../logs/`.
2. In the navigation pane, click **Configuration Console Settings**.
3. Click the **Modify Configuration Console Authentication** check box.
4. Click the **SQL Authentication** radio button.
5. Click **OK**.
6. Use your database administration tools to specify the Configuration Console (and entity database) login information.

Managing access to the Configuration Console using the password manager utility

You can manage access to the Configuration Console using the password manager utility.

Before you begin

Ensure that no one is logged in to the Configuration Console

Procedure

1. Launch the configuration utility by going to the `<install location>/installer/util/` directory and typing one of the following commands:
   a. For Windows, type `eacfg.bat –i –l ../logs/`.
   b. For Unix, type `eacfg –i –l ../logs/`.
2. In the navigation pane, click **Configuration Console Settings**.
3. Click the **Modify Configuration Console Authentication** check box.

4. Click the **File Authentication** radio button.
5. Click **OK**.

Results

You can now use the password manager utility (pwdmgr.jar) located in the srd-home/console directory, to add or delete users or reset users' passwords in the console_password.properities file.

Viewing a list of users and their status:

You can view a list of users and their status using the password manager command.

Procedure

1. In a command window, change directories to the \srd-home\console directory.
2. Type the following command. `pwdmgr console-passwords.properties console-principals.properties -l`

Example

For example, if you type the command `pwdmgr console-passwords.properties console-principals.properties -l` , the following sample output might be displayed :

```
admin (super-user)
judy (super-user)
allen (super-user)
jose (super-user) *** NEVER LOGGED IN ***
```

If you recently reset a password, a message is displayed showing that the user has not logged in to the Configuration Console yet with the new password.

Adding a new user:

If you are managing access to the Configuration Console, in the console-passwords.properties file, you can add a new user using the password manager command.

Procedure

1. In a command window, change directories to the \srd-home\console directory.
2. Type the following command, `pwdmgr console-passwords.properties console-principals.properties -a username` where *username* is the user name you want to add.

What to do next

A user is added with a default password of the user name that you specified. The new user can now log in to the Configuration Console.

Deleting an existing user:

If you are managing access to the Configuration Console, in the console-passwords.properties file you can delete an existing user using the password manager command.

Before you begin

Make sure you issue the command from the `\srd-home\console\` directory. Also make sure the user you are deleting exists. Trying to delete a user that does not exist will return an error message.

Procedure

1. In a command window, change directories to the \srd-home\console directory.
2. Type the following command, `pwdmgr console-passwords.properties console-principals.properties -d username` where *username* is the user name you want to delete.

What to do next

The user name you just deleted can no longer log in to the Configuration Console.

Resetting a password:

When users forgets their Configuration Console account passwords or a password needs to be changed for security purposes, system administrators can reset the password using the password manager command.

Before you begin

Make sure that you issue the command from the `\srd-home\console\` directory.

Procedure

1. In a command window, change directories to the `\srd-home\console` directory.
2. Type the following command, `pwdmgr console-passwords.properties console-principals.properties -r username` where *username* is the user name of the person whose password you want to reset.

What to do next

The password of the user name that you specified is now reset to match the user name. The next time that users log into the Configuration Console after their password has been reset, the system prompts to reset their password. So after you have reset a password, you might want to suggest that the user log in and change the password as soon as possible to minimize any security concerns or issues.

Password manager command:

Use the password manager command to manage access to the Configuration Console using a properties file. You can add, delete, and list users and reset their passwords..

The syntax for the password manager command is:

```
pwdmgr -option parameter
```

To use a password manager command, issue the command from the `\srd-home\console\` directory.

Options and Parameters

Each option and parameters for the password manager command must be specified as separate commands. If you do not specify an option, the command help is displayed.

-a *username*
: Adds one user at a time.

The name you specify for the user is the default value for the initial password. The user is prompted to change this password when they log in to the Configuration Console for the first time.

If you add a user who already exists, you get an error message.

-d *username*
: Deletes one user at a time.

If you try to delete a user that does not exist, you get an error message. You can display a list of users using the list option to make sure that the user was successfully deleted.

-l
: Displays a list of all users and their status.

-r *username*
: Resets the password for the user you specify to the user ID. For example, judy/sunflower would be reset to judy/judy.

Two files work with the password manager command:

- console-passwords.properties - This file records all user names and the message digest of passwords.
- console-principals.properties - This file is reserved for future use in creating different levels of users. Currently, all Configuration Console users are considered super-users and have access to all areas of the Configuration Console.

These files are located in the srd-home directory. However, do not change them manually. They are used by the product to track user logins, and they are required parameters in some other commands.

Sample password manager commands

To add a new user who's login name and default password are both "judy," type the following command `pwdmgr -a judy`

To delete the existing user named judy and the corresponding password, type the following command `pwdmgr -d judy`

To see a list of current users and their status type the following command `pwdmgr -l`

For example, if you type the command `pwdmgr -l` , the following sample output might be displayed :

```
admin (super-user)
judy (super-user)
allen (super-user)
jose (super-user) *** NEVER LOGGED IN ***
```

If you recently reset a password, a message is displayed showing that the user has not logged in to the Configuration Console yet with the new password.

To reset the password of a user to the user ID, type the following command `pwdmgr -r` *`username`*

For example if you type the command `pwdmgr -r judy`, the password of the existing user named judy is reset to the default password "judy". If the original login/password was judy/sunflower, it is now reset to judy/judy.

Help topics

Configuration Console login window

Use this window to log into the Configuration Console.

User ID
Enter your Configuration Console user ID.

Password
Enter you Configuration Console password.

Edit Configuration
Select this check box to use edit mode.

Login Click to submit user ID and password to gain access to the Configuration Console.

Clear Click to delete user ID and password entries and deselect Edit configuration check box.

Administering System Configuration settings

Modification of system configuration can be done by following these processes:

Deploying patches and hot fixes

You can deploy patches and hot fixes directly to the application server.

Procedure

1. Obtain a `*.ear` file from the IBM Support team or website.
2. Make a backup of the ear file with the same name that is found in `wlp/usr/servers/iiServer/apps` directory
3. Copy the new ear file into that location.
4. Restart the web server to make sure the new application is completely started.

Modifying Database information

Steps to modify database settings for this Identity Insight installation.

About this task

This task will allow you to modify the database port, username, password, and JDBC libraries.

Note: WebSphere Liberty includes a **`securityUtility`** command-line tool that is available in the `wlp/bin` directory. When you run the **`securityUtility encode`** command you can encode an input from the command line or from an input prompt. Copy this value into the password field as required.

Procedure

1. Open the file `bin/pipeline.ini`.
   a. Edit the `Connection` value with the updated values.
2. Open the file `wlp/usr/servers/iiServer/db.xml`.
   a. If a db2 database was configured at install time, you can modify the attributes in the `<properties.db2.jcc>` element
   b. If an oracle database was configured at install time, you can modify the attributes in the `<properties.oracle>` element
   c. To modify the JDBC path, select the `<library>` element, and modify the `dir` attribute.

Modifying HTTP ports

Steps to modify the ports for this Identity Insight installation.

About this task

Steps to modify the ports for this Identity Insight installation.

Procedure

1. Open the file `wlp/usr/servers/iiServer/server.xml`.
   a. Select the `<httpEndpoint>` element
   b. Modify the httpPort or httpsPort values.
2. Open the file `wlp/usr/servers/iiServer/jvm.options`
   a. If httpPort modified above, update the com.srdnet.web.services.port value
   b. If httpsPort modified above, update the com.srdnet.https.port value

Modifying HTTP Transport

How to modify the HTTP Transport used for communication between the web interfaces and the engine.

About this task

The HTTP Transport is used to allow the web applications to communicate with the pipeline engine. This requires a port to be configured in two locations: the pipeline and the configuration for the web applications.

Procedure

1. Open the file `bin/pipeline.ini`.
   a. Edit the `InputTransport` value with the new port number.
2. Open the file `ibm-home/easws/webservices.properties`.
   a. Edit the `pipelineURL` value with the new port number.

Modifying the host name

How to modify the hostname of the server.

About this task

The hostname is used to allow access to the applications

Procedure

1. Open the file `bin/pipeline.ini`.
   a. Edit the `InputTransport` value with the new host name.
2. Open the file `ibm-home/easws/webservices.properties`.
   a. Edit the `pipelineURL` value with the new host name.

Running reports from the Configuration Console

From the Configuration Console, you can generate reports that show summaries of pipeline statistics by data source or a report that lists the current system configuration settings, including entity resolution configuration. The resulting reports display in the web-based BIRT (Business Intelligence Reporting Tool) Report Viewer. Make sure to turn off any pop-up blockers, as they might interfere with displaying the report in the viewer.

Viewing statistical reports

As it processes data, the product tracks statistical information about performance and data for the incoming data source files that were loaded. This information is summarized for you on two reports: the Data Source Summary report and the Load Summary report.

About this task

The statistics on these reports can help you quickly verify that the product is processing all the incoming data records, make operational decisions about product performance, evaluate the quality of the incoming data, and show the number of new identities, new entities, new relationships, and new alerts resulting from processing the data files.

Procedure

1. In the Configuration Console, select **Status** > **Reports**.
2. Required: From the **Report** list, choose a statistical report:
   - **Data Source Summary Report** - This report provides a quick statistical summary by data source of the records loaded and processed. Use it to see the total number of records loaded by data source file, the total number of new identity records processed by data source file, and the total number of new entities based on the data in this data source file. The Data Source Summary report is sorted by load date, load ID, data source, and data source file.
   - **Load Summary Report** - This report summarizes statistics and quality characteristics for one or more data sources. Use the report to see load performance information, quality of the data source file, and summaries of the data values used to resolve entities, detect relationships, and generate alerts. This report can help you determine the quality of the data being loaded from a particular data source. Lower quality data can indicate that the data in this data source requires additional cleansing, either before being loaded into the product or during entity resolution by applying specific DQM (data quality management) rules to the data. The Load Summary report is sorted by Load ID.
3. In the **From Date** field, enter the starting date for the report using `mm/dd/yyyy` format. By default, this field contains the current date.

This field can be left blank, which means that the product reports all data within all other specified criteria beginning with the date the product became operational.

4. In the **Thru Date** field, enter the ending date for the report using `mm/dd/yyyy` format. By default, this field contains the current date.

   This field can be left blank, which means that the product reports all data within all other specified criteria through the current date.
5. Optional: In **Data Source Code**, enter a specific data source code to report on. The data source code you enter must exactly match a configured data source code.

   This field can be left blank, which means that the product reports statistics for all data sources within all other specified criteria.
6. Required: Click **Run Report** to generate the selected report.

Results

The product generates the selected statistical report based on all specified criteria and displays the report in a separate web browser window, entitled **BIRT Report Viewer**. If there is no data to report, based on the criteria that you selected, the **BIRT Report Viewer** window displays the name of the report, the date and time the report was generated, and **Page 1/1** at the top. The data section is blank.

What to do next

Use the statistical information on this report to help tune the product or data files.

Data Source Summary Report

The Data Source Summary report provides a quick statistical summary by data source of the records loaded into the system for processing. From this report, you can see the total number of records processed by load ID. Of those total records loaded, the report shows the number of records represented new identities or new entities and calculates the percentage of records that were new identities, as well as the percentage of records that were newly created entities.

Statistics by load within data source

Date Loaded
Displays the date that this data source file was loaded

Load ID
Displays the system-assigned load ID number.

Data Source
Displays the data source code and description (separated by a dash) for the data source file that was loaded.

UMF Records Loaded
Indicates the total number of identity records in this data source file that were loaded.

New Identities
Indicates the total number of new identities discovered in the data file that was loaded. (This number indicates an identity that has not been processed by the system before.)

New Identity %
Indicates the percentage of total records loaded (New Identities divided by UMF Records Loaded) that represent new identities.

New Entities

Indicates the total number of new entities created from this data load.

New Entities %

Indicates the percentage of total records loaded (New Entities divided by Loaded) that represent new entities.

Statistical charts by data source

Records Loaded by Data Source

Displays a bar chart that graphically shows how many records from each data source were loaded into the system, based on the other specified report criteria. You can see which data sources provided the most records or the least records and compare that to your estimated load numbers.

- The vertical axis shows the data sources by data source code.
- The horizontal axis shows the number of records loaded.

If there are fewer records loaded for a particular data source than expected, you can inspect the data files for this data source. (You might also consider running a Load Summary Report to see the data quality of the files loaded for this data source; data quality directly impacts the number of records loaded.)

New Entities by Data Source

Displays a bar chart that graphically shows which data sources yielded the most number of new entities, based on the other specified report criteria.

- The vertical axis shows the data sources by data source code.
- The horizontal axis shows the number of new entities created.

Load Summary Report

The Load Summary report summarizes statistics and quality characteristics by data source. It contains information about the data source files. Use this report to determine performance load statistics, the number of entities and alerts created by load, general information about the data quality of the data loaded, a summary of the actions about the UMF records by load, and any UMF exceptions that were generated by load. The report is grouped by load ID.

For each load, the report breaks the statistics into sections:

- Load Summary
- Role Alert Summary
- Relationship Summary
- Quality Summary
- UMF Document Summary
- Exception Summary

Load Summary

Use this section to help determine how long it took to process a particular file, as well as to give you a general idea of how useful this data source file is in overall entity resolution and relationship detection.

Date and Time Started

Indicates the date and time that the data load began.

Date and Time Completed

Indicates the date and time that the data source file load ended.

UMF Record Count
Indicates the total number of records loaded from this data source file within the **Date and Time Started** and **Date and Time Completed** range.

The **Date and Time Completed** number minus the **Date and Time Started** number is the number of minutes it took to load this particular data source file, which can give you an idea of system performance. It can also indicate that a larger data source file needs to be split into smaller files for quicker processing.

New Identities
Indicates the total number of new identities loaded within the **Date and Time Started** and **Date and Time Completed** time frame.

New Identity %
Indicates the percentage of total identities in this data load that are new identities (identities that are new to the entity database).

New Entities
Indicates the total number of newly created entities in the **Date and Time Started** and **Date and Time Completed** time frame.

New Entity %
Indicates the percentage of total entities that are newly created entities as a result of this data source load.

The number of new identities and new entities can provide you a general idea of how valuable this data source is in overall entity resolution and relationship detection. If these numbers are low and remain low over time, it might be that this data source is not useful in meeting your company entity resolution goals.

Role Alert Summary

Use this section to see the resolution rules and resolution scores common to the relationships detected that resulted in role alerts. Each row represents the number of role alerts that were generated, based on the criteria listed.

Resolution Rule
Displays the name of the resolution rule used to evaluate the identity and entity during entity resolution and relationship detection.

Alert Description
Displays the name of the role alert rule that triggered the role alert.

Severity
Displays a user-defined indicator to measure the priority or importance of this role alert.

Resolution Score
Displays a resolution score (0-100) for the resolution rule given to the identity and entity involved in the role alert. This score indicates the degree of likeness between the identity and the entity. A score of 100 means the identity record resolved to the entity.

Alert Count
Indicates the total number of role alerts generated based on the role alert rule description, resolution rule, and resolution score.

Relationship Summary

Use this section to see the attributes common to detected relationships that did not generate a role alert. Each row represents the number of relationships that were detected, based on the criteria listed.

Resolution Rule
Displays the name of the resolution rule used to evaluate the incoming identity records and existing entities during entity resolution and relationship detection.

Resolution Score
Displays a resolution score (0-100) for the resolution rule given to the identity and entity during entity resolution. This score indicates the degree of likeness between the identity and the entity. A score of 100 means the identity record resolved to the entity.

Relationship Score
Displays a relationship score (0-100) for the resolution rule given to the identity and entity during relationship resolution. This score indicates the degree of relationship between the identity and the entity.

The higher the relationship score, the more closely related the identity and entity are, based on matching attributes.

Relationship Count
Indicates the total number of relationships that are detected based on the resolution rule, resolution score, and relationship score.

Quality Summary

Use the information in this section to evaluate the quality of the data in each data source file. The section indicates the quality by attribute type within a UMF segment and UMF document type. By reviewing the Quality summary with the UMF exceptions summary, you can see which data source files have quality issues or malformed UMF that need to be addressed. Typically, you can resolve these issues through ETL or DQM/data source configuration before processing the data source file.

In some cases, this section can indicate that a data source is of such poor quality that you might not want to use this data source for entity resolution.

Document Type
Displays the name of the UMF document type that contains the data type listed in Data Type. Typically, this value is UMF_ENTITY.

Table Name
Displays the name of the database table that stores data from similarly named UMF segments. For example, data from the NUMBER segment is stored in the NUMS table.

Data Type
Indicate the data type, as listed in the incoming records attribute type UMF tags. This type corresponds to a UMF segment listed in Table Name. For example, if the Table Name is *ADDRESS* and the Data Type listed is *H*, the quality information is evaluating the address type of *Home*.

If you do not recognize a data type, that can indicate that the data source file is not correctly mapped to the appropriate combination of UMF documents, segments, and tags. Check the Exception Summary section to see if a matching UMF segment and UMF tag caused one or more segment exceptions. If the problem is invalid UMF, the numbers in the Low Quality Count in the Quality Summary section and the Segment Exception Count in the UMF Exception section often match.

Record Count
Indicates the total number of incoming identity records for the given Document Type, Table Name, and Data Type.

Generic Count
Indicates the total number of incoming identity records with the given Document Type, Table Name, and Data Type that contain values which are considered generic.

Low Quality Count
Indicates the total number of incoming identity records with the given Document Type, Table Name, and Data Type that are considered of poor quality. This number can indicate a data entry or ETL transformation problem in the data source file.

Usable Percent
Indicates the percentage of the incoming identity records with the given Document Type, Table Name (of this UMF segment) and Data Type that are usable for entity resolution and relationship detection. (Record Count minus Generic Count minus Low Quality Count) divided by Record Count equals Usable Percent.

Identity Percent
Indicates the percentage of the incoming identity records that contained the given Document Type, Table Name, and Data Type.

Attribute Summary

Use this section to see the attributes in the data source file that helped to detect relationships and generate role alerts. Each attribute maps to a specific UMF segment, and this section shows the number of relationships detected and role alerts generated, based on the data in the incoming UMF segment.

Segment Name
Displays the name of the UMF segment, which directly maps to an attribute.

Data Type
Lists the attribute type (or data type) within the UMF segment corresponding to the Precision Description. The report might list a specific attribute type or list *ALL*, indicating all attribute types in the UMF segment.

Precision Description
Describes the matching threshold between an attribute from an inbound identity and an attribute from an existing entity.

Role Alerts
Indicates the total number of role alerts generated based on this UMF segment, data type, and precision description.

Relationships
Indicates the total number of relationships detected based on this UMF segment, data type, and precision description

UMF Document Summary

You can use this section to validate the total number of incoming records in a data source file, based on what action is to be taken to the record. You can reconcile these numbers to the Record Count in the Load Summary section.

Document Type
Displays the name of the UMF document type. Typically, this value is UMF_ENTITY.

Action
Indicates the type of action for the incoming identity record. Here is a list of the most commonly used actions:

- *A* for add
- *C* for change
- *D* for delete

As part of the ETL process, identity records are typically tagged through UMF to indicate how to act on each incoming record during system processing.

UMF Record Count
Indicates the total number of records processed for each action type within document type.

Percent
Indicates the percentage of the total records loaded that the Record Count represents. (The sum should not exceed 100%.)

Exception Summary

Use this information to help pinpoint bad identity records, such as those with malformed UMF. The exception describes the problem, while the table name and element show which segment and record are bad. The count shows how many of the records in the file contained this bad UMF.

Document Type
Displays the name of the UMF document type. Typically, this value is UMF_ENTITY.

Action
Indicates the type of action for the incoming identity record:

- *A* for add
- *C* for change
- *D* for delete

As part of the ETL process, identity records are typically tagged through UMF to indicate how to act on each incoming record during system processing.

Segment
Displays the name of the UMF segment where the exception occurred.

UMF Tag
Displays the value of the UMF tag that caused the UMF exception.

Exception
Displays the message ID or other exception code to indicate the type of UMF exception that occurred and give information about how to resolve the exception. This information is also available in the UMF_EXCEPT table.

Segment Exception Count

Indicates the total number of this type of UMF exception.

Check the Low Quality Count in the Quality Summary section to see if a matching data type is reported as being of poor or unusable quality. If the problem is incorrect UMF, the numbers in the Low Quality Count in the Quality Summary section and the Segment Exception Count in the UMF Exception section often match for the same UMF segment and UMF tags.

Running the Configuration report

The Configuration report provides a unified overview of all the system settings that you can configure using the Configuration Console. View this report to see the current system configuration settings before changing the current product configuration, when troubleshooting a configuration issue, or when comparing different configuration settings.

Procedure

1. In the Configuration Console, click **Setup** > **Reports**.
2. From **Report**, select **Configuration Report**.
3. Click **Run Report**.

Results

The product generates the selected statistical report based on all specified criteria and displays the report in a separate web browser window, entitled **BIRT Report Viewer**. If there is no data to report, based on the criteria that you selected, the **BIRT Report Viewer** window displays the name of the report, the date and time the report was generated, and **Page 1/1** at the top. The data section is blank.

Configuration report

The Configuration report provides a unified view of the system settings that you configure using the Configuration Console. Use this report to view or print the current system configuration before you change system configuration, when you are troubleshooting a configuration issue, or when you need to compare different configuration settings.

The report lists the current configuration settings by category:

Data Sources

View configuration settings for data sources, including the data source ID, the data source code, the role code associated with the data source, the entity resolution configuration associated with the data source, and the current status of the data source code (active or inactive).

To configure data sources, select **Setup** > **Sources** > **Data Sources**.

Number Types

View configuration settings for number types, including the number type ID, the number type, the minimum and maximum length for the number type, any associated masks for the number type, information about how the number type is used in entity resolution, and whether the number type is active or inactive.

To configure number types, select**Setup** > **Sources** > **Numbers**.

Characteristic Types

View configuration settings for characteristic types, including the characteristic type ID, the name of the characteristic type, the associated data type for the characteristic (such as character or date), information about how the characteristic type is used in entity resolution, and whether the characteristic type is active or inactive.

To configure characteristic types, select **Setup** > **Sources** > **Characteristics**.

Plugin

View configuration settings for attribute and scoring customization, including plugin ID, name, type, version, and Library Short Name.

To configure plugins for attribute and scoring customization, select **Setup** > **General** > **Plugins**.

Event Types

View configuration settings for event types, including the unit of measure associated with the value for this event. Event types are part of Event Manager.

To configure event types, select **Setup** > **Sources** > **Event Types**.

Data Quality Management Rules

View the list of data quality management rules (DQM rules) and their associated parameters that are configured for a specific UMF tag within a UMF segment, including which UMF segment and UMF tag name the DQM rule is associated, the order in which the DQM rule is used on that UMF segment and tag, associated parameters for the DQM rule on that UMF segment and tag, whether the DQM rule corrects incoming data for that UMF segment and tag, and whether the DQM rule is currently enabled on that UMF segment and tag.

To configure a UMF segment and UMF tag to use DQM rules, select **Setup** > **UMF** > **DQM Rules**.

Load Mapping

View the configuration information for how UMF data is mapped to the corresponding tables and table columns in the entity database, including UMF segment name, UMF data path, the name of the entity database table, the field name and type within that entity database table, the data type for that field, and whether the mapping is enabled.

To map data from a UMF segment to a table in the entity database, select **Setup** > **UMF** > **Data Map**.

Entity Resolution Rules

View the configuration settings for each entity resolution rule, including the entity resolution rule ID, the order for the rule, the minimum resolution and relationship scores for the rule, and whether the rule incorporates denials.

To configure entity resolution rules, select **Setup** > **Resolution** > **Resolution Rules**.

Entity Resolution Confirm/Deny

View the settings for the scores that contribute to the confirmation and denial process of entity resolution, including the associated entity resolution ID and configuration ID, the priority of each score, the description and attribute name for each score, and the numeric value of the score.

To configure settings for entity resolution confirmations and denials, select **Setup** > **Resolution** > **Confirms & Denials**.

Entity Resolution Characteristics

View the settings for characteristic types that are configured with confirmation and denial weights used during entity resolution, including the priority, the confirmation weight, and the denial weight.

To configure confirmation and denial weights for characteristic types, select **Setup** > **Resolution** > **Characteristics**.

Role Codes

View the list of configured role codes and their associated settings, including role code ID and description, role code class, and the current status of the role code (active or inactive).

To configure role codes, select **Setup** > **Relationships** > **Roles**.

Role Alert Rules

View the list of configured role alert rules and their associated settings, including role alert rule ID and description, severity, minimum alert threshold, and the role code IDs of the two roles that trigger this role alert rule.

To configure role alert rules, select **Setup** > **Relationships** > **Role Alert Rules**.

Name Manager Configuration

View the configured settings for the Name Manager feature that extends name precision during entity resolution.

To configure settings for Name Manager, select **Setup** > **Resolution** > **Name Manager Match Config**.

Separation Configuration

View the configured settings for the Degrees of Separation feature of the pipeline that can detect relationships one, two, or more degrees of separation.

To configure settings for degrees of separation, select **Setup** > **Relationships** > **Separation Config**.

System Sequences

View configuration settings for sequence numbers that indicate how the system loads and processes data. System sequence numbers assist with system load performance in two ways. First, because they allow each pipeline to issue one query that grabs a sequential set of numbers, and then keep those numbers in cache until the numbers are used up. Secondly, because sequence numbers prevent multiple pipelines that are generating system-generated IDs from using the same ID number for more than one record.

For example, each time the pipeline creates a new entity during entity resolution processing, the system generates a unique Entity ID. Using system sequences, the pipeline can send one query to ask for the next available 1000 Entity ID numbers. So for the next 1000 newly created entities, the pipeline can use the available Entity ID numbers stored in its memory. The alternative (slower) method is for each pipeline to send one query to the entity database asking for a new Entity ID for each new entity created.

To configure system sequences, **Setup** > **UMF** > **Load Sequence**

Generic Thresholds

View the generic thresholds settings configured by attribute, including attribute name, attribute type, and the threshold number that determines when a specific value for that attribute becomes generic.

To configure generic thresholds by attribute type, select **Setup** > **UMF** > **Generic Threshold**.

Table Dictionary
View the dictionary settings by entity database table, including table name, description, and the table type.

To configure the table dictionary, select **Setup** > **UMF** > **Dictionary**.

Look Up Tables
View the settings for the list of tables the system uses as lookup tables during processing, including the table name, the key field name, the ID field name, and whether to load the table in memory during processing.

To configure which tables the system uses as look up tables, select **Setup** > **UMF** > **Lookup**.

Matching Configuration
View the settings for each resolution configuration configured in your system, including the configuration name and ID, match type, and UMF segment name.

To configure match configurations, select **Setup** > **Resolution** > **Candidate Builder**.

Document Types
View the settings for UMF input documents, including document type, whether to perform data quality management on this document type, whether to load the data processed by this document type into the entity resolution database, and the level of entity resolution to perform on this input UMF document type.

To configure UMF input documents, select **Setup** > **UMF** > **Input Documents**.

UMF Output Format
View the format settings for UMF output documents, including the format ID and code, the route direction, and whether the output format setting is enabled.

To configure formats for UMF output documents, select **Setup** > **UMF** > **Output Documents**.

GEM Event Types
View format settings for Event Manager events, including event ID, type, description, category, unit of measure, and date and time created.

To configure event types, select **Setup** > **Sources** > **Event Type**.

System Parameters
View the list of system parameter settings by parameter group, including the value and default value for the system parameter, and the validation type and value for the parameter

To configure system parameters, select **Setup** > **General** > **System Parameters**.

Application Activity Codes
View the list of activity codes configured for the Visualizer by activity type

(role alert, attribute alert, or event alert), including the activity code, valid statuses for the activity code, and whether the activity code is active or inactive.

To configure activity codes used in the Visualizer, select **Setup** > **Visualizer** > **Activity Codes**.

User Groups
View the settings for user groups configured for the Visualizer, including the associated Visualizer user names, the create date and time for the user group, and whether the user group is active or inactive.

To configure activity codes used in the Visualizer, select **Setup** > **Visualizer** > **Codes**, and then select **ANALYZER_GROUP**.

Role Alert Groups
View the settings for configured role alert groups, including assigned application group, the associated role alert rule ID and description, the created date and time for the role alert group, and whether the role alert group is active or inactive.

To configure role alert groups used in the Visualizer, select **Setup** > **Relationships** > **Role Alert Rules**, and then edit the **Alert Group** field.

Users View the settings for the users configured to log into the Visualizer, including user login names, whether to authenticate the Visualizer user using entity database credentials, and whether the Visualizer user is active or inactive.

To configure the users, select **Setup** > **Visualizer** > **Visualizer Users**.

Exporting reports

The BIRT Report Viewer gives you the option of exporting Configuration Console report data into other applications, such as Microsoft Excel, Microsoft PowerPoint, Microsoft Word, or Adobe Acrobat. You can export the entire report or specific data from a report.

Exporting Configuration Console reports

If you want to export an entire report (both data and format) into another application, such as Microsoft PowerPoint, or into another format, such as Adobe Acrobat PDF, use the **Export reports** option in the BIRT Report Viewer. Exporting full reports works well for reports that span multiple pages, and in cases where you do not want to manipulate the data after exporting the report.

About this task

Opening a Configuration Console report exported into a Microsoft Word *.doc file requires using Microsoft Word version 2003 or later.

If you want to make small changes or additions to the exported report, export the report to either Microsoft Word or Microsoft Excel. These applications retain the report formatting, but the data is typically displayed in columns or tables, allowing some data manipulation. Because the exported report is a Read-Only file, save the file using a new name to save your changes.

Procedure

1. After generating the report, from the **BIRT Report Viewer** window, click **Export report**. The Export report icon is the fourth icon from the left on the BIRT Report Viewer icon toolbar.

2. In **Export Report**, select the format or application to export the data:
   - **PDF**
   - **PowerPoint**
   - **Word**
   - **PostScript**
   - **Excel**
3. Select the pages or page range to export.
4. Optional: Select the size of the resulting report: This option is only available if you selected either the PDF, PowerPoint, or PostScript option.
   - **Auto**: Each page of the report becomes a separate page.
   - **Actual size**: All the pages of the report are fit into one, long page.
   - **Fit to whole page**: All the pages of the report pages are decreased to fit on roughly one-third of a single page. If you selected the PowerPoint option, the report is inserted as an image on the page, allowing you to resize the image.
5. Click **OK**.

Results

If you exported the report into PDF or PostScript format, the resulting file is typically placed in the folder location where files download on the client. For example, C:\Documents and Settings\Administrator\My Documents\Downloads.

If you exported into PowerPoint, Word, or Excel, the data is exported into a Read-Only file typically named *reportname.selected_application_extension*.

- *reportname* is the name of the Configuration Console report that you exported.
- *selected_application_extension* is the appropriate file format extension for the selected application.

For example, if you exported the Load Summary report into Word, the file name is typically named LoadSummary.doc. A dialog displays giving you the option to either open the file in the selected application or save the file.

Exporting data from Configuration Console reports

If you want to export report data into a CSV file (comma separated values file) to view and manipulate the data in another application such as Microsoft Excel, use the **Export data** option in the BIRT Report Viewer. You can select a section of the report, which fields to export, and the export data format.

About this task

The BIRT Report Viewer exports one section of data from a report at a time, which means that the viewer creates a separate results set for each section on the report. The data that is exported is raw data without any formatting.

If you want to export the entire report, use the **Export report** option instead. However, that export option exports both the data and the report formatting, which prevents you from manipulating the data after the export.

Procedure

1. After generating the report, from the BIRT Report Viewer, click the**Export data** icon. The **Export data** icon is the third icon from the left on the BIRT Report Viewer icon toolbar.

2. Required: In **Available results sets** in **Export Data**, select the one report section that you want to export. The names of the report sections display by element, such as ELEMENT_2041. You can typically tell which section you are selecting by looking at the column names listed in **Available Columns**.
3. Required: In **Available Columns**, select the columns to export. The column names that apply to the report section that you chose in **Available results sets**display in **Selected Columns**. You might not want to see the data in all the columns available for that report section.
4. Optional: Set the order of the columns in **Selected Columns**. This option allows you to reorder the data by column before exporting the data.
5. Optional: In **Separator**, select a separator, if you want to use a type of separator other than **Comma**, which is the default choice:
   - **Semi-colon**
   - **Colon**
   - **Vertical line**
   - **Tab**
6. Click **OK**. From the dialog that displays, select whether to open the exported data or save the file. Microsoft Excel is the default application to open the file, but you can browse to select any application that can export a CSV file.

Results

The data is exported into a file typically named *reportname*.csv, where *reportname* is the name of the Configuration Console report from which you exported the data.

Administering the Visualizer

To use the Visualizer effectively you must configure the browsers, set up accounts for the appropriate users, and manage access to the Visualizer.

Visualizer

The Visualizer is a graphical user interface that analysts and investigators use to analyze the results of alerts, relationships, and entity resolutions.

The Visualizer is hosted by an embedded version of IBM WebSphere Application Server. You configure the Visualizer through the Configuration Console and through the Visualizer **Preferences** selection on the **File** menu.

Visualizer users can accomplish various analysis tasks:

Analyze and disposition alerts

Alerts generated by entity resolution processing represent relationships or entity resolutions of interest to an organization. Typically, analysts review alerts and decide what action to take, if any, based on the alert information. There are three types of alerts: role alerts, attribute alerts, and event alerts.

The Visualizer displays the alerts, providing analysts with both textual and graphical views of the alerts and the entities involved in the alerts. Analysts can drill down into the details and then set the disposition status of the alert appropriately.

Create and manage attribute alert generators

Using the Visualizer, analysts can create and manage persistent searches through the Attribute Alert Generator feature, and manage how they view and receive attribute alerts. Analysts can create Attribute Alert Generators based on attribute data to locate identities that resolved to entities based on that attribute data. Or analysts can create an Attribute Alert Generator to persistently search the entity database looking for a particular entity.

Find entities

Visualizer users can also find entities for further analysis using several methods:

- By attributes
- By data source account
- By entity ID
- By resolution (how closely the criteria entered matches identities and entities in the entity database, based on minimum resolution score thresholds)

Add entities and disclosed relationships

Analysts can use the Visualizer to add records for entity resolution and relationship detection. They can add a single identity record or load a UMF file containing a few thousand identity records. Just as when identity records are added through acquisition programs, records added through the Visualizer are processed by a pipeline for entity resolution and relationship detection. The results of processing are written to the entity database, and any alerts are published to the Visualizer.

Analysts can also disclose relationships between entities (by identity), when they know of a link between the identities. Examples of disclosed relationships include relating entities based on emergency contacts or references listed on an employment application. The entity disclosed these relationships on the application.

Generate and print reports

The Visualizer also contains several reports that analysts can view and print to help them manage and track their Visualizer work.

User roles and responsibilities

User roles help categorize the typical tasks that must be completed to effectively deploy and use IBM InfoSphere Identity Insight. Many different types of users might use IBM InfoSphere Identity Insight for various purposes; that is, users take on the responsibilities of one or more roles in using the product.

You can define groups of users based on the various user roles and responsibilities.

The most common user roles include these roles:

Analyst

Analyzes the data and reviews entities, relationships, and alerts. The analyst defines what results are most valuable and makes sure that the system returns those results. The analyst works closely with the operator and application administrator.

Operator

Loads data into the system, runs the pipelines, and verifies that the system is running acceptably, providing load-quality reports as necessary. The operator also reviews the results, exceptions, and events. The operator works closely with the analyst, data source administrator, and application administrator.

Data source administrator
Prepares the data for loading it into the system, which includes converting the data to a UMF file and validating the file. The data source administrator works closely with the operators, application administrators, and database administrators.

Application administrator
Configures the application, including the configuration of the data, entity model, and rules. The application administrator works closely with the data source administrators and operators to define the entity model, and coodinates configuration changes with the database administrator, data source administrator, and operators. The application administrator also coordinates and consults with overall system administrators, if they exist.

Database administrator
Ensures that the database is configured and tuned appropriately for use with the application. The database administrator works closely with the operator, data source administrator, and application administrator.

System architect
Sizes and estimates the hardware and software requirements in planning for the deployment of the application. The system architect works closely with the installer, database administrator, data source administrator, and application administrator to ensure the deployment achieves the vision, strategies, and objectives and integrates into your business processes while delivering expected results.

Installer
Manages the installation and initial configuration of the application. The Installer sets up initial users in the system. Frequently, IBM Professional Services works with the system architect to complete these responsibilities.

Programmer
Designs and develops graphical interfaces or customizes graphical interfaces for the various functions, such that the deployment of the application integrates seamlessly into your environment. The programmer works closely with the system architect and the application administrator, often to disseminate alerts to the appropriate people in the most effective manner for your environment.

Security architect
Ensures that the project team adheres to and implements a secure system. The security architect works closely with the system architect, installer, and database administrator.

Optimum browser settings for using the Visualizer

The Visualizer is a Web-accessed, Java-based application that performs best with specific settings for the browser that you use to access it.

To best view the Visualizer, use the following browser settings:

Table 18. Optimum browser settings for the Visualizer

Setting	Value	Description
Text size	Medium	
JavaScript	On	

Table 18. Optimum browser settings for the Visualizer (continued)

Setting	Value	Description
Cookies	On	At a minimum, first-party session cookies must be enabled.
Security - Trusted Web site	HTTP address of the Visualizer	Make sure that the HTTP address of the Visualizer is included on the list of trusted Internet Web sites.
Security - Download options	Enabled	Make sure that all download options for trusted Internet Web sites are enabled.
Pop-up blockers	Allow pop-ups from the HTTP address of the Visualizer	Make sure that the HTTP address of the Visualizer is on the list of Web sites that allow pop-ups.

Logging in to the Visualizer

Before you log in to the Visualizer, you must have a Visualizer user account (user name and password). Your system administrator can provide you with your Visualizer user account information.

Procedure

1. Complete one of the following steps:
   - Double-click the Visualizer icon on your desktop.
   - Or open your Internet browser and enter the uniform resource locator (URL) for the Visualizer in the address line.

   The URL for launching the Visualizer is:

   ```
   http://server:install_port
   ```

   For example, `http://localhost:13510`. When the Visualizer is installed, the default *install_port* is 13510, but the port number can be changed. Check with your system administrator if you are unsure of the correct server name or port number.
2. Log in by entering your user name and password.

   Note: Both user name and password fields are case sensitive. The first time that you log in, use the password assigned to you by your system administrator. After your first successful login, typically you change your Visualizer password to safeguard the security of your Visualizer account.
3. Click **Login**.

Closing the Visualizer

When you are finished using the Visualizer, close the application. By closing the Visualizer, you also log out. If you are taking a break and just want to secure your workstation for a few minutes, you can lock the Visualizer instead.

Procedure

To close the Visualizer and log out:

- Select **File** > **Exit**.
- Or press **Ctrl** + **Q**.

Managing access to the Visualizer

Visualizer users must have a registered account before they can log in to the Visualizer. These user accounts are not the same as the user accounts for the Configuration Console but are specifically authorized to use the Visualizer.

Creating new Visualizer users

In order to access and use the Visualizer, a system administrator must create a Visualizer user account for the user in the Configuration Console.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Visualizer** button.
3. Click the **Visualizer Users** tab.
4. Click the **New** button.
5. From the **Database Login** drop-down list, select one of the following values:
   - Select **Yes**, if the user has a user account that grants access to the entity database, and you want to use that database login information.
   - Select **No**, if you are using the default file login information. This choice means that a system administrator chooses the first password that the user uses to log into the Visualizer, and that a system administrator can reset Visualizer users passwords on-demand.
6. In the **User Name** field, type the user name that you want to add. If you selected **Yes** in the **Database Login** drop-down, this user name must match the entity database user name for this user.
7. In the **Password** field:
   a. If you selected **Yes** in the **Database Login** drop-down list, this value must match the password that is stored in the database login information.
   b. If you selected **No** in the **Database Login** drop-down list, type the initial password for the user.

      Note: For security reasons, encourage your Visualizer users to change their initial password after they successfully login for the first time.
8. Optional: In the **Group** field, from the drop-down list, select the Analyzer Group that this person belongs to.
9. Click the **Save** button.

What to do next

The user can now immediately use this user name and password to log in to the Visualizer.

Deactivating Visualizer users

You can deactivate Visualizer user accounts for users that no longer need access to the Visualizer.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Visualizer** button.
3. Click **Visualizer Users** tab.
4. Click the user name whose user account you want to deactivate.
5. From the **Status** drop-down list, choose **Inactive**.
6. Click the **Save** button.

Results

The user that you deactivated can no longer log into the Visualizer.

Resetting Visualizer passwords

If Visualizer users forget their password, and their login information is configured through the Configuration Console, not through the underlying database login option, you can reset their password in the Configuration Console. Otherwise, you must reset their password using the underlying database login configuration.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Visualizer** button.
3. Click the **Visualizer Users** tab.
4. Click the user name of the user whose password you want to edit.
5. In the **Password** field, type a new password for the user.

   Note: For security reasons, encourage users to change their password after successfully logging in, so that only they know it.
6. Click the **Save** button.

What to do next

The user can immediately use this new password to log into the Visualizer. For security reasons, after you reset passwords, encourage users to change their password after successfully logging in.

Creating groups of Visualizer users

Alerts are assigned to groups of analysts in the Visualizer. If you are adding a new group of analysts to a project, you can use the Configuration Console to create a new analyst group.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **General** button.
3. Click the **Codes** tab.
4. From the **Type** drop-down list, click**ANALYZER_GROUP**.
5. Click the **New** button.
6. In the **Code** field, type the name of the analyst group.
7. From the **Status** drop-down list, select **Active**.
8. Click the **Save** button.

Help Topics

Visualizer Users General tab:

Use this tab to add new Visualizer users or to change existing user passwords.

Database Login
Select an option to determine whether to use the underlying entity database login information (user name and password) for Visualizer access.

- Yes - Only use this setting if this Visualizer user already has a user account that grants the user access to the entity database. If you select this option, use the entity database login user name and password for the Visualizer user name and password. (If the two do not match, the Visualizer user will not be able to login.)
- No - Use the login information entered on this tab.

User Name
Type the user name for this Visualizer user. If this user uses a database login, this user name must match the corresponding entity database user name.

Password
Type the new password for this Visualizer user. If this user uses a database login, this password must exactly match the corresponding database password.

Group Select the Visualizer group that this user belongs to. The Visualizer group determines which alerts and notifications that the user sees in the Visualizer **Alert Summary** window. (For example, if your organization has a Security Visualizer group and a Reservation group, users in each group might see different types of alerts in the Visualizer.).

Status Select a status to indicate whether this Visualizer user is currently active (able to log into Visualizer) or not.

Configuring activity codes for the Visualizer

The Visualizer provides several default activity codes for dealing with alerts. You can add new activity codes and delete existing activity codes through the Configuration Console.

Creating activity codes for searches

The Visualizer provides activity codes for search result alerts. If you need to track additional activities related to alert handling, you can add new activity codes using the Configuration Console.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Visualizer** button.
3. Click the **Activity Codes** tab.
4. From the **Activity Type** drop-down list, click **SEARCH**.
5. Click the **New** button.
6. In the **Activity Code** field, type the name of the activity code.
7. From the **Activity Status Code** drop-down list, select the internally-recognized activity status code that the new activity code corresponds to.
8. From the **Status** drop-down list, select **Active**.

9. Click the **Save** button.

Deleting activity codes for searches

The Visualizer provides activity codes for search result alerts. If you need to delete activity codes related to alert handling, you can delete existing activity codes using the Configuration Console.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Visualizer** button.
3. Click the **Activity Codes** tab.
4. From the **Activity Type** drop-down list, click **SEARCH**.
5. Select the check box next to the activity code you want to delete.
6. Click the **Delete** button. A confirmation window appears.
7. Click **OK**.

Creating activity codes for role alerts

The Visualizer provides activity codes for role alerts. If you need to track additional activities related to alert handling, you can add new activity codes using the Configuration Console.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Visualizer** button.
3. Click the **Activity Codes** tab.
4. From the **Activity Type** drop-down list, click **CONFLICT**.
5. Click the **New** button.
6. In the **Activity Code** field, type the name of the activity code.
7. From the **Activity Status Code** drop-down list, select the internally-recognized activity status code that the new activity code corresponds to.
8. From the **Status** drop-down list, select **Active**.
9. Click the **Save** button.

Deleting activity codes for role alerts

The Visualizer provides activity codes for role alerts. If you need to delete activity codes related to alert handling, you can delete existing activity codes using the Configuration Console.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Visualizer** button.
3. Click the **Activity Codes** tab.
4. From the **Activity Type** drop-down list, click **CONFLICT**.
5. Select the check box next to the activity code you want to delete.
6. Click the **Delete** button. A confirmation window appears.
7. Click **OK**.

Creating activity codes for event alerts

The Visualizer provides activity codes for event alerts generated through event processing, if your system has Event Manager enabled. Event activity codes allow you to track additional activities related to event alert handling. The system provides three pre-defined event activity codes, but you can add new activity codes for event alerts using the Configuration Console.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Visualizer** button.
3. Click the **Activity Codes** tab.
4. From the **Activity Type** drop-down list, select **EVENT**.
5. Click the **New** button.
6. In the **Activity Code** field, type a unique name for the new activity code.
7. From the **Activity Status Code** drop-down list, select the internally-recognized activity status code that the new activity code corresponds to.
8. From the **Status** drop-down list, select **Active** to make this activity code available for use in the Visualizer.
9. Click the **Save** button.

Pre-defined activity codes for event alerts:

Event activity codes are used by analysts in the Visualizer to disposition event alerts. The system includes three pre-defined activity codes for events, after you run the post-installation v4.2 fix pack 1 SQL scripts.

The following activity codes for event alerts are included in the pre-defined set of event alert activyt codes:

ASSIGNED
When analysts assign an event alert to themselves or to another analyst group, the system defaults to the ASSIGNED activity code.

CLOSED
When analysts close an event alert, the system defaults to the CLOSED activity code

PENDING
Before an analyst dispositions event alerts, the system automatically assigns them the PENDING activity, which means the event alert is open for any analyst in the assigned group to review or disposition.

Editing activity codes for event alerts

You can edit existing activity codes used to disposition event alerts in the Visualizer. You cannot rename an existing activity code, but you can change its associated description, activity status code, or status.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Visualizer** button.
3. Click the **Activity Codes** tab.
4. From the **Activity Type** drop-down list, select **EVENT**.
5. Click on the activity code that you want to edit.
6. From the **Activity Codes General** tab, make your changes. For example, if you want to configure an activity code but not display it for selection in the Visualizer, select the **Inactive** status. That way, you do not have to delete the activity code, if you want to activate it later.
7. Click the **OK** button.

Deleting activity codes for event alerts

The Visualizer provides activity codes to disposition event alerts. If you need to delete activity codes related to event alert handling, you can delete existing activity codes using the Configuration Console, including the pre-defined event alert activity codes. Deleting the activity code makes it no longer available in the Visualizer to use when handling event alerts.

About this task

If you only want to change information for the activity code, you can edit the activity code without deleting and recreating it.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Visualizer** button.
3. Click the **Activity Codes** tab.
4. From the **Activity Type** drop-down list, select **EVENT**.
5. Select the check box next to the activity codes that you want to delete.
6. Click the **Delete** button. A confirmation window appears.
7. Click the **OK** button.

Visualizer Activity Codes General tab

Activity codes are used by analysts in the Visualizer to disposition role alerts, event alerts, and searches.

Activity Type
Populated by the system. Select the activity type to view, add, or delete activity codes:

- **CONFLICT** used for role alerts
- **EVENT** used for event alerts
- **SEARCH** used for Visualizer searches

Activity Code
Type the unique name for this activity code.

Description
Type a description of this activity code.

Activity Status Code
Select the internal status code that this user activity code corresponds to:

- **Open**
- **Assigned**
- **Closed**
- **Filtered**

Status Indicates whether this activity code is currently active. For example, you can configure an activity code before implementing the code in the Visualizer by making the activity code inactive. Then, when the time comes to implement the activity code, edit it to make it active.

Administering System Configuration settings

Modification of system configuration can be done by following these processes:

Chapter 5. Configuring the system for data

To use IBM InfoSphere Identity Insight effectively you must configure the entity database, entity resolution, and system parameters.

Configuring data in the system

Before you use IBM InfoSphere Identity Insight, you must first configure the entity database to work with your source data.

Configuring characteristic types

You can configure characteristic types for data that cannot be classified as a name, number, address, or email address type. When new data is added to a data source and you want to classify that data as a characteristic type not already configured in the system, you need to create a new characteristic type for the new data.

Characteristics

Characteristics are user-defined traits or properties that are associated with an identity that is not commonly expressed as a name, number, address, or e-mail address.

This attribute enables users to extend the product by defining customizable entity attributes that are meaningful to their data sources

Characteristic types:

Characteristic types organize and identify data that is stored in the entity database. Examples of default Characteristic types that are already configured in the entity database are date of birth and gender.

If you have data that is not defined by one of the default characteristic types, you must create a new characteristic type for that data.

Example

Second National Banker's Trust has recently added a new kind of data about their types of customers. The data comes to the acquisition node using these UMF tags:

```
<attribute>
 <attr_type>cust_type</attr_type>
 <attr_value>merchant</attr_value>
</attribute>
```

In this instance, you need to set up a new characteristic type called cust_type.

System-created characteristic types:

If a UMF message is processed with an characteristic type that is not configured, the system automatically creates a new characteristic type.

The value of the UMF message is recorded in the database using the newly created characteristic type and a UMF exception is written.

When the system automatically creates a new characteristic, it results in an incomplete database record containing only:

- The new **Type** information based on the UMF message.
- A **Status** value of System Created.

Viewing characteristic types

Characteristic types are for data that cannot be classified as a name, number, address, or email address type. You may want to view existing characteristic types if you are thinking of adding a new one.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Characteristics** tab.
4. Select the characteristic type you want to view.

Creating a characteristic type

Characteristics of entities are organized in the system by type.

Before you begin

Before creating a new characteristic type, review the incoming characteristic data determine if it can be accurately described using any existing characteristic types.

About this task

To effectively use new characteristic data, you must configure a new characteristic type using the Configuration Console. If you create a new characteristic type with a data type value of DATE, you will be given a choice to create a new DQM rule to validate the new characteristic type.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Characteristics** tab.
4. Click the **New** button.
5. In the **General** tab, specify the type, description, data type, class, resolution usage, status, keep history, and display level for this characteristic type.
6. Click the **Save** button. If you create a new characteristic type with a data type value of DATE, and you choose to create a new DQM rule to validate the new characteristic type, you will be redirected to the DQM rule creation page with pre-filled values based on the new characteristic type.

Results

The system can now process data in a UMF file that is specified for <CHARACTERISTIC_TYPE>.

Deleting characteristic types

You may delete an existing characteristic type when it is no longer used by the entity database.

About this task

If you created a DQM rule to go with the characteristic type, you may want to delete the corresponding DQM rule as well.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Characteristics** tab.
4. Select the check box next to the characteristic type you want to delete.
5. Click the **Delete** button.

Help topics

Characteristics - General tab:

Use the **General** tab to specify the details of the characteristic type.

Type Type the name of the characteristic type you want to create.

Description
Type the description of the characteristic type you want to create.

Data Type
From the drop-down list, select the data type of the characteristic type you want to create.

CHAR
Select this field type to specify the characteristic type data type as characters.

CLOB Select this field type to specify the characteristic type data type as CLOB.

CLOB must be used for characteristic types that consist of large amounts of data.

Note: Setting the data type to CLOB may negatively impact performance. Use VARCHAR (LVARCHAR for Informix) if possible to reduce possible performance impact.

DATE Select this field type to specify the characteristic type data type as a date.

`Create DQM rule`
If you create a new characteristic type with a data type value of DATE, you will be given a choice to create a new DQM rule to validate the new characteristic type. You will be redirected to the DQM rule creation page with pre-filled values based on the new characteristic type.

VARCHAR
Select this field type to specify the characteristic type data type as variable characters.

Class From the drop-down list, select the class for the characteristic type you want to create.

LC Select this field type to specify the characteristic type as a life characteristic.

For example, height or weight.

SC Select this field type to specify the characteristic type as a system characteristic.

For example, an airline seat preference or a frequent customer point balance.

Resolution Usage
From the drop-down list, select if this characteristic should be used for entity resolution.

None Select this field type to specify that the characteristic value will not be used for entity resolution.

Confirm/Deny
Select this field type to specify that the characteristic value will be used for entity resolution.

Candidates
Select this field type to specify that the characteristic value will be used to build a candidate list and increment a candidate's score.

Candidates/No Scoring
Select this field type to specify that the characteristic value will be used to build a candidate list, but it will not increment a candidate's score.

Status From the drop-down list, select **Active** to specify that this characteristic is active. Otherwise, select **Inactive**.

Keep History
From the drop-down list, select **Yes** to record the historical status of the characteristic type value. Should only be used for characteristic types whose values do not change frequently. Otherwise, select **No**.

Display Level
From the drop-down list, select if this characteristic should be used for graphs and reports.

None Select this field type to exclude the value of this characteristic type on graphs and reports.

All Select this field type to include the value of this characteristic type on all graphs and reports.

Configuring number types

You can configure numbers types for data that can be classified as numbers. When new data is added to a data source and you want to classify that data as a number not already configured in the system, you need to create a new number type for the new data.

Numbers

Numbers are user-defined traits or properties that are associated with an identity that can be classified as a number.

Number types

Number types organize and identify number data that is stored in the entity database. Examples of default number types that are already configured in the entity database are phone and social security number.

If you have number data that is not defined by one of the default number types, you must create a new number type for that data.

Example

Second National Banker's Trust has number data that includes customer checking account numbers. They want to add this new data to the entity database. The data comes to the acquisition node using these UMF tags:

```
<number>
 <num_type>ca</num_type>
 <num_value>41510155060</num_value>
</number>
```

In this instance, you need to set up a new number type called ca.

Viewing number types

Numbers types are for data that can be classified as numbers. You may want to view existing number types when you plan on adding a new number type.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Numbers** tab.
4. Select the number type you want to view.

Creating number types

You need to create a new number type when new data is added in a source system and you want to classify that data as a number type not already configured.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Numbers** tab.
4. Complete one of the following steps:
   - To create a new number type, click the **New** button.
   - To create a number type based on an existing number type, select a number type from the list, and then click the **Clone** button.
5. In the **General** tab, specify the type, description, class, unique, resolution usage, status, keep history, location confirm weight, location deny weight, and other configuration information for the number type.
6. In the **Format** tab, specify the min length, max length, mask, mask fill, fill character, hash length, and other configuration information for the number type.
7. Click the **Save** button.

Deleting number types

You can delete an existing number type when it is no longer used by the system.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Numbers** tab.

4. Select a number type from the list, and then click the **Delete** button.

Help topics

Numbers - General tab:

Use the **General** tab to specify the details of the number type.

Type Type the name of the number type you want to create.

Description
Type the description of the number type you want to create.

Class From the drop-down list, select the class for the number type you want to create.

CC Select this field type to specify the number type as a credit card.

MISC Select this field type to specify the number type as miscellaneous.

For example, a frequent flyer number.

OTHER
Select this field type to specify the number type as other.

For example, an unknown number in a data source.

PHONE
Select this field type to specify the number type as a phone number.

PID Select this field type to specify the number type as a personal identification number.

For example, a driver's license number or a social security number.

SYSID
Select this field type to specify the number type as a system identification number.

For example, an IP address.

Resolution Usage
From the drop-down list, select if this number type should be used for entity resolution.

None Select this field type to specify that the number value will not be used for entity resolution.

Candidates
Select this field type to specify that the number value will be used to build a candidate list and increment a candidate's score.

Status From the drop-down list, select **Active** to specify that this number is active. Otherwise, select **Inactive**.

Keep History
From the drop-down list, select **Yes** to record the historical status of the number type value. Should only be used for number types whose values do not change frequently. Otherwise, select **No**.

Display Level
From the drop-down list, select if this number should be used for graphs and reports.

None Select this field type to exclude the value of this number type on graphs and reports.

All Select this field type to include the value of this number type on all graphs and reports.

Configuring name data

Name data is what is contained in the <NAME> segment of any incoming UMF document. During the entity resolution process, name data is analyzed, compared against the name data of existing entities in the entity database, and given a score based on how closely the name data matched.

Enhanced name hashing with the IBM Global Name Recognition Name Hasher

The Name Hasher uses IBM Global Name Recognition technology to enhance name hashing by creating variant hashes for each incoming name. The variant name hashes enable entity resolution to use fuzzy name matching during name analysis and scoring.

The following scenarios show the benefits of when you might use the Name Hasher:

- When much of the data can be matched only on the `<NAME>` segment
- When much of the data can be matched only on the `<NAME>` segment and most of the data does not conform to the Anglo culture first name, middle name, and last name notation

Using the Name Hasher with the Name Manager name scoring algorithm provides the ability to classify names for culture and accurately compare and score names on the candidate list in a culturally-sensitive context.

The Name Hasher is not enabled by default. You use the Configuration Console to enable the Name Hasher and the associated DQM functions.

Attention: Contact IBM Services or Support if you are upgrading from Name Hasher from product versions 8.0 or 4.2, and if you are an existing customer enabling Name Hasher for the first time. In both cases, without assistance from IBM Services or Support, entity resolution of new data fails when compared against the existing data in the entity database.

Enabling the IBM Global Name Recognition Name Hasher feature:

Enabling IBM Global Name Recognition Name Hasher feature for UMF `<NAME>` segment data quality processing can improve name parsing, culture classification, and name hash generation.

Before you begin

If you are enabling the Name Hasher on an existing installation for the first time, contact IBM Services or Support for assistance. All existing data from all data sources must be reloaded to prevent the entity resolution of new data from failing against the existing data in the entity database.

About this task

These instructions are summaries of the tasks that must be completed to enable the Name Hasher. All steps are completed using the Configuration Console. Click the link to get the step-by-step instructions for each task.

Procedure

1. Enable the DQM function 282 to create name hashes. This function turns on the Name Hasher functionality within the pipelines. If you used the Name Hasher prior to product version 8.0 fix pack 2, see the instructions for migrating to the upgraded Name Hasher. You might want to reuse some of the parameters used by DQM 282.
2. Enable the DQM function 610, so that the Name Hasher can create composite name hash attributes.
3. Configure the Default w/ Name only candidate builder for enhanced name hashing.
4. Configure each data source for enhanced name hashing.
5. Disable full name parsing in DQM function 252. The Name Hasher creates name hash variants for all name parts, not just the full name.
6. Configure DQM rule 255 for enhanced name hashing By completing this step, you keep the name standardization capability of DQM 255, but you disable the standard name hashing to use the enhanced name hashing of the Name Hasher. You also ensure that pipeline validation check of verifying that DQM 255 is enabled does not fail and shut down the pipelines.
7. Enable the DQM function 260 for the UMF `<NAME>` segment. This DQM function assigns name cultures to incoming name data. The Name Hasher requires name culture to apply multi-cultural expertise to enhanced name hashing. Make sure that Name Manager is turned on. (Typically, Name Manager is on.) If you enable the DQM rule 260 and Name Manager is not turned on, DQM 260 rule fails, and the pipelines shut down.
8. Set the system parameters for the Name Hasher. By completing this step, you configure the necessary system parameters for the pipelines used during enhanced name hashing.

Configuring system parameters for enhanced name hashing:

For the Name Hasher to work properly during entity resolution, the default value of the MM system parameter HASHLESS_NAMES_ARE_GENERIC must be turned off. By turning off this value, Name Hasher functionality applies to all incoming name data.

Disabling full name parsing for enhanced name hashing:

For the Name Hasher to work properly, you must disable the existing DQM 252 rule on the `<NAME>` segment.

Configuring DQM rule 255 for the IBM Global Name Recognition Name Hasher feature:

For the Name Hasher to work properly, you must configure the value of the **`UMF Exclude`** parameter in DQM function 255.

About this task

- Disable the standard name parsing and name hashing functionality of the DQM 255 rule in favor of the enhanced parsing and hashing provided by the Name Hasher
- Ensure that the DQM 255 rule is enabled, satisfying the pipeline validation check that requires the DQM rule 255 is enabled

Configuring candidate builders for enhanced name hashing:

For the Name Hasher to work properly, ensure that the **`Default w/ Name Only`** candidate builder configuration contains a match type of **`Characteristic`**.

Configuring data sources for enhanced name hashing:

If you use enhanced name hashing, you must configure each data source to permit name attribute candidate list building, by setting the candidate builder configuration to the **Default w/ Name Only** candidate builder.

Creating composite name hash attributes:

The DQM 610 function builds new attributes out of different smaller values contained in the incoming UMF document. The Name Hasher uses DQM 610 to create composite name hashes and store those hashes as attributes in both the <NAME> and <ATTRIBUTE> UMF segments.

About this task

The resulting composite name hash attributes always contain an <ATTR_TYPE> of GNR_HASH. By creating these name hash attributes, entity resolution can use fuzzy name matching during name analysis and scoring. Fuzzy name matching capability broadens the range of possible identity and entity matches on name data.

Migrating to the upgraded IBM Global Name Recognition Name Hasher:

If your product used the Name Hasher prior to product version 8.0 fix pack 2, complete these tasks in addition to the standard tasks required to upgrade to the latest Name Hasher functionality.

Procedure

1. Perform the standard product upgrade using the product installation program.
2. In the Configuration Console, disable DQM function 660 on the UMF `<NAME>` segment. Copy or write down the current values for the **`maxVariants`** and **`variantScoreThreshold`** parameters contained with the HTTP URL parameter. Prior to product version 8.0 fix pack 2, enhanced name hashing functionality used a Name Hasher servlet that ran on a web application server. In product version 8.0 fix pack 2 and later, the Name Hasher functionality is built into the pipeline. By disabling DQM function 660 on the `<NAME>` segment, you disable the existing Name Hasher servlet.
3. In the Configuration Console, enable DQM rule 282 (Name hash variants) on the UMF `<NAME>` segment and paste or manually configure the following function parameter values:

   `maxVariants`

   Set this value to the same value previously used on the DQM function 660 **`maxVariants`** parameter.

   `variantScoreThreshold`

   Set this value to the same value previously used on the DQM function 660 **`variantScoreThreshold`** parameter

Note: If the DQM function 660 does not contain values for these parameters in the URL, use the DQM function 282 default values.
By completing this step, you activate the Name Hasher functionality within the pipeline.

4. In the Configuration Console, configure the system parameters for the Name Hasher. By completing this step, you globally configure the necessary parameters that the pipeline uses as part of the Name Hasher functionality.

Alternate name parsing

Creating alternate name parses for an incoming full name enhances the name scoring and matching capabilities of entity resolution

Parsing names into name parts is one of the first steps in name matching. Alternate name parses are possible variations of the name. By generating alternate name parses for incoming name data, you can increase the likelihood that the incoming name is correctly analyzed and scored.

Use DQM function 289 to generate alternate name parses. Be default, this function is not enabled. To generate alternate name parses, you must configure the DQM function 289 on the `<NAME>` segment in the Configuration Console.

There might not be an alternate parse for all names. If an alternate name parse exists for the name and only if that alternate parse is different from the primary name parse, then the DQM function generates a second `<NAME>` segment that includes the alternate parse.

For example, consider the following incoming name data:

```
<UMF_ENTITY>
 <NAME>
  <NAME_TYPE>M</NAME_TYPE>
  <FULL_NAME>ALLEN CRAIG</FULL_NAME>
 </NAME>
....
</UMF_ENTITY>
```

In this example, the full name can have at least two different parses. Both "Allen" and "Craig" might be given names or surnames. By generating alternate parses of this name, the entity resolution process can analyze and score the name against more entities in the entity database.

If DQM function 289 is configured on the `<FULL_NAME>` UMF tag of the `<NAME>` segment, during name processing, an alternate name parse is created and added to the UMF record. The resulting record looks like the following record:

```
<UMF_ENTITY>
 <NAME>
  <NAME_TYPE>M</NAME_TYPE>
  <FIRST_NAME>ALLEN</FIRST_NAME>
  <LAST_NAME>CRAIG</LAST_NAME>
 </NAME>
  <NAME>
  <NAME_TYPE>ALT</NAME_TYPE>
  <FIRST_NAME>CRAIG</FIRST_NAME>
  <LAST_NAME>ALLEN</LAST_NAME>
 </NAME>
....
</UMF_ENTITY>
```

The first <NAME> segment contains the primary name parse and the original <NAME_TYPE> value. The second <NAME> segment contains the generated alternate parse, indicated by the <NAME_TYPE> value of ALT. (This example assumes that the value for the alternate parse name type is the default value.)

Configuring names to create alternate name parses:

You can configure names to create alternate name parses, which can be used to support generating multiple name hashes. If you use the IBM Global Name Recognition Name Hasher feature, creating alternate name parses can enhance the fuzzy name matching capabilities to improve entity resolution on name data.

Before you begin

- Make sure that Name Manager is turned on and the path to the support files is set in system parameters. If this DQM function is enabled without a valid path to the Name Manager support files, the pipeline logs an error and shuts down.
- When you enable the DQM function for the alternate name parse feature, you are changing system configuration. As with any configuration changes, make sure to stop any active pipelines before you change the configuration. Then restart the pipelines to reinitialize them with the configuration changes.

About this task

- For new product installations of V8.0 fix pack 2 or later, this DQM function is already configured and active.
- For upgraded product versions as of V8.0 fix pack 2 or later, this DQM function is configured but is inactive. To generate alternate name parses, change the status of the existing DQM function to **Active**.

Procedure

1. In the Configuration Console, select **Setup** > **UMF** > **DQM Rules**.
2. Select NAME from the **Segment** list.
3. Select the UMF tag name that lists **289 - Alternate Name Parsing** in **Function**.
4. In **Status**, make sure that Active is selected.
5. On the **Parameters** tab, review or set the following parameter values:
   - **Parse Score Threshold**: Set this value to a number between 0 and 100. The higher the score, the fewer alternate parses are created. This value sets the threshold for the minimum confidence score that the name parser will use to determine if an alternate parse for the incoming name is to be created. If no alternate parse with a higher confidence score is found, or if the originally-provided inbound parse already scores above the threshold, then no alternate parse will be created.
   - **Alternate Name Type**: Enter the value for NAME_TYPE to indicate that this name is an alternate parse. This value is the UMF tag that is added to the <NAME> segment for each alternate name parse created. By default, this value is set to ALT. To ensure full attribution of entity resolution, do not set this value to an existing inbound NAME_TYPE configured in the Configuration Console. Specifically, do not set this value to either **M** or **A**.
6. Click **Save**.

Gender determination

When processing incoming name data, sometimes the gender of a personal name can be the determining factor in whether two entities match. Gender adds confirmation or denial weight to entity resolution scoring to determine if two identities are the same entity.

DQM function 258 dynamically identifies the gender of the <NAME> segment in an incoming UMF record, creates a gender characteristic, and adds the gender characteristic to the incoming UMF record. The gender characteristic is added using the <ATTRIBUTE> segment.

- If the incoming UMF record already contains a gender characteristic in its data, DQM function 258 does not generate another gender characteristic.
- If the UMF record contains more than one <NAME> segment, DQM function 258 creates only one gender characteristic for the entire input record. In this case, generating multiple gender attributes can be either redundant or conflicting.

To dynamically determine the gender of a name, ensure that at least one UMF tag in the <NAME> segment is configured to use DQM function 258.

- For new product installations of V8.0 fix pack 2 or later, this DQM function is already configured and active.
- For upgraded product versions as of V8.0 fix pack 2 or later, this DQM function is configured but is inactive. If you want to use this improved gender functionality, you must change the status to **Active**. If you previously assigned gender using the **Gender Characteristic Type** parameter of DQM function 255, reset the value of that parameter to NONE. You can still use DQM 255 on any UMF <NAME> tag to standardize names.

You might also want to check the following configurations in the Configuration Console:

- Ensure that the gender characteristic is configured as a confirmation or denial in entity resolution by data source. You view or configure this setting in the **Resolution Usage** field located by selecting **Setup** > **Sources** > **Characteristics**.
- Ensure that the gender characteristic is configured with the correct adjustment values for entity resolution. You view or configure this setting by selecting **Setup** > **Resolution** > **Characteristics**. Check the values of the confirmation and denial weights assigned to the gender characteristic to be certain that they fit your business needs.

.

Consider the following example <NAME> segment in this incoming UMF record:

```
<UMF_ENTITY>
 <NAME>
  <NAME_TYPE>M</NAME_TYPE>
  <LAST_NAME>RASUL</LAST_NAME>
  <FIRST_NAME>KARIM</FIRST_NAME>
 </NAME>
.....
</UMF_ENTITY>
```

If DQM 258 is activated on the <FIRST_NAME> UMF tag of the <NAME> segment, the incoming UMF record looks like the following record after the gender is analyzed and created:

```
<UMF_ENTITY>
 <NAME>
  <NAME_TYPE>M</NAME_TYPE>
  <LAST_NAME>RASUL</LAST_NAME>
  <FIRST_NAME>KARIM</FIRST_NAME>
 </NAME>
 <ATTRIBUTE>
  <ATTR_TYPE>GENDER</ATTR_TYPE>
  <ATTR_VALUE>M</ATTR_TYPE>
 </ATTRIBUTE>
.....
</UMF_ENTITY>
```

Configuring names to assign gender:

By assigning gender based on a name, you can improve entity resolution. You can set confirmation and denial scores based on whether the gender of the compared entities is the same. You can configure names to dynamically assign gender and add the Gender characteristic to the incoming UMF records.

Before you begin

- Make sure that Name Manager is turned on and the path to the Name Manager support files is set in system parameters. If this DQM function is enabled without a valid path to the Name Manager support files, the pipeline logs an error and shuts down.
- When you enable the gender feature of this DQM function, you are changing system configuration. As with any configuration changes, make sure to stop any active pipelines before you change the configuration. Then restart the pipelines to reinitialize them with the configuration changes.

About this task

- For new product installations of V8.0 fix pack 2 or later, this DQM function is already configured and active.
- For upgraded product versions as of V8.0 fix pack 2 or later, this DQM function is configured but is inactive. To use this improved gender functionality, change the status of the existing DQM function to **Active**. If you previously assigned gender using the **Gender Characteristic Type** parameter of DQM function 255, reset the value of that parameter to `NONE`. You can still use DQM 255 on any UMF `<NAME>` tag to standardize names.

Procedure

1. In the Configuration Console, select **Setup** > **UMF** > **DQM Rules**.
2. Select **NAME** from the **Segment** list.
3. Select the UMF tag name of **FIRST_NAME** that also lists **258 - Name Genderizer** as **Function**. This configuration evaluates only the first name in the incoming record for personal names. If the categorizing names feature of Name Manager is turned on, you must specify the entire name in the `LAST_NAME` UMF tag of the `NAME` segment.
4. In **Status**, make sure that **Active** is selected.
5. In **Rule Filter**, make sure that `NAME_TYPE=M` is the field value. This value ensures that only the main name for each input record is evaluated to assign gender.
6. On the **Parameters** tab, make sure that the **Minimum Gender Confidence Score** is set to a number between 0 and 100. The default score is set to 90, meaning that gender is not assigned unless there is a 90% confidence in the gender assignment. Be cautious about lowering this score, because a minimum score below 90 can impact entity resolution during gender confirmation or denial.
7. Click **Save**.

Name categorization

If the **NAMESIFTER** Name Manager system parameter is enabled, the product categorizes names by type. By categorizing names by type, entity resolution can apply the appropriate linguistic and reference-data resources during name analysis, scoring, and matching:

Names are categorized into either personal or organizational name types.

Personal names
A personal name contains no indicators that suggest it belongs in any other category. (For example: "Linda K. Smith".) Names that are categorized as personal names are parsed into name parts. Name parts are then categorized by culture, which adds precision to the analyzing and scoring process.

Organizational names
A business or organizational name contains some form of a non-personal indicator. (For example, "Smith & Company".) Names that are categorized as organizational names are automatically assigned a culture of "Company"

Unknown names
A name categorized as "Unknown" contain some element that appears to either be a misspelling or some other construct that does not normally appear in either personal or business names. (For example "SMI".)

Categorizing names by type:

One of the Name Manager system parameters (**NAMESIFTER**) includes the ability to categorize names by type. The most common name types are personal and business. Categorizing names can make the name analysis and scoring part of the entity resolution process more precise.

Categorizing personal names by culture:

DQM function 260 was created to determine the culture of the name and append that value to the UMF <NAME> segment. By default, the<NAME> segment configuration includes a DQM 260 rule on the <LAST_NAME> UMF tag. Use these instructions to add the DQM 260 rule to another UMF tag in the <NAME> segment, or to update the existing rule on the <LAST_NAME> UMF tag

Overview of Name Manager

Name Manager

Name Manager increases name precision for advanced name confirmation issues, such as multiple name transliterations, misspellings within cultures, spelling variations within cultures and from culture-to-culture, and names that use patronymic or honorific designations. It uses the IBM InfoSphere Global Name Recognition component libraries, which contain a knowledge-base of over 1,000,000,000 multicultural names and unique linguistic information, which adds culture-specific name matching capabilities.

Name Manager scores names using the following process:

- Categorizes names by name type (personal or business)
- Parses personal names into name parts
- Classifies names by culture (supports more than 20 cultures, including Afgani, Arabic, Farsi, Han, Japanese, Korean, Thai, Vietnamese, and Yoruban)

- Normalizes personal names (if the name is classified as either Anglo, Arabic, Chinese, French, German, Hispanic, Indian, Korean, Russian, or Thai)

Configuring Name Manager

By default, Name Manager name scoring is already enabled and configured when you install IBM InfoSphere Identity Insight. However, you can use the Configuration Console to review or change Name Manager configuration settings, including the following settings:

- Name Manager system parameters, including the support path to Name Manager component libraries (global parameters that the pipeline uses to perform entity resolution)
- Name Manager name scoring thresholds used during name matching (confirmations and denials)

Configuring system parameters for the Name Manager:

By default, the Name Manager name scoring system parameters are configured when you install the product. But you can update the default system parameters, when needed. For example, you might need to change the location of the Name Manager support libraries.

About this task

You set the path to the Name Manager support libraries and enable categorizing names by type through Name Manager system parameters. You also set the `CROSSCHECKCULTURE` system parameter to configure name processing between different name cultures.

Procedure

1. In the Configuration Console, select **Setup** > **General** > **System Parameters**.
2. From the `Parameter Group` list, select the **NAMEMANAGER** parameter group.
3. From the left pane, select the Name Manager system parameter to configure:

Name Manager system parameter	Description
SUPPORTPATH	Indicates the location of the Name Manager support files. The default value is ./data, which is a path relative to the top-level product directory. If the support files are moved to a different location during installation, modify this value to the absolute path of the new location.
NAMESIFTER	Indicates whether the name categorization by name type (personal or organizational names) functionality is turned on. To enable categorizing names by type (Name Sifter functionality), enter 1 (new installation default) in **Current Value** To disable categorizing names by type (Name Sifter functionality), enter 0 (upgrade default) in **Current Value**

Name Manager system parameter	Description
CROSSCHECKCULTURE	Indicates whether to perform Name Manager name scoring between name cultures when the name cultures are different. To check only the inbound name culture before scoring both names, enter `0` in **Current Value**. To check name culture values before scoring them (new installation default), enter 1 in **Current Value**.

Attention: The `CROSSCHECKCULTURE` system parameter affects how entity resolution handles name scoring by culture in the pipelines. Before changing this system parameter from its current value, consult IBM Services or Support.

4. Click **Save**.

Configuring Name Manager thresholds for confirmations and denials:

You can set the name score thresholds that Name Manager uses during entity resolution by resolution rule. After the candidate list is built, entity resolution compares the Name Manager name score, based on name part and the culture determined for each name part, to these thresholds. If the Name Manager score meets or exceeds the configured threshold score for the name part, the names are considered a match.

About this task

Important: By default, the Name Manager name part scoring thresholds are configured for optimum Name Manager scoring and performance. Changing the default values is an advanced configuration task, because these values can negatively affect entity resolution for rules that include name scoring. Before changing these default values, consult IBM Services or Support.

Procedure

1. In the Configuration Console, select **Setup** > **Resolution** > **Resolution Rules**
2. Select the resolution configuration from the **Resolution Config** list.
3. Select the resolution rule.
4. Click **Confirm/Deny Thresholds**.
5. Under **Name Manager**, enter the minimum score for each name part threshold, based on a score from 0.0 to 1.0. The higher the score, the more exact the name parts must be to match. Typically, a score below 0.7 is not considered suitable to match name parts.

Name Manager name scoring:

The Name Manager algorithm scores incoming name data based on grouping the name into name parts and then determining the culture for each name part. The algorithm then scores each name part, and the resulting scores are used during entity resolution.

While the Name Manager algorithm is separate from the Name Comparator algorithms (NC1 and NC2), you must still select either NC1 or NC2. During the entity resolution process, names are first scored based on the selected Name Comparator algorithms. If the name scores an exact match, entity resolution skips the Name Manager scoring, because the exact name match satisfies the name score portion of the resolution rule. If the incoming name scores less than an exact match, however, the entity resolution process scores the name using the Name Manager algorithm.

First, the algorithm parses the name into name parts (given name, surname, and full name), and then the algorithm determines the culture for each name part. Finally, the algorithm assigns each name part a score, and compares the scores against the configured Name Manager score thresholds to determine how closely the names matched. The higher the score threshold is set, the closer the name parts from the incoming name data must match the name parts from the existing entity in the entity database.

Selecting cultures for Name Manager name scoring:

You can configure which name scoring methods are used by culture during the name scoring process of entity resolution. Name Manager can only determine name culture and score names for the cultures configured to use Name Manager name matching.

About this task

By default, each supported culture is already configured based on the most recent best practices for typical name scoring. Changing the default values is an advanced task that can negatively affect entity resolution for rules that include name scoring. Consult IBM Services or Support before changing the default configuration values.

Procedure

1. In the Configuration Console, select **Setup** > **Resolution** > **Name Manager Match Config**.
2. Select a Name Manager culture.
3. In **Use Name Manager Name Matching**, select `Yes`
4. Click **Save**.

Configuring DQM rules

You can configure DQM rules to repair or clean up data that does not meet minimum data quality standards. DQM rules are applied to a specific UMF tag in a specific UMF segment.

About this task

DQM Rules can be viewed and modified by using the Console on the **DQM Rules**tab.

DQM rules

DQM rules are configured system-defined repair, clean up, and standardization functions applied to incoming identity data values in a specific order.

DQM rules define how the system processes the incoming data and are designed to properly format numbers, identify and correct clerical or transposition errors, and identify and correct intentional inaccuracies introduced by those intent on trying to conceal their identities. DQM rules can perform a variety of repair, clean up, and standardization functions on incoming identity data values.

To configure a DQM rule, you select a specific UMF segment (such as NAME) and UMF tag (such as NAME_TYPE), then you select a system-defined DQM function to apply to the incoming data, and finally you specify the associated parameters for that function, including any default values the system should apply. You also choose the order in which to apply this DQM rule on the selected UMF segment, since the product supports multiple DQM rules for each UMF segment.

Viewing DQM rules

DQM rules repair or clean up data that does not meet minimum data quality standards.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **UMF** button.
3. Click the **DQM Rules** tab.
4. From the **Segment** drop-down list, select the UMF Segment containing the DQM Rules to be viewed.

Creating DQM rules

You create DQM rules to repair or clean up data that does not meet minimum data quality standards.

About this task

DQM rules are applied to a specific UMF tag in a specific UMF segment. DQM rules can also be cloned to create the basis for a new rule.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **UMF** button.
3. Click the **DQM Rules** tab.
4. From the **Segment** drop-down list, select the UMF Segment for which to create a DQM rule.
5. Complete one of the following steps:
   - To create a new DQM rule, click the **New** button.
   - To create a DQM rule based on an existing DQM rule, select a DQM rule from the list, and then click the **Clone** button.
6. In the `General` tab, specify the order UMF tag name, function, rule filter, UMF exclude, correctable, status, and other configuration information for the DQM rule.
7. In the `Parameters` tab, specify the parameters for the DQM rule.
8. Click the **Save** button.
9. Validate the DQM rule.

Deleting DQM rules

When a DQM rule is no longer required, you should delete the DQM rule.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **UMF** button.
3. Click the **DQM Rules** tab.

4. From the **Segment** drop-down list, select the UMF segment for which to delete a DQM rule.
5. Select the check box(es) next to the DQM rule(s) you want to delete.
6. Click the **Delete** button.

Validating DQM rules

When you add or edit a DQM rule, you should validate it before applying the DQM rule to source data.

About this task

The validate function is used to validate all rules, in relation to each other, for a whole segment. Validation that can be performed on a single rule is automatically performed when you save the rule.

When you log into the configuration console, an automatic validation check is performed to see if the DQM rules are valid. If an error is found, a header message will display at the top of the configuration console screen. Click the **Review the errors** link to open a new window describing the errors.

Procedure

1. Click the **Setup** button.
2. Click the **UMF** button.
3. Click the **DQM Rules** tab.
4. From the **Segment** drop-down list, select the UMF segment for which to validate a DQM rule. If a segment is not selected, then validation will be performed on all segments.
5. Click the **Validate** button.

Turning off DQM rules

You can turn off a DQM rule that is no longer required.

Procedure

1. Click the **Setup** button.
2. Click the **UMF** button.
3. Click the **DQM Rules** tab.
4. From the **Segment** drop-down list, select the desired UMF Segment containing the DQM rule to be turned off.
5. Click on the DQM rule to be turned off.
6. In the **General** tab, set the status field to **Inactive**.
7. Click the **Save** button.

Help topics

DQM Rules- General tab:

Use the **General** tab to specify the details of the DQM rule.

Segment
Type the UMF segment name where to apply the DQM rule. This field will usually be read-only. The only time it can be edited is when the **Segment** drop-down list was left blank when creating a new DQM rule. The segment name must be entered in uppercase.

Order Type the order number in which the DQM rule will be applied.

UMF Tag Name
Type the UMF tag name where to apply the DQM rule. The UMF tag name must be entered in uppercase

Function
From the drop-down list, select the DQM function you want to base the DQM rule on.

Function Description
The function description field is a read-only field that describes what the DQM rule does.

Rule Filter
If you want the DQM rule to only be applied if the UMF tag contains a specific value, enter an equation that includes the UMF tag name and the required value to run the DQM rule.

For Example: `NAME_TYPE=m`
This example setting only applies the DQM rule if the UMF tag `NAME_TYPE` value is `m`.

UMF Exclude
If you want the DQM rule to not apply to specific UMF input documents, enter a comma delimited list of UMF input documents that this rule should not be run for.

For Example: `UMF_QUERY, UMF_DISCLOSED RELATION`
This example setting only will not apply the DQM rule to `UMF_QUERY` or `UMF_DISCLOSED_RELATION` UMF input documents.

Correctable
From the drop-down list, select **Yes** to adjust invalid and substandard values. Otherwise, select **No**.

The parameters of each DQM Rule determine how substandard data values are adjusted.

Status From the drop-down list, select **Active** to specify that this DQM rule is active. Otherwise, select **Inactive**.

Configuring lookup codes

Lookup codes are default values used by various features of the application.

Lookup codes are classified by code types. DQM rule 190 can be used to validate that incoming lookup codes are part of a defined code type, or optionally add it to that code type if it is missing.

Viewing Lookup Codes

Lookup codes are default values used by various features of the application.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **General** button.
3. Click the **Codes** tab.
4. From the **Type** drop-down, select the type of lookup code values you want to view.

Creating lookup codes

Lookup codes are default values used by various features of the application.

About this task

You can create a new lookup code, or you can create a lookup code that is based on an existing lookup code.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **General** button.
3. Click the **Codes** tab.
4. From the **Type** drop-down, select the type of lookup code values you want to create. To create a completely new code type, leave the value as is.
5. Complete one of the following steps:
   - To create a new lookup code, click the **New** button.
   - To create a lookup code based on an existing lookup code, select a lookup code from the list, and then click the **Clone** button.
6. In the `General` tab, specify the type (will be a read-only field if it was already specified in the **Type** drop-down), code, description, status, and other configuration information for this lookup code.

Deleting lookup codes

You can delete user-created lookup codes that are no longer used.

About this task

You should not delete system default lookup codes as they are required by various components of the product.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **General** button.
3. Click the **Codes** tab.
4. From the **Type** drop-down, select the type of lookup code values you want to delete.
5. Select a lookup code from the list, and then click the **Delete** button.

Turning off lookup codes

You can turn off a lookup code that is no longer required.

Procedure

1. Click the **Setup** button.
2. Click the **General** button.
3. Click the **Codes** tab.
4. From the **Type** drop-down, select the type of lookup code values you want to turn off.
5. Select a lookup code from the list.
6. In the `General` tab, set the status field to **Inactive**.
7. Click the **Save** button.

Help topics

Lookup Codes - General tab:

Use the **General** tab to specify the details of the lookup code.

Type Type the lookup code type to group the lookup code under. This field will be read-only once specified. It can only edited if the **Type** drop-down was left unspecified when creating a new lookup code.

Code Type the value to be available as a default value of the lookup code. It is typically a value that is actually used in UMF tags and stored in database tables. When editing existing lookup codes, this field will be read-only.

Description
Type the description of the lookup code.

Status From the drop-down list, select **Active** to specify that this lookup code is active. Otherwise, select **Inactive**.

Lookup Codes - Type field:

Use the **Type** field to specify the type to group lookup code under.

ADDR_STAT
This lookup code type is used for address status values. These values can be used to mark particular addresses with information like whether it is a deliverable address.

ADDR_TYPE
User definable classifications for addresses. These are the valid values for the `ADDR_TYPE` UMF tag.

ANALYZER_GROUP
This lookup code type is used by role alert rules and the visualizer. Any new lookup code with a type of ANALYZER_GROUP is an available option in the **Alert Group** drop-down of the **Setup** > **Relationships** > **Role Alert Rules** > **General** tab and the **Group** drop-down of the **Setup** > **Visualizer** > **Visualizer Users** > **General** tab.

ATTR_CLASS
User definable classifications for characteristic types. Values entered here will appear as options in the **Class** drop-down under the **Setup** > **Sources** > **Characteristics** > **General** tab. Characteristics that use the LINK lookup code for their attribute class can be displayed as an HTML link in the Visualizer if the characteristic's value follows this format:

Link Display Text=URL

ATTR_MATCH_LEVEL
This lookup code type is deprecated.

CONF_LEVEL
This lookup code type is deprecated.

DENSITY_LOG_LEVEL
This lookup code type is deprecated.

DOC_TYPE
This lookup code type is deprecated.

DSRC_ACTION
This lookup code type is used by the system and should not be modified.

EX_CLASS
This lookup code type is used by the system and should not be modified.

EX_SEVERITY
This lookup code type is used by the system and should not be modified.

LOG_LEVEL
This lookup code type is used by the system and should not be modified.

ER_LEVEL
This lookup code type is used by the system and should not be modified.

ER_LOG_LEVEL
This lookup code type is used by the system and should not be modified.

LDR_MESSAGE_TYPE
This lookup code type is deprecated.

MM_STAT
This lookup code type is deprecated.

NAME_TYPE
This lookup code type is used to store user definable classifications for names. These are the valid values for the `NAME_TYPE` UMF tag.

NS-FGEN
This lookup code type is used by the system and should not be modified.

NS-LGEN
This lookup code type is used by the system and should not be modified.

NS-PREFIX
This lookup code type is used by the system and should not be modified.

NS-SUFFIX
This lookup code type is used by the system and should not be modified.

NUM_CLASS
This lookup code type is used to store user definable classifications for number types. Values entered here will appear as options in the **Class** drop-down in the **Setup** > **Sources** > **Numbers** > **General** tab.

REC_STAT
This lookup code type is used by the system and should not be modified.

SEARCH_REASON
This lookup code type is used by the visualizer for a list of drop-down options for the attribute alert search reason field. Users can add there own list of valid reasons for an attribute alert here.

SYS_DELETE_STAT
This lookup code type is used by the system and should not be modified.

UNIQUE_FLAG
This lookup code type is deprecated.

USABILITY_LOG_LEVEL
This lookup code type is used by the system and should not be modified.

Configuring generic data values

You can configure data values to be generic if they exceed a configured number of occurrences in the entity database.

Generic values

Generic values describe data values that repeatedly occur in the entity database and, as a result, are no longer used by the system to resolve entities.

Data values are considered generic after they exceed a certain threshold. The threshold is a configured maximum number of occurrences of entities in the entity database that can share the data value.

Generic values are organized and configured by attribute and attribute type. The generic data value of a specific attribute type will override the generic data value of the parent attribute. Standard data elements whose values can be considered generic are:

- Address
- Characteristic
- Email
- Name
- Number

Example

If the generic threshold for phone numbers is set to 25, once a phone number value (like 555-555-5555 for example) is the phone number value of more than 25 entities, from that time forward that specific value is not used to resolve entities.

Note: When considering how high to set generic thresholds, consider that setting a threshold too high might result in the system performance eventually being overwhelmed by a flood of data that should be generic. In contrast, setting a generic threshold too low might result in important alerts not being generated because key criteria is considered generic.

Viewing generic data values

Generic data values are generics thresholds for each data element you want to consider generic. You may want to view existing generics when adding a new data source.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **UMF** button.
3. Click **Generic Threshold** tab.

Configuring generic data values

To have generic values ignored during entity resolution, you must configure the generic threshold for the data element.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **UMF** button.
3. Click the **Generic Threshold** tab.
4. Complete one of the following steps:
   - To create a new generic data value, click the **New** button.
   - To create a generic data value based on an existing generic data value, select a generic data value from the list, and then click the **Clone** button.

5. In the **General** tab, specify the attribute, attribute type and threshold value of the generic value.
6. Click the **Save** button.

Deleting generic data values

Generic data values are generics thresholds for each data element you want to consider generic. You may want to delete existing generics when it no longer is relevant to incoming data.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **UMF** button.
3. Click the **Generic Threshold** tab.
4. Select the check box next to any existing element name you want to delete.
5. Click the **Delete** button.

Help topics

Generic Threshold - General tab:

Use the **General** tab to specify the details of the generic data value.

Attribute Name
From the drop-down list, select the attribute you want to apply the generic data value to.

Attribute Type
From the drop-down list, select the attribute type you want to apply the generic data value to.

This drop-down list will only have multiple options if the **Attribute Name** field is set to Name or Characteristic.

Threshold
Type the number of entities that can share a UMF value of the configured type before it is considered generic.

Configuring roles

You can configure roles to classify entities in the entity database. Roles can be assigned to data sources or entities. Conflicting roles generate alerts.

About this task

Roles can be viewed and modified by using the Console, on the **Data Sources** tab.

Roles

A role is a classification of an identity that defines the focus or purpose of that identity. You can associate one or more roles with an identity. As identities are resolved into entities, entities inherit all associated roles.

You use roles to configure role alert rules, which define relationships of interest and generate alerts.

Every identity is assigned a role in one of two ways:

By incoming data source

When you configure a new data source, you associate a role with that data source, which will assign that role to all identities containing that data source code.

By UMF

When you transform the data source into Universal Message Format (UMF), you can directly assign roles as part of the UMF record using the <SEP_ROLES> UMF segment with the <ROLE_CODE> UMF tag. If you configure by UMF, DQM rules and a lookup table will need to be added.

Examples of useful roles might include employees, vendors, customers, or watch list.

Example of assigning roles using UMF

To assign the role of employee to an identity record using UMF, you would enter the following <SEP_ROLES> UMF segment and <ROLE_CODE> UMF tag for the identity record:

```
<SEP_ROLES>

  <ROLE_CODE>Employee</ROLE_CODE>

</SEP_ROLES>
```

Viewing roles

A role defines how an entity is classified or known within the system. You may want to view existing roles if you are planning to add a new role.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Relationships** button.
3. Click **Role Codes** tab.
4. Select the role you want to view.

Creating roles

To define how entities relate to other entities, create roles in the system.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Relationships** button.
3. Click **Role Codes** tab.
4. Complete one of the following steps:
   - To create a new role, click the **New** button.
   - To create a role based on an existing role, select a role from the list, and then click the **Clone** button.
5. In the **General** tab, specify the role code, description, class, status, and other configuration information for the new role.
6. Click the **Save** button.

What to do next

You can use this role when defining role alert rules.

Deleting roles

A role defines how an entity is classified or known within the system. You may want to delete an existing role if it is not longer valid.

About this task

You cannot delete a role that is being used by a role alert rule or a data source.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Relationships** button.
3. Click **Role Codes** tab.
4. Select the check box next to any existing role you want to delete.
5. Click the **Delete** button.

Help topics

Roles - General tab:

Use the **General** tab to specify the details of the role.

ID Type the unique integer to identify the role ID.

The ID value is automatically populated with the next sequential number not in use.

Role Code
Type a unique value to identify this role.

Description
Type a description for this role.

Role Class
Type a role class for this role.

Status From the drop-down list, select **Active** to specify that this role is active. Otherwise, select **Inactive**.

Configuring role alert rules

You can configure role alert rules to define a combination of roles that, when detected, generate alerts.

About this task

Role Alert Rules can be viewed and modified by using the Console, on the **Role Alert Rules** tab.

Role alert

A role alert is defined in the system by a role alert rule which represents relationships that are used to generate alerts.

Role alert rules define a combination of roles that, when detected in a relationship or entity, indicate some form of conflict. For example, a role alert rule might indicate that whenever an entity with the Employee role knows an entity with the Vendor role, a role alert exists. This role alert rule can be described as "Employee knows Vendor." When the system finds role alerts in entities or relationships, alerts are created that may be published to the enterprise and seen in the Analyst Toolkit applications.

Though most role alert rules specify a combination of two different roles that indicate a conflict, it is also valid to have a role alert rule where an entity of one role knows another entity of the same role. For example, you might want to know about any relationships among your customers and build a role alert rule that generates a role alert any time one customer entity relates to another customer entity. This role alert rule can be described as "customer knows customer."

Role alert rules are based on existing role codes. Roles must be defined before conflict rules can be created around those roles.

Viewing role alert rules

A role alert rule is used to generate alerts when a relationship between two defined roles is detected. You may want to view existing role alert rules when you plan to add a new role alert rule.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Relationships** button.
3. Click **Role Alert Rules** tab.
4. Select the role alert rule you want to view.

Configuring role alert rules

You configure role alert rules to generate role alerts or relationships between two roles or identities.

Before you begin

Before defining a role alert rule, you must first configure the roles that you want to use in the role alert rule. For example, if you want to configure a role alert rule where an employee cannot be a vendor, your system must contain the roles of "Employee" and "Vendor".

Procedure

1. Click the **Setup** button.
2. Click the **Relationships** button.
3. Click **Role Alert Rules** tab.
4. Complete one of the following steps:
   - To create a new role alert rule, click the **New** button.
   - To create a role alert rule based on an existing role alert rule, select a role alert rule from the list, and then click the **Clone** button.

   The **Role Alert Rule ID** field is automatically filled with the next unique ID. You can change this to any unique ID number.
5. Click the **New** button.
6. On the **General** tab, specify the ID, description, severity, role codes, alert group and minimum alert threshold for this role alert rule.

7. On the **Filters** tab, optionally specify the identity filter, data change filter, and path strength adjustment (only shown if the data change filter field is set to Path Strength Adjustment). If both filters are set, only one filter has to be met to generate a role alert.
8. Click the **Save** button.

Deleting role alert rules

A role alert rule should be deleted when a defined role in the role alert rule is to be deleted, or the role combination in the role alert rule is no longer of interest.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Relationships** button.
3. Click **Role Alert Rules** tab.
4. Select the check box next to any existing role alert rule you want to delete.
5. Click the **Delete** button.

Help topics

Role Alert Rules- General tab:

Use the **General** tab of the **Role Alert Rules** window to configure the details of role alert rules. Roles are associated with data sources. Each identity coming into the system from a data source is assigned a role, based on how the data source is configured. Role alert rules define when to generate a role alert, based on a conflict between the roles assigned to incoming identities and identities associated with entities in the entity database.

Role Alert Rule ID
The ID value is automatically populated with the next sequential number not in use.

Description
Type a description for this role alert rule. This text displays in the Visualizer whenever a role alert is generated based on this role alert rule.

Severity
A user-defined one-character code used to categorize the importance of alerts generated from this rule.

Match the severity of the role alert to its importance. This code displays with role alerts generated from this role alert rule in the Visualizer. Analysts use it to prioritize which alerts to review first, so the one-character code should be meaningful to Visualizer users. For example, a role alert rule that generates an alert whenever a passenger matches someone on a No Fly list might be more critical to review than a role alert rule designed to generate an alert whenever an employee knows a customer.

Examples of severity codes include the following: C for critical, N for neutral, I for interesting, H for high, or L for low.

Role 1 From the drop-down list, select the first role for comparison in this role alert rule.

The role options that display are the existing, configured roles. If you do not see the role that you want to select, configure the role on the **Roles** tab first.

Role 2 From the drop-down list, select the second role for comparison in this role alert rule.

The role options that display are the existing, configured roles. If you do not see the role that you want to select, configure the role on the **Roles** tab first.

Alert Group
From the drop-down list, select the Visualizer analyzer group that will analyze the role alerts that are generated from this role alert rule. For example, you could direct all Passenger-No Fly List role alerts to a security desk, and all Employee-Vendor role alerts to human resources.

The group options that display are the active, configured Visualizer analyzer groups with the code type of `ANALYZER_GROUP`. If you do not see the group that you want to select, configure a new ANALYZER_GROUP code on the **Setup - General - Codes** tab first.

This is a required field, so even if your organization does not use the Visualizer, you must configure and select an alert group code.

Role Alert Rules tab:

If both filters are set, only one filter has to be met to generate a role alert.

Identity Filter
From the drop-down list, select a filter to restrict role alert generation when new identities are added to the entities involved in the role alert.

This filter only affects re-alerting behavior. The first time the role alert rule is satisfied for a given set of entities, a role alert is always generated. This filter can prevent further generation of the same role alert when changes are made to the entities involved.

Off Select this field type to turn off role alert restriction when new identities are added to the entities involved in the role alert.

New Identity
Select this field type to only realert when a new data source code is introduced amongst the identities in the entities involved in the role alert.

New Data Source Code
Select this field type to alert when a new data source code is introduced amongst the identities.

Data Change Filter
From the drop-down list, select a filter to restrict role alert generation when new attribute data is added to the entities involved in the role alert.

This filter only affects re-alerting behavior. The first time the role alert rule is satisfied for a given set of entities, a role alert is always generated. This filter can prevent further generation of the same role alert when changes are made to the entities involved.

Off Select this field type to turn off role alert restriction when new attribute data is added to the entities involved in the role alert..

New Attribute Data
Select this field type to only realert when new attribute data is added to the entities involved in the role alert.

Path Strength Adjustment
Select this field type to only realert when new attribute data is added that causes a change in path strength equal to or greater than the **Path Strength Adjustment** value.

Path Strength Adjustment
This field only displays if the **Data Change Filter** drop-down is set to Path Strength Adjustment. Type an adjustment value (-100 to 100) to use when the **Data Change Filter** is set to Path Strength Adjustment. This allows regeneration of role alerts only when new attribute data is added that causes a change in path strength equal to or great than the Path Strength Adjustment value. Specifying zero is the same as turning the filter off.

Configuring entity types

You can configure entity types to identify the exact nature of the entity.

About this task

When new identity data is added to a data source and you want to classify that data as an entity type not already configured in the system, you need to create a new entity type for the new data.

Entity Types can be viewed and modified by using the Console, on the **Entity Types**tab.

Entity Types

Entity types are user-defined traits or properties that are associated with an entity to identify the exact nature of the entity.

Impersonal awareness uses entity types to link entities that would otherwise not have a 1-degree relationship.

For example, if you wanted to find impersonal relationships using telephone calls, you would create a new entity type of *Phone call* and adjust your acquisition node to correctly tag each telephone call record with the *Phone call* entity type.

When the telephone records are ingested into the pipelines, the entity and relationship resolution process finds a 1-degree relationship between the *Phone call* entity and the calling entity (*Person*). It also finds a 1-degree relationship between the person called and the *Phone call* entity. By itself, the system does not finds a 1-degree relationship between the persons.

```
<UMF_ENTITY>
<DSRC_CODE>100</DSRC_CODE>
<DSRC_ACCT>123abc</DSRC_ACCT>
<DSRC_REF>1</DSRC_REF>
<ENTITY_TYPE>PHONE</ENTITY_TYPE>
<NUMBER>
<NUM_TYPE>PH</NUM_TYPE>
<NUM_VALUE>702-555-1212</NUM_VALUE>
</NUMBER>
</UMF_ENTITY>
```

Impersonal awareness

Impersonal awareness is a product feature that extends the traditional relationship resolution process to find and analyze impersonal relationships. The relationship detection process finds relationships between entities based on attribute values associated with those entities. Sometimes, it is important to find relationships between entities based on activities or other impersonal identifiers. These relationships between entities based on activities or other impersonal identifiers are referred to as *impersonal* relationships, and activities or impersonal identifiers that relate people are called *relating facts*.

Impersonal relationships always exist at two or more degrees of separation, because the relating fact is, itself, an entity. So to enable impersonal awareness and find impersonal relationships, configure your data sources to use the Degrees of Separation feature, which extends entity and relationship resolution to detect relationships at more than two degrees of separation.

For example, a telephone transaction contains data about telephone numbers - both the calling number and the receiving number. Even though a person placed the telephone call to another person, from the telephone transaction alone, no common data can be attributed to the persons. Often, the relating fact (the telephone call) is known before any other information about the related entities (the two people involved in the telephone call) is known. Since these relating facts cannot be attributed to a person, they must be represented as separate entities that are not people, but relate to people. However, impersonal awareness recognizes that a relationship between two persons exists as a consequence of the phone call.

UMF includes an entity type functionality, which allows you to define relating facts as entity types. Using this functionality, relating facts become separate entities in the entity database and can be used to detect relationships between Person entities. By configuring new entity types, specifying the appropriate entity type in UMF, and creating new resolution configurations, these relating facts may be used to automatically find impersonal relationships and conflicts between entities.

Entities of differing entity types never cross-resolve, even if the resolution rules allow it, and even if the data supports the resolution. This means that an entity type of Phone call never resolves to an entity type of Person .

The Analyst Toolkit graphs and reports impersonal relationships and any associated alerts, just as it does personal relationships and associated alerts.

Impersonal awareness example

For example, if you wanted to find impersonal relationships using telephone calls, you would create a new entity type of Phone call and adjust your acquisition node to correctly tag each telephone call record with the *Phone call* entity type.

When the telephone records are ingested into the system, standard entity and relationship resolution detects a one degree relationship between the Phone call entity and the calling entity (Person). It also finds a one degree relationship between the person called and the Phone call entity. By itself, the system does not detect a relationship between the persons.

However, when Degrees of Separation is configured, it continues the analysis and detects the two degree impersonal relationship between the caller and person called. An impersonal relationship exists, based on the telephone numbers that are attributes of the Phone call entity type. Degrees of Separation then analyzes the impersonal relationship and generates an alert if a conflict is found.

Viewing entity types

Entity types are user-defined traits or properties that are associated with an entity to identify the exact nature of the entity. You may want to view existing entity types if you are thinking of adding a new one.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Entity Types** tab.
4. Select the entity type you want to view.

Creating entity types

Entity types are user-defined traits or properties that are associated with an entity to identify the exact nature of the entity. You may want to add a new entity type in the system if you will be adding a new type of data to your system.

Before you begin

Before creating a new entity type, review the incoming identity data determine if it can be accurately described using any existing entity types.

About this task

Impersonal awareness uses entity types to link entities that would otherwise not have a 1-degree relationship.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Entity Types** tab.
4. Click the **New** button.
5. In the `General` tab, specify the id, type, description, entity resolution configuration, generic contributor, role alert contributor, search type, and allow resolve for this entity type.
6. Click the **Save** button.

Results

The system can now assign entity types to data and use impersonal awareness to link entities that would otherwise not have a 1-degree relationship.

Deleting entity types

You may delete an existing entity type when it is no longer used by the entity database.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Entity Types** tab.
4. Select the check box next to the characteristic type you want to delete.
5. Click the **Delete** button.

Help Topics

Entity Types - General tab:

Use the **Entity Types** tab to specify the details of the entity type.

ID Type the ID number of the entity type you want to create.

The ID is an auto-incrementing numeric code. While the product provides the next available number in sequence, you can set the code to be any unique numeric value by typing that value in the ID field.

Type Type the name of the entity type you want to create.

For example, an entity type of Phone call would be used to describe entities that are actual records of phone calls between two identities.

Description
Type the description of the entity type you want to create.

Entity Resolution Configuration
From the drop-down list, select the resolution configuration this entity type will use when loading.

Resolution configurations are set on the **Setup** > **Resolution** > **Resolution Configs** screen.

Generic Contributor
From the drop-down list, select **Yes** to allow this entity type data to go generic. Otherwise, select **No**.

Role Alert Contributor
From the drop-down list, select **Yes** to allow this entity type data to go generate role alerts. Otherwise, select **No**.

Search Type
From the drop-down list, select **Yes** to allow this entity type data to be used for searching. Otherwise, select **No**.

Allow Resolve
From the drop-down list, select **Yes** to allow this entity type data to be used for resolving entities. Otherwise, select **No**.

Overview of Degrees of Separation

The Degrees of Separation feature extends the relationship matching capabilities of IBM Relationship Resolution.

The default behavior of IBM InfoSphere Identity Insight identifies high-interest relationships and matches entities at one degree of separation from an inbound identity resolved to an entity. The enablement of the Degrees of Separation feature extends these capabilities to an almost limitless range of user-defined degrees of separation from an inbound identity resolved to an entity.

The Degrees of Separation feature uses separation configurations, roles, role alert rules, and relationship scores, to make real-time link analysis against very large data sets.

When an inbound identity is resolved to an entity, an entity graph is created using the one degree relationships that IBM InfoSphere Identity Insight detects. The entity graph uses the one degree relationships to build multi-degree relationship chains stemming from the entity the inbound identity was resolved to. A role alert chain can then be created by linking two multi-degree relationship chains, each stemming from the entity the inbound identity was resolved to. The role alert chain can then be used to find a relationship between the entities at the end and inclusive of each multi-degree relationship chain.

Degrees of Separation reduces work by evaluating all paths that connect two entities and using the strongest path strength in reporting relationships. Degrees of Separation can be configured to report one role alert for each configured role alert rule per entity the inbound identity was resolved to.

The Degrees of Separation configuration can be set in the Console by using the **System Configuration** tab, Degrees of Separation value.

Degrees of separation example

This example walks you through one relationship path and shows how the degrees of separation configuration factors into determining role alerts.

Degrees of separation example

After processing incoming data, Identity Insight reports the following relationship path:

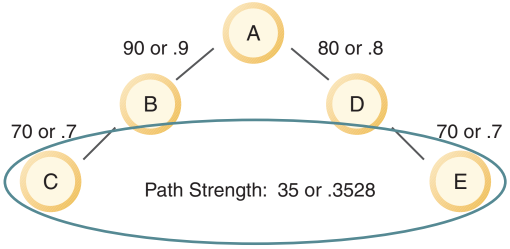
- Entity A knows entity B.
- Entity B knows entity C.
- Entity A knows entity D.
- Entity D knows entity E.

A *relationship path* is the chain of entities and attributes that link one entity to another entity.

As part of the relationship and role alert processing, Identity Insight determines the strength of the relationship path. The path strength is the product of the relationship score decimal conversions of every entity in the chain, converted to an integer.

Using our example, the product calculates the relationship scores and converts the scores into decimals:

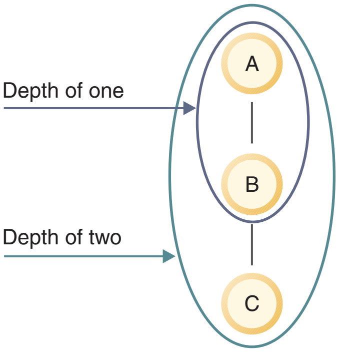
- The relationship score for entity A knows entity B is 90. 90 is converted to the decimal 0.9
- The relationship score for entity B knows entity C is 70. 70 is converted to the decimal 0.7
- The relationship score for entity A knows entity D is 80. 80 is converted to the decimal 0.8
- The relationship score for entity D knows entity E is 70. 70 is converted to the decimal 0.7



The relationship scores in the relationship path are multiplied. The result of the calculation is a relationship path strength of .3528, which is converted to the integer 35.

The product then compares the calculated path strength against the configured **`path strength threshold`** degrees of separation parameter. If the relationship path strength meets or exceeds the configured path threshold, the product generated role alerts. If the relationship path strength is below the configured path strength threshold, the product does not generate role alerts.

The product then uses the configured **`max depth`** degrees of separation parameter to calculate the degrees of separation between the entities in the relationship chain. The max depth setting determines the maximum number of degrees of separation in a multi-degree relationship path that can be considered as part of role alert detection.



Typically, the **`max depth`** parameter is set to two.

In this example, the **`max depth`** parameter is set to 6. Entity C and entity E have conflicting roles and are separated by 6-degrees, so a role alert is generated.

Viewing separation configurations

Because the product allows multiple separation configurations, use these instructions to view the settings for a specific separation configuration.

Procedure

1. In the Configuration Console, click **Setup** > **Relationships** > **Separation Configuration** .
2. Select the separation configuration.

Creating new separation configurations

You define separation configurations to determine whether relationship resolution detects one, two, or multiple degrees of separation between entities.

Procedure

1. In the Configuration Console, click **Setup** > **Relationships** > **Separation Configuration**.
2. Click **New**.
3. On the **General** tab, specify the settings for this separation configuration.
4. Click **Save**.

Editing separation configurations

Edit a separation configuration to change the settings that determine how many degrees can separate two entities and still be considered a relationship.

Procedure

1. In the Configuration Console, click **Setup** > **Relationships** > **Separation Configuration**.
2. Select the **separation configuration** to edit and make your changes.
3. Click **Save**.

Help topics

Separation Configuration - General tab:

Use the **General** tab to specify the details of the separation configuration.

ID Type the unique integer to identify the separation configuration.

The ID value is automatically populated with the next sequential number not in use.

Code Type a unique value to identify this role.

Description
Type a description for this separation configuration.

Max depth
The maximum number of degrees of separation of one multi-degree relationship chain in an entity graph considered for role alert detection.

Path strength threshold
The calculated **`path strength threshold`** of a role alert chain. A role alert chain whose path strength is below this threshold will not generate role alerts.

The path strength is the product, converted to an integer, of the relationship score decimal conversions of every entity in the role alert chain. The default setting for this parameter is 15.

Degrees of separation evaluates all paths that connect two entities and will use the strongest path strength in reporting relationships.

Configuring UMF documents

To successfully use Unified Messaging Format (UMF) documents, they must be known and configured.

Viewing the default UMF input documents

UMF input documents are the collection of UMF segments that structure the incoming data to load, modify, or query data in the entity database.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **UMF** button.
3. Click the **Input Documents** tab.

Configuring output documents

You must configure the enabled status of an output document format code if it is used.

About this task

UMF output documents format UMF result data.

Procedure

1. Click the **Setup** button.
2. Click the **UMF** button.
3. Click the **Output Documents** tab.
4. Click any link in the row that contains the UMF output document format code you want to edit.
5. From the **Enabled** drop-down list, select the appropriate status of the UMF output document format code.
6. Click the **Save** button.

Configuring the data source

You must configure a data source when there is a new data source you want to load into the entity database.

Before you begin

In order to configure a data source, you must first set up roles.

About this task

Data Sources can be viewed and modified by using the Console, on the **Data Sources**tab.

Data sources

Data sources contain the identities that you want to process for entity resolution and load into the entity database. Data sources contain identifying data (unique, personal identifiers for an identity) and non-identifying data (other attributes and data points for an identity). The identity records in the data source must be exported as Universal Message Format (UMF) before they can be processed by the system or loaded into the entity database. Examples of data sources include, but are not limited to, employee lists, watch lists, customer lists, and vendor lists.

Data sources contain vital information, such as the information about the original source (because the original data was transformed into UMF) or the external reference for the data source. These details make each data source unique in the system.

During entity resolution, if two entities are unresolved, the system uses the data source information to determine which information belongs with which entity.

Data source locations and source systems

You can organize incoming data sources by creating source locations and source systems and associating them with your data sources. You can use source locations and source systems to distinguish among similar types of data sources.

For example, if you are processing reservation data and human resource data from more than one location, you can use a data source location to distinguish which location is contributing the data:

- Property X Reservation data

- Property X Human Resource data
- Property Y Reservation data
- Property Y Human Resource data

Configurations by data source

To maximize the results of entity resolution and relationship detection, configure each data source using these settings:

Roles Because data sources are groupings of the same type of data, you can automatically assign the same role to every identity record in the same incoming data source. For example, by associating the Employee role to a human resources data source, all incoming records from the employee list are automatically assigned the Employee role.

Load levels
You can determine whether to load all the data in an incoming data source or only the data that resolves or relates to one or more entities.

Relationship resolution settings
You can configure the level of relationship detection by data source. For example, you can turn off relationship resolution for a data source or select the number of degrees of separation for detecting relationships within that particular data source.

Viewing data sources

A data source contains the data loaded into the entity database. You may want to view existing data sources if you are planning on adding a new data source.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Data Sources** tab.
4. Select the data source you want to view.

Configuring a data source

To successfully load data into the entity database, you must configure the system to recognize each data source.

Before you begin

Before you can load data in to the system, the data source must use the UMF standard.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Data Sources** tab.
4. Click the **New** button.
5. In the **General** tab, specify the ID, description, and other configuration information for the data source.
6. Click the **`Entity Resolution`** tab.
7. In the **`Entity Resolution`** tab, specify the resolution configuration information for the data source.

8. Click the **Relationships** tab.
9. In the **Relationships** tab, specify the relationship configuration information for the data source.
10. Click the **Save** button.

Configuring the Name Manager name match level

You configure the Name Manager match level by data source, because your name data can vary by source. The match level you select is a comparison parameter that determines how strict the matching is for incoming names from this data source.

Procedure

1. In the Configuration Console, select **Setup** > **Sources** > **Data Sources**.
2. Select the data source.
3. Click **Entity Resolution**.
4. In **Name Manager Match Level**, select the match level. For most situations, use the **Default** value, which is strict enough to yield good name matches.

Configuring data sources for enhanced name hashing

If you use enhanced name hashing, you must configure each data source to permit name attribute candidate list building, by setting the candidate builder configuration to the **Default w/ Name Only** candidate builder.

Deleting data sources

A data source contains the data loaded into the entity database. You may want to delete an existing data source if the data source no longer exists, or is no longer relevant to the entity database.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Data Sources** tab.
4. Select the check box next to the data source you want to delete.
5. Click the **Delete** button.

Creating a data source location

To select a location to classify a data source, that location must be configured in the system.

About this task

Data source locations are created using the configuration console. This is an optional primarily used if your data source gathers data from multiple physical locations. For example, a hotel system database that gathers data from multiple physical hotel locations.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **General** button.
3. Click the **Locations** tab.
4. Click the **New** button.

5. In the **General** tab, specify the location code, location name, district, company, latitude, longitude, status, and other configuration information for the data source location.
6. Click the **Save** button.

What to do next

You can now apply the newly configured location to data sources within the system.

Help topics

Data Sources - Entity Resolution tab:

Use the **Entity Resolution** tab to specify the entity resolution details of the data source.

Entity Resolution Configuration
From the list, select the resolution configuration this data source uses when loading data.

Candidate Builder Configuration
From the list, select the appropriate candidate builder configuration used during entity resolution processing when loading data from for this data source.

Default
Select this setting to use the default candidate builder configuration.

Default w/ Name Only
Select this setting to use the default candidate builder configuration with the addition of name only matching.

To use the Name Hasher to process name data for this data source, select this candidate builder configuration. (Make sure that the system parameters for the Name Hasher are set.)

Characteristic Confirmation
From the list, select **Yes** to specify that characteristic confirmations are processed when loading data from this data source. Otherwise, select **No**.

Perform Detach
This setting is typically used only for hotel systems.

From the list, select **Yes** to specify that the pipeline can match data without the data source account. If the match is not successful, the delete date is set on prior data. Otherwise, select **No**.

Name Manager Match Level
From the list, select the value for the comparison level to use when scoring incoming name data from this data source.

Default
Select this value to use the most common name match comparison level.

Loose Select this value when you want to produce more name matches from this data source. This value loosens the name comparison match level, so that the comparison is less strict than the Default value.

Tight Select this value when you want to produce fewer name matches from this data source. This value tightens the name comparison match level, so that the comparison is more strict than the Default value.

Allow Unresolve
The unresolve feature is the process of separating resolved identities into two separate entities, based on new information from incoming data. From the list, make the appropriate selection for this data source:

- Select **Yes** to permit entity resolution to separate identities into separate entities, if warranted, when loading accounts for this data source.
- Select **No** to prevent entity resolution from separating identities into separate entities when loading accounts for this data source.

Data Sources - General tab:

Use the **General** tab to specify the details of the data source.

ID Type the ID number of the data source you want to create.

The ID is an auto-incrementing numeric code. While the product provides the next available number in sequence, you can set the code to be any unique numeric value by typing that value in the ID field.

Code Type the code of the data source you want to create.

This is the value of the `DSRC_CODE` UMF tag. The data source code value can be alphanumeric and is used to further identify a data source. This value must be unique and cannot be changed once the record has been saved.

Description
Type the description of the data source you want to create.

Location
From the drop-down list, select the location code for the data source you want to create.

This field is for reference use only.

Source System
From the drop-down list, select the source system code for the data source you want to create.

This field is for reference use only.

Status From the drop-down list, select **Active** to specify that this data source is active. Otherwise, select **Inactive**.

Trust Action
From the drop-down list, select **Yes** to specify that you can rely on the accuracy of the `ACTION` UMF tag from your data source. Otherwise, select **No** to determine the action by examining the entity database. Selecting **No** will introduce a performance hit.

For Searching
From the drop-down list, select **Yes** to specify that this data source is used for loading searches. Otherwise, select **No**.

Transliterate
From the drop-down list, select **Yes** to specify that transliteration should occur for this data source. This allows support for the Latin 1 character set. Otherwise, select **No**.

Note: If you enable the transliterate setting for any data source, you must also enable the transliterate configuration setting for data source ID 1589 (Search). The 1589 data source is used by the product to input searches to the pipeline, and by default, assumes ASCII character inputs. Enabling this configuration ensures that names that are part of a search are also properly transliterated to provide the most accurate search results.

Data Sources - Relationships tab:

Use the **Relationships** tab to specify the relationship details of the data source.

Role Select the role code to assign this data source.

Data Source Class
Select the appropriate data source class for this data source.

Full Load
Select this field type to load the data into the database.

This setting resolves any identities that can be resolved, updates the entity, detects any possible relationships, and generates the user-defined role alerts.

Fully Passive
Select this field type to not load the data into the database.

If you load fully passive, then no data is stored. Visualizer cannot display the alert.

Load if Resolve/Relate
Select this field type to load the data into the database if they resolve or relate to existing records in the entity database.

This setting resolves any identities that can be resolved, updates the entity, detects any possible relationships, and generates the user-defined role alerts.

Load if Selective Resolve/Relate
Select this field type to load the data into the database if they resolve or relate to existing records in the entity database, only if this data source is configured in the SELECTIVE_PASSIVE_CONFIG table.

This setting resolves any identities that can be resolved, updates the entity, detects any possible relationships, and generates the user-defined role alerts.

Load if Selective Resolve
Select this field type to load the data into the database if they relate to existing records in the entity database, only if this data source is configured in the SELECTIVE_PASSIVE_CONFIG table.

This setting will also resolve any identities that can be resolved, update the entity, detect any possible relationships, and generate the user-defined role alerts.

Separation Level
From the drop-down list, select the appropriate separation level for this data source.

Load Data
Always select this field type. It is currently the only option.

DoS Configuration
From the drop-down list, select the appropriate degrees of separation configuration for this data source.

The separation configurations are set on the **Setup** > **Relationships** > **Separation Config** screen.

Locations - General tab:

Use the **Locations** tab to specify the details of the data source location.

Location Code
Type the location code to assign to this data source location.

An alphanumeric value that cannot be changed once the record is saved.

This value is required.

Location Name
Type the location name to assign to this data source location.

District
Type the district to assign to this data source location.

This value is required.

Company
Type the company name to assign to this data source location.

Latitude
Type the latitude of this data source location in the following format:

DD:MM:SS

Longitude
Type the longitude of this data source location in the following format:

DD:MM:SS

Status From the drop-down list, select **Active** to specify that this data source location is active. Otherwise, select **Inactive**.

Turning off relationship detection

If your business requirements specify that you only need to know who is who and not who knows who, you can reduce the amount of processing required for each new record and speed overall system performance by configuring relationship resolution to only perform entity resolution and not detect relationships between entities.

Before you begin

Make sure you selected **Edit Configuration** when you logged in to the current session of the Configuration Console.

Procedure

1. Turn off role assignments for each data source.
   a. Click **Setup**.
   b. Click **Sources**.
   c. On the **Data Sources** tab, click the data source you want to edit.
   d. Click the **Relationships** tab.
   e. From the **Role** drop-down list, choose **— Select One —**.

   f. From the **Separation Level** drop-down list, choose **Alerts Only**.
   g. Click **Save**.
2. Disable the Default Role Assignments data quality management rule.
   a. Click **Setup**.
   b. Click **UMF**.
   c. On the **DQM Rules** tab, from the **Segment** drop-down list, choose **ROOT**.
   d. Click any link in the row that contains the DQM 551, Default Role Assignment function.
   e. On the **General** tab, from the **Status** drop-down list, choose **Inactive**.
   f. Click **Save**.
3. Delete all resolution rules that are not set to resolve entities.
   a. Click **Setup**.
   b. Click **Resolution**.
   c. Click the **Resolution Rules** tab.
   d. From the **Resolution Config** drop-down list, choose **DEFAULT**.
   e. Click the check box next to any resolution rule that has a **No** value displayed in the **Triggers Resolve** column.
   f. Click **Delete**.
   g. Click **OK** to confirm that you want to delete the selected resolution rules.
4. Finally, delete all conflict rules.
   a. Click **Setup**.
   b. Click **Relationships**.
   c. Click the **Conflict Rules** tab.
   d. Click the check box next to each conflict rule.
   e. Click **Delete**.
   f. Click **OK** to confirm that you want to delete the selected conflict rules.

What to do next

Now, the system is configured to resolve entities without detecting relationships.

Configuring event types

You configure event types to define and categorize events that are processed by Event Manager. However, before the system processes incoming data containing event types, you must enable event processing in the Event Manager system parameters, configure the business rules in the Eclipse-based complex event processor tool, and format the incoming event data using the UMF EVENT data segment definitions.

Event Types can be viewed and modified by using the Console, on the **Event Types**tab.

Event types

Event types categorize events and define the unit of measure for the value associated with events in Event Manager. Examples of event types include wire transfer, account opening, or credit card transaction.

Event types are required for event processing, because the user-defined business rules that the event processor uses call a specific event type. If the event type does not exist, the event processor cannot process the event.

Creating event types

When you want to add a new event scenario to process events, you may need to create a new event type to define the types of transactions or activities that are included in that event scenario, as well as the unit of measure associated with this category of event.

Before you begin

Event Manager must be enabled for your IBM InfoSphere Identity Insight system.

About this task

Event types are called by the complex event processor, while it processes events according to the user-defined business rules. Before an event type is used, you must also create at least one business rule that uses that event type.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Event Type** button.
4. Click the **New** button.
5. Required: On the **General** tab, specify the name and description of the event type, the unit of measure associated with this event type, and the status of this event type (active or inactive).
6. Optional: You can also specify additional information, such as category, sub-category, and notes about this event type.
7. Click the **Save** button.

Editing event types

Edit an event type when you want to change the description, the unit of measure, or the additional information associated with the event type. You can also edit an event type to inactivate it, so that it can no longer be used. You cannot edit the event type name.

About this task

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Event Type** button.
4. Select the event type you want to edit.
5. On the **General** tab, make your changes.
6. Click the **Save** button.

What to do next

Deleting event types

You may want to delete an event type when it is no longer used for event processing. If you want to keep the event type, but simply inactivate it, you can edit the event type status instead of deleting it.

Before you begin

About this task

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Sources** button.
3. Click the **Event Type** button.
4. Select the check box next to the event type that you want to delete.
5. Click the **Delete** button.

What to do next

Help topics

Event Types - General tab:

Use this tab to define or edit an event type. Event types define and categorize events, and are used during event processing, if Event Manager is enabled for your system.

Type Enter a unique name for this event type. For example, you might create an event type named `Wire Transfer`.

Description
Enter a description of the event type.

Unit of Measure
Enter an abbreviation for the unit of measure for the value that is associated with the event type. For example, you might enter `USD` for U.S. dollars.

Status From the drop-down list, select the status for the event type, either **Active** or **Inactive**. (You can use the **Inactive** status to remove the event type from event processing, but keep the configuration for the event type.)

Category
Enter an optional category name for the event type.

Sub-Category
Enter an optional sub-category name for the event type.

Memo Heading 1
Enter an optional memo 1 heading for the event type.

Memo Heading 2
Enter an optional memo 2 heading for the event type.

Configuring entity resolution

Entity resolution is the process that finds relationships in data. Entity resolution configuration settings are organized by groupings called resolution configurations. Five components make up a resolution configuration: resolution rules, confirmations & denials, attributes, Name Manager match configurations, and the candidate builder.

Entity resolution

Entity resolution is the process that resolves entities and detects relationships. The pipelines perform entity resolution as they process incoming identity records in three phases: recognize, resolve, and relate.

Configuring resolution configurations

All entity resolution settings are maintained in a resolution configuration, two of which are provided by default.

Resolution configurations

Entity resolution settings are organized by a group of resolution configurations that are defined using the Configuration Console's **System Configuration** tab System load resolution rule value.

The default installation of Relationship Resolution includes two resolution configurations

- **DEFAULT** - default resolution settings used whenever new data comes into the system from a defined data source.
- **SEARCH** - resolution settings used by the resolved search process whenever a user submits a fully-resolved search request.

You can create your own set of resolution settings and identify them using a newly created resolution configuration. This process should start by cloning the **DEFAULT** resolution configuration and using it as a starting point for your new resolution configuration.

Different resolution configurations can be assigned to specific data sources. If you choose to apply multiple resolution configurations across multiple data sources, you must consider that entity resolution always uses the resolution configuration assigned to the incoming identity when generating alerts. This can result in different alert results based on which of the compared identities is the incoming identity and which of the identities already exists in the entity database. For example, Identity #123, from the Customer data source is assigned the DEFAULT resolution configuration which contains a resolution rule for name and address with a name threshold of 80 and an address threshold of 5. Identity #456 from the Vendor datasource uses the NEW resolution configuration that has the same resolution rule but the name threshold is set at 95 and the address threshold is set at 7. When Customer 123 is the incoming identity and is compared to existing Vendor 456, the name score between them is calculated at 85 and the address score is 5, which results in an alert. If the order of processing is reversed with Customer 123 already in the system and Vendor 456 coming into the system, they still generate the same resolution scores of 85 for name and 5 for address. However, in this case there will be no alert because the resolution scores do not meet the resolution thresholds in the NEW resolution configuration where name is set at 95 and address is set at 7.

Note:

Using a resolution configuration other than the default entity resolution configuration should be done with care and planning. The default entity resolution settings, such as resolution rules and scoring settings, are the result of hundreds of person-years of analysis and study of real-world data. Changes to these defaults are typically only necessary when data or business rules demand specific and non-standard behaviors from the system.

Viewing resolution configurations

A resolution configuration is used to specify a collection of entity resolution settings. You might want to view existing resolution configurations when you plan to make changes to your entity resolution settings or if you want to create a new set of entity resolution settings.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Resolution** button.

Cloning and customizing the default resolution configuration

The ideal way to create a new entity resolution configuration is to clone (make a copy of) the default resolution configuration and use it as a starting point for the new resolution configuration. By maintaining the default configuration in an unmodified state, you can always return to the default configuration if you need to, without having to reinstall the product.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Resolution** button.
3. On the **Resolution Configs** tab, select the check box next to the DEFAULT resolution configuration.
4. Click the **Clone** button.
5. On the **General** tab, in the **Code** field, type the new name for your resolution configuration.
6. In the **Description** field, type a new description of the cloned Resolution Configuration.
7. Click the **Save** button.

What to do next

When you make changes to the entity resolution settings, such as configuring resolution rules, confirmations and denials, or candidate builder, you can select your new resolution configuration.

Deleting customized resolution configurations

I you are no longer using a customized resolution configuration, you can delete it. Do not delete the DEFAULT resolution configuration; by maintaining the default configuration in an unmodified state, you can always return to the default configuration if you need to, without having to reinstall the product.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Resolution** button.
3. On the **Resolution Configs** tab, select the check box next to the resolution configuration that you want to delete.
4. Click the **Delete** button.
5. In the confirmation window, click **OK** to delete the resolution configuration.

What to do next

You can no longer select this resolution configuration when making changes to the entity resolution settings. Also, the entity resolution settings related to this resolution configuration can no longer be applied to the entity resolution process.

Help topics

Resolution Configs window:

Use this window to view a list of available entity resolution configurations. Entity resolution settings are organized in to groups called resolution configurations. Different resolution configurations can be assigned to individual data sources. Each data source can only have one resolution configuration applied to it at a time.

Code Name of the resolution configuration.

Description
Description of the resolution configuration.

Configuring resolution rules

To define how compared entities resolve and relate, you must configure resolution rules, including candidate thresholds and confirm/deny thresholds.

About this task

Resolution Rules can be viewed and modified by using the Console, on the **Resolution Rules** tab.

Resolution rules

Resolution rules are a set of criteria that the system uses to define how compared entities resolve (if they are or are not the same entity) and relate (if entities are not resolved to the same entity, how many attributes they share).

When defining resolution rules, you must specify thresholds that contribute to the total resolution score, which determines whether an incoming identity resolves into an existing entity:

- Candidate thresholds specify which attribute data values are compared to determine whether an identity and an entity will be resolved into one composite entity. The threshold is the minimum score at which a particular attribute value must match between the incoming identity and an existing entity to satisfy the resolution rule.
- Confirm/deny thresholds specify how much scoring weight (positive or negative) is given to matching or conflicting attribute data values when you enable the use of denials.

You can also specify how conflicting values for the same attributes affect the resolution score. These conflicting values are called denials. You can configure resolution rules that specify that the rule is not met if there are any conflicts (denials) in the attribute values. You can also adjust the thresholds for a resolution rule to create automatic denials, based on the comparison scores not meeting one or more specified threshold scores. The higher a threshold score is set, the more exact the match must be in order to satisfy the resolution rule.

Candidate Thresholds

Candidate thresholds are the first parts of a resolution rule used to determine if an incoming identity actually represents an existing entity or represents an entirely new entity.

Candidate thresholds are configured using the Console and are an integral part of a resolution rule. For example, if a resolution rule has a unique number candidate threshold, that resolution rule can be described as requiring a matching unique number.

Candidate thresholds are only applied to existing entities in order to place that entity on the candidate list as part of the entity resolution process. The actual threshold is the minimum level at which a particular data type must match between an incoming identity and an existing entity for the entity resolution process to add the existing entity to the candidate list.

Address precision:

Address precision is the scoring process used by entity resolution to determine whether two compared addresses represent the same address.

Address precision has been divided into nine distinct levels (1-9). Most addresses contain fundamental components that can be compared, like street (including number), city, state, postal code, postal+4. When comparing these components, address precision starts with a matching street component and assigns a precision level of 5. That precision level is then adjusted up or down based on whether additional components match or differ. Each matching component increases the precision level by 1 and each differing component decreases the precision level by 1. If a component value is present in one address but no value is present for the same component in the other address, no precision adjustment occurs.

By default, entity resolution considers all compared addresses with an address precision level of five or greater as candidates for matching addresses.

Table 19. Address Precision Levels

Level	Description
1	Street match with all parts, postal+4 different. This means that there must be an address that matches all parts but the postal +4 is different. For example, 123 N Water St. Las Vegas, NV 89123-1234 and 123 S Water St. Las Vegas, NV 89123-5433.
2	Street match with all parts differing. This means that only the street address matches and City, State, Postal, Country are all different or missing. For example, 123 Main St. Orlando, FL 32555 and 123 Main St. Las Vegas, NV
3	Street match with -2 difference modifier. This means that the street address matches, but the calculation added up to a -2. For example, 123 Main St. Las Vegas, NV 89111 and 123 Main St. Las Cruces, NM.
4	Street match with -1 difference modifier. This means that the street address matches, but the calculation added up to a -1. For example, 123 Main St. Las Vegas, NV 89111 and 123 Main St. Las Vegas, NM 54633.
5	Street match with 0 modifier (baseline). This means that the street address matches, but the calculation added up to a 0. For example, 123 Main St. Las Vegas, NV 89111 and 123 Main St.

Table 19. Address Precision Levels (continued)

6	Street match with +1 matching modifier. This means that the street address matches, but the calculation added up to a +1. 123 Main St. Las Vegas, NV 89111 and 123 Main St. Las Vegas
7	Street match with +2 matching modifier. This means that the street address matches, but the calculation added up to a +2. For example, 123 Main St. Las Vegas, NV 89111 and 123 Main St. Las Vegas, NV.
8	Street match with all parts, postal +4 missing. This means that all parts of the address match except that the postal +4 is not present. For example, 123 Main St. Las Vegas, NV 89111 and 123 Main St. Las Vegas, NV 89111
9	Exact match (street with all parts). This selection means that all parts of the address match including the postal +4. For example, 123 Main St. Las Vegas, NV 89111-1234 and 123 Main St. Las Vegas, NV 89111-1234 **Note:** This does not work on international postal codes where postal +4 are not used.

Precision level 1

Each of the precision levels, from one through nine, represents an increasing level of precision with the exception of level 1. Level 1 represents a special case where address information might be the same with the exception of a North/South or East/West street designation, such as 456 North Main Street Sometown, Nevada and 456 South Main Street Sometown, Nevada. In this case the addresses might be the same but the postal+4 is definitely different. On the surface, these addresses might seem to require a resolution. However, they should not be resolved to each other because they are, in fact, different addresses. Because this seemingly strong case for address resolution is in fact a strong case for not resolving the addresses to each other, the value assigned to the precision level of this scenario is at the bottom of the scale (level one) to prevent the addresses from being resolved.

Level 1 might also indicate an intentional address error. Some customers take an interest in intentional patterns of address errors - people who deliberately alter an address in order to deceive. For that reason, resolution rules' order can be configured to consider a low address precision level such as level 1.

Note: If level 1 is of interest for resolving entities - for example, if you want to know if someone gives conflicting address information at the postal+4 level - you must create a separate resolution rule. That rule must precede the default resolution rule that considers all precision levels of five and greater. Because of the complexities involved with properly creating new resolution rules, you should only do so with sufficient expertise or with the help of IBM.

Address precision detailed examples:

The following examples represent the data as it is compared along with the resulting address precision scores.

The first address represents the existing address in the entity database, and the second address is the incoming address.

Precision Level 1 - Street match with all parts, postal+4 different.

This case shows two addresses that are on the same street, but are distinct addresses. One address in on the north end of the street, and the other is on the south end. The only differences between these two addresses are the Zip+4 values.

STREET	CITY	STATE	POSTAL
123 N Main St	Fairmount	IN	46928-1655
123 S Main St	Fairmount	IN	46928-1924

Note: Precision level 1 represents a special case where address information might be the same with the exception of a North/South or East/West street designation. On the surface, these addresses might seem to require a resolution. However, they must not be resolved to each other because they are in fact different addresses. Because this seemingly strong case for address resolution is in fact a strong case for not resolving the addresses to each other, the value of this scenario is placed at the bottom of the scale (1) to prevent the addresses from being resolved.

Precision Level 2 - Street match with all parts differing.

This example shows two addresses with the same street information but different city, state and postal information. The second address is obviously a mistake (perhaps intentional) because postal codes in Nevada all start with 89.

STREET	CITY	STATE	POSTAL
123 E Main St	Fairmount	IN	46928
123 S Main St	Las Vegas	NV	46999

Precision Level 3 - Street match with -2 difference modifier.

In this example, only the street information matches. No state information is provided in the incoming address, and the city and postal information are conflicting.

STREET	CITY	STATE	POSTAL
123 E Main St	Delphi	IN	46923-1522
123 E Main St	Fairmount		46928

Precision Level 4 - Street match with -1 difference modifier.

This example shows two addresses with the same street and state information but conflicting city and postal information.

STREET	CITY	STATE	POSTAL
123 E Main St	Delphi	IN	46923-1522
123 E Main St	Fairmount	IN	46928-1924

Precision Level 5 - Street match with 0 modifier (baseline)

In this example, only the street information is provided in the incoming address. Even though it does not contain city, state, or postal information, the match receives the baseline address precision score (5). The precision score reflects the missing parts (not to be confused with conflicting parts as missing parts are not scored).

STREET	CITY	STATE	POSTAL
220 JEFFERSON	BUFFALO	IA	
220 Jefferson St.			

Precision Level 6 - Street match with +1 matching modifier.

This example shows an incoming street address without state or postal information but the same street and city information. The incoming address is likely the correct address but is missing data.

STREET	CITY	STATE	POSTAL
220 Washington	Syracuse	NY	
220 Washington Sq.	Syracuse		

Precision Level 7 - Street match with +2 matching modifier

This example shows matching street, city, and simple postal information, but no state information is provided in the incoming address.

STREET	CITY	STATE	POSTAL
220 JEFFERSON	BUFFALO	IA	52728
220 Jefferson St.	Buffalo		52728

Precision Level 8 - Street match with all parts, postal +4 missing

Here two addresses are the same, but address hygiene was unable to validate the addresses so they did not receive a Zip+4.

STREET	CITY	STATE	POSTAL
220 JEFFERSON	BUFFALO	IA	52728
220 Jefferson St.	Buffalo	IA	52728

Precision Level 9 - Exact match (street with all parts). This selection means that all parts of the address match including the postal +4

In this example, two addresses share the same street address, city, state and Zip+4. As a result, the compared addresses receive the highest address precision score.

Note: This does not work on international postal codes where postal +4 are not used.

STREET	CITY	STATE	POSTAL
123 W Main St	Camden	IN	46917-9997
123 W Main	Camden	IN	46917-9997

Name precision:

Name precision is the scoring process used by entity resolution to determine whether two compared names represent the same name.

Name precision scoring is based on the use of one of two possible algorithms.

- Name Comparator 1.0
- Name Comparator 2.0

Each algorithm has its own set of name matching criteria that are available for configuration as part of configuring resolution rules.

Either of these algorithms works with the Name Manager feature. Name Manager is a separately configurable feature that extends name matching to include additional matching capabilities based on unique cultural considerations.

Comparison considerations

Name Comparator 1.0 is the default setting for upgraded installations from version 3.9.0 and earlier. Name Comparator 2.0 is the default setting for upgraded installations from version 3.9.1 and later and for new installations.

When considering which algorithm is best suited to your needs, consider the advantages each algorithm offers.

Name Comparator 1.0:

- Requires less CPU usage resulting in faster performance
- Allows more precise understanding of why names match

Name Comparator 2.0:

- Handles names consisting of more than three words better
- Matches out-of-order words better
- Performs fuzzy matching better
- Matches names of organizations better
- Handles initials better

Name Comparator 1.0:

This name matching algorithm is designed to work primarily with names consisting of two or three words. It is the default name matching setting for upgrades from version 3.9.0 and earlier.

Name Comparator 1.0 compares two names and then ranks their likeness according to 15 distinct levels of similarity.

Table 20. Name Comparator 1.0 - Levels of precision

Level	Description
1	Only First or Last Name Partial Match EXAMPLE: John Jacob Smith = Joe Smithson
2	Only First or Last Name Exact Match EXAMPLE: John Jacob Smith = Jonathan Henry Smith

Table 20. Name Comparator 1.0 - Levels of precision (continued)

3	Close Hash Match EXAMPLE: Joe Smith = Joe Snith
4	Only the Last Names are Different, but Out of Order EXAMPLE: Bob Jacob Smith = Jacob Bob Jones
5	Only the Last Name are Different EXAMPLE: Bob Jacob Smith = Bob Jacob Jones
6	Standardized Name Match with Some Differences EXAMPLE: John Jacob Smith = Jonathan Henry Smith
7	Standardized Names Match EXAMPLE: Joe W Anderson = Joseph Andersen
8	Standardized Match with Exact Last Names, Middle Initial Match, but Out of Order EXAMPLE: J Bob Smith = Robert J Smith
9	Standardized Match with Exact Last Names, Middle Initial Match EXAMPLE: Joe W Anderson = Joseph W Anderson
10	Standardized Match with Exact Last Names, but Out of Order EXAMPLE: Bob Smith = Robert Smith
11	Standardized Match with Exact Last Names EXAMPLE: John Jacob Smith = Johnny Jake Smith
12	Raw Names Match with middle Initial Match, but Out of Order EXAMPLE: Joe W. Brown = Will Joe Brown
13	Raw Names Match with Middle Initial Matchces EXAMPLE: Joe W Anderson = Joe W Anderson
14	Raw Names Match, but Out of Order EXAMPLE: John Bob Smith = Bob John Smith
15	Raw Names Match EXAMPLE: Joe William Anderson = Joe William Anderson

Name Comparator 2.0:

This name matching algorithm is designed to tokenize compared names – it breaks the group of words in the name string into individual names, or tokens. Then the algorithm compares the tokens and creates a score for each token. It is the default name matching setting for upgraded installations from version 3.9.1 and later and for new installations.

Name Comparator 2.0 groups names into three categories which it then compares and scores:

- Given name (first name and middle – or all words but the last name)
- Surname (last name)

- Full name (all words)

These three scoring categories allow you to tune name matching for specific resolution rules to meet your name matching requirements. Scores are integer-based, from 0-100, with 0 being the lowest score and 100 being the highest score. The higher the score in a category, the closer the names matched in that category.

Configuration considerations - scoring guidelines

Whenever you edit or change name matching settings for Name Comparator two, use these scoring guidelines to assist you in setting up the resolution rules name thresholds. These guidelines are also helpful when interpreting the scoring results of this algorithm's scoring categories.

Full Name Score

Based on the 0-100 score, here are guidelines to assist you in determining the level of matching the Full Name score:

- 100 = exact match
- 90 = very good match (suitable for name and DOB resolution)
- 80 = good match (suitable for most resolution rules)
- 70 = average match (suitable when unique numbers are also present)
- Below 70 = not suitable for matching

Given Name Score

Based on the 0-100 score, here are guidelines to assist you in determining the level of matching the Given Name score:

- 100 = exact match
- 90 = very good match (might indicate given-surname swap)
- 85 = minimum acceptable match
- Below 85 = not suitable for matching; might be useful when combined with Given Name or Full Name to ensure some similarity

Surname Score

Based on the 0-100 score, here are guidelines to assist you in determining the level of matching the Surname score:

- 100 = exact match
- 90 = very good match (might indicate given-surname swap)
- 85 = minimum acceptable match
- Below 85 = not suitable for matching; mighty be useful when combined with Given Name or Full Name to ensure some similarity

Name Manager name scoring:

The Name Manager algorithm scores incoming name data based on grouping the name into name parts and then determining the culture for each name part. The algorithm then scores each name part, and the resulting scores are used during entity resolution.

While the Name Manager algorithm is separate from the Name Comparator algorithms (NC1 and NC2), you must still select either NC1 or NC2. During the entity resolution process, names are first scored based on the selected Name Comparator algorithms. If the name scores an exact match, entity resolution skips the Name Manager scoring, because the exact name match satisfies the name score portion of the resolution rule. If the incoming name scores less than an exact match, however, the entity resolution process scores the name using the Name Manager algorithm.

First, the algorithm parses the name into name parts (given name, surname, and full name), and then the algorithm determines the culture for each name part. Finally, the algorithm assigns each name part a score, and compares the scores against the configured Name Manager score thresholds to determine how closely the names matched. The higher the score threshold is set, the closer the name parts from the incoming name data must match the name parts from the existing entity in the entity database.

Date of birth precision:

Date of birth precision is the scoring process used by entity resolution to determine whether two compared dates of birth represent the same date.

This comparison takes into account various measures of similarity of the date of birth strings including: integer positions, transpositions, and deltas of day, month, and year values. The measures are analyzed to determine an likeness score between 2 and 100. You can configure date of birth precision settings based on four categories of similarity:

- Exact - 100 point match
- Tight - >= 90 point match
- Medium - >= 85 point match
- Loose - >= 80 point match

Configuration considerations

The system provides a pre-configured setting of Tight as the minimum level of likeness for a resolution rule to consider two compared dates of birth to be the same. Changing this setting will affect the number of matches and can affect the number of entity resolutions performed by the system. Carefully consider changes to this setting and be sure to test any changes before implementing them in a production environment.

Date of birth precision detailed examples:
The following examples represent the data as it is compared along with the resulting date of birth precision scores. The first date of birth represents the existing date of birth for an entity in the entity database, and the second date of birth is for an incoming date of birth identity.

Precision level: Exact (100 point)

This case shows two exact dates. The algorithm will generate a 100 point match.

DATE OF BIRTH	STATUS
1963/12/01	Existing
1963/12/01	Incoming

Precision level: Tight (90 point)

This case shows two dates whose precision score is greater than or equal to 90 points. The example shows two date of birth values with the same year and day value, but the month value is different by one month.

DATE OF BIRTH	STATUS
1963/12/01	Existing
1963/11/01	Incoming

Precision level: Medium (85 point)

This case shows two dates whose precision score is greater than or equal to 85 points. The example shows two date of birth values with the same month and day value, but the last two digits of the year value are transposed.

DATE OF BIRTH	STATUS
1963/12/01	Existing
1936/12/01	Incoming

Precision level: Loose (80 point)

This case shows two dates whose precision score is greater than or equal to 80 points. The example shows two date of birth values with the same month and day value, and the third digit of the year value incorrect (which still is a rational value for a date of birth).

DATE OF BIRTH	STATUS
1963/12/01	Existing
1933/12/01	Incoming

Viewing resolution rules

Before adding or deleting resolution rules, you can view the current set of resolution rules.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Resolution** button.
3. Click the **Resolution Rules** tab.
4. From the **Resolution Config** drop-down list, choose a resolution configuration.
5. To view the details of specific resolution rules, click the link in the row that contains the resolution rule that you want to view.

Creating resolution rules

After carefully considering your business requirements and reviewing the existing resolution rules, you might decide to create new resolution rules for your data.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Resolution** button.

3. Click the **Resolution Rules** tab.
4. From the **Resolution Config** drop-down list, choose a resolution configuration.
5. Click the **New** button.
6. On the **General** tab, specify the values to use when comparing the data of two entities
7. Click the **Candidate Thresholds** tab.
8. On the **Candidate Thresholds** tab specify threshold values for the data.
9. Click the **Confirm/Deny Thresholds** tab.
10. On the **Confirm/Deny Thresholds** tab, specify threshold values for the data.
11. Click the **Save** button.

Deleting resolution rules

To remove a resolution rule from consideration during the entity resolution process, delete the rule.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Resolution** button.
3. Click the **Resolution Rules** tab.
4. From the **Resolution Config** drop-down list, choose a resolution configuration.
5. Select the check box next to the resolution rules you want to delete.
6. Click the **Delete** button.
7. In the confirmation window, click **OK** to delete the resolution configuration.

Help topics

Resolution Rules window:

Use this screen to view resolution rules contained within a resolution configuration. Resolution rules are processed in the order listed. After a resolution rule is satisfied, the assigned resolution scores are applied and if the rule is configured to trigger a resolution, then the incoming identity is resolved to the existing entity and no more entity resolution rules are considered for that particular comparison.

Order Order in which the resolution rules are applied to the compared incoming identity and existing entity

Description
Description of the resolution rule

Resolution Confidence
Resolution score applied to the comparison if the rule is satisfied

Relation Confidence
Relation score applied to the comparison if the rule is satisfied

Triggers resolve
Whether the rule automatically resolves the incoming identity to the existing entity if the rule is satisfied

Resolution Rules - General tab:

Use this tab to configure a new resolution rule or to see the details of an existing resolution rule.

Order Enter a unique number specifying the order in which to process the rule.

Description
Enter the description for the rule.

Resolution Confidence
Enter a likeness confidence percentage if this rule succeeds. Only 100% will be considered for resolution.

Relation Confidence
Enter a relation confidence percentage if this rule succeeds. Only 100% will be considered for resolution.

Triggers Resolve
Select "Yes" to resolve the incoming identity and existing entity if resolution and relation confidence are 100%.

Denials Enabled
Select "Yes" to enable processing of confirmations/denials. Otherwise no denial processing will occur.

Characteristic Denials Enabled
Select "Yes" to enable processing of characteristic confirmations/denials. Otherwise no characteristic denial processing will occur.

Resolution Rules - Candidate Thresholds tab:

Use this tab to specify the candidate threshold settings of new resolution rule or to see the candidate threshold details of an existing resolution rule. These settings define the resolution rule **description** entered on the Resolution **General** tab.

Address Precision Threshold
Select the minimum address ranking required for the rule to be considered satisfied.

Approximate Address Threshold
Select the minimum number of approximate address value matches required for the rule to be considered satisfied.

Proximity Threshold
Select the minimum number of addresses within the area defined in the quality rule required for the rule to be considered satisfied.

Unique Number Threshold
Select the minimum number of unique number matches for the rule to be considered satisfied.

Non-Unique Number Threshold
Select the minimum number of non-unique number matches for the rule to be considered satisfied.

Characteristic Threshold
Select the minimum number of characteristic matches for the rule to be considered satisfied.

Email Threshold
Select the minimum number of E-Mail matches for the rule to be considered satisfied.

Summary Data Threshold
Select the minimum sum of Unique Number, Other Number, Address, Characteristic and Email matches for the rule to be considered satisfied.

Summary Threshold
Select the minimum sum of Address Proximity, Approximate Address, Close Number and DOB matches for the rule to be considered satisfied.

Resolution Rules - Confirm/Deny Thresholds tab:

Use this tab to specify the confirmation and denial threshold settings of a new resolution rule or to see the confirmation and denial threshold details of an existing resolution rule.

Close Number Threshold
Select the minimum number of close number matches for the rule to be considered satisfied.

Date of Birth Threshold
Select the minimum date of birth match score for the rule to be considered satisfied.

Name Comparator settings
These settings determine the name precision requirements for entity resolution. These settings work by themselves or with the Name Manager settings.

Given Name Score Threshold
Enter the score threshold for the given name from 0 to 100.

Surname Score Threshold
Enter the score threshold for the surname from 0 to 100.

Full Name Score Threshold
Enter the score threshold for the full name from 0 to 100.

Name Manager settings
Name Manager extends the standard name precision to include important cultural considerations. These settings only apply if Name Manager is configured.

Given Name Score Threshold
Enter the minimum given name score for the rule to be considered satisfied.

The threshold must be an integer value between 0 to 100. The higher the score, the more exact the match. Typically, a score below 70 is not suitable for matching, but might be useful when combined with Surname or Full Name to ensure some similarity.

Surname Score Threshold
Enter the minimum surname score for the rule to be considered satisfied.

The threshold must be an integer value between 0 to 100. The higher the score, the more exact the match. Typically, a score below 70 is not suitable for matching, but might be useful when combined with Given Name or Full Name to ensure some similarity.

Full Name Score Threshold
Enter the minimum full name score for the rule to be considered satisfied.

The threshold must be an integer value between 0 to 100. The higher the score, the more exact the match. Typically, a score below 70 is not suitable for matching.

Customizing the candidate builder

You can change candidate builder settings by using candidate builder configurations. Changes to the candidate builder feature are made using the Configuration Console.

Candidate builder

The candidate builder feature defines criteria the system uses to add an existing entity to the candidate list as part of the entity resolution process.

Typical candidate builder settings include address, unique numbers, and other numbers. These are the data types that the system compares to determine which existing entities might resolve to an incoming identity. When a new identity record enters the system, if an existing entity has a matching value for any of the data types identified by the candidate builder, that entity is added to the candidate list.

Candidate builder configurations

Candidate builder settings are organized by groups called candidate builder configurations. Only one candidate builder configuration can be used within a resolution configuration.

Candidate builder configurations included with the product are:

- **Default** - this setting includes address, unique number, other number as criteria for including an entity in the candidate list.
- **Default with name only** - this setting includes names as a criterion for including an entity in the candidate list. This setting is designed to be used when your entity data might contain only names or names and very few other types of data.

Configuration considerations

Generics directly affect whether a value is considered as part of the Candidate builder process. After a value is considered a generic value, it is no longer used to generate candidate lists.

Candidate builder settings directly affect system performance. When the system uses index lookups to compare an incoming identity to each and every entity in the entity database, it is only comparing data types that are configured in the candidate builder feature. This allows candidate lists to be generated very quickly. As the entity database grows and includes more entities there is more for the candidate builder to compare. For example, if your entity database contains 100,000 entities and the candidate builder is set to compare three data types when creating the candidate list, then whenever a new identity enters the system, the system can make up to 300,000 comparisons just to generate the candidate list. If our entity database contains 1,000,000 entities and the candidate builder is set to compare three data types when creating the candidate list, then whenever a new identity enters the system, the system can make up to 3,000,000 comparisons just to generate the candidate list. If you add a single candidate builder criteria, the system can make up to an additional 1,000,000 comparisons just to generate the candidate list. That is up to 1,000,000 additional comparisons per identity record loaded into the system. If the candidate lists are too large because they consider too many types of data, the entity resolution process will run much slower than if the candidate builder settings only contain the data types necessary to build effective candidate lists.

When considering whether to use the **Default** or the **Default with name only** configuration setting, remember that if you choose **Default with name only**, you are adding comparisons at an order of magnitude greater than those required by the **Default** configuration.

Candidate lists

Candidate lists are the lists of entities that have the potential to match the incoming identity record. The candidate list is built by retrieving those entities that share attributes with the incoming identity, based on the attributes that are specified in the candidate builder configuration.

The entity resolution process only uses the entities on the candidate list for resolving entities and resolving relationships.

Because entity resolution and relationship detection are determined based on attributes, you want to carefully consider the attributes in your data sources to determine which attributes create the strongest candidates.

After the candidate list is generated, the entity resolution process compares the incoming identity to the first candidate on the list using the configured resolution rules. The system uses the resolution rules, in order, to compute a resolution score that represents how closely the incoming identity attributes match the attributes of the candidate entity. If the incoming identity attributes meet or exceed the resolution score for that rule, the incoming identity record is resolved into the candidate entity.

If the resolution score does not meet or exceed the resolution score set for that resolution rule, the system goes to the next resolution rule until the incoming identity record has been resolved into a candidate entity or all resolution rules have been exhausted.

If the incoming identity record is not resolved into an existing entity, the system resolves the record into a new entity and stores the new entity in the entity database.

Creating candidate builder configurations

You can use the Configuration Console to create new groups of candidate builder settings. These candidate builder configurations are useful as an easy way to apply a variety of configured candidate builder settings by changing only one setting.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Resolution** button.
3. Click the **Candidate Builder** tab.
4. Make sure the **Candidate Builder Config** drop-down list is displaying **- - - Select One - - -**, then click the **New** button.
5. In the **Candidate Builder Config** field type the name of the new candidate builder configuration.
6. In the **Match Type** field, choose the first type of data you want to use as a candidate criterion for resolution.

7. In the **Segment Name** field, type the name of the UMF segment where the match type data can be found.
8. Click the **Save** button.

What to do next

Now, the candidate builder configuration you just created is displayed in the **Candidate Builder Config** drop-down list allowing you to add criteria to this new configuration.

Adding criteria to candidate builder configurations

You can use the Configuration Console to add data types to existing candidate builder configurations which specify certain data types as criteria for adding an existing entity to the candidate list as part of the entity resolution process.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Resolution** button.
3. Click the **Candidate Builder** tab.
4. Choose a configuration from the **Candidate Builder Config** drop-down list.
5. Click the **New** button.
6. Choose a data type from the **Match Type** drop-down list.
7. In the **Segment Name** field, type the name of the UMF segment where the match type data can be found.
8. Click the **Save** button.

What to do next

Now, the system will consider the data type you just specified when building candidate lists as part of the entity resolution process.

Deleting candidate builder configurations

You can delete a candidate builder configuration using the Configuration Console. You might want to delete a candidate builder configuration you created that you decide you do not want to use.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Resolution** button.
3. Click the **Candidate Builder** tab.
4. Select a configuration from the **Candidate Builder Config** drop-down list.
5. Select the check box next to any match type you want to delete.
6. Click the **Delete** button. A confirmation box appears stating `The selected records will be deleted`.
7. Click **OK** to confirm the deletion of the candidate builder configuration.

What to do next

The collection of candidate builder settings you just deleted can no longer be used for generating candidate lists as part of the entity resolution process.

Help topics

Candidate Builder window:

Use this window to see a list of candidate builder settings. Candidate builder settings are grouped by candidate builder configurations.

Candidate Builder Config: field
Select the candidate builder configuration whose settings you want to see.

Match Type
Type of data that must match between an incoming identity and an existing entity for that existing entity to be added to the candidate list for entity resolution.

Segment Name
Name of the UMF segment where the match type data can be found.

Match Sequence
Group number of the order in which candidate list criteria are compared.

Candidate Builder - General tab:

Use this tab to configure a new candidate builder criterion or to see the details of an existing candidate builder criterion.

Candidate Builder Config
Candidate builder configuration that this criterion belongs to

Match Type
Select the type of data you want to match for the existing entity to be considered a candidate for resolution.

Segment Name
Type the name of the UMF Segment where the match type data can be found: Unique & Other Number = NUMBER; Address = ADDRESS; Characteristic = ATTRIBUTE; Name = NAME; Email = EMAIL_ADDR

Configuring confirmations and denials

You can adjust confirmation and denial settings to change the resolution scores of compared entities.

About this task

Confirmations and Denials can be viewed and modified by using the Console, on the **Resolution Rules** tab.

Confirmations and denials

After a candidate list has been created, and the basic resolution criteria have been compared, entity resolution compares additional criteria to strengthen or weaken a resolution score. These additional criteria are confirmations and denials.

Confirmations and denials compare the following data types:

- Date of Birth
- Unique Number
- Generation
- Characteristics

- You can specify any characteristic to be used as part of confirmations and denials.

Confirmation weight is the value used to apply more weight to two compared entities’ baseline resolution score. Denial weight is the value (typically a negative value) used to apply less weight to two compared entities' baseline resolution score.

Example

A resolution configuration can have a date of birth confirmation value of +10 and denial value of -20. If the inbound record shares a common date of birth with a candidate entity, a value of 10 will be added to the resolution score. If they have different dates of birth, a value of 20 will be subtracted from the resolution score.

Note: The date of birth confirmation and denial weights apply to the resolution score assigned by a specific resolution rule. They are not the same as the `DOBConfThreshold` parameter configured in the pipeline configuration file.

Viewing characteristic confirmations and denials

Before you create new confirmations and denials, you can review the current list of characteristic types that are used during entity resolution.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Resolution** button.
3. Click the **Characteristics** tab.
4. From the **Resolution Config** drop-down list, select a resolution configuration.

Creating characteristic confirmations and denials

You can specify any characteristic type as a criterion for entity resolution by adding it to the list of characteristic confirmations and denials.

Before you begin

You must have configured the resolution usage of the characteristic type to be confirm/deny when configuring the resolution settings of the characteristic type.

Procedure

1. In the Configuration Console, Click the **Setup** button.
2. Click the **Resolution** button.
3. Click the **Characteristics** tab.
4. From the **Resolution Config** drop-down list, select a resolution configuration.
5. Click the **New** button.
6. On the **General** tab, in the **Group Number** field, type the number of the group you want to apply to this characteristic.
7. In the **Description** field, type a description of characteristic type being configured.
8. From the **Characteristic Type** drop-down list, select the characteristic type to configure.
9. In the **Confirm Weight** field, type the value (on a scale of 1-100) to add to the likeness score (if the compared entities meet the confirmation requirements).

10. In The **Denial Weight** field, using a minus sign (-) type the negative value (on a scale of 1-100) to subtract from the likeness score (if the compared entities meet the denial requirements).
11. Click the **Save** button.

Deleting characteristic confirmations and denials

To remove a characteristic type from consideration as a criterion for entity resolution, remove it from the list of characteristic confirmations and denials.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **Resolution** button.
3. Click the **Characteristics** tab.
4. From the **Resolution Config** drop-down list, select a resolution configuration.
5. Select the check box next to the characteristic types that you want to delete.
6. Click the **Delete** button.
7. In the confirmation window, click **OK** to delete the resolution configuration.

Help topics

Confirms & Denials window:

Use this window to configure the entity resolution confirmation and denial process. You can specify confirmation and denial scores to be added to the resolution score as well as the order in which confirmations and denials are processed. After a confirmation or denial is satisfied, the corresponding score is applied and any remaining confirmations and denials are not processed. Confirmations apply a positive score, and denials apply a negative score.

Order Current processing order

Description
Description of the confirmation or denial

Score Enter a positive or negative score modifier for the given confirmation/denial.

Reorder
Click the arrows (up or down) to move the confirmation or denial one position in the corresponding direction. Because processing stops after the first confirmation or denial is satisfied, selecting the right order is important as it can have a significant impact on the entity resolution process results.

Characteristics window:

Use this window to view a list of entity characteristics whose comparisons are configured to affect entity resolution scoring. This only affects entity resolution scoring if the **Characteristic Denials Enabled** value on the **Resolution Rules General** tab is set to Yes.

Description
Name of the characteristic being compared

Characteristic Type
System name of the characteristic type being compared

Confirm Weight
Value added to the entity resolution scoring process if the compared characteristic values are the same

Denial Weight
Value added to the entity resolution scoring process if the compared characteristic values are different.

Resolution - Characteristics - General tab:

Use this tab to configure a new characteristic confirmation/denial or to see the details of an existing characteristic confirmation/denial.

Group Enter a number specifying the order to process the characteristic confirmation/denial.

Description
Enter the confirmation/denial description.

Characteristic Type
Select the characteristic type for the confirmation/denial.

Confirm Weight
Enter the score to add to the entity resolution score if the compared characteristic values are the same.

Denial Weight
Enter the negative score to add to the entity resolution score if the compared characteristic values are different.

Configuring system parameters

You can configure certain features of the Identity Insight System.

Configuring system parameters for name scoring

You can configure the name scoring algorithm that you want to use while generating a candidate list as part of the entity resolution process.

Procedure

1. In the Configuration Console, select **Setup** > **General** > **System Parameters**.
2. From the `Parameter Group` list, select the **NAME_MATCHING** parameter group.
3. Select the **ALGORITHM** system parameter.
4. In **Current Value**, specify the integer value of the Name Comparator algorithm to use. To return this system parameter to its default value, enter the value that is displayed in **Default Value**, in the **Current Value** field.

   Note: Name Comparator 2 is the default name scoring algorithm for product versions 3.9.1 and later.
5. Click **Save**.

Configuring system parameters for the Name Manager

By default, the Name Manager name scoring system parameters are configured when you install the product. But you can update the default system parameters, when needed. For example, you might need to change the location of the Name Manager support libraries.

About this task

You set the path to the Name Manager support libraries and enable categorizing names by type through Name Manager system parameters. You also set the **CROSSCHECKCULTURE** system parameter to configure name processing between different name cultures.

Procedure

1. In the Configuration Console, select **Setup** > **General** > **System Parameters**.
2. From the **Parameter Group** list, select the **NAMEMANAGER** parameter group.
3. From the left pane, select the Name Manager system parameter to configure:

Name Manager system parameter	Description
SUPPORTPATH	Indicates the location of the Name Manager support files. The default value is ./data, which is a path relative to the top-level product directory. If the support files are moved to a different location during installation, modify this value to the absolute path of the new location.
NAMESIFTER	Indicates whether the name categorization by name type (personal or organizational names) functionality is turned on. To enable categorizing names by type (Name Sifter functionality), enter 1 (new installation default) in **Current Value** To disable categorizing names by type (Name Sifter functionality), enter 0 (upgrade default) in **Current Value**
CROSSCHECKCULTURE	Indicates whether to perform Name Manager name scoring between name cultures when the name cultures are different. To check only the inbound name culture before scoring both names, enter 0 in **Current Value**. To check name culture values before scoring them (new installation default), enter 1 in **Current Value**.

Attention: The **CROSSCHECKCULTURE** system parameter affects how entity resolution handles name scoring by culture in the pipelines. Before changing this system parameter from its current value, consult IBM Services or Support.

4. Click **Save**.

Configuring system parameters for the database

You can configure the maximum size of any candidate list IN clause created by the pipeline during entity resolution.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **General** button.

3. Click the **System Parameters** tab.
4. From the **Parameter Group** drop-down list, select the **DB_CONFIG** parameter group.
5. Click the **MAX_IN_CLAUSE** system parameter.
6. In the **Current Value** field, type the maximum number of characters that you want to include in an IN clause when generating a candidate list as part of the entity resolution process. Valid values are any integer from 0 to 1000. To return this system parameter to its default value, type the value that is displayed in the **Default Value** field, in this field.

   Note: This value affects the performance of your database. Based on the size of your database and the capabilities of your system hardware, carefully consider the value you specify for this parameter.
7. Click **Save**.

Configuring system parameters for the logs

You can configure the logging level that you want to use for specific entity resolution tables within the database.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **General** button.
3. Click the **System Parameters** tab.
4. From the **Parameter Group** drop-down menu, select the **LOG_LEVEL** system parameter.
5. Click the name of the parameter that you want to configure.
6. In the **Current Value** field, type the log level you want to apply to this parameter code. Valid values are listed and described in the **Parameter Description** field. ITo return this system parameter to its default value, type the value that is displayed in the **Default Value** field, in this field.

   Note: This value affects the performance of your database and components such as the Visualizer. Based on the size of your database and the capabilities of your system hardware, carefully consider the value you specify for this parameter. For example, setting the LOG_LEVEL below 4 for some tables can cause the Visualizer to stop working, including:
   - ER_DETAIL
   - ER_ENTITY_SCORE
   - ER_ENTITY_STATE
   - ER_RELOCATION
7. Click **Save**.

Configuring system parameters for confirmation and denial

You can specify whether you want to perform every confirmation and denial comparison that is configured. Or you can specify that those comparisons are made in the order configured until one of the confirmations or denials is satisfied. Using the second option can result in faster processing times.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **General** button.

3. Click the **System Parameters** tab.
4. From the **Parameter Group** drop-down menu, select the **MM** parameter group.
5. Click the **MULTICONFIRMATION** system parameter.
6. n the **Current Value** field, type 1 to have all confirmations and denials processed, and for all of them whose condition is met, apply the sum of changes in their score to the resolution rule that is being processed. Or, type 0 to have confirmations and denials processed in their specified order, stopping on the first one whose condition is met and applying the change in its score to the resolution rule that is being processed. To return this system parameter to its default value, type the value that is displayed in the **Default Value** field, in this field.
7. Click **Save**.

Configuring system parameters for role alerts

You can configure whether you want to report every role alert generated by an entity resolution rule for an inbound entity or you want to report only the strongest role alert generated by an entity resolution rule for an inbound entity.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **General** button.
3. Click the **System Parameters** tab.
4. From the **Parameter Group** drop-down menu, select the **MM** parameter group.
5. Click the **REPORT_SAME_CONFLICTS** system parameter.
6. n the **Current Value** field, type 1 to report all role alerts generated by each resolution rule for an inbound entity. Or, type 0 to report only the strongest role alert generated by each resolution rule for an inbound entity. To return this system parameter to its default value, type the value that is displayed in the **Default Value** field, in this field.
7. Click **Save**.

Configuring system parameters for attribute alert generators

You can configure the default number of days that a new attribute alert generator will be active before it expires.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **General** button.
3. Click the **System Parameters** tab.
4. From the **Parameter Group** drop-down menu, select **PERSISTENT_SEARCH**.
5. Click the **SEARCH_EXPIRATION_TIME** system parameter.
6. In the **Current Value** field, type the default number of days that you want a new attribute alert generator to be active before it expires. Visualizer users can specify another expiration date, but this value provides the default number of active days of a new attribute alert generator.
7. Click **Save**.

Configuring the system parameters for concurrency

If your pipelines are configured for parallel pipeline processing, you can set the default number of parallel pipeline threads that start when you start a pipeline.

Before you begin

Make sure you selected the **Edit Configuration** check box when you logged into the Configuration Console. This selection enables you to add, change, and delete system configuration, including system parameters.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **General** button.
3. Click the **System Parameters** tab.
4. From the **Parameter Group** drop-down menu, select the **CONCURRENCY** parameter group.
5. Select the **DEFAULT_CONCURRENCY** system parameter.
6. In Current Value, enter the number that represents the default number of pipeline processing threads to start whenever a pipeline is started.

Configuring system parameters for data quality management

You can configure the default date delimiter that the Configuration Console uses when formatting dates.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **General** button.
3. Click the **System Parameters** tab.
4. From the **Parameter Group** drop-down menu, select the **DQM** parameter group.
5. Click the **SYSTEM_DATE_DELIMITER** system parameter.
6. In the **Current Value** field, type / or - to specify which delimiter you want the system to use when formatting dates. To return this system parameter to its default value, type the value that is displayed in the **Default Value** field, in this field.
7. Click **Save**.

Configuring system parameters for product options

You can configure which additional product options you want to enable.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **General** button.
3. Click the **System Parameters** tab.
4. From the **Parameter Group** drop-down menu, select the **CONSOLE_CONFIG** parameter group.
5. Click the **PRODUCT_OPTIONS** system parameter.
6. In the **Current Value** field, type the code provided by IBM which corresponds to the product feature you want to enable. You must use all capital letters. You can type a space-delimited list of all features you want the system to enable. To return this system parameter to its default value, type the value that is displayed in the **Default Value** field, in this field.
7. Click **Save**.

Configuring system parameters for Event Manager

You can enable Event Manager event processing and configure system parameters for event processing, including the universal resource indicator (URI) of the event processor.

Procedure

1. In the Configuration Console, click the **System Configuration** tab.
2. From the left pane, select the Event Manager system parameter to configure:
   a. **Enable event processing** indicates whether event processing through Event Manager is enabled or disabled.
   b. **Event processor timeout** indicates the number of seconds that the pipeline waits for a response from the external event processor before timing out with an error. The default value is 60 seconds.
   c. **Event processor URI** indicates the universal resource indicator (URI) to connect to the external event processor, In **Current Value**, enter the URI, including the port number, even if it is the default port number. For example: `http://localhost:13510/gem`
   d. **Event history window** indicates the number of days of event history that the pipeline sends to the external event processor when evaluating a new inbound event. (The default number of days is 180.)
3. Click the **Save** button.

Configuring system parameters for the Visualizer

The system parameter for the Visualizer gives you the ability to enable individual Visualizer users to see all alerts, including those that are lower than the **Minimum Alert Threshold** setting defined in each role alert rule. You might change this setting to allow Visualizer users more flexibility in viewing alerts.

Procedure

1. In the Configuration Console, click the **Setup** button.
2. Click the **General** button.
3. Click the **System Parameters** tab.
4. From the **`Parameter Group`** drop-down list, select the **VISUALIZER** parameter group.
5. Click the **ALLOW_ALERT_THRESHOLD_OVERRIDE** system parameter.
6. Choose one of the following options:
   - To enable Visualizer users to override the defined alert threshold on the **Role Alert Rules - Filters** tab in the Configuration Console, type 1 in the **Current Value** field.
   - To disable (or not allow) Visualizer users the ability to override the system defined alert threshold on the **Role Alert Rules - Filters** tab in the Configuration Console, type `0` .
   - To return this system parameter to its default value, type the value that is displayed in the **Default Value** field, in the **Current Value** field.
7. Click the **Save** button.

Setting the default path for Centrifuge

If you use the optional Centrifuge Desktop from Centrifuge Systems to visualize and display entity graphs, you must specify the Centrifuge Desktop file path in the Visualizer preferences.

About this task

Default path settings are configured for each Visualizer client. By specifying a default path using this task, you only set the path on the Visualizer you are currently logged in to.

Procedure

1. In the Visualizer, click **File** > **Preferences** > **System Preferences**.
2. Under the **File Paths** section in **Centrifuge path**:
   - Enter the file path or URL (uniform resource locator) to the Centrifuge Desktop application in the field.
   - Or browse to the Centrifuge Desktop application and open it.
3. Click **Submit**. A confirmation message informs you that you must restart the Visualizer before your changes take effect.
4. In the confirmation message, click **OK**.
5. Close the Visualizer, reopen the Visualizer, and log in again.

Results

After the path is configured, the **Centrifuge** button displays on the **Role Alert Detail** and **Entity Resume** screens in the **Research** window. Click the button to launch your Centrifuge Desktop application directly from the Visualizer.

Setting the default path for UMF files

If you regularly load identity records in UMF data files for processing through the Visualizer, setting the default path can save you a step.

About this task

Default path settings are configured for each Visualizer client. By specifying a default path using this task, you only set the path on the Visualizer you are currently logged in to.

Procedure

1. In the Visualizer, select **File** > **Preferences** > **System Preferences**.
2. In **Default path for File Load**, do one of the following:
   - Enter the full path of the directory to use.
   - Or browse to select the directory.
3. Click **Submit**. A confirmation message informs you that you must restart the Visualizer before your changes take effect.
4. In the confirmation message, click **OK**.
5. Close the Visualizer, restart the Visualizer, and log in again.

Results

Whenever you load a UMF file, the default path is the directory that you specified.

Attribute and scoring customization

IBM InfoSphere Identity Insight provides functional enhancements for configuring attribute data and integrating scoring algorithms. These changes expand the size and types of identity data that can be compared and scored and enable the addition of new scoring algorithms in the entity resolution process. These capabilities are generally referred to as attribute and scoring customization.

Entity resolution technology allows you to use identity matching and scoring algorithms to compare and resolve common identity data such as names, addresses, phone numbers, credit card numbers, tax identification numbers, and license numbers and indicate potential matches. We refer to the data elements that describe an account or entity as attributes. Attributes can include characteristics or traits that describe a person, organization, place, or item. With the addition of the attribute and scoring customization, you can add new types of identifying data and associate scoring algorithms that have been developed as product scoring plug-ins. For example, you can add identity data derived from fingerprints, retina scans, or DNA tests and compare and score them using a scoring plug-in that includes an appropriate comparison algorithm.

These attribute and scoring enhancements enhance the entity resolution process by allowing you to:

- Store and compare larger-sized attribute data using ATTR_VALUE (expanded to 8 KB) and ATTR_LARGE_DATA for storing even larger data.
- Apply provided scoring algorithms to a broader range of attribute types and more easily configure those attributes with more control.
- Integrate the results from customized attribute comparison and scoring using Visualizer reporting and alert functions.
- Apply a plug-in model for adding user-created scoring algorithms.
- Integrate custom scoring plug-ins using the Configuration Console.

Storing large attribute data

In order for the system to store and process larger attribute data with scoring plugins, metadata must be converted to Universal Message Format (UMF) and stored in the appropriate columns.

About this task

Procedure

1. Using the entity model you have created for the system, analyze your incoming data to see how it fits with the UMF standard. You should have a clear understanding of existing UMF segments and tags before proceeding to the next step.
2. Configure the ETL tool to produce UMF records that match your entity model.
3. Run the ETL tool.

What to do next

After converting your data to UMF, you can send the UMF records to the pipeline for processing.

Large attribute data storage parameters

In order for the system to store and process large attribute data for scoring, metadata must be converted to Universal Message Format (UMF) and stored in the appropriate columns.

Use the ATTR_VALUE and ATTR_LARGE_DATA columns to store large or unstructured attribute data for custom attribute and scoring applications.

Column and UMF tag name	Data type and size	Required	Explanation
ATTR_VALUE	varchar(255) (default) resizable up to 8k	**Yes**	Data used as one of the attributes in an ETL process with the base scoring plugins. In cases where the data is larger than 8k and in binary format, store the data in the ATTR_LARGE_DATA column and create a unique identifier for that data in the ATTR_VALUE column. That ATTR_VALUE identifier is used for comparison and scoring. For example, create an MD5 (Message-Digest algorithm 5) one-way hash that can be compared and displayed in the visualizer and reports. Max column size is database dependant. For any binary data bigger than 255/3 to be stored in ATTR_VALUE, the column must be resized. For performance reasons you should consider re-tuning the database cache because it is likely that far fewer rows will fit in the cache.

ATTR_LARGE_DATA	Character large object (CLOB), use for data larger than 8k.	**No**	Store as character data. For example, use Base64 encoding of binary data. Use this column to store attribute data that is too large for the ATTR_VALUE column. ATTR_LARGE_DATA is of type CLOB (character large object) column that can handle data of unlimited size. This data is available to entity resolution. The structure of the data must be known to the author of the customized comparison plugin. The visualizer will not display this data because the format is non-standard and will be different for various types of systems. A CLOB will not perform as well as a varchar column because a CLOB cannot be cached and requires a disk read, which is why ATTR_VALUE is preferable. If increasing the size of ATTR_VALUE will cause very little attribute data to be cached, it may be better to just use ATTR_LARGE_DATA for data smaller than 8k to ensure that other non-large attributes like gender and DOB are well cached. This is left to the architect's discretion. Consider consulting with your database administrator. When ATTR_LARGE_DATA is used, ATTR_VALUE must be populated with some value. If there is a way to make a meaningful search key from the data that fits in ATTR_VALUE, this should be created and put into ATTR_VALUE. If there is no way to create a meaningful search key, something else unique to the value must be put in ATTR_VALUE or the pipeline will not function properly and will likely fail with DQM errors. A unique key can be generated automatically by setting up a DQM rule to create a MD5 hash of the data(600 rule), or a custom hash based on configured rules(615 rule). It is important that this value be fairly unique especially if the attribute type is going to be setup for persistent searches as the ATTR_VALUE is used in the determination of generic-values. **Note:** The shipped 'binaryAttributeScoring' plugin does not compare ATTR_VALUE at all. It only examines and scores the ATTR_LARGE_DATA segment.

Example

Here is an example of an MD5 hash output of large binary data:

```
<ATTRIBUTE><ATTR_TYPE>BIOMETRIC-1</ATTR_TYPE>
<ATTR_VALUE>214b21fc3e040f844a07710b1bb451a0
</ATTR_VALUE><ATTR_LARGE_DATA>
<![H4sICBRTqkgAA2Zvby50eHQAK0ktLuHlAgDkTqoPBgAAAA==]>
</ATTR_LARGE_DATA></ATTRIBUTE>
```

Actual ATTR_LARGE_DATA values are likely to be much larger than this example.

Configuring source characteristics for large attribute data

Use the Configuration Console to configure source characteristics for large attribute data.

About this task

The Configuration Console allows you to configure new types of attribute data for customized scoring plugins in the same way you configure data for base plugins.

Procedure

1. From the Plugins tab of the Configuration Console, click the selection box for the customized plugin.
2. Click the Characteristics tab.
3. Click the General tab and fill in the fields as appropriate.
4. Select the appropriate Data Type. Data Type can be one of: CHAR, DATE or CLOB. Note the requirements for Data Type found in “Large attribute data storage parameters” on page 165.
5. Select an appropriate Class.
6. Select the a value for Resolution Usage.
7. Select the name of the Scoring Plugin you are configuring.
8. Select an appropriate value in the Display Level field. Choose "Type only without value" to prevent the Visualizer from displaying the contents of the ATTRIBUTE.ATTR_VALUE or ATTRIBUTE.ATTR_LARGE_DATA columns. The ATTR_VALUE column is not typically used when the large object column (CLOB) is used. Additionally, the ATTR_LARGE_DATA (CLOB) column would normally contain Base-64-encoded data that would not be relevant or useful to display in the Visualizer.
9. Click Save.

Results

The Characteristics tab under Sources displays the new Type and related information.

Configuring resolution characteristics for large data

Use the Configuration Console to configure the resolution characteristic for large attribute data and custom scoring plugins.

About this task

Confirmation and denial information for a new characteristic type is configured last.

Procedure

1. From the Configuration Console, click the Setup button.
2. Click the Resolution button.
3. Click the Characteristics tab.
4. Select an appropriate Resolution Configuration such as DEFAULT from the dropdown menu and then click the New button.

5. Select the General tab and enter values into the displayed fields. See “Resolution characteristics and options” for descriptions of the field options and recommendations.
6. Click Save.

Results

The overview screen displays a summary table with the values you created for resolution configuration.

Resolution characteristics and options

Use the general tab view of Resolution Characterics to configure actions and options for large data types and custom scoring plugins.

If you are configuring a character type that has a Resolution Usage field and select with a "Confirm/Deny" value, additional fields will dynamically appear.

Field	Required	Field selections and descriptions
Group	**Yes**	Type the number of the group you want to use to identify this characteristic.
Description	**Yes**	Enter a short description of this default resolution configuration. Leaving this field blank can cause an error.
Characteristic Type	**Yes**	Select the Type you are working on. The list will include all of the types you have configured for Sources.
Confirm Weight	**Yes**	Any value 0-100. Affects the likeness score.
Plugin Confirm Threshold	**No**	Free-form text field. This field is displayed when a Characteristic Type is specified whose Resolution Usage field is set to “Confirm/Deny," such as when the type is for a custom plugin. If the Characteristic Type is scored by a scoring plugin during the confirmation and denial part of the entity resolution process, then specify a confirmation threshold value. When the score assigned by the plugin is equal to or above this value, the match is considered a confirmation, which causes the Confirm Weight field value to be added to the Resolution Confidence score.
Denial Weight	**Yes**	Any value 0-100. Affects the likeness score.
Plugin Denial Threshold	**No**	Free-form text field. This field is only displayed when a Characteristic Type is specified whose Resolution Usage field is set to “Confirm/Deny," such as when the type is for a custom plugin. If the Characteristic Type is scored by a scoring plugin during the confirmation and denial part of the entity resolution process, then specify a denial threshold value here (to be interpreted by the plugin). When the score assigned by the plugin is equal to or below this value, the match is considered a denial, which causes the Denial Weight field value to be added to the Resolution Confidence score.

Configuration reports for attribute and scoring customization

The Configuration report in the Configuration Console also includes elements for attribute and scoring customization.

Additions to the Configuration report include:

- The "Characteristics Types" section of the report has a new column called "Scoring Plugin", which displays the value of the respective plugin characteristics type.
- A new Plugin report section to show records configured. Column header labels include: ID, Name, Type, Version and Library Short Name.
- The “Entity Resolution Characteristics” section has two new columns added to display both the “Plugin Confirm Threshold” and "Plugin Denial Threshold” values.

Configuring custom scoring plug-ins

Use the Configuration Console to configure custom scoring plug-ins.

Before you begin

Ensure that the new plug-in has been correctly adapted for IBM InfoSphere Identity Insight. See Developing custom scoring plug-ins for IBM InfoSphere Identity Insight.

About this task

The Configuration Console allows you to configure scoring plug-ins that have been added to your system.

Procedure

1. In the Configuration Console, click the Setup button.
2. Click the General button.
3. Click the Plugins tab.
4. To configure a new plug-in, click the New button.
5. To edit an existing plug-in, click the select the plug-in you want to configure from the list in the Plugins column. Only the customer plug-ins are editable.
6. On the General tab fill in the fields as appropriate:

Field name	**Required**	**Description**
Plugin	**Yes**	Name of plug-in which will be displayed in the “Scoring Plugin” menu options.
Short Library Name	**Yes**	This name in this field is used in the LIBRARY_NAME column of the Plugin table. The Library Short Name field is used to construct the name of the software library file called by the pipeline code. It is recommended that you match the case used in the actual library file that is being called by the pipeline. This is because some systems are case sensitive. This name is prefixed and/or suffixed by EAS as per OS.
Version	**Yes**	This field is used to track the version number of the software library.

7. Click Save.

Results

The Plugin tab displays the updated plug-in name and related information.

Developing custom scoring plug-ins for IBM InfoSphere Identity Insight

IBM InfoSphere Identity Insight allows you to create custom scoring plug-ins and include additional types of attribute data in the entity resolution process.

To create a scoring plug-in for IBM InfoSphere Identity Insight, you must include several basic elements and build a shared library. Custom plug-ins should be installed to a directory that is specified in the library load path.

Scoring plugin development interface

Custom scoring plugins require a standard interface.

Use primitive objects to eliminate a dependence on library versions and compiler options. This allows plugins to be used with multiple pipeline versions without having to rebuild the plugin when the pipeline changes library, compiler versions, or other options. You must include the following C or C++ interface prototypes:

```
#ifdef _WIN32
#define _DLEXPORT __declspec(dllexport)
#else
#define _DLEXPORT
#endif

extern "C"
{
  _DLEXPORT const int initPlugin(const char *configInfo,
                                 const uint configSize,
                                 char *errorStr,
                                 const uint maxStrSize);
  _DLEXPORT const char *getVersion();
  _DLEXPORT const int score(const char *thresholdStr,
                            const uint thresholdSize,
                            const char *inboundStr,
                            const uint inboundSize,
                            const char *candidateStr,
                            const uint candidateSize,
                            char *result,
                            const uint resultSize);
};
```

getVersion

Custom scoring plugins require the getVersion function.

Example

You must include the following:

```
const char *getVersion();
```

`return char *` contains a null terminated string that describes the plugin version.

Implement this function by storing the plugin version number in a static string and return a pointer to the string's base pointer.

`myPlugin.h` includes the following:

```
class MyPlugin
{
public:
  static const std::string mVersion;

};
```

```
myPlugin.cpp includes the following
const std::string MyPlugin::mVersion = std::string("1.0");

const char *getVersion ()
{
  return MyPlugin::mVersion.c_str();
}
```

initPlugin

Custom scoring plugins require an initPlugin function.

Example

initPlugin allows the plugin to load and save configuration information that it will need for scoring. The database connection string and the .ini file name are provided in the **configInfo** string. initPlugin will be called once for each attribute type that uses a plug-in. These are shared objects. To support plugin use for more than one attribute type, configuration information must be saved for each attribute type. This way when score is called, it can look up the configuration information for the appropriate attribute type.

```
const int initPlugin(const char *configInfo,
                     const uint configSize,
                     char *errorStr,
                     const uint maxStrSize);
```

configSize
is the length of the string contained in configInfo Error must be in the following format.

errorStr
is a pre-allocated memory buffer to copy a null terminated string. The string contains XML that describes any initialization errors. The error must be in the following format:

```
<ERROR>error text</ERROR>
```

maxStrSize
is the size of the pre-allocated memory buffer that errorStr points to. The size of error string cannot exceed this value.

The following is a pseudocode example of a score function:

```
const int initPlugin(const char *configInfo, const uint configSize,
char *errorStr, const uint maxStrSize)
{
  //create string out of configInfo
  //parse string with XML parser
  //extract DB_CONNECTION and CONFIG_FILE
  //connect to database
  //select config info from database
  //open CONFIG_FILE
  //read config info from .ini file

  //if there was an error create null terminated error string and
  //strcpy into errorStr. Return -1.
  //if no error, return 0.
}
```

The initPlugin should return -1 if an error is encountered.

score

Custom scoring plugins require a score function.

score contains the following parameters:

```
const int score(const char *thresholdStr,
                const uint thresholdSize,
                const char *inboundStr,
                const uint inboundSize,
                const char *candidateStr,
                const uint candidateSize,
                char *result,
                const uint resultSize);
```

thresholdStr
contains the confirm and deny thresholds. These thresholds are not required.

thresholdSize
is the size of the string contained in thresholdStr.

inboundStr
contains the attribute from the inbound entity being scored.

inboundSize
is the size of the string contained in inboundStr.

candidateStr
is a pointer to a string containing the attribute from the candidate entity being scored.

candidateSize
is the size of the string contained in candidateStr.

result
is a pre-allocated memory buffer to copy a null terminated string containing xml that describes the scoring results. In the case of an error, the results will be a description of the error. Format of this return string is defined as follows:

```
<SCORE_RESULT>
  <MATCH_SCORE>integer 0-100</MATCH_SCORE>
  <CONFIRMATION>TRUE/FALSE</CONFIRMATION>
</SCORE_RESULT>
```

In the case of an error, the result format is as follows:

```
<ERROR>error text</ERROR>
```

resultSize
is the size of the pre-allocated memory buffer that result points to. The result string cannot exceed this size. The result document is quite small, so this should not be an issue except with extremely long error messages.

The following is a pseudocode example of a score function:

```
const int score(const char *thresholdStr,
       const uint thresholdSize,
                const char *inboundStr,
       const uint inboundSize,
             const char *candidateStr,
       const uint candidateSize,
       char *result,
       const uint resultSize)
{
  //create strings out of thresholdStr, inboundStr, and candidateStr
  //create XML documents out of thresholdStr, inboundStr, and candidateStr
  //parse thresholds out of threshold xml doc if thresholds are used
  //parse values out of inbound xml doc
  //parse values out of candidate xml doc

  //check for any errors such as attr type mismatches, bad data, etc.
  //un-encode attr_value and attr_large_data data fields if necessary
  //apply scoring algorithm to attribute data
  //scale score into 0-100 range
  //determine confirmation or denial (possibly using thresholds)

  //if there was an error, create null terminated error string and
  //strcpy into result. Return -1.
  //if no error, create null terminated result document and strcpy into
  //result.  Return 0.
}
```

The score function should return -1 if an error is encountered.

Data formats

Custom scoring plugins require a specified data format.

Example

Threshold data format

```
<THRESHOLDS>
  <CONFIRMATION_THRESHOLD>string</CONFIRMATION_THRESHOLD>
  <DENY_THRESHOLD>string</DENY_THRESHOLD>
</THRESHOLDS>
```

The thresholds are free-form strings. They are loaded from the MATCH_MERGE_ATTR table and need to conform to the format the plugin is expecting. The format is defined by the plugin author and can vary from plugin to plugin.

Attribute data format

```
<ATTRIBUTE>
  <ATTR_TYPE_ID>unsigned int</ATTR_TYPE_ID>
  <ATTR_VALUE>string</ATTR_VALUE>
  <ATTR_LARGE_DATA>string</ATTR_LARGE_DATA>
</ATTRIBUTE>
```

ATTR_LARGE_DATA can be an empty string depending on the attribute type and the ETL process. ATTR_LARGE_DATA is optional and should only be used when an attribute's data is too large to fit in the ATTR_VALUE column. This must be determined during system configuration so the UMF can be correctly created and plugins can be written to use to use the correct fields.

ATTR_LARGE_DATA can be encoded to conform to XML's valid character set. Base64 encoding is recommended, but this is done in the ETL process. The plugin might require unencoding the data in ATTR_LARGE_DATA. The string should also be UTF-8 encoded. If the string was base64 encoded in ETL, then the UTF-8 string will be identical to the ASCII7 string.

The following is a pseudocode example of a score function:

```
const int score(const char *thresholdStr,
       const uint thresholdSize,
              const char *inboundStr,
       const uint inboundSize,
            const char *candidateStr,
       const uint candidateSize,
       char *result,
       const uint resultSize)
```

```
{
  //create strings out of thresholdStr, inboundStr, and candidateStr
  //create XML documents out of thresholdStr, inboundStr, and candidateStr
  //parse thresholds out of threshold xml doc if thresholds are used
  //parse values out of inbound xml doc
  //parse values out of candidate xml doc

  //check for any errors such as attr type mismatches, bad data, etc.
  //un-encode attr_value and attr_large_data data fields if necessary
  //apply scoring algorithm to attribute data
  //scale score into 0-100 range
  //determine confirmation or denial (possibly using thresholds)

  //if there was an error, create null terminated error string and
  //strcpy into result. Return -1.
  //if no error, create null terminated result document and strcpy into
  //result.  Return 0.
}
```

The score function should return -1 if an error is encountered.

Building the plugin object

The plugin object must be build into a shared library.

About this task

Build the object into a shared library (.dll on windows, .so on linux/unix). All libraries should be statically linked. This will prevent possible library version mismatches and unresolved symbols.

Chapter 6. Managing pipelines

Pipelines are the heart of the system. They are where the processing takes place: where entities are resolved, where relationships are detected, and where alerts are generated. Pipelines are the primary way that data is loaded into the entity database. Managing pipelines is an ongoing operational task that involves configuring pipelines, starting and stopping pipelines, monitoring pipelines, and routing messages from pipelines to other pipelines, nodes, or external systems.

Pipelines

Pipelines are the components that perform name and address hygiene standardization, data quality management, and entity resolution. The pipelines also perform relationship resolution and generate alerts, based on the system configuration.

Pipelines perform three core processes:

- Recognize, which involves optimizing incoming data by performing data standardization, hygiene, enhancement, and quality checks
- Resolve, which involves resolving entities
- Relate, which involves detecting relationships and generating alerts

Pipelines are hosted by pipeline nodes.

You can configure pipelines for parallel processing, so that one pipeline command spawns multiple parallel pipeline processing threads, which enables the system to concurrently process multiple data requests. This feature can help improve system performance, reduce data processing time, and mitigate hardware memory constraints.

The parallel pipeline processing feature is configured in two places:

- The global concurrency setting is controlled by the `Pipeline default concurrency` parameter on the **System Configuration** tab in the Configuration Console. The value here determines the number of parallel processing threads started from a pipeline start command. The default value for this parameter is 1, meaning that unless this parameter is edited, only one pipeline processing thread starts.
- A local concurrency setting (by pipeline node) can be configured in the pipeline configuration file. If you specify a concurrency parameter and value in the pipeline configuration file by pipeline node, that value overrides the global system parameter. When you issue a pipeline start command on that pipeline node, you start the same number of concurrent pipeline processing threads as specified in the pipeline configuration file.

Pipeline configuration check

The system performs a pipeline configuration check before starting a new pipeline process and at frequent intervals for each running pipeline to be certain that the configuration for the pipeline is valid.

During the pipeline configuration check, the system checks to determine if the pipeline has a valid configuration:

- Is the configuration for this pipeline the same as the configuration in the Configuration Console?
- Are there a reasonable number of records for each configuration table that pipeline use?
- Are there standard values in specific configuration tables?
- Are configuration identifiers and values set in specific configuration tables?

If these configuration checks do not pass, depending on the severity of the discrepancy, the system either logs a warning in the log files or automatically shuts down the pipeline (or does not start the pipeline) and logs an error.

Pipeline nodes

Pipeline nodes are the physical machines that host one or more pipeline processes.

The pipeline node is where you install and start the pipeline executable that runs the pipeline processes. You configure and maintain the pipeline configuration file for all pipelines that are hosted by this machine. The system also writes the pipeline messages to the log files on the pipeline nodes.

Pipeline nodes connect pipeline processes to these components of the product architecture:

Acquisition programs
As part of the extract, transform, and load (ETL) process, acquisition programs use transports to send UMF data into pipelines for processing. You use the transport method appropriate to the type of acquisition program to connect to the pipelines. For example, if you use the UMF file utility as an acquisition program, you use the file transport.

Entity database
The entity database contains entity information. Pipelines access entity information while processing incoming records for entity and relationship resolution. The pipeline node must have the appropriate database client installed and configured, so that the pipelines can access the entity database.

Queues
If your system uses queues as transport methods to send data to the pipelines for processing, you must install and configure the appropriate message queuing software on each pipeline node.

Address hygiene servers
If your system uses address hygiene products from other companies for additional address cleansing, each pipeline node must be configured to connect to the address hygiene servers.

Web services
You must use an HTTP transport to connect the pipeline processes on the pipeline node to the Web services.

Starting pipelines

Before a pipeline can receive and process data, it must be started. It is common to run multiple pipelines to increase data throughput or process different types of source data. Use these steps to start a pipeline or re-start a pipeline that is down.

Before you begin

- The pipeline node hosting this pipeline must have the pipeline executable installed.
- There must be at least one pipeline configuration file configured for use with the pipeline that you want to start. You can specify the pipeline configuration file to use as part of the start pipeline command. If you do not specify the name of configuration file as part of the pipeline command, the pipeline configuration file must be located on the pipeline node, and it must match the name of the executable (pipeline name specified). For example, pipeline.ini.
- The database environment variables must be set. See Setting the environment variables.
- If you use a script to start pipelines, make sure the script is located in the same directory where you start the pipeline.
- If the *DEFAULT_CONCURRENCY* system parameter value is set to greater than 1 or if you configured the *concurrency* parameter in the pipeline configuration file for the pipeline node, you can start multiple parallel pipeline processing threads using a single start pipeline command.

About this task

There are three steps to starting a pipeline:

Procedure

1. Each pipeline must have a name unique to its pipeline node, so make sure there are no other pipelines running with the same name as the pipeline you want to start. (The default pipeline name is `pipeline`.) To verify this, type the following command at a command prompt: `pipeline -n` *`pipelinename`* `-l`

   where *`pipelinename`* is the name you want to use to start the new pipeline. Make sure that this name matches the name registered in the Configuration Console for this pipeline.
2. At a command prompt, start one or more pipelines by specifying the type the appropriate pipeline command options and parameters using this format:

   ```
   pipeline -option parameter
   ```
3. Verify that the command worked, and the pipeline is started and active.
   a. If your system is running on a Microsoft Windows platform and you are using the services pipeline option, you can see the status of the pipeline in the Microsoft Windows Services control panel.
   b. If your system is running on a UNIX platform and you are using the daemons pipeline option, you can type the following command to check for running processes:

      ```
      ps -fu userid
      ```

      where *userid* is the identification of the user starting the pipeline.
   c. Or at a command prompt, type the following command:

   ```
   pipeline -n pipelinename -l
   ```

   where *pipelinename* is the name of the pipeline you just started. If the pipeline is active, the command prompt returns `Running`.

Stopping pipelines

Stopping a pipeline means changing its status from active and open for processing data to inactive and closed to incoming data. You can manually stop one pipeline at a time. Use these instructions to stop a pipeline after you make changes to the system configuration (then re-start the pipeline for the configuration changes to take effect), if you are installing a hot fix or an upgrade release, or if you are making configuration changes to the pipeline node that hosts the pipeline.

Procedure

1. Verify that the pipeline that you want to stop is currently running. To verify this: `pipeline -n` *`pipelinename`* `-l` where *`pipelinename`* is the name of the pipeline that you want to stop. The command prompt returns `Running` if the pipeline is active.
2. On a command line, type the pipeline stop command: `pipeline -e -n` *`pipelinename`* where *`pipelinename`* is the name of the pipeline that you want to stop.

   Note: If you started the pipeline using the debug pipeline command option, you can stop the pipeline by pressing **Ctrl** + **C** at a command line.
3. Verify that the command worked, and the pipeline is stopped: `pipeline -n` *`pipelinename`* `-l` where *`pipelinename`* is the name of the pipeline you just stopped. The command prompt returns `Stopped` if the pipeline is stopped.

Configuring pipelines

When a pipeline starts up, it checks for a pipeline configuration file to get its initial startup variables and configuration information necessary to process incoming data. By default, when a pipeline is installed on the pipeline node, the system also installs a default pipeline configuration file, named pipeline.ini, that can be used by all pipelines on that pipeline node. But some sections of this default file must be configured specifically to the pipelines running on the pipeline node so the pipeline has the proper connections and access to the entity database. Use these instructions to configure the pipeline configuration file.

Before you begin

- You must know the exact name of the entity database and the login credentials necessary to access the entity database.
- If your system connects to external address correction software, you must know the name of the address correction software host machine and be able to select the appropriate settings for this software.
- For configuration file changes to take effect, you must stop any running pipelines on this pipeline node, and then restart the pipelines after completing the changes.

About this task

The pipeline.ini configuration file is a standard ASCII text file. You can use any ASCII text editor to edit the file.

Procedure

1. Make a copy of the default pipeline.ini configuration file and save the original file in a safe location. If you save a copy of the original file, you can revert to that file, if needed.

2. Open the copy of the pipeline.ini configuration file in the text editor of your choice.
3. Update the file to reflect the appropriate configuration for the pipelines running on this pipeline node. Typically, the default values in the default pipeline configuration file are adequate; usually, you only need to enter or update the database connection information under the [SQL] heading and any address correction information under the [OAC] section, if your system uses external address correction software.
4. Save the updated pipeline configuration file. The file should be saved to the directory where the pipeline executable command resides. (Otherwise, you must specify the pipeline configuration file name and full path location every time you start a pipeline on this pipeline node.)

What to do next

If you stopped all running pipelines on this pipeline node before making the changes, you can restart the pipelines. If you did not stop all running pipelines before making these changes, you should stop and restart them now. Pipelines that are running do not apply the pipeline configuration file changes until they have been restarted. Changing pipeline configuration information without stopping pipelines may cause pipeline errors, including pipeline shut down, due to incorrect pipeline configuration file values.

Pipelines registration

Before you can monitor status or route results for pipelines, you must first register the pipelines in the Configuration Console. Registering pipelines is not the same as installing or configuring a pipeline; it means adding the pipeline to the **Pipelines** tab in the Configuration Console.

The system uses the information registered on the **Pipelines** tab to uniquely identify the pipeline. This information is used by the application monitor to report status and statistics of monitored pipelines or to route communications and results between the pipelines and other systems. The name you register for a pipeline is the same, exact name (including case) that you must use when you start the pipeline. If you use another name or do not match the case of the registered pipeline, the application monitor does not recognize the pipeline and will not route or monitor it.

Once a pipeline has been registered on the **Pipelines** tab, you can configure routing rules for the pipeline on the **Routing** tab, monitor the status and statistics of the pipeline through the **Pipeline Status** tab, or both. To monitor the status and statistics of a pipeline, when you register it, you must indicate that you want the system to monitor the pipeline.

Once a pipeline is registered, you cannot edit the name of the registered pipeline, but you can update the other information about the pipeline. For example, if the name of the pipeline node changes, or if you want to start monitoring the status and statistics for the pipeline, you can edit this information.

Registering pipelines

There are three reasons to register a pipeline: to use the application monitor to monitor pipeline status and statistics, to configure routing rules for the pipeline, or both. You can add a new pipeline registration or base the registration on an existing registered pipeline.

Before you begin

You must know the unique name of the pipeline and the name of the pipeline node hosting the pipeline. The pipeline does not have to already be installed and configured on the pipeline node before you register it. (But it must be installed and configured before the system can monitor or route to the pipeline.)

About this task

Tip: If you are adding multiple pipelines that all run on the same pipeline node, you might choose to register the first pipeline, and then clone the other pipelines from the first one you added.

Procedure

1. Click the **Setup** button.
2. Click the **General** button.
3. Click the **Pipelines** tab.
4. Complete one of the following steps:
   - To register a new pipeline, click the **New** button.
   - To register a new pipeline based on an existing pipeline, click the **Clone** button.
5. On the **General** tab, specify a unique pipeline name, description, pipeline node name, and whether to monitor the pipeline status and statistics.

   Note:

   - The pipeline name you enter is the same name that you must use when starting this pipeline. This name is case sensitive, so when you start the pipeline, you must enter the exact match of this registered pipeline name. If you do not enter an exact match (or case match), neither the routing rules configured for this pipeline nor the application monitoring for this pipeline will work.
   - If you want to monitor the status and statistics for this pipeline on the **Pipeline Status** tab in the Configuration Console, select **Yes** in the **Monitored** field.
6. Click the **Save** button.

What to do next

If the pipeline was successfully added, it displays in the list on the left side of the screen. You can now set up routing rules for this pipeline or use the system to monitor the pipeline. However, keep in mind that to successfully route to or monitor the pipeline, it must be started using the exact same name, including case, as registered in the **Pipeline Name** field.

Viewing registered pipeline details

You can view the details of a pipeline registered in the Configuration Console to make sure that the registration information is current. You register pipelines to enable the system to monitor pipeline performance and statistics, to route to pipelines, or both.

Before you begin

- The pipeline must be registered in the Configuration Console.

Procedure

1. Click the **Status** button.
2. Click the **Status** button.
3. Click the **Overview** tab.
4. Click the registered name of the pipeline.

Results

In the **Detail** window, review the details of the selected pipeline.

Editing pipeline registrations

Edit the information about a registered pipeline whenever a key component of the pipeline registration has changed, such as the name of the pipeline node. The name registered to a pipeline is the only information you cannot change. If you need to change the registered name for a pipeline, either delete the pipeline registration and then re-add it with the correct information or add another pipeline registration.

About this task

If the pipeline you want to edit is active (currently running), it is a good idea to stop the pipeline before editing its registration, especially if you are changing the monitoring status.

Procedure

1. Click the **Setup** button.
2. Click the **General** button.
3. Click the **Pipelines** tab.
4. Select the pipeline(s) you want to edit and then click the **Edit** button.
5. Change the information.

   Note: Keep in mind that to monitor the status and statistics of the pipeline on the **Pipeline Status** tab, the **Monitored** field must be set to **Yes**.
6. Click the **Save** button.

Results

You can view your changes on the **Pipelines** tab.

What to do next

If you stopped the pipeline, re-start it.

Deleting pipeline registrations

Deleting a pipeline registration in the Configuration Console does not physically delete the pipeline from the system, it removes the pipeline from the **Pipelines** tab, the **Routing Rules** tab, and the **Pipeline Status** tab. These deleted registrations can no longer route information using routing rules or provide status and statistic monitoring information. You cannot edit a registered pipeline name. If you need to change the name of a registered pipeline, either delete the pipeline registration and then re-add it with the correct information or add another pipeline registration.

About this task

If the pipeline you want to delete is active (currently running) and is being monitored by the system on the **Pipeline Status** tab, it is a good idea to stop the pipeline before deleting it. Also, it is a good idea to check the **Routing Rules** tab to see if there are any routing rules associated with this pipeline; if there are, you may want to re-route those routing rules to another pipeline or add a new pipeline that uses those routing rules before deleting this pipeline.

Procedure

1. Click the **Setup** button.
2. Click the **General** button.
3. Click the **Nodes** tab.
4. Select the pipeline(s) you want to delete and then click the **Delete** button.

What to do next

The pipeline you deleted no longer display on the **Nodes** tab or the **Routing Rules** tab . The deleted pipeline no longer report status on the **Pipeline Status** tab. The system no longer routes any of the routing rules assigned to the deleted pipeline on the **Routing Rules** tab.

Help topics

Pipelines tab

Use the **Pipelines** tab to register a pipeline or to edit, delete, or view registered pipelines. When pipelines are registered on this tab and an SNMP agent is installed and configured on the pipeline node running the registered pipeline, you can use see pipeline status, statistics, and performance on the **Status** tab. You can also use the **Routing Rules** tab to configure and route results from a registered pipeline to other databases and external systems.

Pipeline Name
Lists the names of each node registered for application monitoring in the Configuration Console, in alphabetical order.

Description
Provides additional text, which can help to further describe and distinguish this node from other system nodes.

Host name
Displays the name of the pipeline node hosting this pipeline. (If you plan to monitor this pipeline, this is also the server where an SNMP agent should be installed and running.)

Monitored
Displays whether the status and statistics for this pipeline are being monitored and reported on the **Status** tab. (This is not the same as current status of the pipeline; this column indicates how this pipeline is currently registered.)

- Yes indicates that this registered pipeline is being monitored by the application monitor.
- No indicates that this pipeline is not configured for application monitoring but may be configured for routing.

Pipelines - details tab

Use this tab to register a pipeline or to see the details of an existing registered pipeline. You must register a pipeline before configuring routing rules for the pipeline on the **Routing** tab or monitoring its statistics and status on the **Status** tab.

All fields on this tab are required to successfully register a pipeline. After a pipeline is registered, you can change everything, except the name of the pipeline. For example, if you need to change the name of the pipeline node (**Host Name** field), edit that name. However, if you want to change the pipeline name, first delete the incorrect pipeline name registered here and then re-add the pipeline with the correct information.

Pipeline name
Enter a unique name for the pipeline that is 15 characters or less. If you want to monitor or route to or from the pipeline, the name you must exactly match this registered pipeline name when you start the pipeline, including case.

The list to the left displays the names of all pipeline already registered.

Description
Enter a description for the pipeline to distinguish it from the other pipelines, using 50 characters or less. For example, use the description to indicate what the system is used for or the type of data sources the system processes.

Host Name
Enter the name of the pipeline node that runs this pipeline.

Monitored
Select whether the application monitor reports status for this pipeline.

- **Yes** indicates that you want to monitor status and statistics for this pipeline. When this pipeline has been properly registered in the Configuration Console and the SNMP agent is running on the pipeline node, the status and statistics for this pipeline display on the **Status** tab.
- **No** indicates that you want to register this pipeline for routing but not for monitoring. No status or statistics for this pipeline will display on the **Status** tab, but you can configure routing rules for the registered pipeline.

Configuring routing rules

Routing rules allow you to route the results of pipeline processing or acquisition program to a database, a pipeline, or an external system. You configure routing rules in the Configuration Console on the **Routing Rules** tab, but you can only route from pipelines or acquisition programs that have been registered with the application monitor. You can either configure a new routing rule from scratch or based on an existing routing rule.

Before you begin

- The pipeline or acquisition program you want to route from must be registered with the application monitor.
- You must know the exact unique name that was used to register the pipeline or acquisition program.
- You must know which transport method to use and the specific transport URI syntax you must use to route to the destination.

Procedure

1. Click the **Setup** tab.
2. Click the **General** tab.
3. Click the **Routing Rules** tab.
4. Do one of the following:
   - To configure a new routing rule, click the **New** button.
   - To configure a new routing rule based on an existing routing rule, select the check box next to the routing rule you want to base the new rule on and click the **Clone** button.
5. Required: In the **From Pipeline** field, enter the registered name of the pipeline or application program that you want to route from. The name you type must be an exact match to the name registered on the **Pipelines** tab.
6. Required: In the **Order** field, enter a number from 0-999 that represents the order in which the system should use this routing rule. The system defaults this field to 0, which is the first routing rule processed for any pipeline or acquisition program. The number in this field must be unique for this pipeline or acquisition program, especially if there are multiple routing rules already configured for the pipeline or acquisition program.

   Note: Look in the left pane on this tab at the list of pipelines or acquisition programs that have existing routing rules configured. If you see this pipeline or node in the list, look for the highest number following the colon after the pipeline or acquisition program name, and then enter the next highest number. For example, if you are configuring a new routing rule for PIPE08, and you see PIPE08:0 in the left pane list, you must enter the number 1 or higher in the Order field.
7. Required: In the **Destination** field, enter the transport URI for the destination of the routed information. This informs the system on how to route to the intended pipeline, database, or external system destination.

   Note: For routing to be successful, the destination process must be accessible using the same transport URI that is specified. For example, if the destination is a pipeline, the pipeline must be started using the same transport URI.
8. Required: In the **Document** drop-down list, select the UMF document type to indicate the message type to route to the destination.
9. Optional: In the **Route Filter** field, enter a filter to apply to the information to be routed, so that the system only routes particular information to the destination. Filters are an advanced routing rules feature. You type a filter expression `MODDIST(`*`UMF_tag_name`* , where (*`UMF_tag_name`* indicates the name of the UMF tag the system uses to distribute the records.
10. Required: In the **Enabled** drop-down list, select **Yes** to enable this routing rule.
11. Required: Click the **Save** button.

Example

What to do next

The name of the pipeline or acquisition program displays in the **Routing Rules** tab with the details of the routing rule you just configured. The system begins routing information from the pipeline or acquisition program to the destination, using the configured routing rule.

Routing rules

Routing rules instruct the application monitor to send messages from an acquisition program to a pipeline or from a pipeline to a database or external system. Routing rules can only be configured for pipelines that have been registered with the application monitor, but the results can be routed to any destination using the proper transport Universal Resource Indicator (URI) syntax.

Routing rules have many uses, including these common uses:

- Balancing the data load from an acquisition programs (such as the UMF database utility) to multiple pipelines for data processing.
- Directing the results of pipeline processing (such as alerts) to an external system or a reports database for additional investigation or reporting purposes

UMF documents and routing rules

Routing rules are configured to route messages using one or more UMF document types. Your choice depends upon the information that results from the pipeline or system node that you want to route from. For example, a UMF_ALERT is a UMF document type representing alerts generated from processing identity and entity records through a pipeline. You could route any alerts generated from a specific pipeline to an external system, such as to a user interface used by analysts investigating alerts produced by the system.

You can configure a routing rule to route all UMF document types or a specific UMF document type, including any custom UMF document types configured for your system.

Filters

You can filter the information that is routed to the destination by specifying a filter expression when you configure a routing rule. Filters specify that only particular information is routed to the destination.

You construct a routing filter using the `MODDIST(`*`UMF_tag_name`*`)` expression, where

`MODDIST`
is the expression indicating a modulus distribution.

(*`UMF_tag_name`***)**
identifies the UMF tag that indicates to the system how to distribute the records. Using the identified UMF tag, the system sums up the ASCII values of all the characters in that tag to determine the number of routes necessary to balance the data processing load.

If you wanted to route all records from data source code “datasource5” to a separate reports database, you could configure a routing rule using the filter expression `MODDIST(datasource5)` , where `datasource5` is the data source code.

Routing process

When a pipeline or acquisition program has a configured routing rule, the following explains how the application monitor completes the routing process:

1. When the pipeline or acquisition program starts, it sends a request to the application monitor using a UMF message.

2. The application monitor receives the request and looks for all active routing rules that pertain to the requesting pipeline or acquisition program.
3. If the application monitor locates an active routing rule for the requesting pipeline or acquisition program, it builds a UMF document containing the routing instructions and sends that UMF document back to the requesting pipeline or acquisition program.
4. The requesting pipeline or acquisition program interprets the UMF document message and creates a routing file with an *.RTE file extension (where * is the requesting pipeline or acquisition program name). If the pipeline or acquisition program cannot communicate with the application monitor upon startup, it looks for the routing file for instructions.
5. The requesting pipeline or acquisition program opens the transports necessary to communicate with the destination configured in the routing rule.
   - If the pipeline or acquisition program can successfully open the transport and locate the destination, it routes the appropriate UMF document messages to the destination as long as it is started and actively processing data.
   - If the pipeline or acquisition program cannot open the transport or if the destination cannot be located, the pipeline or acquisition program stops with an error.

Help topics

Routing Rules tab

Use this tab to view or delete existing routing rules and configure new routing rules for pipelines that have been registered on the **Pipelines** tab. Once a routing rule is configured, it cannot be edited, only deleted.

From Pipeline
Displays the name of the pipeline that is configured with a routing rule.

Order
Displays the order that this routing rule is processed for the pipeline in the **From Pipeline** column. Order is useful when there are multiple routing rules; often, the order is set to 0.

Destination
Displays the transport URI of the receiving pipeline, database, or external system.

Document type
Displays the UMF document type that this routing rule sends. This is the document type for the results processed by the pipeline in the **From Pipeline** column. This selection may be a specific UMF document type or an * (asterisk), which indicates that this routing rule routes all UMF document types.

Enabled
Indicates whether or not this routing rule is active or not:

- **Yes** indicates the routing rule is enabled. Whenever the pipeline or node displayed in the **From Pipeline** column processes results for the specified document type, the system routes the data associated with that UMF document type to the destination indicated in the **Destination** column.
- **No** indicates the routing rule is not enabled.

Routing Rules details tab

Use this tab to configure a new routing rule or to see the details of an existing routing rule. Routing rules are usually configured to publish specific types of processed results from a pipeline to another database or to an external system. You can only configure routing rules for pipelines that are registered on the **Pipelines** tab.

All fields except the **Route Filter** field are required to successfully configure a new routing rule. Once a routing rule is configured, it cannot be edited; if you need to change the routing rule, you must delete it and then re-add it with the corrected information.

From Pipeline
Enter the unique name of the pipeline that you want to route results from. The name of this pipeline must exactly match what is registered on the **Pipelines** tab and is case-sensitive; if the name does not match, the system displays an error message indicating that the specified pipeline does not exist.

Order Enter a number from 0 to 999 that indicates the order in which the system applies this routing rule to the registered pipeline in the **From Pipeline** field. This field defaults to 0, which indicates the system processes this routing rule first. If this pipeline already has one or more configured routing rules, enter a number greater than the highest order

Check the left pane to see the order set for any existing routing rules already configured for this pipeline, indicated by the sequential number following the colon after the pipeline name. (For example: PIPE08:0 indicates that the pipeline PIPE08 already has one configured routing rule, currently set to process first. If you configured a new routing rule for PIPE08, set the order to 1.)

Destination
Enter the transport URI to the destination pipeline, database, or external system to route the processed results to. Make sure to use the syntax appropriate to the type of transport that you are using.

Document Type drop-down list
From the drop-down list, select the UMF document type to route from the registered pipeline to the destination. If you want to route all processed results to the destination, select the asterisk character *.

Route Filter
If you want to specify that only particular information be routed to the destination, type the expression for the system to use to filter the UMF values routed by this routing rule. (For example, if you wanted to route only the identity or entity records from a specific data source, you might type a filter of DSRC_CODE=x, where *x* is the unique data source code for the data source you want to filter for.)

Filters are an advanced routing rules feature.

Enabled drop-down list
Select an option from the drop-down list:

- **Yes** means that the application monitor routes information from the pipeline to the destination according to this routing rule.
- **No** means that the application monitor does not route information from the pipeline according to this routing rule.

Deleting routing rules

Once a routing rule is configured, it cannot be edited; if you need to correct or update information, you must delete the old routing rule and configure a new one. You might also need to delete a routing rule that is no longer needed or used. You can delete one or more configured routing rules from the **Routing Rules** tab in the Configuration Console.

Procedure

1. Click the **Setup** tab.
2. Click the **General** tab.
3. Click the **Routing Rules** tab.
4. Select the check box next to each configured routing rule that you want to delete.
5. Click the **Delete** button.

What to do next

The system deletes the selected routing rules, and no longer routes information using the deleted routing rules.

Pipeline status and statistics

Monitoring the status, statistics, and performance is important to keeping pipelines running, balancing pipeline data loads, and spotting potential pipeline problems before they occur.

Before you can view status and statistics about a pipeline, the following must be successfully completed:

1. The pipeline is installed and configured on its pipeline node.

Note: (Windows platforms only) If you start the pipeline as a service, you can view additional status information in the Windows Event Viewer that you cannot view elsewhere.

Status and statistic information

Once a pipeline begins processing data, you can see UMF exception information in the Configuration Console:

- UMF exceptions on the **UMF Exceptions** tab

SNMP agents

Simple Network Management Protocol (SNMP) is a standard protocol that is used for monitoring systems and network devices. SNMP agents request periodic status and statistics from each registered pipeline in the system. The information the SNMP agent gathers about each registered pipeline is displayed in the **Pipeline Status** tab.

Before SNMP agents can monitor pipelines:

- An SNMP agent must be installed and configured on the pipeline node that is running the pipelines that you want to monitor.
- Each pipeline that you want to monitor must be registered in the Configuration Console and configured for monitoring.

- The SNMP agent must be started and running on the pipeline node, using the same port number as configured during pipeline installation. This SNMP agent port number is system-wide, not per pipeline node. The default SNMP port number is 13516, but you can locate the SNMP agent port number configured in the server.xml file located on each pipeline node.

SNMP agents are services and can be stopped and started as needed.

Example of using an SNMP agent

Company ABC monitors all their pipelines using the application monitor. They have added another pipeline node (EAS-2) to host three new pipelines: Pipeline300, Pipeline310, and Pipeline320. To monitor these pipelines, Company ABC Operators must complete the following tasks:

- Install and configure one SNMP agent on the pipeline node EAS-2.
- In the Configuration Console, register each new pipeline (Pipeline300, Pipeline310, and Pipeline320) on the **Pipelines** tab.
- Start the SNMP agent on pipeline node EAS-2. Make sure that the SNMP agent is using the system-wide port number that was configured when you installed the pipelines on this pipeline node.
- Start each registered pipeline for processing. Make sure to type the exact name that was registered to the pipeline, since registered pipeline names are case sensitive.

Once the new pipelines are running, Company ABC Operators can now monitor their status and statistics through the Configuration Console.

Starting SNMP agents

To monitor the status and statistics of one or more pipelines in the Configuration Console, you must start an SNMP agent on the pipeline node where those pipelines are running.

Before you begin

- An SNMP agent needs to be installed and configured on the pipeline node where the pipelines are running.
- Pipelines must be registered on the **Pipelines** tab in the Configuration Console and configured for monitoring.

Procedure

1. From a command line on the pipeline node, use the **`Change directory`** command to go to the home directory.
2. Enter the following command: **`java -jar SNMPAgent`**-p *port number* where *port number* is the system-wide port number configured during pipeline installation for SNMP agents. The default port number value is 13516.

   Note: You can find the configured SNMP agent port number in the server.xml file on the pipeline node.

Results

The SNMP agent starts.

What to do next

In the Configuration Console, select the **Pipeline Status** tab to verify that the SNMP agent is running. If it is, the SNMP agent will be reporting status and statistics on all pipelines running on this pipeline node. You do not need to restart the SNMP agent if you add more pipelines, as long as the .SHM files are in the same directory, which is usually the directory where the SNMP agent is started.

Stopping SNMP agents

Stop an SNMP agent on a pipeline node whenever you need to make changes to pipeline node, such as configuration updates.

Before you begin

An SNMP agent must be currently running on the pipeline node. It is also a good idea to stop any pipelines running on this pipeline node that are being monitored by the application monitor.

Procedure

In the window running the SNMP agent, press the **Ctrl** + **C** keys.

What to do next

- The SNMP agent stops.
- In the Configuration Console, the **Pipeline Status** tab displays a `STOPPED` status for all pipelines on this pipeline node.

Checking pipeline status in the Configuration Console

Keeping track of the current status of pipelines is important, because if a pipeline is down, part of the system is down. You can quickly see the latest pipeline status and performance statistics on the **Pipeline Status** tab of the Configuration Console. The application monitor receives the information from active SNMP agents, polling the SNMP agents and then refreshing the **Pipeline Status** tab every 60 seconds.

Before you begin

- An SNMP agent must be installed and configured on the pipeline node running the pipelines to be monitored.
- The SNMP agent must be started using the system-wide port number configured during pipeline installation. (You can see this configured port number in the server.xml file.)
- The pipeline must be registered on the **Pipelines** tab on the Configuration Console and configured for monitoring.
- The pipeline must have been started using the exact same name and case of the pipeline name that is registered on **Pipelines** tab.

About this task

If you are not able to view the status through the Configuration Console, you can use a command line to check pipeline status.

Procedure

1. Click the **Status** button.
2. Click the **Pipeline Status** button.

3. Look in the **Pipeline Name** column to locate the name of the pipeline that you want to check. (Pipelines are listed in alphanumeric order by name.) Then look at the information in the status and transaction statistics columns on the same row as that pipeline name.

What to do next

You can also view other information about this pipeline by clicking on one of the other buttons. For example, if you wanted to see the last time this pipeline was started, click the **Events** tab.

Checking pipeline status using the command line

Keeping track of the current status of pipelines is important, because if a pipeline is down, part of the system is down. Many organizations check pipelines through the **Pipeline Status** tab in the Configuration Console, because it displays the latest pipeline status and statistics based on an automatic system poll every 60 seconds. But you can use a command line to check the status of a particular pipeline or all pipelines on a particular pipeline node. (The command line check provides only pipeline status, not pipeline performance statistics.)

Before you begin

- An SNMP agent must be installed and configured on the pipeline node running the pipeline.
- The SNMP agent must be started and running on the pipeline node, using the same port number as configured during pipeline installation. This SNMP agent port number is system-wide, not per pipeline node. The default SNMP port number is 13516, but you can locate the SNMP agent port number configured in the server.xml file located on each pipeline node.

Procedure

1. From a command line on the pipeline node, complete one of the following steps:
   - To check the status of all pipelines on this pipeline node, type the following command**`pipeline -l`**
   - To check the status of a particular pipeline on this pipeline node, type the following command **`pipeline -n pipelinename -l`**

   where *pipelinename* is the unique name of the pipeline you want to check.

   Note: The name you type must match the name used to start the pipeline.
2. Press **Enter**.

Results

The system returns one of the following statuses for each pipeline:

- Running for each currently active pipeline.
- Stopped for each currently inactive pipeline.

Example

For example, to check the status of pipeline08, you would type the following command: **`pipeline -n pipeline08 -l`**

What to do next

If the status of a pipeline is unexpectedly listed as Stopped, you may want to use the troubleshooting topics to discover the reason.

Viewing application monitor events

Application monitor events occur whenever a message is exchanged between the application monitor and the registered pipelines on the **Pipelines** tab in the Configuration Console. These messages include information ranging from when a pipeline starts or stops to when the system logs errors or warnings, except for Universal Message Format (UMF) exceptions. This information can help you troubleshoot errors occurring on a specific pipeline.

Before you begin

- The pipeline must be registered on the **Pipelines** tab on the Configuration Console.
- The pipeline must have been started on the pipeline node registered on **Pipelines** tab using the same registered pipeline name that displays on **Pipelines** tab.

About this task

If the pipeline is registered on the **Pipelines** tab in the Configuration Console, you can view current or historical events on the **Events** tab in the Configuration Console.

Procedure

1. Click the **Status** button.
2. Click the **Events** button.
3. Optional: In the **From Date** field, type the starting date in mm/dd/yyyy format that you want to see application monitor events for. If you leave this field blank, the system displays all application monitor events since the first operational system date that meet the other specified criteria. If you type a date in this field, you do not have to type a date in the **Thru Date** field.
4. Optional: In the **Thru Date** field, type the ending date in mm/dd/yyyy format that you want to see application monitor events for. If you leave this field blank, the system displays all application monitor events through today's date that meet the other specified criteria. If you type a date in this field, you do not have to type a date in the **From Date** field.
5. Optional: In the **From Pipeline** field, type the registered name of the specific pipeline you want to see application monitor events for. If you leave this field blank, the system displays all application monitor events for all pipelines by registered name that meet the other specified criteria.
6. Optional: From the **Max Count** drop-down list, select the maximum number of application monitor events to display. The system displays only up to and including that number of application monitor events that meet all other specified criteria. If there are more exceptions than the number specified, the system does not display them. If there are fewer exceptions than the number specified, all application monitor events that meet all other specified criteria display.
7. Required: Click the **Search** button.

Example

For example, if you want to see the last 500 application monitor events that occurred today for pipeline08, you would specify the following criteria:

- In the **From Date** field, you would type today's date.
- In the **Thru Date** field, you would type today's date.
- In the **From Pipeline** field, you would type `pipeline08`.
- You would select **500** from the **Max Count** drop-down list.

What to do next

You can take a closer look at the details of a application monitor event by clicking on it. The information displayed is what was reported about the event when it occurred.

Viewing UMF exceptions

UMF exceptions indicate problems with incoming data being processed by a pipeline. They occur when the structure of the incoming data cannot be parsed. Typically, UMF exceptions do not contribute to the pipeline error limit count, so the system logs the UMF exception and the pipeline usually continues processing. This information can help you troubleshoot incoming data for a particular pipeline.

Before you begin

- The pipeline must be registered on the **Pipelines** tab on the Configuration Console.
- The pipeline must have been started on the pipeline node registered on **Pipelines** tab using the registered pipeline name as displayed on **Pipelines** tab.

About this task

If the pipeline is registered on the **Pipelines** tab in the Configuration Console, you can view current or historical UMF exceptions on the **UMF Exceptions** tab in the Configuration Console.

Procedure

1. Click the **Status** button.
2. Click the **UMF Exceptions** button.
3. Optional: In the **From Date** field, type the starting date in mm/dd/yyyy format that you want to see UMF exceptions for. If you leave this field blank, the system displays all UMF exceptions since the first operational system date that meet the other specified criteria. If you type a date in this field, you do not have to type a date in the **Thru Date** field.
4. Optional: In the **Thru Date** field, type the ending date in mm/dd/yyyy format that you want to see UMF exceptions for. If you leave this field blank, the system displays all UMF exceptions through today's date that meet the other specified criteria. If you type a date in this field, you do not have to type a date in the **From Date** field.
5. Optional: In the **From Pipeline** field, type the registered name of the specific pipeline you want to see UMF exceptions for. If you leave this field blank, the system displays all UMF exceptions for all pipelines by registered name that meet the other specified criteria.

6. Optional: From the **Max Count** drop-down list, select the maximum number of UMF exceptions to display. The system displays only up to and including that number of UMF exceptions that meet all other specified criteria. If there are more exceptions than the number specified, the system does not display them. If there are fewer exceptions than the number specified, all application monitor events that meet all other specified criteria display.
7. Required: Click the **Search** button.

Example

For example, if you want to see the last 50 UMF exceptions that occurred today for pipeline08, you would specify the following criteria:

- In the **From Date** field, you would type today's date.
- In the **Thru Date** field, you would type today's date.
- In the **From Node** field, you would type `pipeline08`.
- In the **Max Count** field, you would select `50`.

What to do next

You can take a closer look at the details of an UMF exception by clicking on it. The information displayed is what was logged about the exception when it occurred.

Viewing new identities

The **New Identities** tab in the Configuration Console displays new identities processed by the system pipeline over the last seven days. You can use this tab to check the incoming data volumes and be sure the numbers are appropriate to the amount of incoming data or the number of active pipelines. You can also spot check the data sources that are being loaded into the pipeline, to see which sources are feeding data into the system.

Procedure

1. Click the **Status** button.
2. Click the **New Identities** button.

Results

The system displays the list of all new identities processed over the last seven days.

Help topics

Pipeline Status tab

Use this tab to review current status, statistics, and performance information for registered pipelines that are configured for monitoring through the application monitor and the SNMP agent. The system collects status and statistics from the SNMP agent once each minute and refreshes the **Pipeline Status** tab.

Note: Each SNMP agent running on each pipeline node must use the same system-wide port number. This port number is configured when installing pipelines on a pipeline node. The default SNMP agent port number value is 13516, but you can locate the configured SNMP port number in the server.xml file.

Total Pipelines
Displays the total number of pipelines that are registered for application

monitoring in the Configuration Console. (Total Pipelines equals Active Pipelines plus Stale Pipelines plus Pipelines Down.)

Active Pipelines
Displays the total number of registered pipelines that are configured for monitoring in the Configuration Console that are currently running.

Stale Pipelines
Displays the total number of pipelines whose configuration has been modified since the pipeline was started. These pipelines need to be stopped and restarted, so the new configuration changes can take effect.

Pipelines Down
Displays the total number of registered pipelines that are configured for monitoring in the Configuration Console, but that are not currently running or reporting statistics. Each pipeline that is currently down is included in this total, so if a pipeline node is not running, all pipelines configured for monitoring on that server will be counted as down.

TPM Displays the average of the total number of transactions that are being processed each minute for all active pipelines configured for monitoring in the Configuration Console. This number indicates overall system performance; the higher the number, the better each active pipeline is performing. This number is refreshed and recalculated once each minute, based on the information received from each SNMP agent running on each pipeline node where active pipelines are running. (Total TPM equals TPM for Active Pipelines divided by the Total Active Pipelines.)

TPS Displays the average of the total number of transactions that are being processed each second for all active nodes configured in for monitoring in the Configuration Console. This number indicates overall system performance; the higher the number, the better each active node is performing. This number is refreshed and recalculated once each minute, based on the information received from each SNMP agent running on each host machine where active nodes are running. (Total TPS equals TPS for Active Nodes divided by Total Active Nodes.)

Pipeline Name
Lists the names of each pipeline that is registered for monitoring in the Configuration Console, in alphanumeric order.

Host Name
Displays the name of the pipeline node registered to this pipeline. If the status of this pipeline is unexpectedly displayed as `Down`, you can use the pipeline node name to help troubleshoot the problem. (For example, if all pipelines on a particular pipeline node are listed as unexpectedly `Down`, the pipeline node is a good place to start troubleshooting.)

Status Displays the last known status of this pipeline: `Active` (running) or `Down` (not running). The system updates the status information once per minute, depending upon the information received from the SNMP agent running on the pipeline node.

TPM Displays the average number of transactions that are being processed each minute for this pipeline. If the pipeline is in the `Down` status, the system displays `Not Available`. This number indicates pipeline performance; the higher the number, the better the pipeline is performing.

TPS Displays the total number of transactions that were processed each second for this pipeline. If the pipeline is in the `Down` status, the system displays

`Not Available`. This number indicates pipeline performance; the higher the number, the better the pipeline is performing.

UMF Exceptions tab

Use this tab to view the UMF exceptions logged from the data loaded by the pipelines registered for application monitoring. First, you generate an on-screen report of UMF exceptions to view. Then you can select a specific UMF exception and take a closer look at its details; this information may be helpful in resolving the UMF exceptions in the data files. Once you resolve one of these errors, you can safely reprocess the corrected records in that file.

UMF exceptions are data-driven errors. They occur when there are problems with the UMF data structure in an incoming data source file being processed by a pipeline. By default, UMF exceptions do not contribute to the error limit count for the pipeline (set in the pipeline configuration file); so by themselves, UMF exceptions typically do not shut down the pipeline. You can find a full listing of UMF exceptions in the UMF_EXCEPT table or in the *pipeline_name*.msg log, even UMF exceptions for pipelines that are not registered for application monitoring.

On-screen report criteria

Use these fields to specify the criteria for the on-screen UMF exceptions report. Once you specify the criteria, click the **Search** button to generate the report.

From Date

Starting date to report on UMF exceptions within the other criteria specified. (This field is optional and can be blank. Leaving the field blank means display UMF exceptions starting with the first day the system became operational within the other criteria specified.)

This field defaults to today's date. Type the date using the `mm/dd/yyyy` format.

Thru Date

Ending date to report on UMF exceptions within the other criteria specified. (This field is optional and can be blank. Leaving the field blank means display UMF exceptions through today within the other criteria specified.)

This field defaults to today's date. Type the date using the `mm/dd/yyyy` format.

From Node

Name of the registered pipeline to view UMF exceptions for. (This field is optional and can be blank. Leaving the field blank means display UMF exceptions for all registered pipelines within the other criteria specified.)

Keep in mind that you can only see UMF exceptions on this tab for pipelines that are registered for application monitoring. If you want to see all UMF exceptions, refer to the UMF_EXCEPT table or the *pipeline_name*.msg log.

Data Source Code

Data source code (exact) to view UMF exceptions for. (This field is optional and can be blank. Leaving the field blank means display UMF exceptions for all data sources within the other criteria specified.)

Max Count

Drop-down list containing options for the maximum number of

UMF exceptions to display within the other criteria specified. Only the number of UMF exceptions up to and including the maximum number display on-screen. If there are more UMF exceptions that meet the criteria, the system does not display them.

Search button
By clicking this button, the system executes the search - it finds and display all records matching the criteria entered.

On-screen report results display
This section of the window displays the on-screen UMF exceptions report, based on the criteria you entered. The list is sorted by `UMF ID` number.

UMF ID
Displays the sequential, system-assigned number associated with this UMF exception. The UMF ID maps directly to the UMF_EXCEPT table, where UMF exceptions are logged.

From Pipeline
Displays the name of the pipeline that was processing the record when the UMF exception occurred.

Created On
Displays the date that the UMF exception occurred.

Output Document
Displays the type of UMF output document associated with this UMF exception.

Data Source Code
Displays the data source code associated with the incoming data file where the UMF exception occurred.

External Reference
Displays the external reference for the specific data record where the UMF exception occurred. This information can help pinpoint the record inside the data file that needs to be corrected.

Action
Displays the action associated with the incoming data record where the UMF exception occurred. (This action is coded into the UMF for the data record.)

- `A`: Add
- `C`: Change
- `D`: Delete

Events tab

Use this tab to view messages exchanged between the application monitor and the pipelines registered for monitoring or routing. Typically, these messages are also logged to system log files, depending on how your system is configured for logging. First, you generate an on-screen report of application monitor events to view. Then you can select a specific event and take a closer look at its details; this information can be informational or may be helpful in resolving pipeline errors or warnings.

Application monitor events typically include messages or errors exchanged during pipeline processing - such as pipeline start, pipeline stop, or warnings or errors generated during pipeline processing. The only type of errors and warnings not included on this tab are UMF exceptions, which are data-driven exceptions rather than processing information or exceptions.

On-screen report criteria

Use these fields to specify the criteria for the on-screen application monitor events report. Once you specify the criteria, click the **Search** button to generate the report. By default, this tab displays application monitor events that occurred today for pipelines that are registered for application monitoring.

From Date

Starting date to report on application monitor events within the other criteria specified. (This field is optional and can be blank. Leaving the field blank means display application monitor events starting with the first day the system became operational within the other criteria specified.)

Thru Date

Ending date to report on application monitor events within the other criteria specified. (This field is optional and can be blank. Leaving the field blank means display application monitor events through today within the other criteria specified.)

From Pipeline

Name of the registered pipeline to view application monitor events for. (This field is optional and can be blank. Leaving the field blank means display application monitor events for all registered pipelines within the other criteria specified.)

Keep in mind that you can only see application monitor events on this tab for pipelines that are registered for application monitoring.

Max Count

Drop-down list containing options for the maximum number of application monitor events to display within the other criteria specified. Only the number of application events up to and including the maximum number display on-screen. If there are more application events that meet the criteria, the system does not display them.

Search button

By clicking this button, the system executes the search by finding and displaying all application event monitor records that match the entered criteria.

On-screen report results display

This section of the window displays the on-screen application monitor events report, based on the criteria you entered. The list is sorted by ID number.

ID Displays the sequential, system-assigned number associated with this application monitor event.

From Pipeline

Displays the registered pipeline that is affected by or that is involved in the application monitor event. This is the pipeline that you may need to troubleshoot.

Date/Time

Displays the date and time stamp when the application monitor event occurred.

Event Displays the type of application monitor event that occurred. The **Event Description** and **Error Level** columns contain more information about this event and indicate the severity of the event type. There are currently two possible application monitor event types:

- `NODE-INFO` is a note or other type of informational event that occurred in the affected pipeline. This event type typically displays when the affected pipeline is started or stopped.
- `NODE-ERROR` is an error that occurred in the affected pipeline. Check the **Error Level** column to see if this requires immediate action. Typically, you should take a closer look at the information about this application monitor event; it may assist you in resolving a problem with this pipeline.

Event Description
Provides up to 30 characters of more information about the application monitor event.

Error Level
Displays the type of error level of the application monitor event. There are currently two possible event types:

- `NOTE` is the error level associated with the `NODE-INFO` event. Usually, this error level type is informational, so it generally requires no user action.
- `ERR` is the error level associated with the `NODE-ERROR` event. This type of error level usually indicates that you should take a closer look at the details of this application monitor event to resolve the error. You can see the full details of the event by clicking on it.

Events details tab

When you select a specific application monitor event from the **Events** tab, a new screen displays the details of that selected event. These details are taken directly from the system logging files, except for the UMF exceptions log file. (This log file has its own **UMF Exceptions** tab for you to view.) The detail here may assist you in troubleshooting a pipeline error.

ID The sequential number assigned by the system to this application monitor event.

Pipeline
Lists the name of the pipeline where this application monitor event took place.

Date/Time
Displays the date and time of this CME event in `Month, DD, YYYY HH:MM:SS A/PM Time Zone` format. This date and time corresponds with the date and time the event was logged in the log file.

Event Displays the type of application monitor event:

- `NODE-INFO` is a note or other type of informational event that occurred in the affected pipeline. This type of event typically displays when the affected pipeline is started or stopped.
- `NODE-ERROR` is an error that occurred in the affected pipeline. Check the **Error Level** column to see if this requires immediate action. Typically, you should take a closer look at the information for this event; it may assist you in resolving a problem with this pipeline.

Event Description
Displays the first several characters of the application monitor event, as

logged in the log file. This description is intended to provide more specific information about what triggered the event type.

Error Level

Displays the type of error level of the application monitor event:

- `NOTE` is the error level associated with the `NODE-INFO` event. Usually, this error level type is informational, so it generally requires no user action.
- `ERR` is the error level associated with the `NODE-ERROR` event. This type of error level usually indicates that you should take a closer look at the details of this event to resolve the error. You can see the full details of the event by clicking on it.

New Accounts tab

Use this tab to review data loads from the last seven days. At a glance, you can verify which data sources have contributed files for processing, and the number of new identities resulting from that processing. These statistics can give you an idea of the processing volume, to quickly see if the incoming data volumes are appropriate to the amount of expected incoming data.

When you click on this tab, the last seven days display. If there are more records that fit on the visible page, use the scroll bar to view the other records. The **New Accounts** tab is sorted in alphanumeric order by data source code.

Data Source Code

Displays the data source code associated with this new identity record. This information is based on the data source code Universal Message Format (UMF) tag in the incoming file that was processed.

Note: You can see a full list of all data source codes in the Configuration Console by clicking on the **Setup** tab and then the **Sources** tab.

Description

Displays the data source description, as configured for this data source in the Configuration Console. The description should provide more information to help you identity the data source from which these identity records came.

Load Date

Displays the data this data source file was processed and contributed the number of new identities in the **Record Count** column. The date is represented in `Month DD, YYYY` format.

Record Count

Displays the total number of new identities processed from this data source code on the date indicated in the **Load Date** column. This is the number that can indicate the processing volume.

Chapter 7. Loading data

To use IBM InfoSphere Identity Insight you must convert data to Universal Message Format (UMF) format and load it into the system.

Adding a new data source

You must add a new data source when you have a new source of data for the entity database.

About this task

All results are a product of quality data. Getting data of high quality into the entity database is therefore one of the most fundamental tasks, but to do so requires significant analysis of the data and configuration .

Procedure

1. Identify the source of the data. It is critical to know where to go for resolution of the data problems.
2. Analyze the metadata. Every configured data source in the entity database must have a unique identifier on its records, so that the entity database can fully attribute all data back to its original source. Find the field that will provide record uniqueness, and ensure that it is indeed unique.
3. Use an acquisition program to transform the data from its native format into UMF.
4. Configure the data.
   a. Define a role for the data source.
   b. Configure the data source.
   c. Create any required number types.
   d. Create any required characteristic types.
   e. Review the resolution configuration, and customization if necessary.
   f. Configure new DQM rules.
   g. Validate the new DQM rules.
   h. Configure the role alert rules.
5. Verify the data.
   a. Check that the pipeline started.
   b. Verify that the pipeline was able to use configured transports and that it received UMF from the acquisition program.
   c. Verify that the acquisition node produced well-formed XML messages by examining the `.bad` file.
   d. Verify that no UMF exceptions occurred as a result of invalid mapping or configuration.
   e. Check for expected results by viewing the data source and load summary reports.
   f. Search for one or more of the entities resolved using the visualizer.
   g. If applicable, examine role alerts.

Converting data to UMF

In order for the system to process incoming data it must be converted to Universal Message Format (UMF). The process of converting incoming data to UMF can be accomplished using a variety of tools, including the basic utilities that are provided with the product or with standard XML transform products.

Procedure

1. Using the entity model you have created for the system, analyze your incoming data to see how it fits with the UMF standard. You must have a clear understanding of existing UMF segments and tags before proceeding to the next step.
2. Configure your conversion utility to produce UMF records that match your entity model.
3. Run the conversion utility.

What to do next

After converting your data to UMF, you can send the UMF records to the pipeline for processing.

Acquisition programs

An acquisition program contains the tools and programs that acquire data, transform it into Universal Message Format (UMF), and then submit the transformed data to the pipeline for processing.

You can use the acquisition program utilities provided with the product to transform data into UMF, or you can use extract, transform, and load (ETL) tools, such as WebSphere QualityStage, as your acquisition programs.

Transferring UMF files to a queue

You can transfer UMF files to a queue using the queue utility.

Procedure

1. Make sure the data you want to send is in wide format (one record per line).
2. Specify configuration settings in the configuration file.
3. Run the queue utility.

Queue utility

IBM provides a queue utility that manages the transfer of UMF data from a process or a file to a queue.

Although its main job is to move data to one or more queues, you can also use the queue utility, to:

- Create queues
- Remove records from a queue
- View queue status
- View the records in a queue

The queue utility expects data in a certain format:

- Wide-format UMF, which means one line for each record

- One new line at the end of each record
- No other new lines within a record

You must use one of the following queue managers to use the queue utility.

Microsoft Windows Server x86
Microsoft Message Queuing, a component of Microsoft Windows Server 2003 or 2008.

IBM Websphere MQ 6.0

Microsoft Windows Server x86_64
Microsoft Message Queuing, a component of Microsoft Windows Server 2003 or 2008.

IBM Websphere MQ 7.0

Solaris Operating Environment
IBM Websphere MQ 6.0

Linux IBM Websphere MQ 6.0

AIX IBM Websphere MQ 6.0

When a pipeline runs in queue mode, the queue manager is always required and must be installed and running. When a pipeline runs in file mode, the queue manager must be installed, but it does not have to be running for Windows and AIX platforms. It does not need to be installed or running on Solaris or Linux.

Configuration file for the queue utility

You can use a configuration file to send records to multiple queues, with the queue utility.

When delivering one set of data to many queues, you must tell the queue manager how to set up its distribution. The idea is to create a type of distribution in which the first queue gets one record, then the next queue gets one, and so on.

The queue utility configuration file is named qutil.ini and should be in the same directory as the queue utility executable file.

Parameters

[sectionname]
Name of the section. You can specify multiple groups of configuration settings within a single configuration file, and then refer to those settings on the command line by specifying that section name. For example, you can name sections CFG1 (configuration 1) or CFG2 (configuration 2) and refer to those sections when issuing queue utility commands.

MessageCountMax
Maximum number of records allowed in each queue at any given time. When a queue is full, the utility stops processing records.

FullCountMax
Specifies the total number of records that can be in all the queues, as opposed to just one queue. When all the queues are full, the utility pauses the flow of data and waits for records to move into pipelines for processing, freeing up space in the queues. Works with FullPause.

FullPause

The number of milliseconds that the queue utility pauses the flow of data, allowing data in the queues to be processed when FullCountMax is reached.

Qout*n*=qname

The names of the output queues for this section. The names of the output queues can be whatever makes sense, however the parameter must be Qout*n* where n is an integer starting with 0. The value of *n* must be sequential from 0 to *n* where n is the last queue defined. This format is required. Change only the number of the Qout*n* identifier and the qnames.

Example

The following example shows that you have two sets of instructions (one using 2 queues and one using 4 queues). A maximum of 2,500 records in each queue at any time, the most records in all queues is 10,000, and the queue utility pauses for 3 seconds before it tries to load more records in any queue after FullCountMax has been reached. Then, it lists the names of the 4 queues to use.

```
[CFG1]
MessageCountMax=2500
FullCountMax=10000
FullPause=3000
Qout0=qnameA
Qout1=qnameB
[CFG2]
MessageCountMax=2500
FullCountMax=10000
FullPause=3000
Qout0=qnameA
Qout1=qnameB
Qout2=qnameC
Qout3=qnameD
```

Command syntax of the queue utility

Queue utility commands are made up of operations and modifiers.

The basic syntax of a queue utility command is:

```
qutil -operation qname -modifier
```

qname is the name of the queue.

Command operations

Operations define the various functions of the queue utility. You can add only one operation to a qutil command.

-C Creates a new queue.

Requires a unique name for qname.

Must be an upper case C.

-f Copies stdin to the queue.

Requires a qname.

-i Copies stdin to many queues.

Requires section name as defined in the qutil.ini file. Specifies a section from the qutil.ini to load for delivering messages to many queues.

-k Purge count for each record.

Requires a *qname*.

Can be used in conjunction with -c modifier to limit the number of records processed.

-p Peek count for each record.

Does not remove records from the queue.

Requires a *qname*.

Writes to stdout.

Can be used in conjunction with -c modifier to limit the number of records processed.

-r Read count for each record.

Removes records from the queue.

Requires a *qname*.

Writes to stdout.

Can be used in conjunction with -c modifier to limit the number of records processed.

-s Queue status.

Requires a *qname*.

-x Delete *qname*.

Requires a *qname*.

Command modifiers

Modifiers configure additional parameters for a queue utility operation. You can use more than one modifier on a qutil command.

-T Specifies whether a queue is transactional.

By default, all new queues are non-transactional, unless specified upon creation as transactional with a -T modifier.

Transactional queues must not be used when a queue might receive routing information from an application monitor.

Transactional queues in Microsoft Message Queueing do not allow messages to be prioritized or to be processed in any order other than the order in which they were received.

-c Specifies to stop after a count of records have been processed.

Requires an integer.

Must be a lower case c.

-l Specifies the priority level for each record.

Requires an integer.

Valid integer values are:

0-7
Microsoft Message Queueing
the priority levels are 0 - 7, where 0 is the lowest priority, and 7 is the highest priority.

3 is the default.

0-9
IBM Websphere MQ
The priority levels are 0 - 9, where 0 is the lowest priority, and 9 is the highest priority.

The default value depends on a queue property. You can change this property in the IBM Websphere MQ manager.

-m Specifies the queue manager.

AIX, HP-UX, Linux, and Solaris only

-o Specifies the number of seconds before a message expires.

Requires an integer.

-q Specifies the queue type.

Microsoft Windows only

Valid values are:

mq IBM WebSphere MQ

msmq Microsoft Message Queueing (MSMQ)

-t Specifies the number of milliseconds to wait between each record.

Requires an integer.

Command Operation and Modifier relationships

Certain modifiers are only recommended for use with certain operations. The following table describes the relationship of each operation to its potential modifiers:

Table 21. Queue Utility command operation and modifier relationships

Operation	Valid Modifiers
`-C`	`-T, -q` *EXAMPLE:* qutil -C *qname* -T -q mq
`-f`	`-c, -t, -l, -o, -q` *EXAMPLE:* qutil -f *qname* -c 50 -t 20 -l 4 -o 10 -q msmq
`-i`	NONE *EXAMPLE:* qutil -i configsection
`-k`	`-c` *EXAMPLE:* qutil -k *qname* -c 50
`-p`	`-c` *EXAMPLE:* qutil -p *qname* -c 50

Table 21. Queue Utility command operation and modifier relationships (continued)

-r	-c *EXAMPLE:* qutil -r *qname* -c 50
-s	NONE *EXAMPLE:* qutil -s *qname*
-x	NONE *EXAMPLE:* qutil -x *qname*

Converting UMF files to appropriate formats

You can use the UMF formatting utility to switch the UMF records between wide and tall formats.

UMF formatting utility

You can use the UMF formatting utility to convert UMF records to and from wide and tall formats. The UMF formatting utility can also extract UMF data that is defined by a given tag.

UMF records can be displayed as a single line (wide format) or as a series of indented lines where each line contains an XML element and value (tall format).

Example: wide format

```
<name><name_type>M</name_type><first_name>John</first_name>
<last_name>Smith</last_name></name>
```

Example: tall format

```
<name>
 <name_type>M</name_type>
 <first_name>John</first_name>
 <last_name>Smith</last_name>
</name>
```

Command syntax of the UMF formatting utility

The UMF formatting utility uses a variety of commands to format and extract data.

The basic syntax of an UMF Formatting Utility command is:

```
xutil -o[switch] option
```

Parameters

-o **Out** Sends output to stdout. Required parameter. Parameter switches are:

- **w** Defines the format of output. All the UMF for one record is on one line. Removes all returns and line feeds.
- **t** Defines the format of output. The UMF for one record is on many lines. Puts one tag per line and places tabs in the document to make it more readable.

-t **Tagname:** Filters records based on a tag name. Only records framed with this tag are output to stdout. Any errors are sent to stderr.

Use the tagname parameter when you want to filter records. For example, you may have a file with mixed records: entities and activities. It is a good idea to process entities before activities so the activities have existing entities for matching.

Examples

This command filters output for entities only, using mixedlist.xml as the input source and entity.xml as the output file.

```
xutil -ow -t UMF_ENTITY < mixedlist.xml > entity.xml
```

This command directs the output of the UMF Formatting Utility process to a pipeline or to the queue utility.

```
xutil -ow < file.xml |qutil -f qname
```

Extending the entity model

An entity model is a set of data that defines what you consider to be an entity. Use these instructions to extend the default entity model. This is not a common task, but you can extend the entity model for your environment.

Universal Message Format (UMF)

Universal Message Format (UMF) is an extensible XML dialect used for structuring data source files. UMF contains standard tags that represent key pieces of identities, relationships, and activities. Before data can be processed by the pipelines, it must be converted into UMF and follow the UMF specification.

UMF consists of these hierarchical components:

UMF documents
The collection of UMF segments that structure the data and indicate the type of data source record.

UMF segments
The part of the UMF document that structures the data for the data source.

UMF elements
XML tags and values that define the data within a UMF segment of a UMF document.

The UMF specification lists the specific types of UMF documents, the UMF segments within each UMF document type, and the valid UMF elements within each UMF segment.

Analyzing source data

The first task to getting your source data into the entity database is to analyze your source data for mapping to UMF.

Procedure

1. Identify data you want to load into the entity database.
2. Ensure the data is consistent and complete.
3. Identify width of incoming UMF segment element values versus the width of their corresponding database table columns.
4. Identify invalid characters in the source data.

Results

The results of your analysis might present several options such as:

- Using DQM rules to correct data with invalid characters.

- Using DQM rules to truncate data with a larger width than their corresponding database table columns.
- Asking external data source providers to supply more complete data.
- Loading only fields with valid data.

Reviewing the default UMF specification

You should review the default UMF specification to assist in creating your customized UMF specification and entity model. These items map out data transfer from data sources to UMF tags which will be ingested by the entity database.

Mapping UMF segments to the entity database

Whenever your data requires new UMF segments, you must create new data mappings for the data in those UMF segments. Without a valid data mapping, you cannot successfully load data into the entity database.

Risks in modifying the entity database

Modifying the entity database has risks and should not be performed without sufficient experience or expertise.

- Tables should not be added to the entity database without sufficient experience or expertise
- Adding fields to database tables is a process that involves more than just the table in question. It is recommended to use existing tables and fields to classify new data if at all possible.
- Database table indexes should not be modified. Modifying the indexes on the database tables may cause unpredictable and undesirable results, such as the Visualizer hanging.
- It is recommended that DQM changes only be performed with sufficient expertise or with the help of IBM.
- Always use a test database when verifying new configurations before applying those new configurations to your production environment.

Adding tables to the entity database

You may need to add a new database table when adding a new data source.

About this task

Adding tables to the entity database does not allow resolution on the new data, it is just a place to store data.

It is recommended that you should use a test database when verifying new configurations before applying those new configurations to your production environment.

It is recommended to use existing tables and fields to classify new data if at all possible.

Adding a new table will accommodate expected data that is not yet configured in the system. You should create the new database table to be consistent with your current data model.

Be sure to include the required relevant fields:

- ENTITY_ID
- DSRC_ACCT_ID

- HIST_STAT - required if using sequential history tracking.
- SYS_CREATE_DT
- SYS_DELETE_DT
- SYS_LSTUPD_DT
- SYS_LSTUPD_US

Procedure

1. Create the new table in the entity database.
2. Create the data mapping for the new table.
3. Add new database tables to the dictionary.
4. Define the data mappings for the new table.
5. Determine the appropriate DQM rules to apply to the new segment and configure those rules through the console.
6. Verify the new configuration by running known test data through a pipeline and checking the resultant log files.
   a. Verify the test runs without errors.
   b. Check for UMF Exceptions in console.
   c. Check log files `nodename.Sql.Err.log` and `nodename.err` for errors.
   d. Verify test results match the expected results.
   e. Check the UMF_LOG table to ensure all records load properly.

Adding fields to entity database tables:

You may need to add a new field to an existing entity database table to accommodate a new data.

About this task

A new field can be added to an existing table when a new UMF segment does not require an entirely new table.

Adding fields to an existing entity database table does not allow resolution on the new data, it is just a place to store data.

It is recommended that you should use a test database when verifying new configurations before applying those new configurations to your production environment.

It is recommended to use existing tables and fields to classify new data if at all possible.

Procedure

1. Add the new field to the appropriate database table.
2. Create the data mapping for the new field in the console.
3. Determine the appropriate DQM rules to apply to the new field and configure those rules through the console.
4. Verify the new configuration by running known test data through a pipeline and checking the resultant log files.
   a. Verify the test runs without errors.
   b. Check for UMF Exceptions in console.
   c. Check log files `nodename.Sql.Err.log` and `nodename.err` for errors.

d. Verify test results match the expected results.
e. Check the UMF_LOG table to ensure all records load properly.

Adding new database tables to the dictionary:

When your data (and UMF) requires that you create a new database table, you must add this table to the dictionary of database tables that the system uses. If the table does not exist in the dictionary, you cannot create a data mapping for the UMF and the table.

Before you begin

The user must be granted appropriate access to read and store data in the database table.

Procedure

1. Click the **Setup** button.
2. Click the **UMF** button.
3. Click the **Dictionary** tab.
4. Click the **New** button
5. In the **Table Name** field, type the name of the new database table.

Defining data mappings

You must create a data mapping for new UMF segments and tags. When new source systems are added to the product, new UMF segments and tags are sometimes created as a result. A data mapping maps the data in a UMF to the corresponding tables and table columns in the entity database.

Data mappings:

A data mapping maps the data in a UMF file to the corresponding tables and table columns in the entity database.

Without a valid data mapping, you cannot successfully load data into the entity database. Whenever your data requires new UMF segments, you must create new data mappings for the data in those UMF segments.

Example

Finn's Auto Service has recently started collecting insurance company data for their customers. For example, the UMF data for a new insurance company might use these UMF segments:

```
<ATTRIBUTE>
<INSURANCECOMPANY>Mooninite Casualty Company</INSURANCECOMPANY>
</ATTRIBUTE>
```

You must create a new data mapping for the <ATTRIBUTE><INSURANCECOMPANY> UMF data path to the appropriate table column in the entity database. The XPath value for the UMF Data Path is `./ATTRIBUTE/INSURANCECOMPANY/`

Viewing data mappings:

A data mapping maps the data in a UMF file to the corresponding tables and table columns in the entity database.

Procedure

1. Click the **Setup** button.
2. Click the **UMF** button.
3. Click the **Data Map** tab.
4. From the **Segment** drop-down, select the UMF segment that you want to view.
5. From the **Table** drop-down, select the UMF segment table whose mapping you want to view.

Creating data mappings:

A data mapping maps UMF data to the corresponding tables and table columns in the entity database. A new data mapping is required when incoming data with new UMF tags will be processed by the system.

Before you begin

If this data mapping maps data to multiple tables, it is necessary to check that the tables will be inserted in the correct load sequence during pipeline operations. If the table does not exist in the dictionary, you must add the new table to the dictionary to be able to create a data mapping for the UMF and the table.

Procedure

1. Click the **Setup** button.
2. Click the **UMF** button.
3. Click the **Data Map** tab.
4. From the **Segment** drop-down, select the UMF segment where you want to add a new data mapping to a table.
5. From the **Table** drop-down, select the UMF segment table where you want to add a new data mapping.
6. Complete one of the following steps:
   - To create a new data mapping, click the **New** button.
   - To create a data mapping based on an existing data mapping, select a data mapping from the list, and then click the **Clone** button.
7. If this is a new segment, type the name of the UMF segment into the **Segment** field.
8. Select the desired database table from the **Table** drop-down list
9. In the **Table Column** field, type the name of the database table column to which you want to map the UMF data path to.
10. From the **Field Type** drop-down, select the appropriate field type that represents the field type of the table column in the database.
11. From the **Data Type** drop-down, choose the appropriate data type that represents the value of the data.
12. In the **UMF Data Path** field, enter the UMF tag.
13. From the **Update Method** drop-down , choose the appropriate update method to determine which value, between the inbound value and the previously stored value, will be retained.
14. In the **Status** field, from the drop-down list, choose the appropriate status of the data mapping.
15. Click the **Save** button.

Deleting data mappings:

A data mapping maps UMF data to the corresponding tables and table columns in the entity database. You can delete a data mapping no longer used by the system.

Procedure

1. Click the **Setup** button.
2. Click the **UMF** button.
3. Click the **Data Map** tab.
4. From the **Segment** drop-down, select the UMF segment where you want to select a table to delete a data mapping.
5. From the **Table** drop-down, select the UMF segment table where you want to delete a data mapping.
6. Select a data mapping from the list, and then click the **Delete** button.

Help topics:

Data Mappings- General tab:

Use the **General** tab to specify the details of the data mapping.

Segment
Type the name of the segment you want to create a data mapping for. The segment name must be entered in uppercase.

Table From the drop-down list, select the table for the data mapping you want to create.

Table Column name
Type the name of the table column you want to create.

Table column type
From the drop-down list, select the table column type for the table column name..

Unique ID
The table column is an automatically incrementing unique key generated by the database engine. Only one table column can be configured with this value.

Entity Key
If selected, the table column is always set to the ENTITY_ID.

Business Key
The table column along with other designated business key table columns make up a composite lookup key to determine the existence of the same record

Attribute
The table column is used to simply store data and does not have any functional effect on the insert/update/delete of the table.

Key Attribute
The table column value is used to determine if there is an existing record with the same value. The database keeps track of changes to these values through time. For example: if you want to keep a version of the record if the ADDR1 value changes, you would identify the ADDR1 value as a key attribute.

This value has nothing to do with indexes.

History Sequence
The table column is used to determine which record provided by a given source is the most recent current and which are historical.

History Sequence is always assigned to the HIST_STAT table column.

Delete Timestamp
The table column is used to store the last date/time the record was deleted.

Update Timestamp
The table column is used to store the last date/time the record was updated.

Data type
From the drop-down list, select the data type for the table column.

CHAR
Character data (alphanumeric).

INT Integer data.

DATE Date data. For example: yyyy-mm-dd or mm-dd-yyyy.

DATE/TIME
Date/time data. For example: yyyy-mm-dd hh:mm:ss or mm-dd-yyyy hh:mm:ss.

UMF Data Path
Type the XPath location of the UMF tag.

Update Method
From the drop-down list, select the update method for the data mapping you want to create. The update method determines which value, between the inbound value and the previously stored value, will be retained.

Never If a value for the UMF element exists in the database table, that value cannot be updated.

Always
If a value for the UMF element exists in the database table, that value can be updated.

Maximum Value
The greater value, incoming or stored, will be retained or updated.

Restricted to Table Column Data Types equal to INT, DATE, or DATE/TIME.

Minimum Value
The lesser value, incoming or stored, will be retained or updated.

Restricted to Table Column Data Types equal to INT, DATE, or DATE/TIME.

Status From the drop-down list, select the status for the data mapping you want to create.

Active The data mapping is active.

Inactive
The data mapping is inactive.

Address standardization with IBM InfoSphere QualityStage and AddressDoctor

Address hygiene and standardization is a pipeline process that allows you to correct and standardize address information for optimal entity resolution processing. This new IBM® InfoSphere™ Identity Insight feature enables the use of an industry standard address data standardization solution that includes AddressDoctor®, IBM InfoSphere Information Server, IBM InfoSphere DataStage®, and IBM WebSphere® QualityStage™.

Support for an address standardization module provided by AddressDoctor eliminates the dependencies and limitations of other modules such as Worldwide Address Verification and Enhancement System (WAVES). The AddressDoctor address standardization module can be used for Identity Insight entity resolution by using DataStage and the QualityStage Address Verification interface (QS-AVI). QualityStage is a component IBM Information Server.

AddressDoctor® has the following advantages:

- Supports over 240 countries and territories.
- Has better coverage at the street level.
- Is Unicode enabled and supports all of the major character sets.
- Provides transliteration.
- Provides a validation status on how deliverable is the address.
- Provides formats to the local Postal Standard.

Implementing AddressDoctor with QS-AVI is not a trivial task. It is recommended that you contact your IBM representative for assistance.

QS-AVI address cleansing requirements and task overview

The detailed process steps for Using IBM QualityStage and AddressDoctor interface (QS-AVI) to do Identity Insight address cleansing are described in a techdoc at ibm.com. This topic provides an overview of the process, requirements, and a link to the detailed information.

Before you begin

The following products are required:

- IBM InfoSphere Information Server including IBM InfoSphere DataStage and IBM InfoSphere QualityStage Version 8.0.1
- QS-AVI Data Quality stages
- AddressDoctor(R) Database for the required country.

About this task

The process follows this general sequence:

Procedure

1. Define a QS-AVI stage job in DataStage and QualityStage Designer.
2. Import the “AddressValidateWS.dsx” file to the stage. (This is a predefined Address cleansing job and has been designed for EAS and QS-AVI integration.) The file can be found on the fix pack installation disk: `<RR_INSTALL>/srd-home/qsavi/AddressValidateWS.dsx`

3. Modify the Address Verification stage Enable the DataStage job for Information Services.
4. Define the DataStage job as a service in the Information Server Console.
5. Verify the deployment using WebSphere Information Services Director (WISD) to generate and examine a Web service definition language (WSDL) document for this new service.
6. Test the Service in an environment such as WebSphere Integration Developer.
7. Activate the QSAVI feature, by changing the AddrConnection under the OAC section of the pipeline.ini file to following format:

   [OAC]
   AddrConnection=qsavi://host:port/?timeout=ms

   host is the host name or IP address of the Infoserver.

   port is the port number. The default port is 9080.

   timeout
   is an optional parameter. You can set the connection timeout parameter externally. The default connection timeout is 10000 ms (10 seconds).

What to do next

The detailed steps for this process are described in: QS-AVI address cleansing as a Web process for IBM InfoSphere Identity Insight.

QS-AVI troubleshooting

QS-AVI returns ‘valstatus_qsav' which describes the quality of address cleansing and enables troubleshooting for related problems.

Exceptions

An exception is generated based on the handle value status:

```
// handle value status
// V  - Validated
// C  - Corrected
// P3 - Not corrected - Deliverability High
// P2 - Not corrected - Deliverability Fair
// P1 - Not corrected - Deliverability Small
// N1 - Not checked - Country not recognized
// N2 - Not checked - Country DB not found
// N3 - Not checked - Country not unlocked
// N4 - Not checked - Validation not called
// N5 - Insufficient information
// Q1 - No suggestions
// Q2 - Suggestions incomplete
// Q3 – Suggestions
```

QS-AVI also returns ‘resultstatus_qsav' which describes the address cleansing probability:

```
// handle delivery probability
// 0 - Empty
// 1 - Not checked
// 2 - Not checked, but standardized
// 3 - Checked and corrected
// 4 - Validated, but changed
// 5 - Validated, but standardized
// 6 - Validated and unchanged
// 7 - No value given because of multiple matches
```

Error messages

6301E - Invalid response.

6302E - Cannot connect to InforServer server

This messages is generated when when EAS fails to connect to InfoServer. It is also generated with a ‘soapenv:Fault' response from InforServer, which is treated as an invalid response.

6303E - Error, failure to connect to the server : {0}", __serverName

This messages is generated when when EAS fails to connect to the correct InfoServer server.

Chapter 8. Analyzing data

The Analyst Toolkit provides a set of application development and customization capabilities to Identity Insight. These are a set of user interfaces and reports that can be modified as needed or referenced by other applications.

Analyzing data using the Visualizer

You can use the Visualizer to accomplish various analysis tasks: review and disposition alerts, find entities, view entity data, view graphs of entities and their relationships to other entities, create and manage attribute alert generators, add a single entity or a small file of entities, disclose relationships between entities, and print reports.

Setting up the Visualizer

To successfully use the Visualizer, you need to know how to access the Visualizer and how to customize the way that the Visualizer displays information to suit your preferences.

Visualizer

The Visualizer is a graphical user interface that analysts and investigators use to analyze the results of alerts, relationships, and entity resolutions.

The Visualizer is hosted by an embedded version of IBM WebSphere Application Server. You configure the Visualizer through the Configuration Console and through the Visualizer **Preferences** selection on the **File** menu.

Visualizer users can accomplish various analysis tasks:

Analyze and disposition alerts

Alerts generated by entity resolution processing represent relationships or entity resolutions of interest to an organization. Typically, analysts review alerts and decide what action to take, if any, based on the alert information. There are three types of alerts: role alerts, attribute alerts, and event alerts.

The Visualizer displays the alerts, providing analysts with both textual and graphical views of the alerts and the entities involved in the alerts. Analysts can drill down into the details and then set the disposition status of the alert appropriately.

Create and manage attribute alert generators

Using the Visualizer, analysts can create and manage persistent searches through the Attribute Alert Generator feature, and manage how they view and receive attribute alerts. Analysts can create Attribute Alert Generators based on attribute data to locate identities that resolved to entities based on that attribute data. Or analysts can create an Attribute Alert Generator to persistently search the entity database looking for a particular entity.

Find entities

Visualizer users can also find entities for further analysis using several methods:

- By attributes
- By data source account

- By entity ID
- By resolution (how closely the criteria entered matches identities and entities in the entity database, based on minimum resolution score thresholds)

Add entities and disclosed relationships

Analysts can use the Visualizer to add records for entity resolution and relationship detection. They can add a single identity record or load a UMF file containing a few thousand identity records. Just as when identity records are added through acquisition programs, records added through the Visualizer are processed by a pipeline for entity resolution and relationship detection. The results of processing are written to the entity database, and any alerts are published to the Visualizer.

Analysts can also disclose relationships between entities (by identity), when they know of a link between the identities. Examples of disclosed relationships include relating entities based on emergency contacts or references listed on an employment application. The entity disclosed these relationships on the application.

Generate and print reports

The Visualizer also contains several reports that analysts can view and print to help them manage and track their Visualizer work.

Configuring the Visualizer

You can configure Visualizer settings to tailor how information displays in your Visualizer sessions.

Setting Visualizer display options:

You can customize the Visualizer display by changing the background color, font, and other display options on the **Window Preferences** tab.

About this task

Visualizer display options are configured for each Visualizer client. By using these instructions, you only change the display for the Visualizer client that you are currently logged on to.

Procedure

1. In the Visualizer, select **File** > **Preferences** > **Window Preferences**.
2. Choose the look and feel display options to use. You can only change the settings in the **Theme**, **Font**, and **Size** drop-down lists if you select the *Metal* option in **Look and Feel**.
3. Click **Submit**. A confirmation message informs you that you must restart the Visualizer before your changes take effect.
4. Click **OK**.
5. Close the Visualizer. Start the Visualizer and log in again.

Results

The Visualizer now displays using the new window display options that you selected.

Setting the default path for UMF files:

If you regularly load identity records in UMF data files for processing through the Visualizer, setting the default path can save you a step.

About this task

Default path settings are configured for each Visualizer client. By specifying a default path using this task, you only set the path on the Visualizer you are currently logged in to.

Procedure

1. In the Visualizer, select **File** > **Preferences** > **System Preferences**.
2. In **Default path for File Load**, do one of the following:
   - Enter the full path of the directory to use.
   - Or browse to select the directory.
3. Click **Submit**. A confirmation message informs you that you must restart the Visualizer before your changes take effect.
4. In the confirmation message, click **OK**.
5. Close the Visualizer, restart the Visualizer, and log in again.

Results

Whenever you load a UMF file, the default path is the directory that you specified.

Setting the default path for Centrifuge:

If you use the optional Centrifuge Desktop from Centrifuge Systems to visualize and display entity graphs, you must specify the Centrifuge Desktop file path in the Visualizer preferences.

About this task

Default path settings are configured for each Visualizer client. By specifying a default path using this task, you only set the path on the Visualizer you are currently logged in to.

Procedure

1. In the Visualizer, click **File** > **Preferences** > **System Preferences**.
2. Under the **File Paths** section in **Centrifuge path**:
   - Enter the file path or URL (uniform resource locator) to the Centrifuge Desktop application in the field.
   - Or browse to the Centrifuge Desktop application and open it.
3. Click **Submit**. A confirmation message informs you that you must restart the Visualizer before your changes take effect.
4. In the confirmation message, click **OK**.
5. Close the Visualizer, reopen the Visualizer, and log in again.

Results

After the path is configured, the **Centrifuge** button displays on the **Role Alert Detail** and **Entity Resume** screens in the **Research** window. Click the button to launch your Centrifuge Desktop application directly from the Visualizer.

Setting minimum threshold score values for Visualizer queries:

When you look for an entity using either the Find By Resolution feature or an attribute alert generator in the Visualizer, you select a minimum likeness score as part of the criteria. Your choice determines the strength of entity and relationship resolution the system uses to find and return entities. You can change the default values for one or more of these thresholds on the Visualizer **System Preferences** tab.

About this task

These settings are configured for each Visualizer client. By using this task, you only change the minimum score threshold for the Visualizer you are currently logged in to.

Procedure

1. In the Visualizer, click **File** > **Preferences** > **System Preferences**.
2. In the **Minimum Score Values** section, specify the lowest likeness score to use to determine which search results are displayed. The higher the number, the more entity data that must match the search criteria, which can reduce the number of results that are returned.
3. Click **Submit**. A confirmation message informs you that you must restart the Visualizer before your changes take effect.
4. In the confirmation message, click **OK**.
5. Close the Visualizer, reopen the Visualizer, and log in again.

Setting default Alert Summary window filter options:

Use the **Alert Display Filter Settings** tab of the **System Preferences** screen to customize the default settings for the filter options on your **Alert Summary** window.

About this task

These settings control the following default values in the Visualizer:

- The maximum number of alerts to display in the **Alert List**
- The minimum relationship score for role alerts to display
- The number of days of alert summaries to display (from the current date backwards)

The values you set here determine the default filter values that your Visualizer instance uses each time you open a new **Alert Summary** window.

Procedure

1. In the Visualizer, select **File** > **Preferences** > **System Preferences**.
2. Under the **Alert Display Filter Settings** section, in **Maximum alerts to display in alert list**, enter a number that represents the maximum number of alerts to display in the **Alert List** table. The default setting is 100, which means that when you select an alert summary, the first 100 associated alerts display in the **Alert List**. You might want to change the default setting to display fewer alerts.
3. In **Minimum relationship score**, enter the lowest relationship score to use as a threshold for displaying role alerts. The higher the relationship score, the fewer the number of role alerts and role alert summaries that display.

4. In **Number of days of alerts to display (including today)**, enter a number 1-99 that indicates the number of days of alerts to see. The number starts with the current date and counts backwards, so if you enter 1, you only see alerts generated on the current day. If you enter 10, you only see alerts for a total of 10 days – the current day and 9 days prior. The default value is 99.
5. Optional: If your system administrator has enabled alert threshold override in the Configuration Console, the **Include filtered role alerts** check box displays.
   - Select the **Include filtered role alerts** check box to display all role alerts and role alert summaries on the **Alert Summary** window with relationship scores outside the minimum alert threshold defined in the role alert rule.
   - Clear the **Include filtered role alerts** check box to display only those role alerts and role alert summaries on the **Alert Summary** window with relationship scores that meet the minimum alert threshold.
6. Click **Submit**. A confirmation message informs you that you must restart the Visualizer before your changes take effect.
7. In the confirmation message, click **OK**.
8. Close your Visualizer session, restart the Visualizer, and log in again.

Setting Visualizer log options:

You can turn Visualizer client logging on or off by configuring Visualizer log options. By default, Visualizer client logging is turned off. Generally, you only turn on Visualizer client logging to assist you and your administrator with troubleshooting.

About this task

These settings are configured for each Visualizer client. By using this task, you only change the logging options for the Visualizer you are currently logged in to.

Procedure

1. In the Visualizer, click **File** > **Preferences** > **Log and Link Settings**.
2. Do one of the following actions in the **Turn on logging** check box:
   - Select the check box to turn on Visualizer client logging.
   - Clear the check box to turn off Visualizer client logging.
3. If you turned on logging, specify the type of logging by selecting an option in **Log detail level**. If you are not sure which level to select, consult your system administrator. Because generally, you only turn on Visualizer client logging to troubleshoot a problem, you typically select the debugging level. The debug level logs every action that you take in the Visualizer and every message (error, warning, or informational) that occurs. This logging level quickly fills the Visualizer log file, which means you might need to occasionally delete the file.
4. In **Log file directory path**:
   - Enter the path to store Visualizer log files.
   - Or browse to the directory and select it.
5. Click **Submit**. A confirmation message informs you that you must restart the Visualizer before your changes take effect.
6. In the confirmation message, click **OK**.
7. Close the Visualizer, restart the Visualizer, and log in again.

Setting Visualizer hyperlink options for displaying custom attributes:

If your organization includes links to files or pictures in other systems as part of identity record attributes, the Visualizer can display hyperlinks to those files. You click the hyperlink to launch your Web browser or application to display the selected file or picture. Use Visualizer system preferences to choose which browser or program opens the files when you click a hyperlink.

About this task

These settings are configured for each Visualizer client. By using this task, you only change the hyperlink options for the Visualizer you are currently logged in to.

Procedure

1. In the Visualizer, select **File** > **Preferences** > **Log and Link Settings**
2. Under **Hyperlink Handling Settings**, select one of the following options:
   - **Default browser setting**
   - Or **Use Program** and specify a browser or program to use to open hyperlinks.

   Note: You might only need to specify a Web browser or other program to open links that are stored on secure Web sites (https://).
3. Click **Submit**. A confirmation message informs you that you must restart the Visualizer before your changes take effect.
4. In the confirmation message, click **OK**.
5. Close the Visualizer, restart the Visualizer, and log in again.

Setting Visualizer graph options:

You can customize the graph settings you see in the Visualizer by changing the color or thickness of lines on the **Graph Preferences** tab.

About this task

Visualizer graph display settings are configured for each Visualizer client. By using these instructions, you only affect the settings for the Visualizer client that you are currently logged in to.

Procedure

1. In the Visualizer, click **File** > **Preferences** > **Graph Preferences**.
2. Select the line thickness and color to use.
3. Click **Submit**. A confirmation message informs you that you must restart the Visualizer before your changes take effect.
4. In the confirmation message, click **OK**.
5. Close the Visualizer, reopen the Visualizer, and log in again.

Results

The Visualizer now displays graphs using the new display options that you selected.

Help topics:

Window Preferences tab:

Use this tab to configure the way the Visualizer displays background colors, fonts, and navigational icons for your Visualizer sessions. Configuring preferences on this tab only affects the settings for the local Visualizer client. If you change any of these settings, exit, reopen, and log in to the Visualizer to see the changes.

Look and Feel
Choose a pre-formatted group of display settings. Group display settings control the selections available in **Theme**, **Font**, and **Size**.

Note: Most display settings do not allow you to select any of the other fields. Currently, **Metal** is the only option that allows you to choose other display settings.

The default group display setting is **EAS Visualizer**.

Theme
Choose a pre-formatted screen color combination for the group display setting that you selected in **Look and Feel**.

Font Choose a display font.

Size Choose a font size.

Sample
Shows you an example of what your Visualizer display will look like, based on your selections.

Background color
Click this button to choose a background color. This field is only available if you selected **Metal** in the **Look and Feel** field.

Control color
Click to choose a control outline color.

Text color
Click to choose a text color.

System Preferences tab:

Use this tab to configure system preferences for your Visualizer sessions. Configuring preferences here only affects the system settings for your local Visualizer client. If you change any of these settings, exit, reopen, and log in to the Visualizer to see the changes.

File Paths section
Specify the default file paths that the Visualizer uses to load UMF files and open the Centrifuge Desktop graphing tool. If you use the Centrifuge Desktop application to visualize entity graphs and entity data, enter the full path to the application. By entering the full path here, you can access Centrifuge directly from the Visualizer.

Minimum Score Values section
Define the values for the minimum likeness score thresholds that you can select from when you create a Find by Resolution query or an attribute alert generator.

By default, this section contains the recommended values for each of these thresholds. These recommended values are conservative values, intended to return fewer false positives. You can redefine the values to suit your goals.

Generally, the higher you set the value of a minimum score threshold, the fewer the results that are returned. The lower the value is set, the more results that are returned.

Is Entity

Enter the lowest resolution score that defines when the search entity defined in a Find by Resolution query or an attribute alert generator and an entity in the entity database are the same entity.

The default value is 100. This default means that when the search entity and an identity are compared, if the resolution score is 100, the entity returned is the same as the search entity.

Close Entity Match

Enter the lowest resolution score that defines when there is a "close match" between the search entity defined by a Find by Resolution query or an attribute alert generator and an entity in the entity database.

The default value is 85. This default means that when the search entity and an entity from the entity database are compared, if the minimum resolution score is equal to or greater than 85, but less than the **Is Entity** score, the entity returned is a close entity match to the search entity.

Good Relationship

Enter the lowest score that defines when there is a close or strong relationship between the search entity defined by a Find by Resolution query or an attribute alert generator and an entity in the entity database. The value represents the strength of the relationship.

The default value is 35, which means that when the search entity and an entity from the entity database are compared, if the minimum resolution score is equal to or greater than 35, the relationship between the two is good.

Any Relationship

Enter the lowest score that defines when there is any relationship between the search entity defined by a Find by Resolution query or an attribute alert generator and an entity in the entity database. (The value represents the strength of the relationship.)

The default value is 1. This default means that when the search entity and an entity from the entity database are compared, if the minimum resolution score is equal to or greater than 1, the two have a relationship.

Alert Display Filter Settings section

Use this section to configure the default alert filter settings that affect which alert summaries display in the **Alert Summary** window. Each time you open a new **Alert Summary** window, the system uses these default settings.

Maximum alerts to display in alert list

Enter a number that represents the highest number of alerts to display in the **Alert List** table of the **Alert Summary** window.

The default filter value is 100, which means that, by default, only the first 100 alerts of any selected alert summary display.

Minimum relationship score

Enter the lowest relationship score to filter unassigned role alert summaries lower than that score from the display in your **Alert Summary** window.

For example, to only see role alert summaries where the relationship score between the two compared entities is equal to or greater than 50, enter `50` in this field.

The default is 0, meaning that all alert summaries for your Visualizer Analyst group that are currently in the "Unassigned" status display by default.

Number of days of alerts to display (including today)

Enter a number 1-99 that indicates the number of days of alerts from the current date to display. Keep in mind that this "day" is a full calendar day, which starts at 0:00:00 and ends at 23:59:59.

The number starts with the current date and counts backwards. If you want to see alerts generated within the last 90 days (the current day and 89 days prior), enter `90`.

The default value is `99`, meaning that you see alerts generated today and 98 calendar days before today.

Include filtered role alerts check box

(Optional) Select this check box to display all unassigned role alerts generated, even those alerts below the minimum alert threshold that is specified in the role alert rule configuration. This check box only displays if your system administrator has enabled this feature.

The default selection is cleared, meaning that only role alerts currently unassigned that meet or exceed the minimum alert threshold (as defined in the role alert rule) display in the Visualizer.

Miscellaneous Settings section

Use this section to enable hover help and the exit confirmation window.

Enable Tooltips

If ToolTips are enabled, whenever your cursor hovers over a toolbar icon or over an area where extra information is available, hover help displays. By default, ToolTips are enabled.

Show Exit Confirmation Dialog:

This option determines whether the system displays a confirmation dialog when you exit the Visualizer.

- Select this check box to confirm your choice to exit the Visualizer every time. The default setting is selected.
- Clear this check box to exit the Visualizer without displaying the **Exit Confirmation** dialog every time you choose to exit and log out of the Visualizer.

Log and Link Settings tab:

Use this tab to configure Visualizer client logging and hyperlink settings. Configuring preferences here only affects settings for the local Visualizer client. If you change to any of these settings, exit, reopen, and login to the Visualizer to see the changes.

Log Settings

Select the check box to turn on Visualizer client logging or clear the check box to turn off client logging. Typically, you only enable Visualizer client logging if you are working with your system administrator to resolve an error message or problem that occurred during your Visualizer session. By default, Visualizer client logging is turned off.

Log detail level

Select the level of logging detail, only available if Visualizer client logging is on. The level of detail controls how much information is collected in the Visualizer log as you use the Visualizer. Consult your system administrator before making a selection. Typically, you turn on logging to troubleshoot a problem in the Visualizer, so usually you select the debugging level, which is the highest level of logging detail. The debugging level logs every action and message that occur while you use the Visualizer. But this log level also fills up the Visualizer client log file very quickly, so you might need to occasionally clear the log file. This is the reason that you usually turn off logging when the problem is resolved.

Log file directory path

Specify the file and directory location of the Visualizer client log files. Typically, you only need to review log files when you are troubleshooting a message or problem. Log files can fill up with information quickly, especially at the debugging level. If Visualizer client logging is turned on, you might occasionally need to purge the log files to prevent the files from becoming too large.

Hyperlink Handling Settings

Select an option to determine what program or browser the Visualizer uses to open and display hyperlinks. Incoming identity records can contain hyperlinks, which can direct you to other files, Web sites, or systems that contain identity or entity information relevant to your analysis. Hyperlinks are part of the identity record and display on the entity resume and the entity resolution graph as attributes.

If you experience problems or issues when clicking a hyperlink, select the **Use program** option and specify which browser or program to use to open hyperlinks. For example, if your organization stores fingerprint files on a secure Web site (https://), use this option to specify your Web browser or other program to open links that go to the fingerprint files secure site.

Graph Preferences tab:

Use this tab to specify the display properties of the lines that connect entities on Visualizer graphs. Configuring preferences here only affects the settings for the local Visualizer client. If you change any of these settings, exit, reopen, and login to the Visualizer to see the changes.

Line Thickness

Select a line thickness. The default line thickness is 2 pixels.

Line Color
Select a line color. The default line color is a medium blue.

Sample Line
Displays a sample graph line, based on your selections.

Starting the Visualizer

To use the Visualizer to view entities and entity data from the entity database, you must first start the Visualizer and log in.

To start the Visualizer, the default system version of Java processes a Java Web Start JNLP (Java Network Launch Protocol) file that the product application server downloads to your workstation client. The JNLP file can be accessed many different ways. But to successfully open the Visualizer, the required client version of Java Web Start must open the JNLP file.

If you have multiple versions of Java installed on your client machine, the default system version of Java Web Start could be set to a version other than the required client version. You can still successfully open and run the Visualizer, but first you need to configure your Web browser to use the required client version of Java Web Start .

Note: The client Java version required to open and run the Visualizer may not be the latest version of Java.

Logging in to the Visualizer

Before you log in to the Visualizer, you must have a Visualizer user account (user name and password). Your system administrator can provide you with your Visualizer user account information.

Procedure

1. Complete one of the following steps:
   - Double-click the Visualizer icon on your desktop.
   - Or open your Internet browser and enter the uniform resource locator (URL) for the Visualizer in the address line.

   The URL for launching the Visualizer is:

   ```
   http://server:install_port
   ```

   For example, `http://localhost:13510`. When the Visualizer is installed, the default *install_port* is 13510, but the port number can be changed. Check with your system administrator if you are unsure of the correct server name or port number.
2. Log in by entering your user name and password.

   Note: Both user name and password fields are case sensitive. The first time that you log in, use the password assigned to you by your system administrator. After your first successful login, typically you change your Visualizer password to safeguard the security of your Visualizer account.
3. Click **Login**.

Setting your Web browser to use the required client version of Java Web Start:

If your workstation contains multiple versions of Java and you are having difficulty opening the Visualizer, set your Web browser preferences to select the required client version of Java Web Start. By doing so, your Web browser automatically uses the required client version of Java Web Start to open the Visualizer opens successfully every time.

Setting Microsoft Windows Internet Explorer to use the required Java Web Start:

Microsoft Internet Explorer uses the default file associations defined for the Microsoft Windows operating system to determine how to handle JNLP (Java Network Launch Protocol) files. By defining or modifying the default file application associated with processing JNLP files, you can redirect Internet Explorer to use the correct Java Web Start version. If you have multiple versions of Java installed, modifying this setting can prevent problems with opening the Visualizer.

About this task

This procedure directs Internet Explorer to use the Java Web Start version to open all Web applications. If you run other Web Start applications that require later versions of Java, use the direct launch approach instead.

Note: There are a couple of known issues with Java version 1.6 to keep in mind:

- Java version 1.6 sometimes overrides the default Windows file association for JNLP files. If you use Java version 1.6 as your system JVM (Java Virtual Machine) and these steps do not allow you to successfully start and open the Visualizer, either try using another Web browser to launch the Visualizer, or use the direct launch approach instead.
- If your workstation uses Java version 1.6, you may also need to configure the JRE (Java Runtime Environment) to accept automatic downloads. If your workstation has this issue, when you try to start the Visualizer, you will see an error message indicating that the application has requested a version of JRE that is not locally installed.

Procedure

1. From the **Windows Control Panel**, do one of the following steps:
   - From the Category View, double-click **Performance and Maintenance**. From the **See Also** navigation pane in the upper-left corner of the window, select **File Types**.
   - From the Classic View, double-click **Folder Options**.
2. From the **Folders Options** dialog, click the **File Types** tab.
3. Under the Extensions column, locate and select the **JNLP** entry. Entries are alphabetized by extension.

   Note: If the JNLP entry does not exist, click **New** to create the entry.
4. Click **Change**.
5. From the **Open With** dialog, make sure that **Java WebStart Executable** is selected. Click **Browse** to navigate to your installed Java directory.
6. Select the executable file named `javaws` and click **OK**.
7. Click **OK** to close the **Folder Options** dialog. (You can also close the**Control Panel** window.)

Results

Internet Explorer now uses the associated Java Web Start file to successfully process and open the Visualizer.

Setting Mozilla Firefox to use the required Java Web Start:

By setting or modifying the way that Mozilla Firefox handles JNLP (Java Network Launch Protocol) files, you can direct Firefox to automatically use the required client Java Web Start version to start the Visualizer. If you have multiple versions of Java installed, modifying this setting can prevent problems with opening the Visualizer.

About this task

This procedure directs Firefox to use the Java Web Start version to open all Web applications. If you run other Web Start applications that require later versions of Java, use the direct launch approach instead.

Procedure

1. Launch Mozilla Firefox.
2. Select **Tools** > **Options**.
3. Select **Applications**
4. Under **Content Type**, locate the entry for **JNLP File**.

   Note: If you do not see an entry for **JNLP File**, close the **Options** dialog. From the Visualizer Web Start page, try to start the Visualizer by clicking the **Click here to start the IBM Identity Insight Visualizer** link. Then start again at Step 1.
5. Select the **JNLP File** entry.
6. Under **Action**, select the **Use Other** option.
7. From the **Select Helper Application** dialog, click **Browse**, navigate to the directory where the required client Java version is installed, and select the `javaws` executable file.
8. Click **OK** to close the **Select Helper Application** dialog.
9. Click **OK** to close the **Options** dialog.

Results

Mozilla Firefox now uses the selected Java Web Start file to handle all JNLP file types. The Visualizer successfully opens.

Starting the Visualizer directly from the Java Web Start executable:

If you want to start the Visualizer without changing Java or other system settings, you can use the direct launch approach. This approach launches the Visualizer directly from the Java Web Start executable. You might want to use the direct launch approach if you have multiple versions of Java installed on your workstation and you use other Web Start applications besides the Visualizer.

Before you begin

Locate the path to the required Java Web Start executable file (`javaws`) on your workstation.

About this task

You can also create a shortcut on your desktop to the Java Web Start executable file by selecting the `javaws` file and entering the URL to the Visualizer in the **Target** field.

Procedure

1. From your desktop, open a DOS command window.
2. On the command line, enter the direct launch command: *`path_to_Java_installationpath_to_javaws_exe_file`*`>javaws.exe` *`URL for the Visualizer`* For example: **`C:/IBM/Java60/jre/bin>javaws.exe http://localhost:13510/docs/rrmdi.jnlp`**

   Important: Note the space between the Java Web Start executable file extension and the URL.

Results

The Visualizer successfully opens.

Configuring Java v1.6 for running the Visualizer on Microsoft Windows workstations:

If you try to start the Visualizer and you see an error message indicating that the application has requested a version of JRE that is not locally installed, try changing the automatic download settings for Java. This error message is a known issue for Microsoft Windows workstations that have Java version 1.6 installed.

Procedure

1. From the **Windows Control Panel**, select one of the following options:
   - For IBM installations of Java, select **IBM Control Panel for Java**.
   - For Sun installations of Java, select **Java**.
2. On the **Advanced** tab, expand the **JRE Auto-Download** setting. If you do not see this option and you have multiple versions of Java installed on this workstation, close the **Java Control Panel** and select the other entry.
3. Make sure the **JRE Auto-Download** setting is set to either **Always Auto-Download** (recommended) or **Prompt User**. The **Never Auto-Download** setting prohibits opening the Visualizer and Configuration Console.
4. Click **Apply**.
5. Click **OK**.
6. Close the **Control Panel** window.

Closing the Visualizer

When you are finished using the Visualizer, close the application. By closing the Visualizer, you also log out. If you are taking a break and just want to secure your workstation for a few minutes, you can lock the Visualizer instead.

Procedure

To close the Visualizer and log out:

- Select **File** > **Exit**.
- Or press **Ctrl** + **Q**.

Locking the Visualizer

If you are taking a quick break or walking away from your workstation for a few minutes, instead of closing and logging out of the Visualizer, you can lock it. Locking the Visualizer secures your work by acting like a secured screen saver. When you lock the Visualizer, the **Login** window displays. You get back into your Visualizer session by entering your user password.

Procedure

To lock the Visualizer:

- Select **File** > **Lock application**.
- Or press **Ctrl** + **L**.

Results

Your Visualizer session is now securely locked.

What to do next

To resume using the Visualizer, enter your password and click **Unlock**.

Changing your Visualizer password

Changing your Visualizer password regularly is a good way to protect the security of your Visualizer user account.

Before you begin

You must be logged in to the Visualizer to change your password.

About this task

There are no minimum number of characters required for Visualizer passwords. You can use any combination of letters (uppercase or lowercase), special characters, and numbers. The password is case sensitive, so when you log in, the password that you enter must match your Visualizer account password. For example, if your password is `PASSw0rd`, and you try to log in using `passw0rd`, the passwords do not match and the system displays an error message.

Procedure

1. In the Visualizer, click **File** > **Change password**.
2. In **Current password**, enter the password you used to log in to this Visualizer session. If your password was assigned to you or reset, this password is the password from your system administrator.
3. In **New password**, enter the new password to be your Visualizer password.
4. In **Repeat new password**, enter the same password that you just entered in **New password**.
5. Click **Change password**.

Results

- If the entries in both the **New password** and **Repeat new password** are the same, the system displays a message that your password has been changed. Click **OK**. Use your new password the next time that you log in to the Visualizer

- If the entries do not match, the system displays an error message indicating that the new passwords do not match. Click **OK**. Your password is not changed. To change the password, start at step 2 again.

Analyzing alerts in the Visualizer

One of the most common tasks that Visualizer users perform is to evaluate alerts to decide which alerts to review and which alerts to transfer to other Visualizer groups.

Alerts display in the **Alert Summary** window of the Visualizer. This window is the starting point for evaluating, assigning or transferring, and reviewing alerts.

Alerts are grouped into alert summaries. Alert summaries contain all alerts of the same alert type with the same description, alert severity, status, resolution rule, relationship score, and resolution (likeness) score. One alert summary typically contains multiple individual alerts, each of which need to be reviewed and analyzed. Part of the review includes assigning a disposition to the alert, so that you and other Visualizer users know the status of the analysis and can see comments indicating your findings.

Remember, your **Alert Summary** window only displays the following:

- Alert summaries for your Visualizer analyst group that contain unassigned alerts
- Alerts that you have already assigned to yourself

You do not see the alerts that other analysts in your Visualizer analyst group have assigned to themselves. Nor do you see alerts that are assigned to other Visualizer analyst groups.

Evaluating alert summaries

How do you decide which alerts to assign to yourself for analysis? Start by reviewing the alert summaries in the **Alert Summary** window. As you look at the alert summaries, compare the importance of the information that makes up that alert summary with your analysis goals. You might need to evaluate one or more pieces of alert information before you can decide.

Tips to Help Prioritize Alert Summaries:

- **Alert severity:** Start by sorting the alert summaries by severity. Click the **Alert Severity** column header. This information might be enough information to help you decide which alerts are most critical or important to begin analyzing. For example, if your organization uses "C" for alerts with a critical severity, you can immediately see which alerts are critical by simply looking at their severity.
- **Alert description:** Severity might not be enough information, alone. The alert description might help you choose which alerts are higher on the priority list, if there are multiple alert summaries with the same alert severity. For example, it might be more important to analyze alerts that are grouped by the "No Fly knows Passenger" description than the "Passenger knows Employee" description.
- **Likeness Score and Relationship Score:** The higher the scores, the more likely it is that there is a relationship of interest or that the identity is the entity. In the "No Fly knows Passenger" example, if both the Likeness and the Relationship scores are 100, then the person on the No Fly list is the Passenger, and you might want to take immediate action. If the likeness score is less than 70 and the relationship score is less than 85, this alert might still be important, but not critical. You might still want to analyze the entities involved in the alert, but you might not need to take immediate action.

As a Visualizer user, you are familiar with your organization's goals, so you can probably add your own personal factors to use when prioritizing alerts. These tips are to get you started.

Assigning alerts

After you know which alerts you want to work on, based on priority, you can assign those alerts to yourself. Assigning alerts allows your Visualizer analyst group to divide and conquer the list of incoming alerts. When an alert is assigned to you, that alert only displays on your Alert Summary window, preventing duplicate work by another Visualizer user on the same alert. You can immediately see the alerts that you alone are currently researching.

If you see one or more alerts in your Alert Summary window that you think might belong to another Visualizer analyst group, you can transfer those alerts. For example, you work as a reservation clerk and evaluate alerts generated by new or changed reservations. You see an alert listed that security handles. You can assign that alert to the Security group, because the alert is under that group's jurisdiction.

Reviewing and dispositioning alerts

When you assign yourself one or more alerts, then you can get down to the business of researching and analyzing those alerts. The Visualizer simplifies the task in the Research window, which displays all the relevant, associated information about the alert into one window. From the Research window, you can do the following tasks as part of your analysis:

- Review the alert details
- Look at the entity resumes of the related entities
- View the associated entity or alert graphs to visualize and explore the commonalities of the entities or attributes that are part of the alert
- Add comments indicating the findings of your analysis
- Change the status (disposition) of the alert as your analysis progresses

Attribute alerts

Attribute alerts are alerts that are produced by attribute alert generators, which create a persistent system query looking for specific attributes or identities in the entity database. Whenever attributes for entities match the criteria of the attribute alert generator, the system creates an attribute alert.

Visualizer users create their own personal attribute alert generators. If you are looking for a specific identity or any identities or entities that match a specific set of attributes, you can create your own personal attribute alert generator that searches for matches until the specified expiration date.

Examples of possible entity attributes you might want to be notified about include:

- Name and unique number (such as a credit card number)
- Name and phone number
- Address
- Name and non-unique number

Attribute alert generators are configured and viewable using the Visualizer. The attribute alert generators that you create are only available to you.

Example of an address attribute alert

You are watching the address of 675 Hickory Street Las Vegas, NV. You can configure an attribute alert generator to create an attribute alert whenever that address is associated with an incoming identity record added to the entity database.

Event alerts

An event alert occurs when one of more complex events meets set criteria over a specified life span. Event alerts are based on complex event business rules and other configurations contained in an event rules file (`cep.xml`). These alerts can indicate situations of interest, such as "Two or more purchases of more than $10,000 U.S. dollars occurred in the last hour at locations 200 miles from each other".

Role alerts

A role alert identifies when one or two entities linked through a relationship that meets or exceeds a configured role alert rule. Role alerts are based on configured roles and role alert rules. They can indicate a warning or a problem (such as a customer knows a bad guy) or simply indicate relationships of interest (such as a customer knows an employee).

You define relationships *of interest* or as *conflicting* by configuring role alert rules that identify which roles should not exist in a single entity or cannot be linked between one or more entities. Use the Configuration Console to configure filters for role alerts, which determine if the system re-alerts when there is new information (such as a new identity or a new data source code).

During entity resolution, the pipeline evaluates relationships between the incoming identity and entities on the candidate list. After determining a relationship exists between the incoming identity and a candidate entity, the system then evaluates whether the roles assigned meet a configured role alert rule. If so, the system generates a role alert.

A role alert identifies entity data at the time the role alert was created. The Role Alert detail screen shows the entity data as it existed at the time the role alert was created. As entity data changes over time, the entity resume contains the latest entity data. If you want to see the current data for a particular entity, view the entity resume.

You can view and work with role alerts in the Analyst's Toolkit components (Cognos reports, the Identity Insight plug-in for i2, and the Identity Insight Explorer).

Displaying alerts

You view alerts in the **Alert Summary** window to evaluate which alerts to analyze and assign to yourself or transfer to another Visualizer analyst group. Then, you can begin researching and dispositioning the alerts that you assigned to yourself.

About this task

Alerts that display in your **Alert Summary** window include the following:

- Alerts that you have assigned to yourself for analysis.
- Unassigned alerts for your Visualizer analyst group
- Attribute alerts generated from one of your attribute alert generators

The unassigned alert summaries are filtered based on the default **Alert Summary** window alert display filter values that are configured on the **System Preferences** tab of the **Configure Screen Preferences** window. You can change one or more alert display filter values in the **Display Filters** group box.

Procedure

1. Select **View** > **Alert Summary**.
2. Then select the type of alerts that you want to display or select **Show All Alert Types**.

Results

From the **Alert Summary** window, you can decide which alerts to work with. You can assign alerts to yourself or transfer alerts to another Visualizer analyst group. You can select the alerts you assigned to yourself to analyze and add comments about your analysis.

Filtering the alert display on the Alert Summary window

While reviewing alerts on the **Alert Summary** window, you can filter which alert summaries display by changing the values in the **Display Filters** group box. The display filters only affect alert summaries currently in the "Unassigned" status.

About this task

The default values for these alert filters are configured on the **System Preferences** tab of the **Configure Screen Preferences** window. When you change the alert display filters on the **Alert Summary** window, you temporarily override those default values. The next time you open a new **Alert Summary** window, the filters revert to their default values.

Procedure

1. From the **Alert Summary** window, open the **Display Filters** group box twistie.
2. Make your changes to one or more of the alert display filters.
3. Click **Apply** to refresh the **Alert Summary** window and apply your specified alert filters.

Assigning alerts to yourself

By assigning an alert to yourself, you take ownership for reviewing, researching, and dispositioning that alert. After you assign an alert to yourself, that alert only displays in your **Alert Summary** window, making it easier for you to identify your alerts.

Procedure

1. From the Visualizer, in the **Alert Summary** window, from the **Alert Summary** table, click an unassigned alert summary. The alert summary contains one or more alerts that are grouped by alert type, and share the same description, status, resolution rule, likeness score, and relationship score.
2. From the **Alert List** table, double-click the alert to assign to yourself.
3. In the **Research** window, click **Set Status**.
4. In **Set Status**, do the following actions:
   a. In **Select the action you want to perform**, select **Set Status**. A corresponding activity code displays in **Select Activity Code**.

b. Required: In **Select Status**, select **Assigned**. If you select another status, the alert is not assigned to you.

c. Optional: To assign a different activity code, select it from **Select Activity Code**. If you do not see the activity code that you want to select, contact your system administrator to arrange to configure the activity code.

d. Enter comments or notes in the **Comments** text box. For example, you might choose to enter comments about why you are changing the status, or include notes about your analysis of this alert.

e. Click **OK** to save your changes.

Results

The alert now reflects the assigned status and only displays in your **Alert Summary** window, after you refresh the window display. Other analysts in your Visualizer analyst group will no longer see this alert, after they refresh their **Alert Summary** window display.

Assigning alerts to other analyst groups

If you determine that an alert needs to be assigned to another Visualizer analyst group, you can transfer that alert. You cannot transfer an alert to a specific Visualizer user, but you can transfer that alert to the Visualizer analyst group that the user belongs to.

Procedure

1. From the Visualizer, in the **Alert Summary** window, from the **Alert Summary** table, click the alert summary that the alert is associated with.
2. From the **Alert List** table, double-click the alert to transfer.
3. In the **Research** window, click **Set Status**.
4. In **Set Status**, do the following:
   a. From **Select the action you want to perform**, select **Transfer Alert**.
   b. In **Transfer Alert to**, select the Visualizer analyst group to transfer the alert to. If you do not see the Visualizer analyst group that you want to select, contact your system administrator to arrange to configure the analyst group. A corresponding activity code displays in **Select Activity Code**.
   c. Optional: To assign a different activity code, select it from **Select Activity Code**. If you do not see the activity status code that you want to select, contact your system administrator to arrange to configure the activity code.
   d. Enter comments or notes in the **Comments** text box. For example, you might choose to enter comments about why you are transferring the alert.
   e. Click **OK** to complete the transfer.

Results

The alert is now transferred to the selected Visualizer analyst group and displays in the **Alert Summary** windows of analysts in that Visualizer analyst group. (Analysts in that group might have to refresh their **Alert Summary** window display first). This alert no longer displays in the **Alert Summary** window for analysts in your Visualizer analyst group, including you, after refreshing the **Alert Summary** window display.

Changing the status of an alert

As you analyze the alerts that are assigned to you or to your Visualizer analyst group, you can use the Visualizer to track your research, comments, and the way that you disposition the alert.

About this task

You can update the alert status for alerts that are assigned to you or to your Visualizer analyst group at any time. You can also add comments to these alerts at any time. However, you cannot edit existing comments.

Procedure

1. From the Visualizer, in the **Alert Summary** window, in the **Alert Summary** table, click the alert summary containing the alert to update.
2. From **Alert List**, double-click the alert to change the status of.
3. In the **Research** window, click **Set Status**.
4. In **Set Status**, do the following:
   a. From **Select the action you want to perform**, select **Set Status**. A corresponding activity code displays in **Select Activity Code**.
   b. Optional: To assign a different activity code, select it from **Select Activity Code**. If you do not see the activity status code that you want to select, contact your system administrator to arrange to configure the activity code.
   c. Enter comments or notes in **Comments**. For example, you might enter comments indicating why you are changing the status or include notes about your analysis of this alert.
   d. Click **OK** to save your changes.

Results

The alert now reflects the new status in the **Alert Summary** window.

The newest status or comments update for an attribute alert display at the top of the **Status Summary** section.

If the status change involved assigning the attribute alert to yourself, this attribute alert now displays only in your **Alert Summary** window, after you refresh the display. Other analysts in your Visualizer analyst group will no longer see that alert in their **Alert Summary** window, after they refresh their display.

Help topics

Alert Summary window:

Use this window to view unassigned alert summaries for your Visualizer analyst group or alerts that you have assigned to yourself.

Use the twisties to expand or collapse the sections of the screen to help you focus on a specific detail.

Display alerts by type
Select an alert type to display or show all alert types.

Display Filters group box
Changes the default filter settings that determine which alert summaries display in your **Alert Summary** window. These filters only change the

display for alert summaries that are currently unassigned and are a temporary change only. If you close the **Alert Summary** window and reopen it at another time, these settings revert to the default filter settings.

The default settings are the alert filter settings that are configured for your workstation. (You can change the default settings on the **System Preferences** tab on the **Configure Screen Preferences** window.)

Alert Summary table

Alerts that share the same alert type, description, severity, status, resolution rule, Likeness Score, and Relationship Score are grouped into alert summaries. The **Count** column shows how many individual alerts are grouped into the summary.

You can sort the table by clicking a column header in the table. The first click sorts the column values in ascending order. The second click sorts the column values in descending order.

By default, the table is sorted by alert type.

Type Type of alert represented by the alert summary.

Description

Description of the alerts in this summary.

For attribute alerts, this description is the case number. For event alerts, this description is the event situation description. For role alerts, this description is the role alert rule description.

Status Current activity status of the alerts in this summary.

Resolution Rule

Name of the resolution rule used to relate the entities within the alerts in this alert summary.

Likeness Score

Score (0-100) which indicates how likely the related entities represent the same entity.

Relationship score

Score (0-100) which indicates how strongly the entities within the alert are related to each other.

Count Number of individual alerts grouped into this alert summary which meet the currently selected **Display Filters** group box criteria.

Alert List table

After you select an alert summary from the **Alert Summary** table, the individual alerts that are part of that summary display in this section. The number of alerts (lines) that display depend upon the total number of alerts in the summary (found in the **Count** column in the Alert Summary table), and the number in the **Maximum Lines in Alert List** field in the **Display Filters** group box. A list count on the **Alert List** table title bar shows how the number of alerts currently displayed fit into the total number of alerts for this summary.

Sort the table by clicking a column header in the table. The first click sorts the column values in ascending order. The second click sorts the column values in descending order.

The fields that display are based on the type of alert summary selected.

Attribute Alert screen:

Use this screen to set or change the analysis status of an attribute alert and review the details that make up the alert.

Use the twisties to expand or collapse the sections of the screen to help you focus on a specific detail.

Status Summary
Summarizes the current analysis status and disposition of the alert.

Alert Summary
Provides the description of the alert summary and the date and time the alert was generated.

Match to Entity section
Contains details about which attributes matched between the search criteria of your attribute alert generator and existing entities in the entity database. Click a specific attribute to highlight the matching information from the identities on the matching entity.

Attribute Alert Generator Details
Summarizes the criteria for the attribute alert generator that generated this attribute alert. Click the data source to highlight all the criteria.

Entity section
Displays information about the entity that matched the attribute alert generator criteria. Click the data source to highlight the data that came in on an identity record from this data source.

Entity Resume button
Click to display the entity resume for the matched entity. You might want to look at the other identities associated with the entity to further your analysis of this alert.

Event Alert screen:

Use the **Event Alert** screen to set or change the analysis status and review the details of an event alert. Event alerts only display if Event Manager is enabled for your system, if activity codes for event alerts are configured, and if one or more event alerts exist.

Use the twisties to expand or collapse the sections of the scree to help you focus on a specific detail.

Status Summary
Summarizes the current analysis status and disposition of the event alert.

Alert Summary
Provides the description of the event alert and the date and time the alert was generated.

Event Alert section
Provides the event details that make up this event alert.

Entity section
Provides a short resume for each entity involved in this event alert.

Report button
Click to create an **Event Alert Detail** report.

Role Alert screen:

Use this screen to view the details of a role alert and to set or change the analysis status of the role alert.

Click the twisties to expand or collapse the sections of the screen to help you focus on a specific detail.

Degrees of Separation
Indicates the number of degrees of separation between the entities in this role alert.

Status Summary
Summarizes the current analysis status and disposition of the alert.

Alert Summary
Provides a description of the alert summary, the alert severity code for this alert, the resolution rule used to match entities within the alert, the resolution score that indicates how alike the two entities are, and the relationship score that indicates the likelihood that these two entities know each other.

Matching Details tabs
Contains details about which attributes matched between the two entities. Click a specific attribute to highlight the matching information from the identities on the matching entity.

Contains details about which attributes matched between the search criteria of your attribute alert generator and existing entities in the entity database.

Report button
Click to create a **Role Alert Detail** report for this role alert.

Entity Resume button
Click to display the entity resume for the selected entity. You might want to look at the other identities associated with the entity to further your analysis of this alert.

Entity Events screen:

Use the **Entity Events** screen to review the events for an entity that occurred within a specific date range. You initially access this screen by clicking **Show Events** from the **Entity Resume** screen.

Event Summary section
Displays a summary of all the events for this entity with the date range indicated. By default, the screen displays all events associated with the entity from the first event date to the current date. Change the date range using the event date filter to see events within a different date range.

On-screen event date filter
Filters the events displayed by the specified date range when you click **Update View**.

From Date
Enter a date or click the calendar control to select the starting date in the date range.

If you choose to type a date, use one of the following date formats:

- MM/dd/yyyy, MM-dd-yyyy, MM.dd.yyyy, or MMddyyyy
- yyyy/MM/dd, yyyy-MM-dd, oryyyy.MM.dd
- January 3, 2008 or January 03, 2008
- January 3, 08 or January 03, 08
- Jan 03, 2008 or Jan 3, 2008
- Jan 3, 08 or Jan 03, 08

The field defaults to the first event date instance.

Through Date
Enter a date or click the calendar control to use as the ending date in the date range.

If you choose to type a date, use one of the following date formats:

- MM/dd/yyyy, MM-dd-yyyy, MM.dd.yyyy, or MMddyyyy
- yyyy/MM/dd, yyyy-MM-dd, oryyyy.MM.dd
- January 3, 2008 or January 03, 2008
- January 3, 08 or January 03, 08
- Jan 03, 2008 or Jan 3, 2008
- Jan 3, 08 or Jan 03, 08

The field defaults to the current date.

Update View button
Click to view events for this entity within the specified date range. This button is disabled until you change the default dates in the date fields.

Report button
Click to generate an **All Events** report for this entity.

On-screen display
This section of the screen summarizes the events for this entity by event type within the specified date range.

Event Type
Describes the event type.

Count Indicates the total number of the events for this entity by event type, within the specified date range. (For example, if the count is 4, four events of the same event type occurred for this entity within the specified date range.)

Value Indicates the total value of the events for this entity by event type, within the specified date range. (For example, if there are four events, this number is the sum total of the value for those four events.)

Quantity
Indicates the total number of units for the events for this entity by event type, within the specified date range.

Unit of Measure
Describes the unit of measure for the event value. The unit of measure is configured by event type in the Configuration Console.

Total Count
Indicates the number that represents the total number of all events for this entity within the specified date range.

Total Value
Indicates the number that represents the total value of all events for this entity within the specified date range.

Event Details section
Select an event row in the Event Summary section to see more details about the individual events that are included in the event type summary. If you double-click any event row in this section, the **Event Details** screen displays to show even more detailed information about the selected event.

Date Indicates the date and time of the event.

Data Source - Description
Describes the data source associated with the event.

External ID
Displays the unique key that identifies the inbound record in the original data source for this event.

Event Reference
Provides additional information about the event in the original data source, if that information is part of the inbound record.

Value Indicates the value amount of the event.

Quantity
Indicates the number of units in the event.

Memo or *Custom label*
Provides additional information about the event, such as notes or comments, that can provide more context for the event transaction.

Users can define a custom label for this column, as one of the options when configuring an event type in the Configuration Console. So instead of **Memo**, you might see a more descriptive, custom label (such as **Wire Transfer Notes**).

Finding entities

You can use several Find By methods in the Visualizer to find an entity in the entity database. If you want to be notified each time the system processes a record containing a particular name, address, number, or e-mail address, you create an attribute alert generator to automatically "find" entities.

Finding entities by attribute

When you are using the Visualizer, and you want to find an entity in the entity database, you can find the entity by entering criteria on the attributes associated with the entity. You specify the attribute criteria, and the Visualizer builds a query based on that criteria. This type of entity query does not go through the entity resolution process to return search results.

Procedure

1. In the Visualizer, do one of the following:
   a. Click **View** > **Find By** > **Attribute**.
   b. From the toolbar, click the (Find) icon.
   c. From the toolbar, click the arrow and select **Attribute**.

   d. From the **Find By** window, select **Attribute** from the **Find By** drop-down list.
2. Enter the criteria for each attribute type that you want to use to find entities.
   a. Click **+** to add a row to specify criteria for another attribute type.
   b. Click **-** to remove the selected query criteria entry.
3. Optional: Click **Show Summary** to view a summary of the Find By Attribute query. The summary is a helpful way to make sure that the query contains the values that you intended. If not, close the summary and correct the query criteria.

   Two query criteria of the same attribute type constitutes an "OR" clause. All other query criteria combine as "AND" clauses.

   The order of the attribute type criteria does not affect the results.
4. Click **Find**.

Results

Entities that match the query criteria display in the **Results** pane.

By default, results displayed for Find by Attribute queries are limited to the first 1,000 matching entities. If there are more than 1,000 matches, the **Results** pane indicates that more results exist. (The number of results displayed can be configured by your system administrator in the Configuration Console by setting the MAX_ENTITIES_RETURNED parameter under system parameters.)

Note: If your system uses an additional address hygiene application, addresses that include special characters might be transliterated. For example, the search result for a German address that contains one or more umlauts in the address might return a result that does not contain matching umlauts.

What to do next

Click an entity to display the entity resume for the selected entity.

Finding entities by data source account

When you know the account number (or external ID) of an identity, and you want to find the entity that contains that identity, use the Find By Datasource Account in the Visualizer. You can also find an entity that you added through the **Add Entity** screen.

Before you begin

You must know the data source description and external ID of the identity (or account). If you are trying to find an entity by name, use the Find By Attribute method.

Procedure

1. In the Visualizer, do one of the following:
   a. Click **View** > **Find By** > **Datasource Account**.
   b. From the toolbar, click the arrow, and select **Datasource Account**.
   c. From the **Find By** window, select **Datasource Account** from the **Find By** drop-down list.
2. In **Enter External ID**, enter the account number for the identity. The account is the way the identity is known in the original data source.

3. In **Data Source**, select the data source code and description.
4. Click **Find**.

Results

If the system finds an entity that contains an identity with the specified external ID and data source criteria, the Visualizer displays the **Entity Resume** for that entity.

Finding entities by Entity ID

When you know the Entity ID number of an entity, use the Find By Entity ID method in the Visualizer to quickly locate the entity and display the entity resume for that entity.

Before you begin

You must know the Entity ID number of the entity you want to find. If you are trying to find an entity by name, use the Find By Attribute method instead.

Procedure

1. In the Visualizer, do one of the following:
   a. Click **View** > **Find By** > **Entity ID**.
   b. From the toolbar, click the arrow and select **Entity ID**.
   c. From the **Find By** window, select **Entity ID** from the **Find By** drop-down list.
2. In **Enter Entity ID**, enter the Entity ID number for the entity to find.
3. Click **Find**.

Results

If the Entity ID matches an entity in the entity database, the Visualizer displays the entity resume for that entity.

Finding entities by resolution

Use Find by Resolution to create a search entity that goes through the entity resolution process to see if any identities in the entity database meet the criteria of the query.

Before you begin

The Find by Resolution feature requires a running pipeline available for the Visualizer server to communicate with. The pipeline is the component where entity and relationship resolution occur.

About this task

To make the best use of the Find by Resolution feature, it is important to understand how entity resolution works and is configured for your system, because entity resolution is used to find the results. For example, if entity resolution is not configured to find matches based solely on name, Find by Resolution does not return results if a search is performed only on a name value. Likewise, because entity resolution does not resolve entities based on a postal code alone, specifying only a postal code returns no results.

Find by Resolution uses the minimum score values defined on the **System Preferences** tab from the **File** menu.

Procedure

1. In the Visualizer, do one of the following:
   a. Click **View** > **Find By** > **Resolution**.
   b. From the toolbar, click the arrow and select **Resolution**.
   c. From the **Find By** window, select **Resolution** from the **Find By** drop-down list.
2. Enter as many attributes as you know about the identity.
   - If you enter anything in the **Name** section, then **Last Name** is required.
   - If you enter any information in the **Address List** section, **Address** is required.
   - If you select a **Type** in the **Number List** section, you must enter a number value in the **Value** field. (**Location** is optional.)
   - If you select a **Type** in the **Characteristic List** section, you must enter a characteristic value in the **Value** field.
   - If you select a **Type** in the **Email List** section, you must enter an e-mail address value in the **Address** field.
3. Click **Search**.

Finding entities by attribute alert generators

When you have an entity that you are watching, you can create attribute alert generators with the criteria for that entity. Whenever identity records or entities contain attributes that match the criteria, the system generates an attribute alert. Each Visualizer user creates and manages personal attribute alert generators for a specific date range.

Because attribute alert generators are submitted through the pipeline, the entity resolution process is performed on those search requests in the same manner as it is on incoming entity data:

- Names and address are standardized
- Partial or fuzzy searches and comparisons are performed so that applicable entities are identified in the subsequent attribute alerts

To make the best use of attribute alert generators, it is important to understand how entity resolution works and is configured for your system, because entity resolution is used to find your attribute alert results. For example, if entity resolution is not configured to find matches based solely on a name, an attribute alert generator configured to search only for a name value returns no results. Likewise, because entity resolution does not resolve entities based on a postal code alone, an attribute alert generator specifying only a postal code returns no results.

When creating an attribute alert generator, use the following guidelines:

- Use **Minimum Score** to filter attribute alert results. The default value for this field is "Any Relationship". This choice allows the most results. Choose a higher level to allow fewer results. These values are configured in the Visualizer system preferences, available from the **File** menu.
- For names: Supply either a last name and first name combination or a last name and middle name combination. Attribute alert generators specifying only a last name, first name, or middle name returns no results.

- For addresses: An address and postal code are both required. Attribute alert generators specifying only city, state, postal code, address, or country returns no results.

Creating attribute alert generators:

To receive an alert whenever a specific attribute value or combination of attribute values are processed by the system, create an attribute alert generator. Attribute alert generators continue to generate alerts until the specified expiration date.

Procedure

1. In the Visualizer, do one of the following:
   a. Select **View** > **Attribute Alert Generator Manager**.
   b. From the toolbar, click (Attribute Alert Generator Manager) icon.
2. From the **Attribute Alert Generator Manager** window, click **Create**.
3. Use the drop-down lists and fields to enter the specific criteria for your new attribute alert, including an expiration date. The default expiration date is set to six months from today's date.
4. Click **Create**.

Results

Whenever data that resolves with the criteria you specified is processed through entity resolution, a new attribute alert display in your **Alert Summary** window. If the information you are looking for is currently in the entity database, you see a new attribute alert in the **Alert Summary** window.

Editing attribute alert generators:

Edit an active attribute alert generator when you want to change the case number, comment, or expiration date.

About this task

You cannot change the attributes or the minimum resolution score for those attributes. If that is what you want to do, create an attribute alert generator. And if the new attribute alert generator replaces an existing one, use these steps to expire the attribute alert generator that you no longer need.

Procedure

1. In the Visualizer, do one of the following:
   a. Click **View** > **Attribute Alert Generator Manager**.
   b. From the toolbar, click (Attribute Alert Generator Manager).
2. Select the attribute alert generator to edit and click **Create**.
3. In the **Attribute Alert Generator Info** window, make your changes.
   - You can change the expiration date, including setting the date to a previous date to expire the attribute alert generator.
   - You can also update the case number and comments.
   - You cannot change the reason code, the attributes that you selected for the attribute alert generator when it was created, or the minimum resolution score.
4. Click **Update**.

Results

The system logs the changes that you made to the attribute alert generator. View or print an **Attribute Alert Generator History** report to see all changes to your attribute alert generators.

Help topics:

Find By Attribute screen:

Use this window to build a query to find entities in the entity database by attributes – name, addresses, numbers, characteristics, and so on. This type of query does not use the entity resolution process to return the query results.

Attribute Type

Type of entity attribute that you want to use as criteria for the query: name, address, numbers, characteristics, e-mail address, data source, or file load date. When you select an attribute type, the window displays query criteria fields appropriate to that type.

The query statement that you build depends on which attribute types you selected to query on:

- In a single query, criteria for more than one of the same attribute type creates an "OR" query statement. For example, "`Bob Hayes`" OR "`Rob Hays`".
- In a single query, criteria for multiple attribute types creates an "AND" query statement. For example, "`Bob Hayes`" AND credit card number "`5252-1010-5252-1010`".

Using this example, if you entered the following two names and a credit card, the query statement looks like the following statement: `Bob Hayes`" OR "`Rob Hays` AND credit card number "`5252-1010-5252-1010`".

Use the **Show Summary** button to see the full query statement.

Value fields

Enter the specific values of the attribute type to use to find entities. Each attribute type has its own set of value fields. If you leave value fields blank, the query looks for all potential values. However, when you enter data in all the value fields, the query runs faster, and the query returns better results.

- Name criteria is required.
- If you enter information in an address or e-mail criteria field, all address fields are required.
- If you select a Number or Characteristic type, the **Value** field is required.

+ button

Adds a new attribute row to the criteria.

- button

Removes the selected attribute row and criteria entry.

Find by Attribute - Results pane

Contains the results of the Find by Attribute query, based on the criteria entries. By default, the display area only shows the first 1,000 records that match the query criteria. (Although this option can be set by your system administrator.)

The results display by entity and represent the most recent information about each entity. If you double-click an entity in the results pane, the Visualizer opens the entity resume for that entity.

Entity ID
Displays the ID of the entity that meets the query criteria.

Name (*count*)
Displays the best name of the entity that meets the query criteria, and a number that represents the number of names associated with this entity. For example, Bob M. Smith (4) indicates that there are four names associated with this entity, Bob Smith.

Address (*count*)
Displays the best address of the entity that meets the query criteria, and a number that represents the number of addresses associated with this entity. For example, 1024 Daisy Lane, Akron, OH 43596 (24) indicates that there are 24 addresses associated with this entity.

Number type: *value*
Displays the best number types and number values of the entity that meets the query criteria.

Characteristic type: *value*
Displays the best characteristic types and values of the entity that meets the query criteria.

Relationships
Displays the number of relationships held by the entity that meets the query criteria.

Alerts Displays the number of alerts associated with the entity that meets the query criteria.

Find by Datasource Account screen:

Use this window to find an entity by account information from the original data source.

Enter External ID
Enter the data source account information that is associated with the entity in the data source specified in **Data Source**.

Data Source
Select the data source that corresponds to the account specified in **Enter External ID**.

Find By Entity ID screen:

Use this Find By method to quickly find an entity by Entity ID in the entity database. If the query locates the entity in the entity database, the Visualizer displays the Entity Resume for that entity.

Find By Resolution window:

Use the **Find by Resolution** window to create a search entity to compare to identities in the entity database.

Data Source Code - Description

Select a data source code and description to associate with the identities found by the Find by Resolution process.

Minimum Resolution Score

Select the minimum resolution score to use when comparing identities to the criteria specified for the Find by Resolution query.

The score you select determines the number and type of results that the query returns.

Find by Resolution criteria section

Specify the attributes to create the search entity that is compared to identities in the entity database. The system returns identities based on the minimum resolution score you specified.

Name List

Enter the name criteria in the name list fields, if you are looking for a specific name. If you enter any of the name fields, **Last Name** is required.

Address List

Enter the address criteria in the address list fields, if you are looking for a specific address. If you enter any of the address fields, **Street** is required.

Number List

Enter the specific number criteria, such as a passport number or a credit card number, in the number list fields. Both **Type** and **Value** are required.

Characteristic List

Enter the specific characteristic criteria, such as gender or date of birth, in the characteristic fields. Both **Type** and **Value** are required.

E-mail List

Enter the specific e-mail address criteria in the e-mail address list fields. Both **Type** and **Address** are required.

Attribute Alert Generator Manager window:

Use this window to view and manage your currently active attribute alert generators. The **Attribute Alert Generator Manager** window does not display expired attribute alert generators.

Expiration Date

Displays the date that the attribute alert generator expires.

Creation Date

Displays the date that the attribute alert generator was created.

Entity ID

The Entity ID of the search entity created by the attribute alert generator criteria.

Reason

The reason code assigned during the attribute alert generator creation process.

Minimum Resolution Score

Displays the minimum resolution score that entities must meet when comparing the attribute alert criteria to existing entities in the entity database before an attribute alert is generated for that entity.

Case Number
Displays the case number assigned during the attribute alert generator creation process.

Create button
Displays the **Create Attribute Alert Generator** window, so that you can create an attribute alert generator.

Edit button
Displays the **Attribute Alert Generator Info** window, so that you can edit the selected attribute alert generator. (Select the attribute alert generator and then click this button.)

Create Attribute Alert Generator window:

Use this window to create an attribute alert generator, which uses specified attribute criteria to persistently search the entity database for entities with matching attribute data.

Data Source Code - Description
Select a data source code and description from the drop-down list to associate with attribute alerts created from this attribute alert generator. The default selection is typically set to "Search".

Minimum Resolution Score
Select the minimum resolution score from the drop-down list to use when comparing identities to the criteria specified for the attribute alert generator.

Reason Code
Select a reason code from the drop-down list to associate with this attribute alert generator.

Case Number
Enter an optional case number for attribute alerts created from this attribute alert generator.

Comment
Enter an optional comment for attribute alerts created from this attribute alert generator.

Expiration Date
Select the date when this attribute alert generator expires or click the calendar icon and select a date using the calendar control. The expiration date defaults to six months from today's date. Because attribute alert generators always run in the background, it is a good idea to set an expiration date.

Attribute criteria section
Specify the attributes that you want to generate an attribute alert whenever the system processes an identity record containing the specified attributes.

Name List
If you are looking for a specific name, enter the name criteria in the name list fields.

Address List
If you are looking for a specific name, enter the name criteria in the address list fields.

Number List
If you are looking for a specific number, such as a passport number or a credit card number, enter the number criteria in the number list fields.

Characteristic List
If you are looking for a specific characteristic, such as gender or date of birth, enter the characteristic criteria in the characteristic list fields.

E-mail List
If you are looking for a specific e-mail address, enter the e-mail address criteria in the e-mail address list fields.

Attribute Alert Generator Info window:

Use this window to edit an existing attribute alert generator. You can only change to the case number, expiration date, and comments.

Reason Code
(Display only) Displays the reason code selected for this attribute alert generator.

Case Number
Displays the optional alphanumeric case number, entered by the user who created the attribute alert generator.

Comment
Displays any comments entered by the user who created the attribute alert generator.

Expiration Date
Displays the current expiration date for the attribute alert generator.

Names Used
(Display only) If name information was entered as criteria for this attribute alert generator, this section lists all name information entered by the user who created the attribute alert generator.

Address
(Display only) If address information was entered as criteria for this attribute alert generator, this section lists all address information entered by the user who created the attribute alert generator.

Numbers
(Display only) If numbers information was entered as criteria for this attribute alert generator, this section lists all number information entered by the user who created the attribute alert generator.

Other Attributes
(Display only) If characteristics information was entered as criteria for this attribute alert generator, this section lists all characteristic information entered by the user who created the attribute alert generator.

Update button
Click to apply your changes.

Analyzing entities

You can use the Visualizer to review, analyze, and graph entities in the entity database using the Visualizer.

Entities

An entity is a collection of one or more identities that represent the same person, organization, place, or item. Entities are stored in the entity database.

Although entities are often thought of as people, entities can also be things such as businesses or vehicles. In fact, you can use the system's extensible configuration to map your organization's data and create any type of entity that you want to resolve or relate.

Entities are often composed of identities that come from several different source systems. Entity resolution determines which identities are really the same entity and creates a composite entity that contains all the identities associated with that composite entity. The system maintains full attribution of the records, identifying the source associated with each identity in the composite entity.

You configure the system to resolve and relate entities in a way that meets the goals of your organization.

Entity resumes

An entity resume is a unified collection of all information in the entity database about a specific entity.

Entities are organized within the entity database using Entity IDs. Each Entity ID has its own entity resume.

You use the Visualizer to view entity resumes. Entity resumes might contain the following types of information:

- Source Document references
- Roles
- Names Used
- Addresses
- Numbers
- Characteristics
- Disclosures
- Related Entities
- Role Alert History
- Event Alert History
- E-mail Addresses

Viewing entity resumes

To see all the information about a specific entity in the entity database, view the entity resume.

About this task

You can access an entity resume from any of the following Visualizer locations:

- Any alert detail window
- Any graph window
- Any **Find By:** window

Procedure

- From a **Role Alert Detail** window, **Attribute Alert Detail** window, or **Event Alert Detail** window, click **Entity Resume**.
- From an entity graph, right-click the **Entity** icon that contains the Entity ID information that you want to see, and select **Entity Resume**.
- From the **Results** section of a **Find By:** window, double-click the row containing the entity whose resume you want to see.

Printing entity resumes

If you want a hardcopy of an entity resume, if you want a PDF version of an entity resume, or if you want to copy entity resume information into another application like a word processor or spreadsheet, there are several different ways to print an entity resume.

Procedure

- To print a snapshot of the **Entity Resume** window, do the following:
  1. From the **Entity Resume** window, click **Print**.
  2. From the print dialog, specify your print settings.
  3. Click **OK**.
- To print the entity resume to a PDF file, in the **Entity Resume** window, click **Report**.
- To copy (print) the entity resume information to paste into another application, do the following:
  1. In the **Entity Resume** window, from the **Edit** menu, **Copy Screen to Clipboard**.

     Note: The **Ctrl** + **C** key combination only copies single field values.
  2. Paste the clipboard contents into the application to use.
  3. Use the print feature of the application to print the entity resume information.

Printing the current window

You can print any window in the Visualizer, including graphs and entity resumes, directly from that window using the print command.

Procedure

1. In the Visualizer, from the window that you want to print, select **Print** from the **File** menu.
2. In the **Print** dialog, specify your print settings.
3. Click **OK**.

Viewing entity graphs

One of the main benefits of the Visualizer is that you can graph entity relationship and role alert information. The graphs provide a visual representation of the information about the selected entity.

About this task

You can access an entity graph from any of the following Visualizer locations:

- **Entity Resume** window
- **Graph** window
- **Event Alert Detail** window

Procedure

- From an **Entity Resume** window, click **Graph**.
- From a **Graph** window, right-click the **Entity** icon that contains the Entity ID whose information you want to see, and select **Show Entity Graph**. To view the entity resume of an entity in a graph, right click the entity and select **Entity Resume**.
- From an **Event Alert Detail** window, click **Graph**.
- Optional: To change the way information is displayed within a graph, right-click any blank space inside the graph, and then:
  1. Select a different **Graph Layout** setting to change the visual organization of the information in the graph.
  2. Select a different **Zoom** setting to change the current zoom level.

  Each time you change graph settings, the new settings are used as default settings for each additional graph you view during the current Visualizer session.

Viewing role alert graphs

If you want to see a graphical representation of how the entities that were identified in a role alert are related, you can view a role alert graph.

Procedure

1. In the Visualizer, from the **Alert Summary** window, double-click the role alert.
2. In the **Role Alert Detail** screen, click **Graph**.
3. Optional: To change the way information is displayed within a graph, right click any blank space inside the graph and then:
   a. Select a different **Graph Layout** setting to change the visual organization of the information in the graph.
   b. Select a different **Zoom** setting to change the current zoom level.

   Each time you change graph settings the new settings are used as default settings for each additional graph you view during the current Visualizer session.
4. Optional: To view the entity resume of an entity in a graph, right click the entity and select **Entity Resume**.

Customizing graph icons

All graphs in the Visualizer use predefined icons to represent entities and the types of attributes, such as addresses, and numbers. You can customize the icons that display on Visualizer graphs or specify an icon to use for a new attribute type.

Before you begin

Keep in mind the following constraints before customizing Visualizer graph icons:

- Custom icons reside on the application server. Only users with administrative privileges to the application server can add or change custom graph icons. All Visualizer clients based on that application server use the same icon set, so the change you make affects which icons display on Visualizer graphs for each of those clients.
- Save custom icons to a separate icons folder on the application server. Installing a new *.EAR file for the Visualizer removes all custom graph icons. After installing a new Visualizer *.EAR file, you can copy the custom graph icons from the icons folder to the designated application server icons folder.

- Icons must be in .GIF format. The recommended size for the image is 24 x 24 pixels.
- The names of the icons must match their corresponding attribute type, in lowercase only. For example, if you add a new attribute type called "Evidence Photo", the file must be named "evidence photo.gif" for the Visualizer to recognize the custom evidence photo. Notice in this example, both the attribute type name and the icon file name contain a space.

About this task

Default Visualizer icon image files are stored on the application server, typically in a folder called `images`.

Procedure

1. Stop the application server.
2. On the application server, find the default Visualizer graph icon folder. Typically, this folder is located at *`IBM Infosphere Identity Insight application server install_path`*`/ was_apps/ibm-is-ii-visualizer.ear/eas-visualizer-client.war/images`.
3. Required: Create a folder named `graph` under the default Visualizer graph icon folder (the `/images` folder) for your custom graph icon image files.

   Note: The folder name must be named `graph`.
4. Save, copy, or move each icon image file to the new folder.

Example

If you created an attribute type named `FINGERPRINT_FILE` and you wanted a custom graph icon to represent that attribute type on the Visualizer graphs, take the following steps:

1. Create or obtain a suitable .GIF image file that is 24 x 24 pixels to represent the `FINGERPRINT_FILE` attribute type. Make sure that the image file name matches the attribute type name and uses all lowercase letters, like this file name: `fingerprint_file.gif`
2. On the IBM InfoSphere Identity Insight application server, locate the `images` folder. For this example, the images folder is located here: `IBM-II_install/ was_apps/ibm-is-ii-visualizer.ear/eas-visualizer-client.war/images`.
3. Under the images folder, create a folder, named `graph`. So the file path looks like this path: `IBM-II_install/ was_apps/ibm-is-ii-visualizer.ear/eas-visualizer-client.war/images/graph`
4. Copy the `fingerprint_file.gif` image icon into the `graph` folder.

What to do next

Restart the application server.

Help topics

Entity Resume screen:

Use this screen to review detailed all known information about an entity, including the attributes of the identities associated with the entity, all related entities, and the history of all alerts associated with the entity.

Use the twisties to expand or collapse the sections of the screen to help you focus on a specific detail.

Data Source Information

Shows the data sources that provided identity records that resolved to this entity. Click a data source to highlight the attributes that make up the identity record that was processed from this data source. Data source information helps you to trace the identity record back to its original source.

When entities have multiple identities, the highlight can help you distinguish one identity from another and the original data source where that identity resides.

Roles

Displays the roles assigned to the identities that resolved to this entity.

Names

Displays the names used by the identities that resolved to this entity.

Addresses

Displays the known addresses used by the identities that resolved to this entity, including the date range each address was valid for the identity (if that information is available).

Numbers

Displays the known numbers used by the identities that resolved to this entity, including the date range each number was valid for the identity (if that information is available).

Characteristics

Displays the known characteristics used by the identities that resolved to this entity, including the date range each characteristic was valid for the identity (if that information is available).

E-Mail Addresses

Displays the known e-mail addresses used by the identities that resolved to this entity, including the date range each e-mail address was valid for the identity (if that information is available).

Disclosures

Displays disclosed relationships that were explicitly added by an analyst or authorized Visualizer user to link two identities. Disclosures create a 100% strength relationship between two identities.

Related Entities

Lists basic information about other entities that are related to this entity. Select a related entity to highlight the information that created the relationship.

Role Alert History

Lists basic information about the role alerts that are associated with this entity.

Event Alert History

Displays information about the event alerts that are associated with this entity.

Print button

Opens the print dialog, so that you can print the entity resume.

Report button

Generates an **Entity Resume** report, which contains all the information from the entity resume.

Entity Relationship Graph screen:

Use this screen to see a visual representation of relationship details for the selected entity, including entity attributes, related entities, and entity events.

Graph area (Canvas)

The body of the graph is referred to as the canvas. It contains the graphic representation of the relationships and shows you which attributes link the entities.

Click the objects (nodes) on the graph to reposition them on the graph. If a hyperlink attribute exists, use **Ctrl** + **Click** to follow the link.

Right-click menu options

Graph Layout
Changes the current layout and position of graph nodes. Each object on the graph is referred to as a node

Experiment with the graph layout settings until you find the setting that works for you. These settings are purely subjective to your tastes and needs while reviewing entity relationships on this graph.

Anneal
Select this setting to distribute nodes evenly. The anneal setting makes the graph edge-lengths uniform, minimizes line-crossings, and keeps nodes from coming too close to the edge of the graph.

Hierarchical
Select this setting to displays nodes according to hierarchy. The hierarchical setting works best on directed graphs that have an overall flow, or graphs that have some starting points, some ending points, and some overall flow between those points.

Organic
Select this setting to distributes the vertices of the graph evenly. The organic setting makes edge-lengths uniform and reflects graph symmetry, but it does not allow you to show related entities.

Self Organizing
Select this setting to create uniformly spaced clusters of linked graph nodes.

Random
Select this setting to randomly scatter graph nodes.

Tilt Select this setting to shift or tilt the graph node placement of the previously selected graph layout.

Circle Select this setting to arrange graph nodes into a circle that has even spacing between neighboring graph nodes.

Zoom Select a setting to change the display size of the canvas within the current screen size.

75% Displays the graph at 75% of its original size.

50% Displays the graph at 50% of its original size.

Show All Attributes
Displays all attributes assigned to that entity.

Hide Attribute
Hides the selected attribute.

Show Related Entities
Displays all other entities that relate to that entity, as well as a graphical representation of how those entities are related. This option is not available if the current Graph Layout setting is **Organic**.

Entity Resume
Opens the Entity Resume window and displays a detailed summary of all information known about that entity.

Entity Events
Opens the Entity Events screen and displays information about the events that are associated with the entity. This option is only available if the selected entity has associated events.

Show Entity Graph
Opens the Entity Graph window and displays a visual representation of information about only that entity

Adjust Graph options

Zoom slider
Move the zoom indicator to resize the canvas.

Layout Constraint
Select a layout bounds constraint for the canvas size.

Properties table

Select a node on the graph, and this table provides the properties of the selected node: Attributes or entities.

Role Alert Graph screen:

Use this screen to see a visual representation of role alert details for the selected entity, including entity attributes, related entities, and entity events.

Graph area (Canvas)

The body of the graph is referred to as the canvas. It contains the graphic representation of the details of the role alert.

Click the objects (nodes) on the graph to reposition them on the graph. If a hyperlink attribute exists, use **Ctrl** + **Click** to follow the link.

Right-click menu options

The right-click menu gives you control over the graph display and provides options to navigate to related entity windows.

Graph Layout
Changes the current layout and position of graph nodes. Each object on the graph is referred to as a node

Experiment with the graph layout settings until you find the setting that works for you. These settings are purely subjective to your tastes and needs while reviewing the role alerts on this graph.

Anneal
Select this setting to distribute nodes evenly. The anneal setting makes the graph edge-lengths uniform, minimizes line-crossings, and keeps nodes from coming too close to the edge of the graph.

Hierarchical
Select this setting to displays nodes according to hierarchy. The hierarchical setting works best on directed graphs that have an overall flow, or graphs that have some starting points, some ending points, and some overall flow between those points.

Organic
Select this setting to distributes the vertices of the graph evenly. The organic setting makes edge-lengths uniform and reflects graph symmetry, but it does not allow you to show related entities.

Self Organizing
Select this setting to create uniformly spaced clusters of linked graph nodes.

Random
Select this setting to randomly scatter graph nodes.

Tilt Select this setting to shift or tilt the graph node placement of the previously selected graph layout.

Circle Select this setting to arrange graph nodes into a circle that has even spacing between neighboring graph nodes.

Zoom Select a setting to change the display size of the canvas within the current screen size.

75% Displays the graph at 75% of its original size.

50% Displays the graph at 50% of its original size.

Show All Attributes
Displays all attributes assigned to that entity.

Hide Attribute
Hides the selected attribute.

Show Related Entities
Displays all other entities that relate to that entity, as well as a graphical representation of how those entities are related. This option is not available if the current Graph Layout setting is **Organic**.

Entity Resume
Opens the Entity Resume window and displays a detailed summary of all information known about that entity.

Entity Events
Opens the Entity Events screen and displays information about the events that are associated with the entity. This option is only available if the selected entity has associated events.

Show Entity Graph
Opens the Entity Graph window and displays a visual representation of information about only that entity

Adjust Graph options

Zoom slider
Move the zoom indicator to resize the canvas.

Layout Constraint

Select a layout bounds constraint for the canvas size.

Properties table

Select a node on the graph, and this table provides the properties of the selected node: Attributes or entities.

Adding data through the Visualizer

Typically, entity data is loaded by UMF data file in batch mode or real-time processes into the pipelines by system operators. However, Visualizer users can use the Visualizer to manually add a single entity, disclose a relationship between two entities (by identity), load and process a UMF data file, or validate a UMF data file before loading it.

Before you begin

Adding data always requires an available, running pipeline to process the data. But Visualizer users do not have to start or run their own pipeline. When the Visualizer adds data, it automatically sends the data through a designated Visualizer pipeline.

Adding a single entity

You can add a single entity to the entity database, without manually creating a UMF record. You can create an entity with as little as name information, but you should enter as much known information about the entity (known addresses, numbers, characteristics, or e-mail addresses) as possible, for optimum entity and relationship resolution.

Procedure

1. In the Visualizer, do one of the following:
   a. Click **View** > **Add** > **Entity**.
   b. From the toolbar, click the (add) icon and select **Entity**.
   c. From the toolbar, click the arrow and select **Entity**.
   d. From the **Add** window, in the **Add** drop-down, select **Entity**.
2. Use the drop-down lists and fields to enter the information about the entity. As you enter data, the screen guides you by highlighting required fields in yellow. A yellow highlighted field indicates that, based on your other selections on this screen, you must enter data into any field highlighted in yellow.
   - **Reference** field: You must enter information in this field. The reference information is an identifier for the identity. For example, enter the data source account number, such as a bank account.
   - Name fields: If you enter any part of a name (first name, middle name, or generation), the last name is required.
   - Address fields: You can add information in the **Address** field without entering city, state, postal code, or country. But you must enter information in the **Address** field if you enter any other part of the address.
   - Number, Characteristic, or Email fields: If you want to enter information in any of these attributes, you must select a type and enter a value for the attribute.

Attention: All information that you enter on this screen becomes part of the entity that you add. You are not indicating relationships with other entities or shared characteristics or numbers. Only enter information that belongs to the entity that you are adding, such as aliases or other names associated with the entity, and addresses, numbers, characteristics, and e-mail addresses that are associated with the entity.

3. Click **Submit**.

Results

The Visualizer creates a UMF identity record that includes all the information you entered for this entity and sends the record to a pipeline, where it is processed for entity and relationship resolution and added to the entity database.

Loading data from a file

Use the **File Load** feature in the Visualizer to load data for multiple identities that are defined in a UMF file. **File Load** will only load `<UMF_ENTITY>` records. When you select a UMF file, the system opens the file, loads the data into the pipeline, and then the pipeline processes the identities in the file, which adds them to the entity database and resolves any entities and identified relationships. Alerts are generated based on the rules that were configured.

About this task

Entity and relationship resolution occur in the pipeline component. To load and process UMF files through the Visualizer, there must be a pipeline that is running and available for the Visualizer server to communicate with.

Before loading a file, you might want to validate the UMF in the file, to be certain that there are no errors in the file.

Procedure

1. In the Visualizer, do one of the following actions:
   a. Click **View** > **UMF** > **File Load**.
   b. From the toolbar, click the (UMF) icon.
   c. From the **UMF** window, in the **UMF** drop-down field, select **File Load**.
2. Click **Load File...** to select the UMF file to load, and then click **Open**. The system loads the selected file into the pipeline, and the pipeline begins processing the data in the file. The **File progress bar** shows you the time elapsed during processing, the number of records processed, and the status of the file load.
   a. To stop the file load and processing, click the (Stop) icon button.
   b. To pause the file load and processing, click the (Pause) icon button.
   c. To resume loading and processing the file after pausing, click the (Continue) icon button.

   As the data in the file is loaded, a pipeline processes the data through entity and relationship resolution. If you see an error, contact your system administrator. The error is most likely a pipeline problem.

   New identities are added to the database, along with resolved entities and relationships. The system generates any alerts related to the data, based on the configured system rules.
3. Optional: After the file is loaded and processed, click **View Results** to display the **File Load Results** dialog, which includes the following information:

- The number of records sent to the pipeline.
- The number of new entities created in the entity database, based on the data from the file that you loaded.
- The number of UMF exceptions that the pipeline encountered while processing the data in this file. (This number can indicate errors in the UMF file or problems in the syntax that prevent the pipeline from fully processing the data.)

What to do next

If the **File Load Results** dialog indicates that there were any UMF exceptions in the file that you loaded, validate the file, using the UMF File Validation feature to help you find the errors in the file, so that you can correct them. Once you correct the errors, reload the data that contained the errors, so that the pipeline can fully process that data.

Validating a UMF file before loading the data

If you plan to use the Visualizer to load and process records in small UMF files, you might want to validate the data in the file first.

About this task

The validation process checks to see if the data meets the minimum requirements for entity and relationship resolution processing. The validation process also provides helpful information about areas of the file to review or correct before loading and processing the data. The better the quality of the data that goes into the system, the better the results.

Procedure

1. In the Visualizer, do one of the following actions:
   - Click **View** > **UMF** > **UMF Validate File**.
   - From the toolbar, click the arrow to the right of the icon and click **UMF Validate File**.
   - From the **UMF** window, in the **UMF** list, select **UMF Validate File**.
2. Click **Validate File...**.
3. Choose the UMF file to validate.

   Note: If you already validated one or more UMF files and kept the **UMF** window open, both the **File to Validate** and **Error/Warning File** fields contain the values from the last UMF file validation.
4. Optional: To change the directory path or file name of the validation process log file in the **UMF Validation Setup** window, choose one of the following actions:
   - Select the directory and filename to use, click **Browse...**, and then click **Open**.
   - Type the full path and file name of the validation error and warning log file. You can either type the name of an existing log file or the name of a new log file.

   Note: If you validate more than one UMF file and keep the **UMF** window open, notice that the log file value in the **UMF Validation Setup** window defaults to the same path and file name as the last validation error and warning log file. Closing the **UMF** window clears the path and log file fields.

5. Click **Validate UMF File** to start the validation process. While the validation process runs, validation statistics display, including dynamic information on the percentage complete, the elapsed time, the number of records processed, and the status of the process. You can pause or stop the validation process at any time.
6. Optional: When you click **Validate UMF File**, if another validation log file exists with the same location and name as the one you typed in step 4, the system displays an informational message to inform you. The message includes the name and location of the file is included in the message. Do one of the following actions:
   - Click **Yes** to use the same validation error and warning log file. This choice overwrites the previous log file.
   - Click **No** to create or use a different validation error/warning log file. The system returns you to the **UMF Validation Setup** window, so that you can manually change the path and filename of the validation error and warning log file.
7. After the validation process completes, click **View Results** if you want to see a summary of the results.

What to do next

Use the information in the **UMF Validation Results View** window to view the results and the information in the error and warning log file.

Disclosing relationships between identities

If you determine that you have data that links two identities (or accounts), you can specify that link to disclose the relationship using the Visualizer.

Procedure

1. In the Visualizer, do one of the following:
   a. Click **View** > **Add** > **Disclosure**.
   b. From the toolbar, click the arrow to the right of the (add) icon and select **Disclosure**.
   c. From the **Add** window, in the **Add** drop-down, select **Disclosure**.
2. Required: In the **Entity ID** fields, enter the Entity ID numbers of the entities that contain the identities to relate.
3. Required: Click **Lookup** for each Entity ID to retrieve its associated identities. Review the list of retrieved identities, to be certain that you entered the intended Entity ID.
4. For each entity, select the option button of the identity (or data source account) that you are disclosing a relationship for.
5. In **Disclosed Relationship Description**, enter a description of how the identities are related.
6. Click **Create**. A confirmation box displays verifying that the disclosed relation was successfully created.

Help topics

Add Entity window:

Use this window to add a single, new identity to the entity database through the Visualizer. All the information that you enter on this screen becomes attributes of the newly created identity. (You create one identity at a time.) After you submit data that you entered for the identity, the system processes the data through the pipeline for entity and relationship resolution, during which the identity can be associated with one or more existing entities.

Data Source Code - Description

Select the data source to associate with the identity that you are adding. The data source must exist in your system. (You cannot add a new data source here. If you do not see the data source code and description that you want to use, contact your system administrator to create the data source for you.)

To add an identity, data source code and description is required.

Reference

Enter an identifier for this data source account, which is used to associate the account with the identity that you are entering. (Examples of reference numbers include case numbers, bank account numbers, or customer rewards numbers.)

To add an identity, reference is required

Name List

Enter the names to associate with the single identity you are adding. To add an identity, name information (at least first name and last name) is required. You can indicate that identity that you are adding has more than one name by entering each name for the identity on a separate line. For example, if you know both the proper name for the identity, as well as one or more aliases ("also known as") names, you can enter all of them on this screen.

Note: Make sure to enter only one name per line.

All the names that you enter in this list are automatically associated with the newly created identity, as attributes of that identity. For example, if you enter "Robert Hays" and "Bob J. Hayes, Jr.", both of these names are associated with the newly created identity.

Address List

Enter one or more addresses that are associated with the identity that you are adding. For example, if you know the current and previous addresses for the identity, enter each full address, one address per line. All addresses that you enter in this list section are automatically associated with the identity that you are adding.

Addresses are not required to add an identity. If there are no known addresses for this identity, you can leave this list section blank.

Address

Typically, this information is the information entered on Address 1 and Address 2 lines. For example: 555 Main Street Building 17 Suite 102-B

If you enter data in any of the address fields, you must enter data in the **Address** field.

From Date

Enter the date that this address information became valid for this identity, if known. For example, if this identity was known to be at this address starting March 15, 1999, enter that date.

You can enter a From Date without a Thru Date.

Thru Date

Enter the date that this address information became invalid for this identity, if known. For example, if this identity was known to leave this address ending June 1, 2001, enter that date.

You can enter a Thru Date without a From Date.

Number List

Indicate one or more numbers that are associated with the identity that you are adding. For example, if you know a credit card used by the identity, a drivers license number, an identification number, a passport number, and a phone number, enter each number on a separate line. All numbers that you enter in this list section are automatically associated with the identity that you are adding.

Numbers are not required to add an identity, so you can leave this list section blank. However, if you enter any number data, then both **Number Type** and **Value** fields are required.

Number Type

Select the number type from the drop-down list of available number types. These number types must exist in your system. (You cannot add a new number type here. If you do not see the number type that you want to use, contact your system administrator to create it for you.)

If you want to associate a number with the identity that you are adding, you must select a number type.

Value

Enter the number value for the selected number type. For example, if you are associating a passport with this identity, enter the passport number here.

If you want to associate a number with the identity that you are adding, you must enter a number value that corresponds with the number type.

Location

Enter the location associated with the number, if known or if it exists. For example, if you are associating a passport with this identity, enter the name of the country that issued the passport here. Or enter the name of the state of issue for a drivers license.

From Date

Enter the date that this number became valid for this identity, if known. You can enter a From Date without a Thru Date.

Thru Date

Enter the date that this number became invalid for this identity, if known. For example, the expiration date for a drivers license, passport, or credit card.

You can enter a Thru Date without a From Date.

Characteristic List

Indicate one or more characteristics that belong to or are associated with the identity that you are adding. For example, if your system collects characteristics such as date of birth, marital status, eye color, or height, you can enter each known characteristic in this list, one per line. All characteristics that you enter in this list section are automatically associated with the identity that you are adding.

Characteristics are not required to add an identity, so you can leave this list section blank. However, if you enter any characteristic data, then all characteristic fields are required.

Type Select a characteristic type from the drop-down list of available types. The characteristic type must exist in your system. (You cannot add a new type here. If you do not see the characteristic type that you want to use, contact your system administrator to create the characteristic type for you.)

If you want to associate a characteristic with the identity that you are adding, you must select a characteristic type.

Value Enter the value of the characteristic. If you want to associate a characteristic with the identity that you are adding, you must enter the characteristic value that corresponds with the characteristic type.

From Date
Enter the date that this characteristic became valid for this identity, if known. You can enter a From Date without a Thru Date.

Thru Date
Enter the date that this characteristic became invalid for this identity, if known. You can enter a Thru Date without a From Date.

Email List
Indicate one or more e-mail addresses that belong to or are associated with the identity that you are adding. Enter each known e-mail address in this list, one e-mail address per line. All e-mail addresses that you enter in this list section are automatically associated with the identity that you are adding.

E-mail addresses are not required to add an identity, so you can leave this list section blank. However, if you enter any e-mail data, then both **Type** and **Address** fields are required.

Type Select a type of e-mail address from the drop-down list of available types. The e-mail address type must exist in your system. (You cannot add a new type here. If you do not see the e-mail address type that you want to use, contact your system administrator to create the e-mail address type for you.)

If you want to associate an e-mail address with the identity that you are adding, you must select a type.

Value Enter the full e-mail address. If you want to associate an e-mail address with the identity that you are adding, you must enter the e-mail address value that corresponds with the e-mail address type.

From Date
Enter the date that this e-mail address information became valid for this identity, if known. For example, if you know the date that this e-mail account was opened, you can enter it here.

You can enter a From Date without a Thru Date.

Thru Date
Enter the date that this e-mail address information became invalid for this identity, if known. For example, if you know the date that this e-mail account was closed, you can enter it here.

You can enter a Thru Date without a From Date.

Submit button
To process the identity through the entity and relationship resolution and add the identity into the entity database, after you enter all the known and relevant information about the identity that you want to add, click **Submit**.

Reset button
To clear the window of all information entered without submitting it, click the **Reset** button. The identity is not processed through the entity and relationship resolution nor added to the entity database.

Add Disclosure window:

Use this window to disclose a relationship between two existing identities. By disclosing the relationship, you create a link between the identities, as well as between the entities containing those identities. Disclosing a relationship indicates that the link between these two identities has not already been detected by entity and relationship resolution, and that you have a specific reason to link the two identities manually.

Entity ID
Enter the Entity ID number of each identity that you want to relate, one in each **Entity ID** field.

Lookup
Click to display the identity information corresponding to the Entity ID that you entered. Do this for both Entity ID numbers. By reviewing the information that displays, you can verify that the Entity IDs correspond to the identities that you intend to relate. Or you can correct the Entity ID for one or both identities before linking them.

Option buttons (next to each identity associated with each Entity ID)
Select one identity for both Entity IDs. These IDs are the two identities that you want to relate.

Note: You might only see one identity for each entity, which means that entity only currently has one identity in the system.

Disclosed Relationship Description
Type a description of how the two selected identities are linked. This description provides helpful information to other Visualizer users, when they view this relationship. It helps those users to understand how and why these two identities are being linked.

Create
Click **Create** to disclose the relationship between the two selected identities. The system sends the information about both identities through the pipeline for processing, and then updates the data for both identities, as well as all entities associated with these identities.

UMF File Load window:

Use this window to load data from a UMF file into the entity database through the Visualizer.

File load status bar
After you select a UMF file to open and load, and then click the **Load File...** button, this status bar shows the progress of processing the data in

the file. The system displays statistics that include the percentage complete, elapsed time since the file started processing, and the status of the system processing.

(Continue) button
If you have paused loading and processing the file using the (Pause) button, click this button to resume loading and processing the remaining unprocessed records in the file. The system continues with the next record in the selected file.

(Pause) button
Click this button if you want to temporarily pause the loading and processing of the file. The file remains in memory, and the system tracks which records have already been processed. Any records in the file that have not yet been processed are not in the entity database until you continue the file load.

This button is only active while the system is loading the file.

(Stop) button
Click this button if you want to stop the loading and processing of the file. The file is cleared from memory. Any records in the file that have not yet been processed are not in the entity database. If you want to continue loading records in this file, you must load the file again. During the file reload, any records that have already been processed are processed again.

This button is only active while the system is loading the file.

View Results button
Click this button to display the **File Load Results** dialog, which includes the following information:

- The number of records sent to the pipeline.
- The number of new entities created in the entity database, based on the data from the file that you loaded.
- The number of UMF exceptions that the pipeline encountered while processing the data in this file. (This number can indicate errors in the UMF file or problems in the syntax that prevent the pipeline from fully processing the data. Contact your system administrator for assistance in fixing UMF exceptions. Your system administrator can review an UMF exceptions log to get more detail.)
- The number of role alerts created, based on the data in the file that you loaded.

Load File... button
Click this button to load the file into the pipeline, and begin processing each record in the file for entity and relationship resolution.

UMF Validate File window:

Use this window to validate data in an UMF file that you want to load and process through entity and relationship resolution. By validating the data first, you can correct potential errors or warnings before loading and processing the file.

Validate... button
Displays the **UMF Validation Setup** window, where you select the UMF file to validate, set the path and file name of the error and warning log file, and initiate the UMF validation process.

If you keep the **UMF Validation Setup** window open and validate another UMF file, when you click **Validate...**, the path and log file fields populate with the locations of the last validated UMF file and the last error and warning log file location. You can either validate the same file again, or you can select a new UMF file to validate.

Closing the **UMF Validation Setup** window clears the path and log file fields.

Running reports from the Visualizer

From the Visualizer, you can view and print reports that show you summaries of statistics by data source and reports that help you view and manage alerts and disclosed relationships.

Viewing and printing reports in the Visualizer

Use the reports in the Visualizer to view statistics and quality summaries of data source files, help manage your assigned alerts, and review disclosed relationships, event alert, or events information. You can view the reports online or print a hardcopy.

About this task

You can access most Visualizer reports from the **View** menu or from the toolbar. But some reports you can only view and print from a specific screen, such as the Entity Resume report or the Event Alert Detail report.

Reports display in your selected Web browser using Adobe Acrobat Reader. You must have Adobe Acrobat Reader version 7.0 or higher installed on your workstation to view and print Visualizer reports.

Note: System-generated date and time stamps printed on reports from a Visualizer client are adjusted for the timezone of the Visualizer application server. The dates display correctly adjusted for the timezone of the Visualizer client when viewed on the screen. For example, an EST Visualizer client connected to a PST Visualizer application server displays a system generated date and time stamp as 8:00 PM on the screen, but it prints from an EST Visualizer client on a report as 5:00 PM.

Procedure

- To view an Attribute Generator History report, an Attribute Generator report, an Attribute Alert report, a Data Source Summary report, a Disclosure report, a Load Summary report, or a Role Alert Status report, do the following:
  1. Click **View** > **Reports**, and then select the report that you want to view or print.
  2. Complete the report criteria.
  3. Click **Run Report** to generate the selected report.
- To view an Entity Resume report, on the **Entity Resume** screen, click **Report**.
- To view a Role Alert detail report, on the **Role Alert ID** screen, click **Report**.
- To view an Event Alert Detail report, on the **Event Alert ID** screen, click **Report**.
- To view an All Events report, on the **Entity Events** screen, click **Report**.

Results

The system generates the selected report based on all specified criteria and displays the report in a separate window. If you want to print the report, click the **Printer** icon button or use your Web browser **Print** function.

Attribute Alert Generator History report:

The Attribute Alert Generator History report lists changes made to attribute alert generators, such as changes in expiration dates, case numbers, comments, or status. The report is sorted by Search Entity ID.

Search Entity
Displays the Entity ID (and name, if supplied) from the attribute alert generator search criteria.

Date and Time Created
Displays the date and time that this attribute alert generator was created.

Status history section
This section of the report shows you each update made to the attribute alert generator, starting with the most recent (last) update.

Comment
Displays comments that were entered by the user making the update.

Date and time updated
Shows the last date and time that this attribute alert generator was modified. If this attribute alert generator has not been modified, the date and time is the same as the **Date and Time Created**.

Date and time of expiration
Shows the date and time that this attribute alert generator is set to expire, or the last date that this attribute alert generator will generate attribute alerts.

Status Indicates whether the attribute alert generator is active or expired.

User Displays the name of the user that made this update.

Analyzer Group
Displays the Visualizer Analyzer Group that the last user to modify this attribute alert generator belongs to.

Min. Resolution Score
Indicates the minimum resolution score and description of the Minimum Score selected as part of the attribute alert generator criteria. This score threshold indicates how closely the attributes must match to generate an alert for this attribute alert generator. So "Is Entity" is the closest possible match, and "Any Relationship" is the least close possible match. You can set the threshold for each of these scores on the **System Preferences** screen on the **Configure Screen Preferences** window.

Reason Code
Displays the user-selected code indicating the reason for the attribute alert generator.

Case Number
Displays the optional alphanumeric case number, entered by the user who created the attribute alert generator.

Attribute Alert Generator report:

Use the Attribute Alert Generator report to manage attribute alert generators. By viewing this report, you can see a quick summary of all attribute alert generators in the system, including the date and time each attribute alert generator was created, its expiration date and time, its status, and the last date and time the attribute alert generator was updated. The report is sorted by Search Entity ID.

Search Entity
Indicates the ID of the Search Entity created by the attribute alert generator.

Date and Time Created
Indicates the date and time this attribute alert generator was created.

Comment
Displays the comment text added by user as part of the attribute alert generator.

Date and time updated
Indicates the last date and time this attribute alert generator was modified. If this attribute alert generator has not been modified, the date and time are the same as the **Date and Time Created**.

Date and time expired
Indicates the date and time that this attribute alert generator is set to expire.

Status Current status of this attribute alert generator, as of the last date and time this attribute alert generator was updated.

User Indicates the last user to modify this attribute alert generator. If the attribute alert generator has never been modified, this user name is the user who created the original attribute alert generator.

Analyzer Group
Indicates the name of the Analyzer Group that the last user to modify this attribute alert attribute generator belongs to.

Min. Resolution Score
Displays the selection in the **Minimum Score** drop-down list when the attribute alert generator was created. This score threshold determines how closely the attributes must match to generate an alert for this attribute alert generator.

You set the threshold for each of these scores on the **System Preferences** tab, which is part of **Configure Screen Preferences** dialog accessed from the **File** menu.

Reason Code
User-selected code indicating the reason for the attribute alert generator.

Case Number
Optional alphanumeric case number, entered by the user who created the attribute alert generator.

Attribute Alert report:

Use the Attribute Alert report to manage individual attribute alerts. By viewing this report, you see a listing of all entities that matched the attribute alert generator criteria, as well as the status and most recent activity on the alert.

The report is sorted by Search Entity ID in ascending order. If there is more than one matched entity per search entity, then the matched entities are sorted by ascending entity ID order.

Search Entity
Displays the Entity ID created by the attribute alert search.

Matched Entity

Displays the ID and name of the entity that matched the Search Entity, based on the attribute alert generator criteria. If an attribute alert has more than one Matched Entities, they display in alphanumeric order by Entity ID. For example, `Entity ID 37` displays before `Entity ID 1003`.

Attribute alert information

This section of the report displays general information about the alert results.

Attribute alert status

Displays the most current status of this attribute alert.

Search result date and time

Displays the date and time that the attribute alert was created.

Attribute search status

Displays the most current status of the attribute alert generator that generated this attribute alert.

Minimum resolution score

Displays the minimum resolution score and description of the Minimum Score selected as part of the attribute alert generator criteria. This score threshold indicates how closely the attributes must match to generate an attribute alert.

Attribute Alert status information

In this section of the report, you see the history of each status on this alert. The status information displays in update order, so the last status update displays first.

Date and Time of Status

Displays the date and time the attribute alert update occurred.

User Displays the name of the user who updated the alert.

Activity Code

Displays the user-defined code that indicates the action taken by a user on this attribute alert. When users update alerts, they select an activity code. Some examples of activity codes include Open, Assigned, Hold, and Closed. Activity codes are configured in the Configuration Console.

Status Displays the disposition status for this alert update, modified on status date and time. Disposition statuses display in update order, so the last status update is listed last.

Comment

Displays comments entered by the user making the update to this alert.

Matched information

This section shows which attributes by data type and value matched between the Search Entity and the Matched Entity.

Data Type

Displays the name of the attribute that matched between the Search Entity to the Matched Entity. The two values of this matched attribute display in both the Match Value and Search Criteria columns.

Search Criteria

Displays the data value belonging to the Search Entity that matched the corresponding value shown in the Matched Entity column.

Match Value
Displays the actual data value belonging to the Matched Entity, which matches the same data type and data value for the Search Entity.

Precision Description
Displays the text that describes the level of precision that the Search Criteria and Match Value matched at. Precision levels are configured during entity resolution configuration, by attribute.

Precision/Max Precision
The first number is the system-generated precision score, which indicates how closely the value in Search Criteria matched the value in Match Value. The second number is the maximum precision score attainable.

By comparing the two numbers, you can determine more about the closeness of the match between the Search Entity and the Matched Entity. You might also use these scores to determine if the attribute alert search criteria needs adjusting.

Score Adjustment
Displays the number associated with this attribute that is used adjust the resolution score up or down during entity resolution. This number is configured as part of the overall entity resolution configuration.

Data Source Summary Report:

The Data Source Summary report provides a quick statistical summary by data source of the records loaded into the system for processing. From this report, you can see the total number of records processed by load ID. Of those total records loaded, the report shows the number of records represented new identities or new entities and calculates the percentage of records that were new identities, as well as the percentage of records that were newly created entities.

Statistics by load within data source

Date Loaded
Displays the date that this data source file was loaded

Load ID
Displays the system-assigned load ID number.

Data Source
Displays the data source code and description (separated by a dash) for the data source file that was loaded.

UMF Records Loaded
Indicates the total number of identity records in this data source file that were loaded.

New Identities
Indicates the total number of new identities discovered in the data file that was loaded. (This number indicates an identity that has not been processed by the system before.)

New Identity %
Indicates the percentage of total records loaded (New Identities divided by UMF Records Loaded) that represent new identities.

New Entities
Indicates the total number of new entities created from this data load.

New Entities %
Indicates the percentage of total records loaded (New Entities divided by Loaded) that represent new entities.

Statistical charts by data source

Records Loaded by Data Source
Displays a bar chart that graphically shows how many records from each data source were loaded into the system, based on the other specified report criteria. You can see which data sources provided the most records or the least records and compare that to your estimated load numbers.

- The vertical axis shows the data sources by data source code.
- The horizontal axis shows the number of records loaded.

If there are fewer records loaded for a particular data source than expected, you can inspect the data files for this data source. (You might also consider running a Load Summary Report to see the data quality of the files loaded for this data source; data quality directly impacts the number of records loaded.)

New Entities by Data Source
Displays a bar chart that graphically shows which data sources yielded the most number of new entities, based on the other specified report criteria.

- The vertical axis shows the data sources by data source code.
- The horizontal axis shows the number of new entities created.

Disclosures report:

Use this report to view and manage disclosed relationships created between identities. Disclosed relationships are relationships that are either manually created by Visualizer users on the **Add Disclosures** screen or by including the disclosed relationship tag pair (<DR> and </DR>) on incoming identity records.

The report is sorted by relationship ID.

Relationship ID
Displays the system-generated number assigned to each disclosed relationship when the relationship is created.

Date and Time Created
Displays the date and time that the disclosed relationship was created.

Relationship description
Displays text that describes the reason for creating the disclosed relationship. This text is typed by the user who created the disclosed relationship.

Date and time updated
Displays the date and time of the last update to this disclosed relationship.

Status Displays the status of this disclosed relationship.

Date deleted
Displays the date and time that the disclosed relationship was manually

deleted. This field is only filled with a date and time if a user determined the relationship was invalid and deleted the disclosed relationship.

Data Source
Lists a data source code and description for both entities (one for each, on separate rows) that are now linked by this disclosed relationship. The data source code points to the original source file.

External ID
Lists an external ID for both entities (one for each, on separate rows) that are now linked by this disclosed relationship. The external ID often points to an account number within the original source file that belongs exclusively to the entity.

Event Alert Detail report:

Use the Event Alert Detail report to view full details about a specific event alert and the entities involved in the alert. This report is useful when you want a hard-copy report of the **Event Alert** tab on the **Research** window.

Alert ID
Displays the description and alert ID for a specific event alert. The alert ID appears before the description in the report header.

Event alert information
This section displays general information for the overall event alert, such as a description of the event alert rule that triggered this alert, and the event alert's status.

Alert date and time
Indicates the date and time this event alert was generated.

Rule ID
Displays an internal number generated by the system when the event alert rule was initially configured. This ID is associated with the event alert rule that triggered this event alert.

Rule description
Displays text to describe the event alert rule, defined by the user who configured the event alert rule.

Status Indicates the current status of this event alert.

Event Details
This section provides more information about the event alert data.

Date and Time
Indicates the date and time the event alert was generated.

Data Source
Displays, for each event, the data source code and description that supplied the event data. This information identifies the original source file.

External ID
Displays, for each event, the external ID associated with the data source code that supplied the event data. This information often identifies an account number for the entity in the original source file.

Event Reference
Displays, for each event, the unique code created by the complex event processor during event processor.

Quantity
Indicates, for each event, the number representing the quantity involved with this event. For example, 1 might mean one wire transfer of the value in the **Value** column.

Value Indicates, for each event, the total value of this event.

Entity information
For the entity involved in the event, this section provides the list of attribute types and their associated values that were involved in the event.

Alert Dispositions
This section provides a summary of the statuses for the event alert.

Activity Code
Displays the event activity code selected by the user who changed the status of this event alert.

Status Displays the status (Active or Inactive) associated with the event activity code.

Status Comments
Displays analyst comments entered about this status update.

User Indicates the user ID of the user who changed the status of this event alert.

Date and Time
Indicates the date and time that the status was changed.

Role Event Alert History section
This section lists all the role alerts that the entity responsible for this event alert is involved in.

Event Alert History section
This section of the report lists the complete history for the entity involved in the main event alert. Use this section to show you the number of event alerts that this entity is involved in.

Date and Time Alerted
Indicates the date and time the event alert was generated.

Alert ID
Displays the ID for this event alert.

Description
Displays text to describe the complex event processing rule that triggered this event alert.

Activity Code
Displays a user-defined code that indicates an action taken by a user on this alert. Activity codes are configured in the Configuration Console and selected from a drop-down list in the Visualizer when an alert is updated. Some examples of activity codes include Assigned, Closed, and Pending.

Status Displays the status for this alert update, modified on the status date and time. Statuses display in update order, so the last status update is listed last.

All Events report:

Use the All Events report to view all of the events associated with a single entity, whether the events generated an event alert or not. The report is useful when you want a hard-copy report of the **Entity Events** screen on the **Research** window. The events that display on the report depend upon the event type and the date range that you selected on that screen.

If you did not select an event type, the report displays events of all types for the given entity within the defined date range. If you selected an event type, only that events of that type within the defined date range display.

Basic report information
This section provides the basic report header information, such as the reporting date range and more about the entity associated with these events.

Report dates: From and Through
Indicates the starting and ending dates for the report. Only events that occurred within the date range for this entity display on the report.

Associated Entity
Indicates the Entity ID of the entity that is associated with these events.

Current name
Indicates the most current name for the entity in the entity database.

Current address
Indicates the most current address for the entity in the entity database.

Event information
This section provides the details of the events associated with this entity by event type.

Event Type
Describes the event type. This description is configured with the event type in the Configuration Console.

Event ID
Displays the system-generated number that identifies this specific event.

Create Date and Time
Displays the date and time that the event occurred.

Data Source
Displays the data source code and data source description associated with the event.

External ID
Displays the unique key that identifies the inbound identity record in the original data source for this event.

Event Reference
Displays additional information about the event, typically the name of the location where the event took place.

Location
Displays the address information for the location where the event took place.

Value Displays the value amount associated with the event.

Quantity
Displays the number of units associated with the event.

Unit of Measure
Indicates the unit of measure associated with the event value. The unit of measure is configured by event type in the Configuration Console. The unit of measure helps you to understand the value. For example, if the unit of measure is U.S. dollars and the event value is 5000, you know that this event involved $5,000.00.

Memo or *Custom label*
Displays additional information about the event, such as notes or comments, that can provide more context for the event transaction.

Users can define a custom label for this column as one of the options when configuring an event type in the Configuration Console. Instead of **Memo**, you might see a more descriptive custom label. For example, **Wire Transfer Notes**.

Additional Memo or *Custom label*
Displays more information about the event, if available.

Users can define a custom label for this column as one of the options when configuring an event type in the Configuration Console. Instead of **Additional Memo**, you might see a more descriptive custom label, For example, **Clerk Comments**.

Load Summary Report:

The Load Summary report summarizes statistics and quality characteristics by data source. It contains information about the data source files. Use this report to determine performance load statistics, the number of entities and alerts created by load, general information about the data quality of the data loaded, a summary of the actions about the UMF records by load, and any UMF exceptions that were generated by load. The report is grouped by load ID.

For each load, the report breaks the statistics into sections:
- Load Summary
- Role Alert Summary
- Relationship Summary
- Quality Summary
- UMF Document Summary
- Exception Summary

Load Summary

Use this section to help determine how long it took to process a particular file, as well as to give you a general idea of how useful this data source file is in overall entity resolution and relationship detection.

Date and Time Started
Indicates the date and time that the data load began.

Date and Time Completed
Indicates the date and time that the data source file load ended.

UMF Record Count
Indicates the total number of records loaded from this data source file within the **Date and Time Started** and **Date and Time Completed** range.

The **Date and Time Completed** number minus the **Date and Time Started** number is the number of minutes it took to load this particular data source file, which can give you an idea of system performance. It can also indicate that a larger data source file needs to be split into smaller files for quicker processing.

New Identities
Indicates the total number of new identities loaded within the **Date and Time Started** and **Date and Time Completed** time frame.

New Identity %
Indicates the percentage of total identities in this data load that are new identities (identities that are new to the entity database).

New Entities
Indicates the total number of newly created entities in the **Date and Time Started** and **Date and Time Completed** time frame.

New Entity %
Indicates the percentage of total entities that are newly created entities as a result of this data source load.

The number of new identities and new entities can provide you a general idea of how valuable this data source is in overall entity resolution and relationship detection. If these numbers are low and remain low over time, it might be that this data source is not useful in meeting your company entity resolution goals.

Role Alert Summary

Use this section to see the resolution rules and resolution scores common to the relationships detected that resulted in role alerts. Each row represents the number of role alerts that were generated, based on the criteria listed.

Resolution Rule
Displays the name of the resolution rule used to evaluate the identity and entity during entity resolution and relationship detection.

Alert Description
Displays the name of the role alert rule that triggered the role alert.

Severity
Displays a user-defined indicator to measure the priority or importance of this role alert.

Resolution Score
Displays a resolution score (0-100) for the resolution rule given to the identity and entity involved in the role alert. This score indicates the degree of likeness between the identity and the entity. A score of 100 means the identity record resolved to the entity.

Alert Count
Indicates the total number of role alerts generated based on the role alert rule description, resolution rule, and resolution score.

Relationship Summary

Use this section to see the attributes common to detected relationships that did not generate a role alert. Each row represents the number of relationships that were detected, based on the criteria listed.

Resolution Rule

Displays the name of the resolution rule used to evaluate the incoming identity records and existing entities during entity resolution and relationship detection.

Resolution Score

Displays a resolution score (0-100) for the resolution rule given to the identity and entity during entity resolution. This score indicates the degree of likeness between the identity and the entity. A score of 100 means the identity record resolved to the entity.

Relationship Score

Displays a relationship score (0-100) for the resolution rule given to the identity and entity during relationship resolution. This score indicates the degree of relationship between the identity and the entity.

The higher the relationship score, the more closely related the identity and entity are, based on matching attributes.

Relationship Count

Indicates the total number of relationships that are detected based on the resolution rule, resolution score, and relationship score.

Quality Summary

Use the information in this section to evaluate the quality of the data in each data source file. The section indicates the quality by attribute type within a UMF segment and UMF document type. By reviewing the Quality summary with the UMF exceptions summary, you can see which data source files have quality issues or malformed UMF that need to be addressed. Typically, you can resolve these issues through ETL or DQM/data source configuration before processing the data source file.

In some cases, this section can indicate that a data source is of such poor quality that you might not want to use this data source for entity resolution.

Document Type

Displays the name of the UMF document type that contains the data type listed in Data Type. Typically, this value is UMF_ENTITY.

Table Name

Displays the name of the database table that stores data from similarly named UMF segments. For example, data from the NUMBER segment is stored in the NUMS table.

Data Type

Indicate the data type, as listed in the incoming records attribute type UMF tags. This type corresponds to a UMF segment listed in Table Name. For example, if the Table Name is *ADDRESS* and the Data Type listed is *H*, the quality information is evaluating the address type of *Home*.

If you do not recognize a data type, that can indicate that the data source file is not correctly mapped to the appropriate combination of UMF documents, segments, and tags. Check the Exception Summary section to see if a matching UMF segment and UMF tag caused one or more segment exceptions. If the problem is invalid UMF, the numbers in the Low Quality Count in the Quality Summary section and the Segment Exception Count in the UMF Exception section often match.

Record Count
Indicates the total number of incoming identity records for the given Document Type, Table Name, and Data Type.

Generic Count
Indicates the total number of incoming identity records with the given Document Type, Table Name, and Data Type that contain values which are considered generic.

Low Quality Count
Indicates the total number of incoming identity records with the given Document Type, Table Name, and Data Type that are considered of poor quality. This number can indicate a data entry or ETL transformation problem in the data source file.

Usable Percent
Indicates the percentage of the incoming identity records with the given Document Type, Table Name (of this UMF segment) and Data Type that are usable for entity resolution and relationship detection. (Record Count minus Generic Count minus Low Quality Count) divided by Record Count equals Usable Percent.

Identity Percent
Indicates the percentage of the incoming identity records that contained the given Document Type, Table Name, and Data Type.

Attribute Summary

Use this section to see the attributes in the data source file that helped to detect relationships and generate role alerts. Each attribute maps to a specific UMF segment, and this section shows the number of relationships detected and role alerts generated, based on the data in the incoming UMF segment.

Segment Name
Displays the name of the UMF segment, which directly maps to an attribute.

Data Type
Lists the attribute type (or data type) within the UMF segment corresponding to the Precision Description. The report might list a specific attribute type or list *ALL*, indicating all attribute types in the UMF segment.

Precision Description
Describes the matching threshold between an attribute from an inbound identity and an attribute from an existing entity.

Role Alerts
Indicates the total number of role alerts generated based on this UMF segment, data type, and precision description.

Relationships
Indicates the total number of relationships detected based on this UMF segment, data type, and precision description

UMF Document Summary

You can use this section to validate the total number of incoming records in a data source file, based on what action is to be taken to the record. You can reconcile these numbers to the Record Count in the Load Summary section.

Document Type
Displays the name of the UMF document type. Typically, this value is UMF_ENTITY.

Action
Indicates the type of action for the incoming identity record. Here is a list of the most commonly used actions:

- *A* for add
- *C* for change
- *D* for delete

As part of the ETL process, identity records are typically tagged through UMF to indicate how to act on each incoming record during system processing.

UMF Record Count
Indicates the total number of records processed for each action type within document type.

Percent
Indicates the percentage of the total records loaded that the Record Count represents. (The sum should not exceed 100%.)

Exception Summary

Use this information to help pinpoint bad identity records, such as those with malformed UMF. The exception describes the problem, while the table name and element show which segment and record are bad. The count shows how many of the records in the file contained this bad UMF.

Document Type
Displays the name of the UMF document type. Typically, this value is UMF_ENTITY.

Action
Indicates the type of action for the incoming identity record:

- *A* for add
- *C* for change
- *D* for delete

As part of the ETL process, identity records are typically tagged through UMF to indicate how to act on each incoming record during system processing.

Segment
Displays the name of the UMF segment where the exception occurred.

UMF Tag
Displays the value of the UMF tag that caused the UMF exception.

Exception
Displays the message ID or other exception code to indicate the type of UMF exception that occurred and give information about how to resolve the exception. This information is also available in the UMF_EXCEPT table.

Segment Exception Count
Indicates the total number of this type of UMF exception.

Check the Low Quality Count in the Quality Summary section to see if a matching data type is reported as being of poor or unusable quality. If the problem is incorrect UMF, the numbers in the Low Quality Count in the Quality Summary section and the Segment Exception Count in the UMF Exception section often match for the same UMF segment and UMF tags.

Role Alert Detail report:

Use the Role Alert Detail report to view full details about a specific role alert, and the entities involved in the alert at each degree of separation. This report is useful when you want to further your analysis of the entities involved in each role alert.

For each degree of separation, the report displays information about the two entities involved in the alert for you to compare and contrast. Then the report shows other alerts associated with each entity, so you get a full picture of each entity and its associated role alerts. Typically, the detail about each role alert spans multiple pages.

Alert ID
Description and alert ID for a specific role alert. The alert ID appears before the description in the report header.

Role alert information
This section displays general information for the overall role alert, such as a description of the role alert rule that triggered this alert, and the role alert's status.

Date and time alerted
Date and time this role alert was generated.

Rule ID
Internal number generated by the system when the role alert rule was initially configured, this ID is associated with the role alert rule that triggered this role alert.

Rule description
Text to describe the role alert rule, defined by the user who configured the role alert rule.

Severity
User-defined code used to indicate the priority or importance of this alert.

Status Current disposition of this role alert.

Relationship confidence
Score that represents how closely related the two entities are that are listed under the Matching Details: Degree *n* section. The higher the score, the more closely they are related. A score of 100 indicates that the inbound entity and the matched entity are the same entity.

The relationship confidence score is generated by the system, as part of the entity resolution process.

Resolution score
Score that represents how closely the two entities match. The higher the score, the more closely they match. A score of 100 indicates that the inbound entity and the matched entity are the same entity.

The resolution score is generated by the system as part of the entity resolution process.

Resolution confidence
Base resolution score configured as part of entity resolution that represents the minimum score to resolve the inbound entity and the matched entity into one entity. Often, the Resolution score and the Resolution confidence score are the same.

Matching Details: Degree *n* section
This section provides matching details for entities involved in the alert and identity information for the respective entities. The two entities are represented as Entity *x* (Inbound Identity) and Entity *y* (Matched Identity).

For each entity and each attribute data type, the report lists the matching data values, as well as the data source and external ID associated with each entity's data values. Then the report displays the precision descriptions and scores for the matching attributes. If one of the matching attributes is Name, the report might also list details about how entity resolution scored the names, depending on which name scoring options are configured for entity resolution.

Data Type
Name of the matching attribute.

Value Data value that matched.

Data Source
For each entity, the data source code and description that supplied the matching attribute and data value. This information identifies the original source file.

External ID
For each entity, the external ID associated with the data source code that supplied the matching attribute and data value. This information often identifies an account number for the entity in the original source file.

Precision Description
Text that describes the level of precision that the entities matched at.

Precision levels are configured during entity resolution configuration, by attribute.

Precision/Max Precision
The first number is the system-generated precision score, which indicates how closely Entity *x* (Inbound Identity) matched Entity *y* (Matched Identity). The second number is the maximum precision score attainable.

By comparing the two numbers, you can determine more about the closeness of the match between the entities, such as the value of exploring the match further. You might also use these scores to determine if the alert search criteria needs adjusting.

Score Adjustment
The resolution score was adjusted by this number. This number is configured during entity resolution configuration.

Name Scoring Details
If one of the matching attributes is the Name data type, the report might also provide details on how the entity resolution process

scored the name matches. For this section of the report to display, one or more of the following name options must be configured as part of entity resolution:

- Name Manager
- Name Comparator 2

Full Name
Score (0-100) that represents how closely the full name of both entities matched. This score is configured as part of entity resolution.

Surname
Score (0-100) that represents how closely the family name of both entities matched. This score is configured as part of entity resolution.

Given Name
Score (0-100) that represents how closely the given name of both entities matched. This score is configured as part of entity resolution.

Entity *x* and *y* Identity Information section
This section of the report lists specific information about each identity.

Data Type
Characteristic name. (For example, Name.)

Value Characteristic value. (For example, SMITH, BRUCE.)

Other Alerts for Entity *x* and *y* section
This section of the report lists the role alert history of all other role alerts and relationships associated with both the inbound entity (Entity *x*) and the matching entity (Entity *y*). It also lists the event alert history of all event alerts associated with both the inbound entity (Entity *x*) and the matching entity (Entity *y*). This information can provide you with a more complete picture of each entity, its associated alerts and relationships to other entities, which can assist your analysis.

Role Alert History
Contains the information from the Entity Resume role alert history.

Alert Date and Time
Date and time the role alert was generated.

Alert ID
Description and alert ID for this role alert.

Description
Text to describe the role alert rule that triggered this alert.

Entity ID
ID number for the entity on this row that matched the entity listed by number in Other Alerts for Entity *n*.

Name Name of the other entity that matched the entity listed by number in Other Alerts for Entity *n*.

Relationships
Number of relationships associated with the related entity.

Relationship Score
Score that represents how closely the two entities are related. The higher the score, the more closely they are

related. A score of 100 indicates that the inbound entity and the matched entity are the same entity.

This score is generated by the system, as part of the entity resolution process.

Activity Code
User-defined code that indicates an action taken by a user on this alert. Activity codes are configured in the Configuration Console and selected from a drop-down in the Visualizer when an alert is updated. Some examples of activity codes include Open, Assigned, Hold, and Closed.

Status Disposition status for this alert update, modified on status date and time. Statuses display in update order, so the last status update is listed last.

Event Alert History
Contains the information from the Entity Resume event alert history.

Date and Time Alerted
Date and time the event alert was generated

Alert ID
Unique system-generated identifier for the event alert.

Description
Description of the event alert, from event configuration in the Configuration Console.

Role Alert Status report:

The Role Alert Status report summarizes the status of all role alerts for a specified time. Use this report to view and manage role alerts.

The report is sorted by role alert ID and alert date and time.

Alert ID - Description
Displays the role alert ID generated by the system, and the description of the role alert, obtained from the associated role alert rule.

Alert Date and Time
Indicates the date and time the role alert was created.

Matching entity information
This section shows the disposition history of the alert, beginning with the most recent status update.

Entity 1 and Entity 2
Displays the Entity IDs and typically the full names of the two entities who matched, based on the criteria for this role alert (by alert ID description).

Activity Code
Displays a user-defined code that indicates an action taken by a user on this alert. Activity codes are configured in the Configuration Console and selected from a drop-down in the Visualizer when an alert is updated. Some examples of activity codes include Open, Assigned, Hold, and Closed.

Status Displays the disposition status for this alert update, modified on status date and time. Statuses display in update order, so the last status update is listed last.

Status Date and Time
Indicates the date and time the alert status occurred.

User Displays the name of the user who updated the alert with this alert status.

Help topics

Attribute Alert Generator History report criteria window:

Use this Visualizer window to specify the criteria to view the Attribute Alert Generator History report. This report can help you to see and audit changes that were made to attribute alert generators, such as changes in expiration dates, case numbers, comments, or status. If you want to see the results of an attribute alert generator, view the Attribute Alert Generator report.

From Date
Type the first date in the date range to view data on the selected report. Use the `MM/DD/YY` format. For example, `01/01/01` represents January 1, 2001. Or click the calendar control and select the date.

The default **From Date** is today's date.

Thru Date
Type the last date in the date range to view data on the selected report. Use the `MM/DD/YY` format. For example, `01/01/01` represents January 1, 2001. Or click the calendar control and select the date.

The default **Through Date** is today's date.

To see data from a single day, use the same date in both **From Date** and **Thru Date** fields.

Status drop-down list
Select a specific status or select **All** to report on all statuses for all attribute alert generators. For example, if you only wanted to see changes made to attribute alert generators that are currently open within the specified date range, you would select **Open** from the drop-down list.

The default status in the **Status** drop-down list is **All**, which displays both active and expired attribute alert generators.

User drop-down list
Select an option to see your attribute alert generators or attribute alert generators created by anyone in your Visualizer user group.

The default option is `My Searches`.

Run Report button
Click this button to generate the report.

Attribute Alert Generator report criteria window:

Use this window to specify the criteria to view the Attribute Alert Generator report from the Visualizer. The Attribute Alert Generator report can be used to manage your attribute alert generators or those analysts in your Visualizer user group. If you want to see the change history of attribute alert generators, use the Attribute Alert Generator History report, instead.

From Date

Type the first date in the date range. Use the `MM/DD/YY` format. For example, `01/01/01` represents January 1, 2001. Or click the calendar control and select the date.

The default **From Date** is today's date.

Thru Date

Type the last date in the date range. Use the `MM/DD/YY` format. For example, `01/01/01` represents January 1, 2001. Or click the calendar control and select the date.

The default **Through Date** is today's date.

To see data from a single day, use the same date in both **From Date** and **Thru Date** fields.

Status drop-down list

Select a specific status or select **All** to report on all statuses for all attribute alert generators. For example, if you only wanted to see attribute alert generators within the specified date range that are currently active, select **Open**.

The default status is **All**, which means that the report displays both expired and active attribute alert generators.

User drop-down list

Make a selection:

- To see only your attribute alert generators, select **My Searches** (the default selection).
- To see all attribute alert generators created users in your Visualizer user group, select **My Group**.

Run Report button

Click this button to generate the report.

Attribute Alert report criteria window:

Use this Visualizer window to specify the criteria to view the Attribute Alert report, which can be used to help you view and manage your attribute alerts.

From Date

Type the first date in the date range to view data on the selected report. Use the `MM/DD/YY` format. For example, `01/01/01` represents January 1, 2001. Or click the calendar control and select the date.

The default **From Date** is today's date.

Thru Date

Type the last date in the date range to view data on the selected report. Use the `MM/DD/YY` format. For example, `01/01/01` represents January 1, 2001. Or click the calendar control and select the date.

The default **Through Date** is today's date.

To see data from a single day, use the same date in both **From Date** and **Thru Date** fields.

Status drop-down list

Select a specific status or select **All** to report on all statuses for all attribute alerts. For example, if you only wanted to see changes made to attribute alerts that are currently open within the specified date range, select **Open** from the drop-down list.

The default status in the **Status** drop-down list is **All**, which displays both active and expired attribute alert generators.

User drop-down list
Select a Visualizer user by user name or select **All** to report on the attribute alerts for all Visualizer users.

The default user in the drop-down list is your user name.

Run Report button
Click this button to generate the report.

Data Source Summary report criteria window:

Use this window to specify the criteria to view the Data Source Summary report from the Visualizer. The Data Source Summary report displays data loaded into the system by data source. Data sources help you to know where the identity data originated.

Data Source drop-down list
Select a specific data source or select **[all]** to view data from all data sources.

From Date
Type the first date in the date range to view data on the selected report. Use the `MM/DD/YY` format. For example, `01/01/01` represents January 1, 2001. Or click the calendar control and select the date.

The default **From Date** is today's date.

Thru Date
Type the last date in the date range to view data on the selected report. Use the `MM/DD/YY` format. For example, `01/01/01` represents January 1, 2001. Or click the calendar control and select the date.

The default **Through Date** is today's date.

To see data from a single day, use the same date in both **From Date** and **Thru Date** fields.

Run Report button
Click this button to generate the report.

Disclosure report criteria window:

Use this Visualizer window to specify the criteria to view the Disclosure report, which can help you see and manage disclosed relationships. Disclosed relationships are not discovered through entity and relationship resolution, but they are manual links between two identities. These manual links are typically created in the Visualizer, but they can also be created by placing the disclosed relationship UMF tag pair (<DR> and </DR>) on identity records loaded and processed by the pipelines.

From Date
Type the first date in the date range to view data on the selected report. Use the `MM/DD/YY` format. For example, `01/01/01` represents January 1, 2001. Or click the calendar control and select the date.

The default **From Date** is today's date.

Thru Date
Type the last date in the date range to view data on the selected report.

Use the MM/DD/YY format. For example, 01/01/01 represents January 1, 2001. Or click the calendar control and select the date.

The default **Through Date** is today's date.

To see data from a single day, use the same date in both **From Date** and **Thru Date** fields.

Run Report button
Click this button to generate the report.

Load Summary report criteria window:

Use this window to specify the criteria to view the Load Summary report from the Visualizer. You can use the Load Summary report to determine general information about the data quality of UMF files that you loaded in the Visualizer, along with helpful information, such as performance statistics and the number of entities resolutions and alerts generated by the file load.

Data Source Code - Description drop-down list
Select a specific data source or select **[all]** to view data loaded from all data sources. For example, if you loaded identity records from multiple UMF files on a single date, you could narrow the data on the report to a single data source by selecting the corresponding data source code.

From Date
Type the first date in the date range to view data on the selected report. Use the MM/DD/YY format. For example, 01/01/01 represents January 1, 2001. Or click the calendar control and select the date.

The default **From Date** is today's date.

Through Date
Type the last date in the date range to view data on the selected report. Use the MM/DD/YY format. For example, 01/01/01 represents January 1, 2001. Or click the calendar control and select the date.

The default **Through Date** is today's date.

To see data from a single day, use the same date in both **From Date** and **Thru Date** fields.

Run Report button
Click this button to generate the report.

Role Alert Status report criteria window:

Use this Visualizer window to specify the criteria to generate the Role Alert Status report, which summarizes the status of role alerts within a specified time period and can be used to manage your role alerts.

From Date and Time
Type the first date in the date range to generate data on the report. Use the MM/DD/YY format. For example, 01/01/01 represents January 1, 2001. Or click the calendar control and select the date.

Using military time, enter the first time in the time range to generate data on the report. Use the HH:MM format. For example, 09:00 represents 9:00 AM, and 20:30 represents 8:30 PM.

The default **From Date** and **Time** is today's date at 00:00.

Through Date and Time

Type the last date in the date range to print data on the report. Use the `MM/DD/YY` format. For example, `01/01/01` represents January 1, 2001. Or click the calendar control and select the date.

The default **Through Date** and **Time** is `today's date` at `23:59`.

To see data from a single day, choose one of the following options:

- Enter the same date in both the **From Date** and **Through Date** fields.
- Enter `00:00` in the **From Time** field, and `23:59` in the **Through Time** field.

Relationship Score Reporting Range

If you want to narrow the results by relationship score, type a range of relationship scores in the **From** and **To** fields.

The default range is from 0 to 100, which is all relationship scores.

Role Alert Rule drop-down list

Select a specific role alert rule to report on.

Role Alert Level drop-down list

Select a specific role alert level or select **All** to report on all role alerts.

Run Report button

Click this button to generate the report.

Analyzing data with the Analyst toolkit

You can use the tools and templates in the Identity Insight Analyst toolkit, to create and customize analysis reports and information in a browser-based application environment.

Reporting on data with the IBM Cognos reports

The Analyst Toolkit provides a set of Cognos reports that can be used to create customized Identity Insight reports.

The integration of IBM Cognos into Identity Insight creates a foundation for the ability to customize Identity Insight reports to suit the information that you require.

The Analyst toolkit includes the following elements to be used with IBM Cognos:

- Cognos Business Intelligence tools for query and application development
- Creation and deployment of an Identity Insight data model (developed with Cognos Framework Manager)
- Template reports for Entity Resume and Role-Alert Detail. These are intended as a starting point for customization and application development.

With Cognos reporting and the framework model, you have the tools necessary to create custom Cognos user-interfaces and reports based on the Identity Insight repository. You can use the included Cognos tools to create custom interfaces and modify EAS-provided templates.

The following terms and concepts are used in this product information:

Analyst toolkit

Identity Insight packaging for the installed Cognos components and sample templates.

EntitySearcher
A thin client browser application that combines the best of find-by-attribute and find-by-resolution search capabilities in a browser-based client.

IBM Cognos Business Intelligence
The general product name of the Cognos component that is included with Identity Insight.

Cognos Report
An XML based output specification that can be rendered as: an interactive UI in the Cognos Viewer, a PDF file, an XML file (for custom rendering), or a number of Excel formats (including CSV).

Active report
Cognos 10 introduced active reports, which are self-contained reports that look and feel more like Web applications than standard Cognos reports.

Cognos Framework Manager
A Cognos tool used to model a data source (typically a database). The Identity Insight data model has been created using Framework Manager.

Cognos Data Model
A logical representation of one or more data sources. Cognos report authors use the data model to create interactive reports.

Cognos Content Store
A separate database used by Cognos to store Cognos objects such as report definitions, data models, and queries. The content store is not used to store your Identity Insight data.

Analyzing data using the EntitySearcher thin client

The EntitySearcher thin client combines the best of find-by-attribute and find-by-resolution search capabilities in a browser-based client.

Identity Insight offers two primary search functions for finding entities. Find-by-resolution search, sometimes referred to as PSearch or pipeline search, uses entity resolution to find results. Find-by-attribute, referred to as EQ or enhanced query, uses a more traditional SQL lookup.

The EntitySearcher combines these two search approaches to help produce optimal results and avoid confusion about which approach to use. The client interface provides the familiar find-by-attribute interface to enter search criteria. One or both search types are called depending on the input criteria and results from each search. The results from both searches are compiled, de-duplicated, ranked, and presented in a search results grid.

An additional search enhancement allows you to search for entities that have a date of birth that falls within a specified date range. This search occurs when the **Expand search by** check box and drop-down list are used. For example, given a date of 6/1/1960 and a range of 30 days, the effective date range used in the search would be 5/2/1960 through 7/1/1960 [6/1/1960 minus 30 days and 6/1/1960 plus 30 days]. The range includes the end points.

You can choose the "Strict Search" option and only the find-by-attribute (EQ) search is used. A strict search is performed by default if any of the following conditions are met.

- A single attribute is entered for the search criteria.
- There are incomplete elements in the attribute search criteria.

- Wildcard characters are used in any attribute search criteria. For example, *.
- DOB (Date of Birth) attribute search criteria includes a date range.

The URL for launching the EntitySearcher is:

`http://server:install_port/EntitySearcher/`

From the search results, you can click through to a Cognos report version of the Identity Insight entity resume, graph component, or any other http-linkable target by having the system administrator configure the URL_ENTITY_DETAIL and URL_ENTITY_GRAPH values of the COMPONENT_CONFIG database table

Searching for entities using the EntitySearcher:

You search for entities based on attribute data and what kind of search you want to perform.

About this task

The EntitySearcher thin client combines the best of find-by-attribute and find-by-resolution search capabilities in a browser-based client. A user-interface for viewing the search results is available once a search is performed.

Procedure

1. Open the EntitySearcher in your browser.

   The URL for launching the EntitySearcher is:

   `http://server:install_port/EntitySearcher/`

   For example, `http://localhost:13510/EntitySearcher/`. The default *install_port* is 13510, but the port number can be changed. Check with your system administrator if you are unsure of the correct server name or port number.
2. On the **Search Entities** pane, enter search criteria. A single attribute is displayed for entity search by default.
   a. From the **Attribute list**, select the attribute type for your attribute search criteria.
   b. Enter the search criteria.

Option	Description
You have additional attribute search criteria.	To the right of the existing attribute, click **+**.
You have no additional attribute search criteria.	Go to the next step. **Note:** A strict search is performed when only a single attribute is entered for the search criteria.

3. Decide if you want to perform a combined search or only a strict search.

Option	Description
Perform a combined search	Combined search is performed by default.

Option	Description
Perform only a strict search.	Select the **Strict search** check box. **Note:** A strict search will be performed by default if any of the following conditions are met. • A single attribute is entered for the search criteria. • There are incomplete elements in the attribute search criteria. • Wildcard characters are used in any attribute search criteria. For example, *. • DOB (Date of Birth) attribute search criteria includes a date range.

4. Click **Search**.

Results

The **Search Results** pane lists any entity search results. The results are ranked according to likeness score and, if available, name score. High scoring (>86) find-by-resolution search results will be ranked first, followed by high scoring find-by-attribute search results. Lower scoring resolution search results follow.

What to do next

View an entity resume for a search result.
From the desired search result row **Entity ID** column on the **Search Results** pane, click the underlined **Entity ID** number value.

Note: The System Administrator may have to configure the URL_ENTITY_DETAIL value of the COMPONENT_CONFIG database table to enable this functionality.

View an entity graph for a search result.
From the desired search result row **Entity ID** column on the **Search Results** pane, click the graphing icon.

Note: The System Administrator may have to configure the URL_ENTITY_GRAPH value of the COMPONENT_CONFIG database table to enable this functionality.

Sample Cognos role alert report

The sample Cognos role alert report displays information about the entities and entity relationships involved in the alert, and is customizable using Cognos tools.

The role alert report makes use of Active Report technology introduced in Cognos 10 and provides a richer user experience.

Alert information is presented with each path of an alert appearing on a separate, dynamically created tab. Role alert summary information is presented at the top of the report, and entity snapshots (the state of the entity at the time the alert was generated) are available if the user wishes to view that information. An expanded matching details section shows identity insight scoring information.

Data access

The role-alert detail report makes extensive use of new Identity Insight database views. This approach allows for more control over data-access. For example, join and query structures are defined by the view SQL and not left to the Cognos engine. It also provides a layer of abstraction from the underlying data tables, which allows the underlying schema to be modified without directly affecting the Cognos reports.

Even though there are new Identity Insight database views to support the Cognos role alert detail screen, data access is provided and controlled by the Cognos server through the model.

Technical notes

The Cognos role-alert detail report makes use of Active Report technology. This means that the only supported output type is HTML. Unlike a standard Cognos report, all of the data used by the report is queried prior to the display of the report. This allows for Active Reports to maintain their interactivity when disconnected from the Cognos server. Active Reports can be distributed as .MHT files (MIME HTML) and are created from the Cognos home page or by accessing a URL for the report from any Web-browser that supports MHT files. Another side-effect of loading all of the report data up-front is that there is no need to reload the page when the user interacts with the UI.

The Cognos role-alert report requires a role-alert ID as a parameter. If the report is accessed directly, the user will be prompted for a role-alert ID. If the report is accessed as a component, the role-alert ID can be passed-in as a URL parameter. The parameter format for passing Cognos parameters via URL is to add a "p_" to the beginning of the prompt name. In the case of the Role Alert report, the parameter it expects is **pAlertID**, so the syntax would be: **p_pAlertID**. For example: **&p_pAlertID=55&**.

The Identity Insight database views created to support Cognos components are named using the prefix COG to make them more easily identifiable.

Firefox 3.x requires additional plug-ins to be installed to enable them to successfully display MHT files.

Sample Cognos entity resume report

The sample Cognos entity resume report provides all known information about an entity, and can be customized using Cognos tools.

In the Cognos resume report, entity data is summarized and the user can decide on which entity details to explore.

Technical notes

The Cognos entity resume makes extensive use of report-defined query objects, as opposed to actual database views. These virtual queries are based on the Cognos data model, and are created by dragging model objects into the report query builder and setting properties. The resume uses a "conditional block" object to display the detail section. Because of the use of a conditional block to make the screen feel more like a user-interface (and not a report), the PDF, text, and Excel output versions of this report do not look and behave like the default HTML output.

The Cognos report server queries only the information it needs to display the visible sections of the report. For example, the role-alert information is only queried when the user chooses to view that information. While this results in faster initial load times and smarter data access, it comes at a cost. The page must be reloaded when the detail section changes. This page reload is automatic and no user interaction is required, but the user must wait for the page to refresh before additional user interaction can take place.

The Cognos resume report requires an Identity Insight entity ID as its only parameter. If the report is run from the Cognos home page, the user will be prompted to enter an entity ID. While it is possible that some users will launch it from the Cognos home page and enter an entity ID, it is more likely that the Cognos resume will be an integrated component and called from another application such as a workflow or case-management tool. In this latter use-case, the Identity Insight entity ID can be passed-in as a URL parameter to the Cognos resume and the entity ID prompt page will not be displayed.

The parameter format for passing Cognos parameters via URL is to add a "p_" to the beginning of the prompt name. In the case of the resume report, the parameter it expects is **`pEntityID`**, so the syntax would be: **`p_pEntityID`**. For example: **`&p_pEntityID=5&`**.

Identifying and Installing Cognos components

You install IBM Cognos components in order to use and modify the IBM Identity Insight Cognos reporting features.

Before you begin

You must install IBM Business Intelligence Reporting before you deploy the IBM Identity Insight Cognos reports.

Note: If you have an existing instance of IBM Cognos Business Intelligence Reporting v10.1.0 or later installed, you can deploy the IBM Identity Insight Cognos reports into it.

To modify the Identity Insight Cognos reports' metadata, you must install IBM Cognos Framework Manager.

Procedure

1. Install IBM Business Intelligence Reporting v10.1.0 or later.
   a. Install the Cognos Reporting component using the detailed Cognos instructions.
2. Install IBM Cognos Framework Manager v10.1.0 or later.
   a. Install the Cognos Reporting component using the detailed Cognos instructions.

What to do next

Deploy the Identity Insight reports into Cognos.

Deploying Identity Insight reports into Cognos:

To enable the IBM Identity Insight Cognos role alert and entity resume reports, you must first deploy them into IBM Cognos Business Intelligence Reporting.

Before you begin

Install IBM Cognos Business Intelligence Reporting.

Procedure

1. Copy the Identity Insight Cognos reports deployment package into the IBM Cognos Business Intelligence Reporting installation. Identity Insight provides two versions of the reports, depending if you want to leverage Cognos' Dynamic or Compatible query mode

Table 22. Identity Insight Cognos reports deployment package locations

Copy file from	Copy file to
<*product installation directory*>ibm-home/cognos/deployment/IdentityInsight_v9.0_CompatibleQueryMode.zip or <*product installation directory*>/ibm-home/cognos/deployment/IdentityInsight_v9.0_DynamicQueryMode.zip	<*Cognos installation directory*>/deployment/

2. Go to the Cognos Connection page in your browser. The page is located at http://<*cognos_server_name_or_IP_address*>:<*cognos_port_#*>:cognos/index.html .
3. Click **Launch** > **IBM Cognos Administration**.

   Note: To access IBM Cognos Administration, you must have the required permissions for the Administration tasks secured feature.
4. Click the **Configuration** tab and click **Content Administration**. On the toolbar, click the **New Import** icon
5. From the list of available deployment packages, select `IdentityInsight_v9.0_Cognos`. When prompted for a password, enter `ISII4YOU`. Click **OK**.
6. In the name and description pane, click **Next**. The name and description pane does not need modification.
7. In the public folder content pane, from the list of available **public folders content**, select the **ISII** folder check box. Click **Next**.
8. In the directory content pane, click **Next**. The directory content pane does not need modification.
9. In the general options pane, click **Next**. The general options pane does not need modification.
10. Review the summary and click **Next**.
11. Select **Save and run once**. Click **Finish** to import the report. Click **Run**. The run options do not need modification.
12. Before closing the dialogue box, choose to view the details of the import. Click **OK**. If the status displays as "`Executing`", click **Refresh**. Upon successful deployment, the status displays as "`Succeeded`". Click **Close**.

What to do next

1. Verify that the reports are deployed.
2. Modify the Identity Insight Cognos report deployment database configuration.

Verifying the Identity Insight report deployment:

After deploying the reports, you must verify the deployment before running the reports.

Before you begin

Deploy the Identity Insight reports into Cognos.

Procedure

1. Go to the Cognos Connection page in your browser. The page is located at http://<*cognos_server_name_or_IP_address*>:<*cognos_port_#*>:cognos/index.html .
2. In the Public Folders tab, verify that the public folder **ISII** exists.
3. Select the **ISII** folder.
4. Verify that a single **Identity_Insight** package object exists. A package object is displayed as a blue folder.
5. Verify that both **ISII_EntityResume** and **ISII_RoleAlertDetailActive** reports exist.

What to do next

Modify the Identity Insight Cognos report deployment database configuration.

Modifying the Identity Insight Cognos report deployment database configuration:

After deploying and verifying the reports, you must modify the Identity Insight Cognos report deployment database configuration. Note: if you are using the dynamic query mode reports, see the Cognos documentation to create a JDBC connection from within Cognos (rather than following the procedure listed below).

Before you begin

Deploy the Identity Insight reports into Cognos.

Procedure

1. Go to the Cognos Administrator page in your browser.
2. From the left side, click **Data Source Connection**.
3. Select the **ISII** Data Source object.
4. Select the **ISII** Data Source Connection object.
5. Select the **ISII** Signon object.
   a. Click **Set properties**.
   b. From the **Signon** tab, click **Edit the Signon...**.
   c. Modify the link to include the Identity Insight database username and password. Click **OK**.
   d. Click **OK**.
6. Click **Set properties** for the Data Source Connection object.
7. From the **Connections** tab, complete the instructions for your Identity Insight database type.

Identity Insight database type	Instructions
DB2	1. Select **IBM DB2** for the type. 2. Select the **Edit the connection string** icon. 3. Modify the DB2 database name value. If a schema is required, add `currentSCHEMA=<schema>` to the DB2 connect string parameter. 4. Click **Test the connection...**. 5. Click **Test**. 6. Verify that the status is `Succeeded`.
Oracle	1. Select **Oracle** for the type. 2. Click **OK** when the `current connection string will be lost` warning appears. 3. Select the **Edit the connection string** icon. 4. Modify the SQL*Net connect string. 5. Click **Test the connection...**. 6. Click **Test**. 7. Verify that the status is `Succeeded`.

8. Click **Close** to close the test results panel.
9. Click **Close** to close the test connections panel.
10. Click **OK** to close the test connections panel.
11. Click **OK** to close the set properties panel.

Analyzing data using the graphing tool

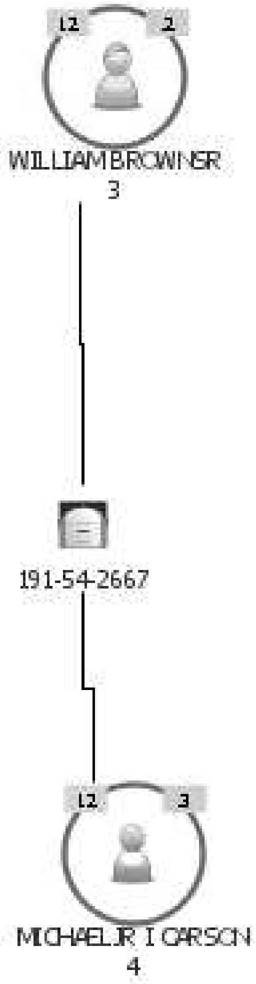
The InfoSphere Identity Insight graphing tool gives users the ability to analyze web-based graphs that visualize Identity Insight alerts, entity relationships, and other entity information.

To render graphs, the graphing tool requires that a product pipeline is up and running in the background.

The graphs rendered by the graphing tool are similar to the graphs rendered in the i2 Analyst Notebook component. However, the benefits of using the graphing tool include being able to embed and launch the graphs within an existing case management tool or other application. Or users can use a URL or web-start page to view and launch the graphs inside a web-browser. Installing and launching the i2 Analyst Notebook is not required to view the graphs rendered by the graphing tool.

Alert graph

The Alert graph produced by the graphing tool displays a specific role alert, based on the alert ID. The Alert graph helps you to visualize the entities involved in the role alert and the attributes that link the entities.



A role alert occurs when one or more entities is linked through a relationship that meets or exceeds a configured role alert rule. Role alerts are based on configured roles and role alert rules and can indicate:

- a warning or a problem, such as a customer is linked to a suspect on a watch list
- relationships of interest, such as a customer is also a vendor or an employee is linked to several customers through a particular phone number

Tips for using the Alert Graph

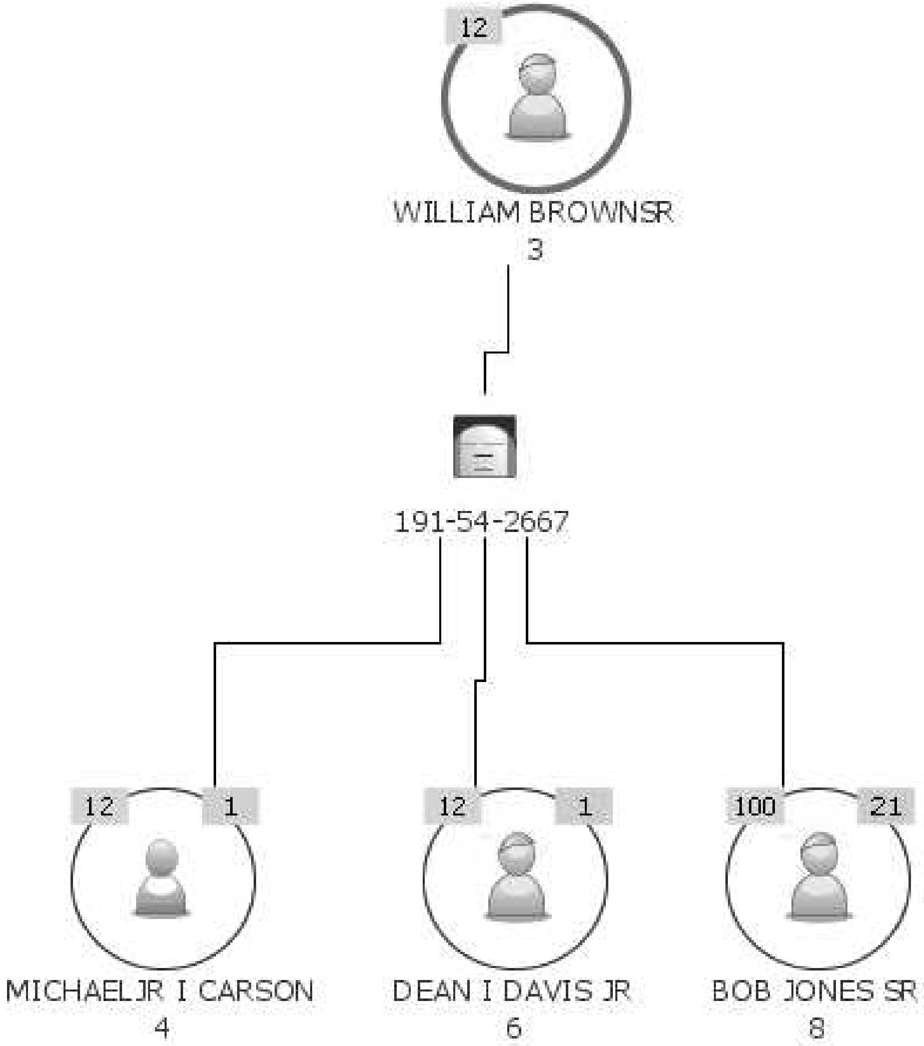
- If you see a related entities indicator for an entity involved in the alert, use the **Show remaining related entities** right-click menu option to show the remaining related entities. The graph redraws to display all the entities related to the chosen entity. The graph also automatically links those remaining entities to any existing entities on the graph to which those remaining entities are also related.
- The Alert graph displays only the attributes for each entity that contributed to the alert. To see all the attributes associated with a particular entity, right-click on the entity and select **Show remaining attributes**.
- To see the entity resume for a particular entity on the graph, right-click on the entity and select **Show resume**. The entity resume provides additional details about the entity, including the identities of that entity and other alerts that the

entity is involved in. This right-click menu option is only available if the link is correctly configured and you have access to the product that generates entity resumes, such as the Analyst toolkit.

Entity graph

The Entity graph produced by the graphing tool helps you to visualize the relationships between the specified entity and all entities related to that entity, based on shared attributes.

The Entity graph depicts the relationships between entities using alternating layers of entities and attributes.



First layer - Main entity

When you initially look at the graph, the first layer contains the main entity. The *main entity* is always the entity that you specified or selected to render the Entity graph. Visually, the line around the main entity node is always thicker, so you can spot the main entity no matter where it might actually display on the graph.

The *top entity* is the entity displayed in the first layer of the graph, at the top. Initially, the main entity is also the top entity. But any entity can become the top entity, simply by using the **Move to top** right-click option.

Second layer (and additional even-numbered layers) - Shared attributes
The second layer consists of the shared attributes that link the top entity to the entities in the third-layer of the graph. Attributes displayed on the graph show both the type and value of the attribute.

If there are additional layers of the graph, the even-numbered layers always contain the shared attributes that link the entities displayed above and below that attribute layer.

Third layer (and additional odd-numbered layers) - Related entities
The third layer of the graph shows entities that are related to the top entity at 1-degree of separation.

If there are additional layers of the graph, the odd-numbered layers always contain entities related to the previous entity layer, based on the shared attributes layer between the two entity layers. Entities displayed in these subsequent entity layers are related to the top entity in corresponding degrees of separation: Entities in the third layer are related to the top entity at 2-degrees of separation. Entities in the fifth layer are related to the top entity at 3-degrees of separation, and so on.

Tips for using the Entity graph

Entity graphs can contain many layers. Here are some tips to help you sort through all the attribute and entity information displayed on an Entity graph.

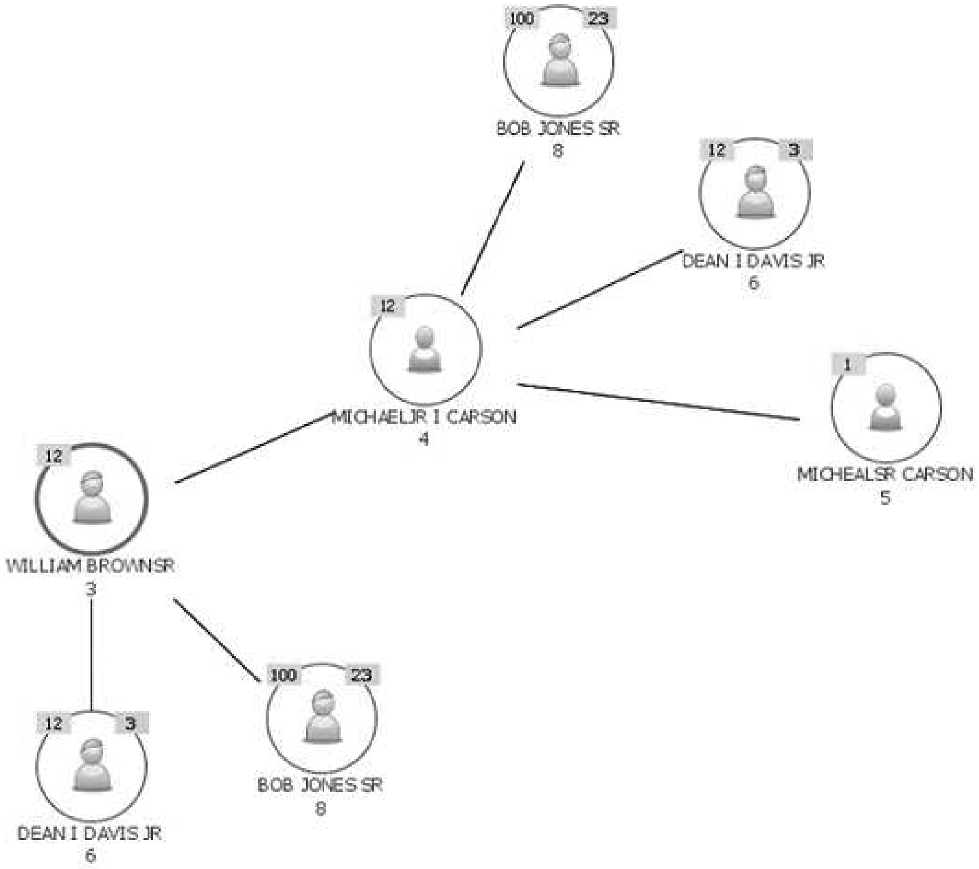
- To see more details about an entity:
  - Use the **Show remaining related entities** right-click menu option to explore the relationships for a specific entity that are not currently displayed on the graph.
  - Use the **Show path to top** Quick Filter to show how an entity or attribute is related to the top entity. This filter temporarily hides the unrelated entities and attributes from the graph.
  - Switch to the Social Network graph to create a graph that helps you to view and focus on the relationships between entities. The Social Network graph does not display shared attributes on the graph, but the shared attributes are listed in the Attribute Explorer. Right-click on the entity you want to create the Social Network graph from, and select **Create new graph - Social Network**.
  - Use the **Show related Entities only** Quick Filter to create a "mini" Social Network graph. This quick filter hides all attributes on the graph and shows only entities related to the chosen entity at 1-degree of separation. (The *chosen entity* is the entity that you right-clicked to apply the quick filter.)
  - Use the **Show related Attributes and Entities only** Quick Filter to highlight the entity and show only those attributes and entities that are related to the chosen entity.
  - Use the **Show entity resume** right-click menu option to see the entity resume of any entity on the graph. The entity resume provides additional details and context about that entity, such as the identities associated with the entity, other alerts that the entity is involved in, and so on. (If the link to the entity resume is not configured, such as to the entity resume in the Analyst toolkit, then this option does not display on the right-click menu.)
- To see the path between two entities on the graph:

- Use the **Show path to top** Quick Filter to visualize how a particular entity on the graph is related to the top entity. This quick filter is especially useful when the graph contains multiple layers.
- Use the **Move to top** right-click menu option to move an entity to the top of the graph and redisplay the existing attributes and entities based on how they relate to the new top entity. No new information is added to the graph

- To see more details about an attribute:
  - Use the **Show Attributes only** Quick Filter to focus the information on the graph to one entity. The filter helps you to see only those attributes for the selected entity.
  - Use the **Show remaining Attributes** right-click menu option to visualize all the attributes for a particular entity, even those attributes are not shared by any other entity on the current graph.
  - Use the **Attribute Explorer** to highlight the entities on the graph that share a particular attribute. The value in the **Entities** column can guide you. The higher the number in the column, the more entities displayed on the graph share that attribute.
- To rearrange the entities and attributes into other patterns and shapes, use the right-click menu to change the graph layout from **Layered** to **Radial**.

Social Network graph

The Social Network graph helps you to visualize the relationships between the selected entity and all entities that the selected entity is linked to. Using this unique graph, you get another way to see "who knows who".



The Social Network graph shows:

- Entity-to-entity links: You see all the entities related to the main (hub) entity. However, the attributes that link the entities do not display on the graph but are accessible by using the Attribute Explorer in combination with the graph.
- Relationship clusters: The Social Network graph is unique in that it displays the related entities in groups or clusters. This graph can help you see all the relationship clusters a particular entity belongs to and look for patterns in among the clusters and relationships.

You can expand the graph to show all the related entities for any entity. Each time you show all entities related to a particular entity, that entity node becomes the hub entity in a new relationship cluster.

To maintain the integrity of each relationship cluster, an entity can be displayed on the graph multiple times in multiple relationship clusters. But each entity displays in each relationship cluster only once. To see every relationship cluster the entity is part of, select the entity by clicking on that node. The interior of the selected entity node changes to blue in each relationship cluster that the entity is part of.

When an entity is the hub entity, the Related Entities indicator does not display, because all entities related to the hub entity already display in the relationship cluster. When the entity is one of the related entities in the relationship cluster and has other relationships that are not displayed in the that cluster, a Related Entities indicator displays.

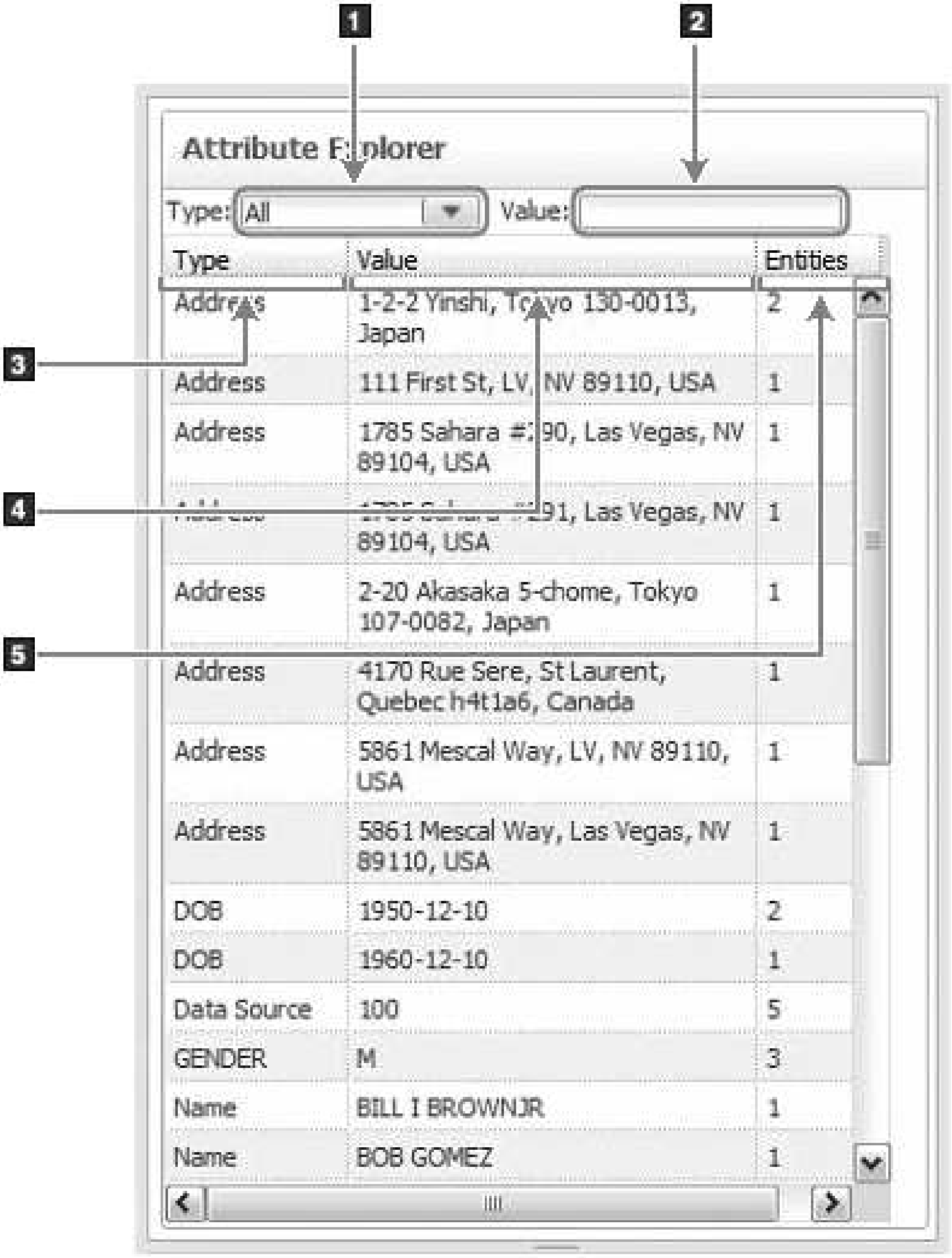
Tips for using the Social Network graph

- Use the **Show remaining related Entities** right-click option to expand the related entities for one or more entities on the graph. Each expansion creates another relationship cluster. Look for patterns between the clusters.
- If multiple relationship clusters are graphed, try zooming out to look for the bigger patterns and context in the clusters. For example, if a particular entity shows up in every cluster or many clusters, then that entity might be a big influencer within a particular sphere. Or that entity might be key to connecting multiple relationship clusters.
- Use the **Attribute Explorer** to see which attributes link the related entities. Select a particular attribute row to highlight every entity on the graph that shares that attribute. The value in the **Entities** column can show you which attributes are shared by the most entities.

Attribute Explorer

A component of the graphing tool, the Attribute Explorer is a table that lists all the attributes by type and value that are associated with all the entities on the currently displayed graph. The Attribute Explorer is automatically docked to the right of the graph canvas.

Parts of the Attribute Explorer



Type	Value	Entities
Address	1-2-2 Yinshi, Tokyo 130-0013, Japan	2
Address	111 First St, LV, NV 89110, USA	1
Address	1785 Sahara #290, Las Vegas, NV 89104, USA	1
Address	1785 Sahara #291, Las Vegas, NV 89104, USA	1
Address	2-20 Akasaka 5-chome, Tokyo 107-0082, Japan	1
Address	4170 Rue Sere, St Laurent, Quebec h4t1a6, Canada	1
Address	5861 Mescal Way, LV, NV 89110, USA	1
Address	5861 Mescal Way, Las Vegas, NV 89110, USA	1
DOB	1950-12-10	2
DOB	1960-12-10	1
Data Source	100	5
GENDER	M	3
Name	BILL I BROWNJR	1
Name	BOB GOMEZ	1

Image call out number	Item	Description
1	Type drop-down list	Select an attribute type to filter the attribute data that displays in the Attribute Explorer. When you use the Type drop-down list, you do not filter the graph, you filter only the data in the Attribute Explorer. For example, you might select **SSN** to filter the data in the Attribute Explorer to show only social security numbers. This drop-down list does not necessarily contain every configured attribute type for the product. The list contains only those attribute types associated with the entities that are currently displayed on the graph.
2	Value text box	Enter data in this field to narrow the attribute information displayed in the table, based on attribute values. The Attribute Explorer looks at each character entered and returns the list of attribute values that exactly match the entry, whether the match is an exact or partial data match. For example, if you enter `123`, the Attribute Explorer filters the list of attributes to only those attribute types that contain `123` somewhere in the attribute value. **Note:** The Attribute Explorer does not recognize wildcard characters. Whatever characters you enter into the text box, the Attribute Explorer looks for an exact, literal match to that character. So if you enter a typical wildcard character, such as an * (asterisk), the Attribute Explorer looks for a literal data value match to the * character.
3	Type column	Displays the attribute types that are currently displayed on the graph. The column items match the descriptions configured for Attribute Types in the Configuration Console. For example, a credit card attribute type might display as **CC** or **credit card**, depending on how it was configured in the Configuration Console. The column does not necessarily contain every configured attribute type for the product. The column contains only those attribute types that are currently displayed on the graph.

Image call out number	Item	Description
4	Value column	Displays the values for the attribute types that are currently displayed on the graph. For example, you might see a value of **04-01-1962** that corresponds to a data of birth attribute type.
5	Entities column	Indicates how many entities displayed on the graph share this attribute type and value. This information can help you to identify the most commonly shared attributes for further exploration.

Tips to using the Attribute Explorer

The Attribute Explorer can assist your analysis of the graphs, especially when the graph contains a lot of information.

- Use the **Entities** column to look for attributes that are only associated with one entity on the graph. Look for a 1 in the column. While the graph displays only the attributes that link entities, the Attribute Explorer displays all the attributes associated with all the entities on the graph. These attributes do not link the entity to any other entity node on the graph, but they might make a particular entity worth further exploration.
- Narrow the information displayed in the Attribute Explorer to one attribute type by selecting a type from the **Type** drop-down list. For example, if you see and select **Phone numbers**, the Attribute Explorer displays only phone number attributes and their values.
- Highlight all the entities on the graph that share the same attribute by selecting an attribute (table row) in the Attribute Explorer.
- Search the data on the existing graph for matching or common attribute values by entering data in **Value**. For example, if you entered `123` , the Attribute Explorer might return any or all of the following matching attributes:

Type	Value
Address	**123** Main Street, Anywhere, California, 1**123**4, USA
Address	97-**123** Rue Sere, St. Laurent, Quebec, H4T1A6, Canada
Phone number	555-222-5**123**
Tax identifier	554-**123**-3**123**

- You can also enter more than one full or partial value at a time in **Value**. The Attribute Explorer then treats the multiple values as an "AND" query. For example, if you enter `dog cat`, the Attribute Explorer searches every row for one that includes both `dog` AND `cat`. The order of the multiple values in the query does not matter. For example, if one of the attribute values in the Attribute Explorer is `her cats and his dogs`, that value is part of the results of the `dog cat` value query.
- Sort the information in the Attribute Explorer by column. Click the column header, and you see an arrow that indicates the sort direction.

Selected Properties

A component of the graphing tool, the Selected Properties table displays the properties of the attribute or entity node that is selected on the graph. The table only displays the properties for one selected node (attribute or entity) at a time.

- If you select an entity, this section displays all the attributes (types and values) associated with the selected entity.
- If you select an attribute, this section displays all the entities that share the selected attribute, including the entity ID for each entity. The third column of this section also shows you the identity ID of the data source, where the attribute data came from.

Navigating and exploring the graphing tool graphs

You can navigate and explore graphs rendered in the graphing tool by using the navigation toolbar or the right-click menu options on each graph.

Navigation toolbar

The navigation toolbar just under the graph title contains icons for standard graphing navigation.

- Selection mode options: Select individual graph items or select multiple graph items (or select a specific area of the graph)
- Reposition the graph on the canvas
- Reset the graph to the default view
- Zoom options: Zoom in or zoom out

Selecting and highlighting

On the Alert and Entity graphs, selecting (left mouse click) a node causes directly related attributes and entities to become highlighted. The selected node appearance changes to display a blue selection rectangle on top of the node. The interior of the highlighted nodes changes to blue.

Table 23. Graphing tool right-click menu option descriptions

When you select this type of node...	On this type of graph...	This data becomes highlighted...
Attribute	Alert graph Entity graph	All entities that share that attribute All related attributes
Entity	Alert graph Entity graph	All entities related to the selected entity at 1-degree The attributes that cause the 1-degree relationship
Entity	Social Network graph	Each time that entity displays on the graph, in each hub that the selected entity is related to. (One entity can display multiple times in multiple hubs on this graph type.)

You can select multiple nodes by using **Ctrl**. You can also move currently selected nodes by dragging and dropping them on the graph.

Right-click menu options

Choose an entity or attribute by pointing your cursor to it and right-clicking.

Table 24. Graphing tool right-click menu option descriptions

This right-click menu option...	Does this action...	Alert graph	Entity graph	Social Network graph
Zoom	Zooms in, zooms out, or fits the graph canvas to the screen size.	X	X	X
Quick Filters (general)	Helps you to focus on interesting graph data by temporarily hiding the less interesting data. Quick filters do not add or remove data from the graph. When a quick filter is turned on, the graph title bar displays [Quick Filter On]. Only one quick filter can be active at a time, but you can select a different quick filter when quick filtering is active. **Note:** When a quick filter is active, the filter displays only the graph data that applies to the currently selected entity or attribute. For example, if you select entity ABC and select the **Show related Entities only** Quick Filter, you see the entities currently displayed on the graph that are related at 1-degree to ABC.	X	X	
Quick Filter - Show Attributes only	Hides entities, so you can see the attributes associated with the entity that you right-clicked.	X	X	
Quick Filter - Show related Entities only	Hides all attributes, including attributes that link the entities to one another, so you can see the entities that are related at 1-degree to the entity that you right-clicked. This quick filter gives you the look of a Social Network graph from an Alerts or Entity graph.	X	X	

Table 24. Graphing tool right-click menu option descriptions (continued)

This right-click menu option...	Does this action...	Alert graph	Entity graph	Social Network graph
Quick Filter - Show related Attributes and Entities	Hides all graph data except for the entities linked at 1-degree to the entity that you right-clicked and the attributes that cause the 1-degree relationship. This quick filter is particularly useful when there is a lot of data on the graph, and you want to remove the excess clutter.	X	X	
Quick Filter - Show path to top	Filters the graph data to show the path that connects the entity or attribute to the top entity. If you right-clicked an attribute, the filter includes all the entities and attributes along the relationship path to the top entity. If you right-clicked an entity, the filter includes all the attributes and entities along the path to the top entity.	X	X	
Quick Filter - Turn off quick filtering	Turns off the current quick filter and redisplays the data filtered from the graph.	X	X	
Move to top	Moves the selected entity to the top of the graph, making that entity the top entity. This option does not add new data to the graph or the Attribute Explorer. But the graph is redrawn to show the data from the perspective of the new top entity.	X	X	

Table 24. Graphing tool right-click menu option descriptions (continued)

This right-click menu option...	Does this action...	Alert graph	Entity graph	Social Network graph
Show remaining Attributes	Displays all the attributes associated with the entity you right-clicked, even if those attributes do not link the entity to any other entity on the graph. The Attribute Explorer always lists all the attributes associated with an entity, so this option does not change any data in the Attribute Explorer. Displaying other attributes for an entity might provide another piece of a puzzle or lead you into further exploration of an entity or attribute.	X	X	
Hide remaining Attributes	Removes from the graph the attributes that do not link any entities displayed on the graph. If all the attributes displayed on the graph link entities that are currently displayed on the graph, this option is not available.	X	X	
Show remaining related Entities	Displays all the relationships not already displayed for the entity that you right-clicked on. The attributes that cause the relationships also display.	X	X	X
Create new Graph	Creates a new graph of the type you selected, that features the entity you right-clicked as the main entity.	X	X	X

Table 24. Graphing tool right-click menu option descriptions (continued)

This right-click menu option...	Does this action...	Alert graph	Entity graph	Social Network graph
Graph Layout	Controls the display of the graph layout: • Layered: Displays graph data in layers, showing alternate rows of attributes and the entities linked to those attributes. This layout is the default layout for both the Alert and Entity graphs. • Radial: Displays the graph data as nodes and connecting lines, randomly spread on the graph canvas. This layout can be helpful if you want to arrange the entities and attributes yourself.	X	X	
Show Resume	Displays the entity resume in a new window, if the link is configured in the `graph.properties` file. The entity resume provides detailed information about the selected entity, including all alerts the entity is involved in and all identities associated with the entity. The resume is a useful analysis tool, especially when used with the graphing tool graphs. This right-click option is only available if the entity resume URL is configured in the `graph.properties` file. For example, if your organization installed the Analyst toolkit, your Identity Insight system administrator can configure the link so that the Cognos based entity resume displays in a new web-browser window. If you do not see this link, contact your Identity Insight system administrator.	X	X	X

Common elements in the graphing tool graphs

Graphs have many common elements: icons, indicators, and line thickness. These common elements provide additional meaning that can help you get a more complete story of each graph and more easily identify areas of interest.

Entity icons

Each entity node is displayed as an icon surrounded by a solid circle.

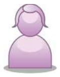
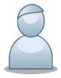
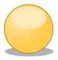
Entities can be defined as people, places, or things (such as organizations, ships, or planes). Typically, entities are people. The most common entity node is a represented as a person icon: male, female, or unknown. The gender displayed by the icon is based on one of two possible gender assignments:

- The gender assigned during the name analysis of entity resolution
- the value from the GENDER attribute that is part of the data on the incoming identity record

If the gender is undetermined, then a generic person entity icon displays.

The following table displays the default person entity icons used in the graphing tool graphs.

Table 25. A sample of the default entity icons used in the graphing tool graphs

This icon...	Represents this type of entity...
 Graphing tool female person entity indicator	Female (person) entity
 Graphing tool male person entity indicator	Male (person) entity
 Graphing tool unknown gender person entity indicator	Unknown gender entity

The main entity on an Entity or Social Network graph always has a thicker circle. Regardless of where the main entity displays on the graph, you can always identify it by the thicker circle.

On an Alert graph, all the entities along the alert path have a thicker circle. Regardless how many entities display on the graph, such as if you choose to show remaining related entities, you can always identify the entities involved in the alert.

Attribute icons

Attribute nodes are depicted on the graphing tool graphs as icons. Each icon represents a specific attribute type. The following table provides a sample of the default attribute icons displayed on the graphing tool graphs.

Table 26. A sample of the default icons displayed on the graphing tool

This icon...	Represents this type of attribute...
🏠 Graphing tool address attribute icon	Address
🏷 Graphing tool name attribute icon	Name
● Graphing tool Social security number attribute icon	Social Security number
● Graphing tool date of birth attribute icon	Date of birth
● Graphing tool other attribute type icon	Other attribute (not assigned to an existing attribute icon)

You can customize the icons that represent attributes on the graphs, either by replacing the default attribute icon or by adding icons to represent attributes that are specific to your organization. See "Adding custom icons to the graphing tool graphs" on page 319 for more information.

Alert indicators

Each entity displays an indicator to show the number of alerts for the entity. The alert indicator is displayed in the upper left corner of the solid circle that surrounds the entity icon.

The alert indicator has a gold background, and the number of alerts displays in black text. For example, this alert indicator 11 on an entity icon shows that this entity has 25 alerts.

Graphing tool alert indicator

Related entities indicators

Entity nodes also have an indicator that shows the number of relationships that belong to this entity, based on shared attributes. These relationships are not yet shown as belonging to this entity.

The related entities indicator has a light blue background, and the number of relationships displays in bold black text. For example, this related entities indicator

11

Graphing tool related entities indicator

shows that there are six additional entities that are not yet shown as having a relationship with the entity.

The related entities indicator behaves differently depending upon the type of graph:

- On the Alert graph: Both entities involved in the alert display a related entities indicator, if the entity is related to more entities that are not currently displayed on the graph. You can expand the graph to show all the related entities to each entity displayed on the graph. In this case, you no longer see a related entities indicator on any entity.
- On the Entity graph:
  - The main entity does not have a related entities indicator. The graph automatically displays all the entities related to that main entity.
  - The other entities on the Entity graph display a related entities indicator, if they are related to other entities that are not already displayed on the graph. You can use the right-click menu to show the remaining entities for such an entity, so that the related entities indicator no longer displays.
  - Like on the Alert graph, you can expand the graph to show all the related entities to each entity displayed on the graph. In this case, no entity shows a related entities indicator.
- On the Social Network graph:
  - The hub entity (at the middle of the cluster) does not have a related entities indicator, because the graph automatically displays all related entities within the cluster formation.
  - Entities that are not the hub entity of a relationship cluster might contain a related entities indicator, if they are related to other entities that are not already linked to the given node.
  - If you expand the graph to include multiple relationship clusters, it is possible for an entity to display on the graph more than once. When the entity is the hub of a cluster, no related entities indicator displays. But when that same entity is part of relationship cluster and not the hub entity, if there are additional related entities for that entity that are not yet present on the graph, the related entities indicator displays. For this reason, you always see some related entity indicators on the graph.

Line indicators

The lines that surround entity nodes and connect entities and attributes can provide additional information:

- Dashed lines connecting attributes indicate a close attribute match.
- A thick line surrounding an entity node indicates the main entity – the entity that was selected or requested when creating this particular graph.

Graphing tool URL syntax and parameters

To access a graphing tool graph, you must link to the appropriate URL. The URL can be embedded within an existing custom application (such as a web-start page, dashboard, or case management tool) or manually entered in a web-browser.

The correct URL syntax and parameters for graphing component graphs looks like the following:

```
http://server:port/graphs/run/graphtype.jsp?height=nnnn&width=yyyy&identifier=xxxx
```

hostserver:port

Indicates the name of the product application server and the port number where IBM InfoSphere Identity Insight is located. Typically, the product application server is the WebSphere server.

The port number defaults to 13510.

/graphs/run

Points to the product directories where the graphing tool files are located. The `/graphs/run` directories are the location where the product installation program installs the graphing tool by default.

***graphtype*.jsp**

Indicates which graph to create:

- For the Alert graph, enter `role-alert.jsp`
- For the Entity graph, enter `entity.jsp`
- For the Social Network graph, enter `social-network.jsp`

? Indicates a URL element.

height=*nnnn*

Indicates the height of the graph canvas - what height to render the graph canvas inside the web-browser window. Enter the number in pixels.

The graph height is determined in the following manner:

- If a height is specified in the URL, that is the default graph height.
- If a value is set in the **`defaultGraphHeight`** property in the `graph.properties` file, that is the default graph height.
- In the absence of any specified graph height in the URL or **`defaultGraphHeight`** property, the default graph height is set to 800 pixels.

.

To approximate a graph canvas that fits within a 1024 x 768 standard web-browser window, set the height to 450 pixels.

? Indicates a URL separator token between parameters.

width=*yyyy*

Indicates the width of the graph canvas - what width to render the graph canvas inside the web-browser window. The Attribute Explorer is not included in this number, since it is a separate component that is docked to the right of the graph canvas in the web-browser window.

The graph width is determined in the following manner:

- If a width is specified in the URL, that is the default graph width.
- If a value is set in the **`defaultGraphWidth`** property in the `graph.properties` file, that is the default graph width.
- In the absence of any specified graph width in the URL or **`defaultGraphHeight`** property, the default graph width is set to 800 pixels.

.

To approximate a graph canvas that fits within a 1024 x 768 standard web-browser window, set the width to 640 pixels.

identifier=xxxx

Indicates the type of ID (entity or alert) and the specific number for that entity or alert. When you use the entity ID, the value of the ID is the main entity on the Entity graph or the hub entity on the Social Network graph. When you use the alert ID, the value is the alert to display on the Alert graph.

- For the Alert graph, enter `alertID=`*`specific_alert_ID_number`*
- For the Entity or Social Network graphs, enter `entityID=`*`specific_entity_ID_number`*

Common administrative tasks for the graphing tool

Some of the tasks for the graphing tool can only be completed by an administrator user.

Adding custom icons to the graphing tool graphs:

The graphing tool includes standard icons that represent the different types of attributes displayed on the graphs. You can change the default icon for one or more attributes or add icons for custom attributes configured in your product. All graphing tool graphs use the same set of image icons on the product application server, so when you customize the attribute icon set, each user views the same attribute icons.

Before you begin

Graph icons start as scalable vector graphic (SVG) files. SVG files can be created using a variety of vector-based drawing tools, or they can be downloaded from various internet sources. It is strongly recommended that the SVG files used for icons be kept to a reasonably small size (to improve readability when the image is scaled).

The graph requires a shape definition stored in Javascript Object Notation (JSON) format. To convert from SVG to JSON requires the use of two separate command utilities: **xlstproc** and **sed**.

If you are on a Unix-based computer, you may have these tools already. If you are on a Windows based computer, you'll have to acquire these Unix-based tools through the use of a Unix emulator (like the free Cygwin application). *Note: if you use Cygwin, be sure to include the libxml2 and libxslt libraries in your installation to obtain the required utilities.*

Finally, you'll need the file svg2gfx.xsl from the free DOJO library (available at https://dojotoolkit.org/download). Once DOJO has been downloaded, the svg2gfx.xsl file can be located in the <install root>/dojox/gfx/resources directory.

Procedure

1. Copy the svg2gfx.xsl file from the DOJO location to the same directory containing the SVG file(s) you want to convert
2. Open a Unix terminal/command-line window and navigate to the directory containing the SVG file(s)
3. Run the following command: `xsltproc ./svg2gfx.xsl <your .SVG file> > <temp_file_name>.json`
4. Run the following command: `sed -e 's/,}/}/g' -e 's/,]/]/g' <temp_file_name.json> > <final name>.json`

5. Locate your Identity Insight installation folder
6. Under the installation folder, navigate to /ibm-home/graphs
7. Create a folder named (case-sensitive): customImages
8. Move the custom icon (.json file) to the customImages folder

Example

If you created an attribute type named `FLIGHT` and you wanted a custom graph icon to represent that attribute type on the graphing tool graphs, take the following steps:

1. Create or obtain a suitable image file to represent the `FLIGHT` attribute type. Make sure that the image file name matches the attribute type name configured in the Configuration Console and uses all lowercase letters, such as this file name: `flight.svg`
2. Make sure svg2gfx.xsl is located in the same directory as `flight.svg`
3. Open a Unix terminal/command-line window, and navigate to the same directory as `flight.svg`
4. Run the following command: `xsltproc ./svg2gfx.xsl flight.svg > flight_tmp.json`
5. Run the following command: `sed -e 's/,}/}/g' -e 's/,]/]/g' flight_tmp.json > flight.json`
6. Copy the `flight.json` icon file into the `/customImages` folder.

Requirements for custom graph icons:

You can customize the attribute icons that display on the graphs. But the new icons must meet the requirements for custom graph icons, so the graphs recognize the icons and display them.

Requirements for the custom icons

For product graphs to recognize and display custom icons, the attribute icons must meet the following requirements:

- File format: Scalable Vector Graphics (SVG)
- Name:
  - The custom icon name must match the name of the corresponding attribute type that is configured in the Configuration Console.
  - The custom icon name must contain lowercase letters only.
- SVG file must be converted to a JSON shape definition (see "Adding custom icons to the graphing tool graphs" on page 319)

For example, if you want to associate an attribute icon with the attribute type FINGERPRINTS that is configured in the Configuration Console, the name of the icon file must be named `fingerprints.svg`.

Name examples

For overriding an existing base type icon, the custom icon must be named one of the following (all lower case):

- address.json
- female.json
- male.json

- name.json
- undetermined_gender.json

For entity numbers, the .json icon file must be named the same as the number type code (NUM_TYPE.NUM_TYPE in the database). For example:

- cc.json
- dl.json
- ff.json
- ssn.json
- pp.json
- ph.json

For entity characteristics, the .json icon file must be named the same as the attribute type code (ATTR_TYPE.ATTR_TYPE in the database). For example:

- dob.json
- died.json
- marital.json
- circa_dob.json
- pop.json
- nat.json
- cit.json

Linking to the entity resume from the graphing tool:

The entity resume provides detailed information about individual entities and is useful when you are analyzing alerts and entity relationships. If you set the URL properties to the web application that generates the entity resume, graphing tool users can open the entity resume from inside one of the graphing tool graphs.

About this task

Setting the link is a global task. After the link is set, all users who view graphing tool graphs have access to the link from the right-click menu. If the link properties are not set, the **Show resume** right-click option does not display.

Procedure

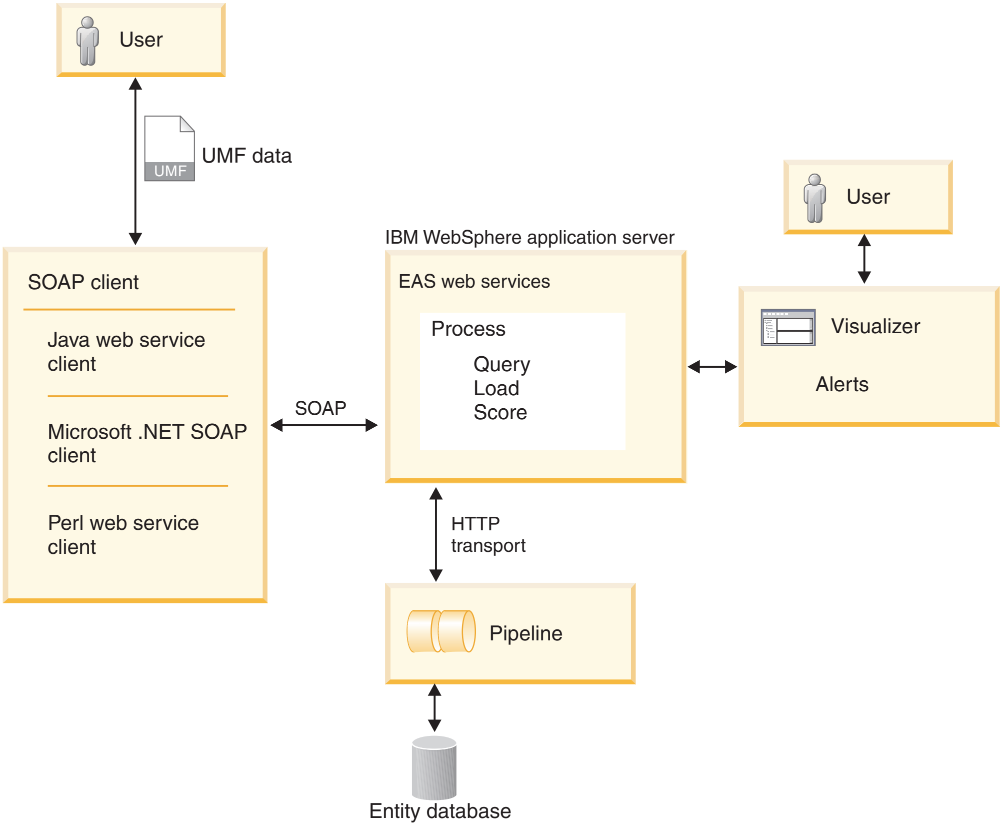
1. Have the system administrator update the COMPONENT_CONFIG table's RESUMESERVER property with a link to the Cognos toolkit report.
   a. Replace the value of this property with the actual URL. Specify the host server, port name, and path to the web application. Refer to the sample value to get an idea of what the path might look like. For example, if your organization installed and is using the Cognos based Analyst toolkit, specify the path to the entity resume generated by the Analyst toolkit.
   b. Make sure that the token **`%ISIIEntityID%`** is within the parameter value. This parameter sends the appropriate entity ID to the web application to generate the correct entity resume.
2. Optional: Test the link.

Chapter 9. Developing

If you have the need to use Web services in your environment, IBM InfoSphere Identity Insight provides a simple XML-based Web service.

Web services

IBM InfoSphere Identity Insight provides a set of Web services that you can use to build external applications that can load Universal Message Format (UMF) data for pipeline processing or search for entities in the entity database. You use the bi-directional HTTP (hypertext transfer protocol) transport method, which is a standard feature in the pipeline.



IBM InfoSphere Identity Insight Web services use four SOAP (Simple Object Access Protocol) methods: process, search, load, and score The product supports SOAP version 1.1.

The product includes several components to help you get started using Web services.

srd.wsdl
: This file contains a Web services description language (WSDL) definition of the product Web services. You can use this file with any SOAP toolkit or

technology to start the Web services. It can be found by starting WebSphere Liberty and loading the file from http://hostname:port/easws/resources/wsdl/srd.wsdl

wsutil.jar
This file is a Web services test client provided for testing your Web services installation and configuration. This utility can be found in the `ibm-home/easws` directory.

Web services software requirements

IBM InfoSphere Identity Insight Web services requires that certain software be installed and operational.

Before using Web services, make sure that the following software is installed and operational:

- IBM InfoSphere Identity Insight Web services must be installed and running.
- The embedded IBM WebSphere Application Server must be running where the IBM InfoSphere Identity Insight Web services are deployed. In most cases, this is the same application server that the Configuration Console and the Visualizer are installed on.
- A Web services pipeline must be started and listening on the appropriate HTTP URL. The application server attempts to send UMF data to this Web services pipeline over the specified HTTP URL, whenever a SOAP request is received.

  Note: The HTTP URL used for communication between the application server and the pipeline is *not* the same as the URL used by Web service clients attempting to send SOAP requests. Sending SOAP requests directly to the Web services pipeline HTTP URL will result in an error.

  For example, if WebSphere Application Server is set up with the default port range, the Port numbers and usage would be:
  - *nnn*0 - HTTP port for Webserver
  - *nnn*1 - HTTPS port for Webserver
  - *nnn*2 - HTTP Administration port
  - *nnn*3 - HTTPS Administration port
  - *nnn*4 - SOAP port
  - *nnn*5 - Application server port.
- The webservices.properties file must be configured with the HTTP URL of the running pipeline, so that the embedded WebSphere Application Server knows where to find the pipeline that will handle the Web service requests. This file is usually located in the following directory: `*product_home*/srd-home/easws`
- A SOAP and WSDL-compatible Web services client, used to invoke IBM InfoSphere Identity Insight Web services, must exist. A sample client, wsutil.jar, is installed with IBM InfoSphere Identity Insight Web services for testing previous release services, but does not apply to enhanced services for version 8.0, fix pack 2.

Starting Web services pipelines

To send and process data sent through a Web service, start the pipelines using the bi-directional HTTP transport. Typically, pipelines used with Web services remain constantly running in the background, listening to the assigned ports for data to process. Use these steps to start a Web services pipeline.

Before you begin

- Make sure that you know the pipeline URL setting that is configured in the webservices.properties file. This setting points to the Web services component running on the embedded IBM Websphere Application Server at the pipeline, and it must match the URL that is used to start Web services pipelines.
- The pipeline node hosting this pipeline must have the pipeline executable installed.
- There must be at least one pipeline configuration file configured for use with the pipeline that you want to start. You can specify the pipeline configuration file to use as part of the start pipeline command. If you do not specify the name of configuration file as part of the pipeline command, the pipeline configuration file must be located on the pipeline node, and it must use the default pipeline configuration file name of pipeline.ini.
- If you use a script to start pipelines, make sure the script is located in the same directory where you start the pipeline.
- If you want to route the results of processing from this pipeline or monitor the statistics and status of this pipeline, register the pipeline in the Configuration Console on the **Pipelines** tab. You must use one of the already registered pipeline names to start this pipeline for monitoring or routing to complete successfully.
- If you are using the application monitor to monitor pipeline status and statistics, make sure the pipeline node has an SNMP Agent installed and running before you start this pipeline.
- If this pipeline routes its results to another system or another database, make sure the routing file for this pipeline is located in the same directory where you start the pipeline.
- If the *DEFAULT_CONCURRENCY* system parameter value is set to greater than 1 or if you configured the *concurrency* parameter in the pipeline configuration file for the pipeline node, you can start multiple parallel pipeline processing threads using a single start pipeline command.

About this task

There are three steps to starting a pipeline:

Procedure

1. Verify that there are no other pipelines currently running on the pipeline node which have the same name as the pipeline you want to start. Each pipeline must have a name unique to its pipeline node. (The default pipeline name is `pipeline`.) There are two ways to verify this:
   a. If you are using the application monitor to check the status of pipelines or route the results to other systems, look at the **Pipeline Status** tab to see if there is another pipeline running that has the same name you want to use.
   b. Or at a command prompt, type the following command:

   ```
   pipeline -n pipelinename -l
   ```

   where *pipelinename* is the name you want to use to start the new pipeline. Make sure that this name matches the name registered in the Configuration Console for this pipeline.
2. At a command prompt, start one or more pipelines by specifying the type the appropriate pipeline command options and parameters using this format:

   ```
   pipeline -option parameter
   ```

Note: If you are using the application monitor for this pipeline and it has been registered in the Configuration Console for monitoring or routing, be sure to use the `-n` option as part of the pipeline start command and specify the registered pipeline name. If the pipeline name specified does not exactly match the registered pipeline name (including case), the pipeline status will not display correctly on the Configuration Console **Pipeline Status** tab and any routing configured for this pipeline will not be successful.

Note: Typically, you use either the `-s` or the `-d` pipeline option to start the pipeline in either service/daemon or debug mode, as appropriate.

3. Verify that the command worked, and the pipeline is started and active.
   a. If you are using the application monitor and this pipeline has been registered in the Configuration Console, check the **Pipeline Status** tab. If the pipeline is active, the status displays as `Active`.
   b. If your system is running on a Microsoft Windows platform and you are using the services pipeline option, you can see the status of the pipeline in the Microsoft Windows Services control panel.
   c. If your system is running on a UNIX platform and you are using the daemons pipeline option, you can type the following command to check for running processes:

      ```
      ps -fu userid
      ```

      where *userid* is the identification of the user starting the pipeline.
   d. Or at a command prompt, type the following command:

   ```
   pipeline -npipelinename -l
   ```

   where *pipelinename* is the name of the pipeline you just started. If the pipeline is active, the command prompt returns `Running`.

What to do next

This pipeline command starts the number of pipeline processing threads equal to the concurrency parameter in the pipeline configuration file. The number of records processed simultaneously is determined by the concurrency parameter included in the HTTP transport option.

Testing Web services

By using the provided test client, wsutil.jar, you can test your IBM InfoSphere Identity Insight Web services installation and configuration.

Before you begin

- Web services must be installed.
- Make sure the embedded WebSphere Application Server is running.
- The application server should have at least one pipeline configuration file configured for Web services pipelines.
- Make sure that the webservices.properties file is configured with the correct pipeline URL setting. This Web services pipeline must be running.
- Create at least one test UMF input document to use during testing.

Procedure

1. On the embedded WebSphere Application Server, go to the directory that contains the wsutil.jar. This file is usually located at `*installation_root*/ewas/webservice/wsutil.jar`

2. At a command line from this directory, enter the wsutil.jar command syntax for the operation you want to perform: `java -jar wsutil.jar —-<SOAP method>=<URI> --input=<URL> —-output=<URI>`

Example of testing the Web services load method

The following wsutil.jar command loads records from a UMF file named "raw_entities.umf "and saves the results to a UMF file named "results.umf":

```
java -jar wsutil.jar --load=http://localhost:13510/easws/services/SRDWebService
 --input=raw_entities.umf --output=results.umf
```

srd.wsdl file

To communicate with IBM InfoSphere Identity Insight Web services, you need a Web services client. When you install IBM InfoSphere Identity Insight Web services, the srd.wsdl file is also installed and contains the SRDWebService methods that are used to communicate with InfoSphere Identity Insight Web services. You can use the srd.wsdl file to build a Web services client for use with IBM InfoSphere Identity Insight Web services.

The srd.wsdl file is accessible through the Web browser by accessing the embedded WebSphere Application Server that is hosting Web services. Typically, this file can be located on the application server at the following root URL:

`http://`*`IBM_WebSphere_Application_Server_host`*`:`*`install_port`*`/easws/resources/wsdl/srd.wsdl`

For example:

```
http://localhost:13510/easws/resources/wsdl/srd.wsdl
```

Note: Make sure that the application server is running before trying to access the srd.wsdl file.

You can also build a Web services wsdl client using any development platform that supports Web services with a SOAP toolkit, such as:

- Java with IBM WebSphere Application Server
- Java with Apache Axis
- Microsoft .NET
- Perl

Refer to your development platform documentation for instructions on how to create a Web services client using a wsdl file.

If you build a Web services client wsdl other than the srd.wsdl Web services client, make sure that the deployment URL correctly points to the wsdl client.

SRDWebService methods

The srd.wsdl file contains the SRDWebService methods that are used to communicate with IBM InfoSphere Identity Insight Web services. SRDWebService includes three methods: one for loading data into the entity database, one for performing a search to query the entity database, and one for processing any pipeline functionality available through UMF.

loadRecord method

```
LoadResult loadRecord(String umfEntity)
```

The LoadResult object returned from the loadRecord() method contains two members:

Member	Description	Type
entityID	ID of the returned entity	Long
merged	Flag indicating if the entity was resolved into an existing entity or was a new entity	Boolean

The umfEntity parameter is an XML string in UMF representing the data for a single entity. Use the UMF specification for instructions on how to properly construct a UMF_ENTITY record, making sure to define the appropriate values for DSRC_ACCT and DSRC_REF.

While the load method enables you to process UMF_ENTITY documents, it does not return the raw UMF output document as a result. Instead, it returns a LoadResult object containing the entity ID, and a flag indicating if this was a new entity or if it was resolved with an existing one. You can use the process method in place of the load method, if you do not mind parsing the UMF output document. The load method simply saves you the work of parsing the resultant UMF output document from the load operation.

basicQuery() method

```
String basicQuery(String umfSearch)
```

The input string to the basicQuery() method must be in the form of a UMF_SEARCH record. The XML string returned from basicQuery() contains the UMF_SEARCH_RESULT from the query.

There are two types of built-in queries: Summary result-set queries and Detail drill-down queries.

Note: This method exists only for backwards compatibility. In this release, the method functions identical to the process method. Use the process method in place of the basicQuery() method for all new client applications.

process() method

```
String process(String umfRequestDocument)
```

Use the process method to process any UMF input document and receive a UMF output document as a result. The process method is intended to handle all requests and responses supported by the pipeline and should be the method of choice for all operations.

This method takes a String parameter and returns a String result.

wsutil.jar

Wsutil.jar is a command-line based Java application that is installed when you install IBM InfoSphere Identity Insight Web services. It is a sample client that you can use to try each of the Web services SOAP methods, to test your Web services installation and configuration.

The wsutil.jar test client should be in the following location:

installation_root/ewas/webservice

wsutil.jar usage syntax

Wsutil.jar is a command-line based Java application provided as a test client to test your IBM InfoSphere Identity Insight Web services installation and configuration. To use wsutil.jar, you specify a wsutil.jar operator with the corresponding input and output modifiers.

The syntax for wsutil.jar usage is based on which Web services operation you want to test:

```
wsutil (unix) or wsutil.bat (win) --operator=URI --input=URI --output=URI
```

help

Displays online help and command line information for the wsutil.jar test client.

```
wsutil (unix) or wsutil.bat (win) --help
```

load=*URI*

Specifies pipeline style UMF records and the Uniform Resource Identifier (URI) for the IBM InfoSphere Identity Insight Web services interface.

```
wsutil (unix) or wsutil.bat (win) --load=URI [--xslt=URI][--
input=URI][--output=URI]
```

This operation loads the UMF records from the specified URI into the Web services pipelines for entity resolution processing. After processing, the operation returns the entity ID and an indicator of whether the inbound entity was merged with an existing entity or caused a new entity to be created.

process=*URI*

Specifies generic XML or UMF records and the Uniform Resource Identifier (URI) for the IBM InfoSphere Identity Insight Web services interface.

```
wsutil (unix) or wsutil.bat (win) --process=URI [--xslt=URI][--
input=URI][--output=URI]
```

Use this operation to process any UMF input document and receive a UMF output document as a result. The process method is intended to handle all requests and responses supported by the pipeline. It is usually the method of choice for all operations.

search=*URI*

Specifies pipeline search style UMF requests and responses with the Uniform Resource Identifier (URI) for the IBM InfoSphere Identity Insight Web services interface.

```
wsutil (unix) or wsutil.bat (win) --score=URI [--xslt=URI][--
input=URI][--output=URI]
```

This operation can either perform a search against the entity database for a specific entity and returns requested information about that entity, or it can query the entity database for entities that match a given attribute and return the list of entities that matched the query.

xslt=*URI*

Specifies the XSLT transform and the XML file the operation will transform into UMF records.

```
wsutil (unix) or wsutil.bat (win) --xslt=URI][--input=URI][--output=URI]
```

Use this operation to transform XML records into UMF before using one of the Web services operations.

wsutil.jar modifiers

Use these modifiers with wsutil.jar operators to specify the input and output methods for the Web services command.

`input=`*`URI`*
: Specifies the input method for UMF records. The default input method is stdin.

`output=`*`URI`*
: Specifies the output method for UMF records. The default output method is stdout. You can use this method to specify a location and file name to save the UMF output to a file.

Example wsutil.jar usage

The following wsutil.jar command on a UNIX system loads records from a file, transforms those records into UMF, and displays the results on the command line interface console:

```
wsutil --load=http://localhost:13510/easws/services/SRDWebService
--input=raw_entities.xml --xslt=transform.xsl
```

The following wsutil.jar command on a Windows system acquires requests from stdin and displays the results on the command line interface console:

```
wsutil.bat --process=http://localhost:13510/SRDWebService
```

Building queries against the entity database

IBM InfoSphere Identity Insight provides several ways to query the entity database. You can build Web services pipeline searches to search the entity database to find entities that match specific attribute search criteria. You can also build Web services pipeline searches to query the database about a specific entity.

Web service pipeline searches

Built into the pipelines is a dynamic search and query interface, that provides a single-point-of-access for Web services to query the entity database. You use UMF input documents to structure the request, and then send the UMF input document through Web services to the pipelines for processing. Once processed, the pipeline returns a UMF output document that contains the results.

Web services pipeline searches provide answers to two types of questions:

Which entities in the entity database match to a particular attribute or set of attributes? (UMF_SEARCH)

This type of Web services pipeline search takes full advantage of entity resolution to recognize and standardize the incoming search criteria, and then to match the search criteria to entities in the database. It is called a summary or result set query, and it returns a list of entities with data values that match to the requested attribute value or list of attribute values.

To perform a summary or result set query, you create a UMF_SEARCH input document that contains the search criteria that the pipeline uses to perform entity resolution. The pipeline responds by returning a UMF_SEARCH_RESULT output document with the query results, which are the list of entities that matched the search criteria.

What does the entity database know about a specific entity? (UMF_QUERY)

This type of Web services pipeline search uses SQL statements and parameters to query the entity database. It is called a detail or drill-down query, and it returns a detailed list of the information about a single entity.

To perform a detail or drill-down query, you create a UMF_QUERY input document that indicates which entity in the entity database you want information about. The pipeline responds by returning a UMF_QUERY_RESULT output document with the detail about the requested entity.

While performing Web service pipeline searches, the pipelines perform all standard pipeline functions, including logging.

Both the input (request) and the output (response) for Web services pipeline searches use UMF documents and structure the information in UMF.

Web service pipeline search formats

The product comes with several built-in formats for each of the Web services pipeline searches:

UMF_SEARCH formats

WS_SUMMARY_TOP10
Returns a list of the top 10 entities in the database that most closely matched the attribute data specified in the search criteria

WS_SUMMARY_TOP100
Returns a list of the top 100 entities in the database that most closely matched the attribute data specified in the search criteria

WS_SUMMARY
Returns a list of all entities in the database that matched the attribute data specified in the search criteria

UMF_QUERY formats

WS_DETAIL
Returns all data from entity database for the requested entity ID

WS_RELATION
Returns a list of all entities in the entity database that are related to the input entity at 1-degree

WS_ALERT
Returns a list of all alerts in the entity database that involve the input entity ID

You indicate which built-in format to use in the FORMAT_CODE tag in the appropriate UMF input document.

Performance considerations

Web services pipeline search requests that contain more search criteria typically mean the system compares against fewer entities in the database. That, in turn, means that the system returns results quicker than requests with fewer search criteria.

Building Web services queries to find a specific entity

Use these instructions to build a UMF_QUERY input document to find a specific entity in the entity database. You send the UMF_QUERY input document through

Web services into a Web services pipeline for processing. After the pipeline processes the query, Web services returns a UMF_QUERY_RESULT output document that contains the details about the requested input entity.

Before you begin

The embedded WebSphere Application Server must be running, and at least one Web services pipeline must be started and running to receive and process the UMF_QUERY input document.

About this task

Because the search request is a UMF input document, the criteria must be formatted using valid UMF tags. You can use any text editor or utility that creates UMF.

Procedure

1. Create a new UMF_QUERY input document.
2. In the ROOT segment, enter the required UMF tags and values:
   a. Enter the data source code in the DSRC_CODE tag. The default data source code for Web services pipeline searches is 1589. If you use a different data source code than the default Web services pipeline search data source code, make sure that it is configured not to resolve entities.
   b. Enter the data source reference code that references the requesting message transaction in the DSRC_REF tag. The data source reference code should be meaningful, because it is returned to the calling application.
   c. Enter the format code to indicate the output format of the results using the FORMAT_CODE tag. The pipelines come with three built-in format codes for a Web services pipeline search using UMF_QUERY:
      - WS_DETAIL, which returns all available entity data for the input entity ID
      - WS_RELATION, which returns a list of all entities related to the input entity ID at a 1-degree relationship
      - WS_ALERT query, which returns all role alerts in the system involving the input entity ID

      If you use a different format code, the format code must be configured in the UMF_OUTPUT_FORMAT table.
   d. In the ENTITY_ID tag, enter the entity ID for the entity you want to return information about.
3. Enter any other query criteria using the other optional UMF segments of `<NAME>`, `<ADDRESS>`, `<EMAIL>`, `<ATTRIBUTE>`, and `<NUMBER>`.
4. Send the UMF_QUERY input document to a Web services pipeline.

Results

A Web services pipeline ingests the UMF_QUERY document, using the criteria specified to find entities in the database that match the query. The pipeline then processes the query, creates normal logging files, and returns the results in a UMF_QUERY_RESULT output document through Web services to the calling application.

Example UMF_QUERY search

This example UMF_QUERY searches for all information about entity ID 1223:

Note: This example is formatted for readability and does not follow the required one line per UMF record format.

```
<UMF_QUERY>
 <DSRC_CODE>1589</DSRC_CODE>
 <DSRC_REF>546</DSRC_REF>
 <FORMAT_CODE>WS_DETAIL</FORMAT_CODE>
 <ENTITY_ID>1223</ENTITY_ID>
</UMF_QUERY>
```

UMF_QUERY input document

The UMF_QUERY input document contains the collection of UMF segments that structure the incoming data to query the entity database, and then find and return information about a specific entity to the calling application. It contains the request and search criteria for a Web services pipeline query.

The information in a UMF_QUERY input document is based on SQL statements. The results of this Web services pipeline search are returned to the calling application in a UMF_QUERY_RESULT output document. UMF_QUERY undertakes an "Enhanced Query / Find by Attribute" query.

These required UMF elements and segments comprise the UMF_QUERY input document:

DSRC_CODE

Data source code UMF tag that is required, because it references and identifies the calling application. As part of normal pipeline logging, this data source code is logged in the UMF_LOG table for each processed UMF_QUERY.

The system is already configured with a data source code, 1589, that can be used for all Web services pipeline searches. This data source code performs entity resolution processing without resolving the incoming search criteria with the entity in the entity database that matches the search. You can create your own data source code for a particular calling application, just be sure that the data source code is set to not resolve entities.

DSRC_REF

Data source reference UMF tag that is required, because it references the requesting message transaction and is returned to the calling application.

FORMAT_CODE

UMF tag that correlates to a UMF output document format that is specified in the UMF_OUTPUT_FORMAT table. IBM InfoSphere Identity Insight comes with three built-in format codes for a Web services pipeline search using UMF_QUERY:

- WS_DETAIL, which returns all available entity data for the requested entity ID
- WS_RELATION, which returns a list of all entities related to the input entity at 1-degree
- WS_ALERT query, which returns all alerts in the system involving the input entity ID

For undertaking EQ (Enhanced Query / Find By Attribute) via this input-document, the following FORMAT_CODE must be specified.

ENHANCED_QUERY_RESULT example:

```
<UMF_QUERY>
<FORMAT_CODE>ENHANCED_QUERY_RESULT</FORMAT_CODE>
 <ATTRIBUTE>
  <ATTR_TYPE>CIT</ATTR_TYPE>
  <ATTR_VALUE>CANADA</ATTR_VALUE>
 </ATTRIBUTE>
</UMF_QUERY>
```

ENTITY_ID

This required UMF tag specifies the entity ID for the entity in the search. The system returns a response with details of the known data about this entity from the entity database, based on the other query criteria.

You then specify the optional search criteria using the other available UMF segments and their valid tags for names, addresses, numbers, characteristics, and e-mail addresses.

NAME

Query for name attributes that define the name of the person, organization, place or item, as defined by the entity model and the incoming identity.

NUMBER

Query for number attributes that are comprised of data that is usually described as a number, such as credit card numbers, phone numbers, and passport numbers.

ADDRESS

Query for address attributes that define a location of the identity and typically contain standard address information: street name and number, unit or building number, city, state, country, and postal code.

ATTRIBUTE

Query for characteristic attributes that define other identity traits or information that is not expressed through the other kinds of attributes.

EMAIL

Query for e-mail attributes that define Internet e-mail addresses.

Example UMF_QUERY search

This example UMF_QUERY uses the WS_DETAIL format code to query the entity database and return all known information about Entity ID 1223:

Note: This example is formatted for readability and does not follow the required one line per UMF record format.

```
<UMF_QUERY>
 <DSRC_CODE>1589</DSRC_CODE>
 <DSRC_REF>546</DSRC_REF>
 <FORMAT_CODE>WS_DETAIL</FORMAT_CODE>
  <ENTITY_ID>1223</ENTITY_ID>
</UMF_QUERY>
```

WS_DETAIL format code:

When you are building a Web services pipeline search to return the details about a specific entity in the entity database, use the built-in format code WS_DETAIL. This format code is specified in the UMF_QUERY input document that contains the criteria for the query.

Example of a Web services pipeline search using the WS_DETAIL format code

This example Web services pipeline search returns all information in the entity database for Joe Franklin, entity ID 87.

Note: This example is formatted for readability and does not follow the required one line per UMF record format.

To request the detail for entity ID 87 (Joe Franklin), create a new UMF_QUERY input document with the request:

```
<UMF_QUERY>
 <FORMAT_CODE>WS_DETAIL</FORMAT_CODE>
 <DSRC_CODE>1589</DSRC_CODE>
 <DSRC_REF>ABC-003</DSRC_REF>
 <ENTITY_ID>87</ENTITY_ID>
</UMF_QUERY>
```

After sending this UMF_QUERY document through Web services for processing by a Web services pipeline, the calling application receives a response in the following UMF_QUERY_RESULT document:

```
<UMF_QUERY_RESULT>
 <DSRC_CODE>1589</DSRC_CODE>
 <ENTITY>
  <ENTITY_ID>87</ENTITY_ID>
  <SOURCE>
  <ACCT>OFAC</ACCT>
  <NAME>
   <NAME_TYPE>MAIN</NAME_TYPE>
   <FIRST_NAME>JOSEPH</FIRST_NAME>
   <LAST_NAME>FRANKLIN</LAST_NAME>
  </NAME>
  <ADDRESS>
   <ADDR_TYPE>H</ADDR_TYPE>
   <ADDR1>5559 W. 4TH ST</ADDR1>
   <CITY>SAN FRANCISCO</CITY>
   <STATE>CA</STATE>
   <POSTAL_CODE>94123-4567</POSTAL_CODE>
   <COUNTRY>USA</COUNTRY>
  </ADDRESS>
  <NUMBER>
   <NUM_TYPE>PHONE</NUM_TYPE>
   <NUM_VALUE>415-555-3325</NUM_VALUE>
  </NUMBER>
  </SOURCE>
  <SOURCE>
  <ACCT>FBI</ACCT>
   <NAME>
    <NAME_TYPE>MAIN</NAME_TYPE>
    <FIRST_NAME>JOEY</FIRST_NAME>
    <LAST_NAME>FRANKLIN</LAST_NAME>
   </NAME>
  <ADDRESS>
   <ADDR_TYPE>H</ADDR_TYPE>
   <ADDR1>392 S.E. MULLENS AVE</ADDR1>
   <CITY>OAKLAND</CITY>
   <STATE>CA</STATE>
   <POSTAL_CODE>94126-1566</POSTAL_CODE>
   <COUNTRY>USA</COUNTRY>
  </ADDRESS>
  <NUMBER>
   <NUM_TYPE>PHONE</NUM_TYPE>
   <NUM_VALUE>415-555-3325</NUM_VALUE>
  </NUMBER>
  <NUMBER>
```

```
   <NUM_TYPE>CC</NUM_TYPE>
   <NUM_VALUE>1111-22-3333</NUM_VALUE>
  </NUMBER>
  </SOURCE>
  <SOURCE>
  <ACCT>A9</ACCT>
  <NAME>
   <NAME_TYPE>MAIN</NAME_TYPE>
   <FIRST_NAME>JOE</FIRST_NAME>
   <LAST_NAME>FRANKLIN</LAST_NAME>
  </NAME>
  <ADDRESS>
   <ADDR_TYPE>B</ADDR_TYPE>
   <ADDR1>392 S.E. MULLENS AVE</ADDR1>
   <CITY>OAKLAND</CITY>
   <STATE>CA</STATE>
   <POSTAL_CODE>94126-1566</POSTAL_CODE>
   <COUNTRY>USA</COUNTRY>
  </ADDRESS>
  <NUMBER>
   <NUM_TYPE>PHONE</NUM_TYPE>
   <NUM_VALUE>415-555-3325</NUM_VALUE>
  </NUMBER>
  <NUMBER>
   <NUM_TYPE>CC</NUM_TYPE>
   <NUM_VALUE>1111-22-3333</NUM_VALUE>
  </NUMBER>
  </SOURCE>
 </ENTITY>
 <FROM_NODE>ABC-003</FROM_NODE>
 <PAGE_NUM>1</PAGE_NUM>
 <FORMAT_CODE>WS_DETAIL</FORMAT_CODE>
</UMF_QUERY_RESULT>
```

From this response, you can see that there are three data sources with information on Joe Franklin: the OFAC list, an FBI list, and the A9 list. Joe is using two different addresses, but in each case, he is using the same phone number and credit card.

WS_ALERT format code:

When you are building a Web services pipeline search to return all role alerts in the entity database involving a specific entity, use the built-in format code WS_ALERT. This format code is specified in the UMF_QUERY input document that contains the criteria for the query.

Example of a Web services pipeline search using the WS_ALERT format code

This example Web services pipeline search returns a list of all role alerts in which Joe Franklin, entity ID 87, is involved.

Note: This example is formatted for readability and does not follow the required one line per UMF record format.

To request the role alerts for entity ID 87 (Joe Franklin), create a new UMF_QUERY input document with the request:

```
<UMF_QUERY>
 <FORMAT_CODE>WS_ALERT</FORMAT_CODE>
 <DSRC_CODE>1589</DSRC_CODE>
 <DSRC_REF>BB123-9003</DSRC_REF>
 <ENTITY_ID>87</ENTITY_ID>
</UMF_QUERY>
```

After sending this UMF_QUERY document through Web services for processing by a Web services pipeline, the calling application receives a response in the following UMF_QUERY_RESULT document:

```
<UMF_QUERY_RESULT>
  <ALERT>
  <CONFLICT_ID>2</CONFLICT_ID>
  <CONFLICT_RULES_DESC>Bad Guy Knows Employee</CONFLICT_RULES_DESC>
  <CONF_ENTITY1>87</CONF_ENTITY1>
  <CONF_ENTITY2>376</CONF_ENTITY2>
  <DEGREE_OF_SEP>1</DEGREE_OF_SEP>
  <INBOUND_ENTITY_ID>87</INBOUND_ENTITY_ID>
  <NAME1>FRANKLIN, JOSEPH</NAME1>
  <NAME2>MILLER, SUSAN</NAME2>
  <PATH_STRENGTH>80</PATH_STRENGTH>
 </ALERT>
 <ALERT>
  <CONFLICT_ID>5</CONFLICT_ID>
  <CONFLICT_RULES_DESC>Bad Guy Knows Vendor</CONFLICT_RULES_DESC>
  <CONF_ENTITY1>87</CONF_ENTITY1>
  <CONF_ENTITY2>10651</CONF_ENTITY2>
  <DEGREE_OF_SEP>1</DEGREE_OF_SEP>
  <INBOUND_ENTITY_ID>87</INBOUND_ENTITY_ID>
  <NAME1>FRANKLIN, JOSEPH</NAME1>
  <NAME2>MARTINEZ, JULIO</NAME2>
  <PATH_STRENGTH>64</PATH_STRENGTH>
 </ALERT>
 <DSRC_CODE>1589</DSRC_CODE>
 <FROMNODE>BB123-9003</FROMNODE>
</UMF_QUERY_RESULT>
```

From this response, you can see that there are two role alerts for Joe Franklin: an alert where employee Susan Miller knows Joe, and an alert where vendor Julio Martinez knows Joe.

WS_RELATION format code:

When you are building a Web services pipeline search to return a list of all entities related to a specific entity at 1-degree, use the built-in format code WS_RELATION. This format code is specified in the UMF_QUERY input document that contains the criteria for the query.

Example of a Web services pipeline search using the WS_RELATION format code

This example Web services pipeline search returns a list of all entities related at 1-degree to Joe Franklin, entity ID 87.

Note: This example is formatted for readability and does not follow the required one line per UMF record format.

```
<UMF_QUERY>
 <FORMAT_CODE>WS_RELATION</FORMAT_CODE>
 <DSRC_CODE>1589</DSRC_CODE>
 <DSRC_REF>ABC-003</DSRC_REF>
 <ENTITY_ID>87</ENTITY_ID>
</UMF_QUERY>
```

After sending this UMF_QUERY document through Web services for processing by a Web services pipeline, the calling application receives a response in the following UMF_QUERY_RESULT document:

```
<UMF_QUERY_RESULT>
 <DSRC_CODE>1589</DSRC_CODE>
 <RELATION>
  <DETAIL>
   <ENTITY_ID>87</ENTITY_ID>
   <INBOUND_VALUE_ABST>415-555-3325</INBOUND_VALUE_ABST>
   <MATCHED_CODE>6</MATCHED_CODE>
   <MATCHED_DSRC_ACCT>6</MATCHED_DSRC_CODE>
   <MATCHED_ENTITY_ID>376</MATCHED_ENTITY_ID>
   <MATCHED_KEY_ID>16</MATCHED_KEY_ID>
   <MATCHED_TYPE>NUMBER</MATCHED_TYPE>
   <MATCHED_VALUE_ABST>415-555-3325</MATCHED_VALUE_ABST>
   <MATCH_PRECISION>EXACT MATCH</MATCH_PRECISION>
   <SIMILARITY_ID>1</SIMILARITY_ID>
  </DETAIL>
  <DETAIL>
   <ENTITY_ID>87</ENTITY_ID>
   <LIKE_CONF>40</LIKE_CONF>
   <MATCH_ID>376</MATCH_ID>
   <RELTO_ID>6</RELTO_ID>
  </DETAIL>
  <DETAIL>
   <ENTITY_ID>87</ENTITY_ID>
   <INBOUND_VALUE_ABST>1111-22-3333</INBOUND_VALUE_ABST>
   <MATCHED_CODE>6</MATCHED_CODE>
   <MATCHED_DSRC_ACCT>6</MATCHED_DSRC_CODE>
   <MATCHED_ENTITY_ID>10651/MATCHED_ENTITY_ID>
   <MATCHED_KEY_ID>16</MATCHED_KEY_ID>
   <MATCHED_TYPE>NUMBER</MATCHED_TYPE>
   <MATCH_PRECISION>EXACT MATCH</MATCH_PRECISION>
   <SIMILARITY_ID>1</SIMILARITY_ID>
  </DETAIL>
  <DETAIL>
   <ENTITY_ID>87</ENTITY_ID>
   <LIKE_CONF>40</LIKE_CONF>
   <MATCH_ID>10651</MATCH_ID>
   <RELTO_ID>6</RELTO_ID>
 </RELATION>
 <FORMAT_CODE>WS_RELATION</FORMAT_CODE>
<UMF_QUERY_RESULT>
```

Building Web services queries to find entities with similar attributes

Use these instructions to build a UMF_SEARCH input document to find entities in the entity database that match the data values of the attributes specified in the search criteria. You send the UMF_SEARCH input document through Web services into a Web services pipeline for processing. After the pipeline processes the query, Web services returns a UMF_SEARCH_RESULTS output document that contains a list of entities that matched the search critiera.

Before you begin

The embedded WebSphere Application Server must be running, and at least one Web services pipeline must be started and running to receive and process the UMF_SEARCH input document.

About this task

Because the search request is a UMF input document, the criteria must be formatted using valid UMF tags. You can use any text editor or utility that creates UMF.

Procedure

1. Create a new UMF_SEARCH input document.
2. In the ROOT segment, enter the required UMF tags and values, as well as any optional UMF tags and values you want to use to specify the search criteria. At a minimum, enter values for these UMF tags:
   a. Enter the data source code in the DSRC_CODE tag. The default data source code for Web services pipeline searches is 1589. If you use a different data source code than the default Web services pipeline search data source code, make sure that it is configured not to resolve entities.
   b. Enter the data source reference code that references the requesting message transaction in the DSRC_REF tag. The data source reference code should be meaningful, because it is returned to the calling application.
   c. Enter the format code to indicate the output format of the results using the FORMAT_CODE tag. The pipelines come with three built-in format codes for a Web services pipeline search using UMF_SEARCH:
      - WS_SUMMARY_TOP10, which returns the top 10 entities that match the search criteria
      - WS_SUMMARY_TOP100, which returns the top 100 entities that match the search criteria
      - WS_SUMMARY query, which returns all entities that match the search criteria

      If you use a different format code, the format code must be configured in the UMF_OUTPUT_FORMAT table.
   d. Enter the minimum resolution score in the MIN_LIKE_SCORE tag to establish the lowest numeric score that is considered a match between the attribute values in the search criteria and the entities in the entity database containing the same attributes. The higher the score, the more exact the match must be. A score of 100 indicates an exact match.
3. Using the other valid UMF input document segments, enter the data values for the attributes that make up the search criteria. These values are the attributes that the Web services pipeline search is looking for to build the list of entities with matching or similar values. The closeness of the match depends upon on the value in MIN_LIKE_SCORE.
4. Send the UMF_SEARCH input document through Web services.

Results

A Web services pipeline ingests the UMF_SEARCH document, using the entity resolution process to find entities in the database using the specified criteria. The pipeline then processes the query, creates normal logging files, and returns the results in a UMF_SEARCH_RESULTS document through Web services to the calling application using the selected format.

Example UMF_SEARCH document query

This example UMF_SEARCH input document uses the WS_SUMMARY_TOP10 format code to query the entity database to look for the top 10 entities that contain social security numbers where the data value of the social security number exactly matches the data value of 555-09-8761:

Note: This example is formatted for readability and does not follow the required one line per UMF record format.

```
<UMF_SEARCH>
 <DSRC_CODE>1589</DSRC_CODE>
 <DSRC_REF>1223</DSRC_REF>
 <MIN_LIKE_SCORE>100</MIN_LIKE_SCORE>
 <FORMAT_CODE>WS_SUMMARY_TOP10</FORMAT_CODE>
 <NUMBER>
  <NUM_TYPE>SSN</NUM_TYPE>
  <NUM_VALUE>555-09-8761</NUM_VALUE>
 </NUMBER>
</UMF_SEARCH>
```

UMF_SEARCH input document

The UMF_SEARCH input document contains the request and search criteria for a Web services pipeline search. It contains the collection of UMF segments that structure the incoming data to search the entity database for entities that contain attribute values that match the search criteria, and then return the list of entities to the calling application. The results of this Web services pipeline search are returned to the calling application in a UMF_SEARCH_RESULT output document. UMF_SEARCH undertakes a full "Find By Resolution" process.

These required UMF elements and segments comprise the UMF_SEARCH input document:

DSRC_CODE

Data source code UMF tag that is required, because it references and identifies the calling application. As part of normal pipeline logging, this data source code is logged in the UMF_LOG table for each processed UMF_SEARCH.

The system is already configured with a data source code, 1589, that can be used for all Web services pipeline searches. This data source code performs entity resolution processing without resolving the incoming search criteria with the entity in the entity database that matches the search. You can create your own data source code for a particular calling application, just be sure that the data source code is set to not resolve entities.

DSRC_REF

Data source reference UMF tag that is required, because it references the requesting message transaction and is returned to the calling application.

SRC_CREATE_DT

Source create date UMF tag that is optional. If this tag contains a value, it is used for logging.

SRC_LSTUPD_DT

Source last updated date UMF tag that is optional. If this tag contains a value, it is used for logging.

SRC_LSTUP_US

Source last updated user UMF tag that is optional. If this tag contains a value, it is used for logging.

MIN_LIKE_SCORE

Minimum resolution (or likeness) score UMF tag that is required to establish the lowest matching value for the other UMF segments and tags specified. This numeric score determines what is considered a match between the attribute values requested and entities in the entity database containing the same attributes. The higher the score, the more exact the match must be. A score of 100 indicates an exact match.

For example, if the search is to find all entities with a specific social security number, the MIN_LIKE_SCORE determines how closely a social security number must match to the social security data value specified in the query before an entity in the database is listed as part of the result-set for this query.

FORMAT_CODE
UMF tag that correlates to a UMF output document format that is specified in the UMF_FORMAT_CODE table. IBM InfoSphere Identity Insight comes with three built-in format codes for a Web services pipeline search using UMF_SEARCH:

- WS_SUMMARY_TOP10, which returns the top 10 entities that match the search criteria
- WS_SUMMARY_TOP100, which returns the top 100 entities that match the search criteria
- WS_SUMMARY query, which returns all entities that match the search criteria

The only difference between these queries is the number of records returned, which is specified in the query name.

You then specify the optional search criteria using the other available UMF segments and their valid tags for names, addresses, numbers, characteristics, and e-mail addresses.

NAME
Search for name attributes that define the name of the person, organization, place or item, as defined by the entity model and the incoming identity.

NUMBER
Search for number attributes that are comprised of data that is usually described as a number, such as credit card numbers, phone numbers, and passport numbers.

ADDRESS
Search for address attributes that define a location of the identity and typically contain standard address information: street name and number, unit or building number, city, state, country, and postal code.

ATTRIBUTE
Search for characteristic attributes that define other identity traits or information that is not expressed through the other kinds of attributes.

EMAIL
Search for e-mail attributes that define Internet e-mail addresses.

Example UMF_SEARCH query

This example UMF_SEARCH query returns the top 5 entities in the entity database that have a social security number that exactly matches the social security number of 555-09-8761. Even if more than entities were found, only the top 5 entities are returned on the list.

Note: This example is formatted for readability and does not follow the required one line per UMF record format.

```
<UMF_SEARCH>
 <DSRC_CODE>1589</DSRC_CODE>
 <DSRC_REF>1223</DSRC_REF>
 <MIN_LIKE_SCORE>100</MIN_LIKE_SCORE>
 <MAX_RETURN_CNT>5</MAX_RETURN_CNT>
 <FORMAT_CODE>WS_SUMMARY</FORMAT_CODE>
 <NUMBER>
  <NUM_TYPE>SSN</NUM_TYPE>
  <NUM_VALUE>555-09-8761</NUM_VALUE>
 </NUMBER>
</UMF_SEARCH>
```

WS_SUMMARY format codes:

IBM InfoSphere Identity Insight comes with three pre-built format codes for use with the UMF_SUMMARY input document: WS_SUMMARY, WS_SUMMARY_TOP10, and WS_SUMMARY_TOP100. These format codes return a list of entities that match the criteria specified in the UMF_SUMMARY input document. The only difference between these format codes is the maximum number of records returned, which is specified in the format code name.

Example of a Web services pipeline search using the WS_SUMMARY_TOP10 format code

This example Web services pipeline search returns the top 10 entities in the entity database that most closely match the following search criteria:

- Name: Joe Franklin
- Phone number: 415-555-3325
- Date of birth: January 2, 1956

It uses the UMF_SEARCH input document to specify this criteria, also specifying the format code of `WS_SUMMARY_TOP10`.

Note: This example is formatted for readability and does not follow the required one line per UMF record format.

```
<UMF_SEARCH>
 <FORMAT_CODE>WS_SUMMARY_TOP10</FORMAT_CODE>
 <DSRC_CODE>1589</DSRC_CODE>
 <DSRC_REF>556</DSRC_REF>
 <MIN_LIKE_SCORE>80</MIN_LIKE_SCORE>
 <NAME>
  <NAME_TYPE>M</NAME_TYPE>
  <LAST_NAME>FRANKLIN</LAST_NAME>
  <FIRST_NAME>JOE</FIRST_NAME>
 </NAME>
 <NUMBER>
  <NUM_TYPE>PHONE</NUM_TYPE>
  <NUM_VALUE>415-555-3325</NUM_VALUE>
 </NUMBER>
 <ATTRIBUTE>
  <ATTR_TYPE>DOB</ATTR_TYPE>
  <ATTR_VALUE>01/02/1956</ATTR_VALUE>
 </ATTRIBUTE>
</UMF_SEARCH>
```

After sending this UMF_SEARCH document through Web services for processing by a Web services pipeline, the calling application receives a response in the following UMF_SEARCH_RESULT document:

```
<UMF_SEARCH_RESULT>
 <DSRC_CODE>1589</DSRC_CODE>
 <ENTITY>
  <MATCHED_ENTITY_ID>38763</MATCHED_ENTITY_ID>
  <ENT_NAME>FRANKLIN, JOEY</ENT_NAME>
  <ENT_PHONE>415-555-3325</ENT_PHONE>
  <ENT_DOB>01/02/1956</ENT_DOB>
  <LIKE_SCORE>90</LIKE_SCORE>
 </ENTITY>
 <ENTITY>
  <MATCHED_ENTITY_ID>87</MATCHED_ENTITY_ID>
  <ENT_NAME>FRANKLIN, JOSEPH</ENT_NAME>
  <ENT_PHONE>415-555-3325</ENT_PHONE>
  <ENT_DOB>02/01/1956</ENT_DOB>
  <LIKE_SCORE>80</LIKE_SCORE>
 </ENTITY>
 <ENTITY>
  <MATCHED_ENTITY_ID>330</MATCHED_ENTITY_ID>
  <ENT_NAME>FRANKLIN, J</ENT_NAME>
  <ENT_PHONE>451-555-3325</ENT_PHONE>
  <ENT_DOB>01/02/1956</ENT_DOB>
  <LIKE_SCORE>80</LIKE_SCORE>
 </ENTITY>
 <FROM_NODE>556</FROM_NODE>
 <FORMAT_CODE>WS_SUMMARY_TOP10</FORMAT_CODE>
 <MIN_LIKE_SCORE>80</MIN_LIKE_SCORE>
 <PAGE_NUM>1</PAGE_NUM>
 <RETURN_CNT>3</RETURN_CNT>
</UMF_SEARCH_RESULT>
```

In this case, there were only 3 entities in the entity database that matched the search criteria with a minimum likeness score of 80.

Chapter 10. Troubleshooting and support

This section provides information about how to troubleshoot a problem with your IBM InfoSphere Identity Insight software, including instructions for searching knowledge bases, downloading fixes, and contacting support.

Troubleshooting overview

Troubleshooting is a systematic approach to solving a problem. The goal is to determine why something does not work as expected and how to resolve the problem.

The first step in the troubleshooting process is to describe the problem completely. Without a problem description, neither you nor IBM can know where to start to find the cause of the problem. This step includes asking yourself basic questions, such as:

- What are the symptoms of the problem?
- Where does the problem occur?
- When does the problem occur?
- Under which conditions does the problem occur?
- Can the problem be reproduced?

The answers to these questions typically lead to a good description of the problem, and that is the best way to start down the path of problem resolution.

What are the symptoms of the problem?

When starting to describe a problem, the most obvious question is "What is the problem?" This might seem like a straightforward question; however, you can break it down into several more-focused questions that create a more descriptive picture of the problem. These questions can include:

- Who, or what, is reporting the problem?
- What are the error codes and messages?
- How does the system fail? For example, is it a loop, hang, crash, performance degradation, or incorrect result?
- What is the business impact of the problem?

Where does the problem occur?

Determining where the problem originates is not always easy, but it is one of the most important steps in resolving a problem. Many layers of technology can exist between the reporting and failing components. Networks, disks, and drivers are only a few components to be considered when you are investigating problems.

The following questions can help you to focus on where the problem occurs in order to isolate the problem layer.

- Is the problem specific to one platform or operating system?
- Is the problem common across multiple servers?
- Is the current environment and configuration supported?

Remember that, even though one layer might report the problem, this does not mean that the problem originates in that layer. Part of identifying where a problem originates is understanding the environment in which it exists. Take some time to completely describe the problem environment, including the operating system, its version, all corresponding software and versions, and hardware information. Confirm that you are running within an environment that is a supported configuration; many problems can be traced back to incompatible levels of software that are not intended to run together or have not been fully tested together.

When does the problem occur?

Develop a detailed timeline of events leading up to a failure, especially for those cases that are one-time occurrences. You can most easily do this by working backward: Start at the time an error was reported (as precisely as possible, even down to the millisecond), and work backward through the available logs and information. Typically, you need to look only as far as the first suspicious event that you find in a diagnostic log; however, this is not always easy to do and takes practice. Knowing when to stop looking is especially difficult when multiple layers of technology are involved, and when each has its own diagnostic information.

To develop a detailed timeline of events, try to answer these questions:

- Does the problem happen only at a certain time of day or night?
- How often does the problem happen?
- What sequence of events leads up to the time that the problem is reported?
- Does the problem happen after an environment change, such as upgrading or installing software or hardware?

Responding to questions like this can help to provide you with a frame of reference in which to investigate the problem.

Under which conditions does the problem occur?

Knowing what other systems and applications are running at the time that a problem occurs is an important part of troubleshooting. These and other questions about your environment can help you to identify the root cause of the problem:

- Does the problem always occur when the same task is being performed?
- Does a certain sequence of events need to occur for the problem to surface?
- Do any other applications fail at the same time?

Answering these types of questions can help you explain the environment in which the problem occurs, and correlate any dependencies. Remember, just because multiple problems might have occurred around the same time, the problems are not necessarily related.

Can the problem be reproduced?

From a troubleshooting standpoint, the "ideal" problem is one that can be reproduced. Typically with problems that can be reproduced, you have a larger set of tools or procedures at your disposal to help you investigate. Consequently, problems that you can reproduce are often easier to debug and solve. However, problems that you can reproduce can have a disadvantage: If the problem is of significant business impact, you do not want it to recur! If possible, re-create the problem in a test or development environment, which typically offers you more flexibility and control during your investigation.

- Can the problem be re-created on a test machine?
- Are multiple users or applications encountering the same type of problem?
- Can the problem be re-created by running a single command, a set of commands, or a particular application, or a stand-alone application?

Troubleshooting IBM InfoSphere Identity Insight

Use the following questions to help you identify and find resolutions for problems that are occurring with IBM InfoSphere Identity Insight.

1. During installation, did the installation program inform you that one or more components were not successfully installed? If so, review the installation log files to determine and fix the problem.
2. Are your service updates at the latest level?
3. Are you receiving an error message?
4. Have you checked the component log files to see if they contain any messages about the problem?
5. Does the problem occur when using one of the following components?
   - the Analyst Toolkit web applications - review the “Analyst toolkit web applications troubleshooting checklist” on page 348
   - the pipelines - review the pipeline troubleshooting checklist
6. Have you reviewed the product knowledge bases for information that might resolve the problem?
7. If you have tried each of these applicable options and your problem is still not resolved, contact IBM Software Support.

Pipelines troubleshooting checklist

If you experience problems with the pipelines, before calling IBM Software Support, review this list of the most common pipeline problems encountered.

1. Pipeline reports a status of Down or is not reporting any status
2. Pipeline shuts down
3. Pipelines do not honor configuration changes made in the Configuration Console
4. Pipelines do not start on AIX
5. Pipeline processes only part of an incoming record
6. Transport does not work
7. Pipeline does not load scientific notations or floating point numbers
8. After starting a pipeline, I receive a warning message that indicates no routes are defined

1. **Pipeline reports a status of "Down" or is not reporting any status at all**
   - Does the pipeline node have an error or is it not running?
   - Did the transport specified in the pipeline command use the correct syntax?
   - Has the pipeline shut down?
2. **Pipeline shuts down or crashes**
   - Did the pipeline encounter too many errors processing incoming data files?
     - Check the log files for more information about the errors. Use that information to resolve the problem.

   - Check the *ErrorLimit* setting in the pipeline configuration file. You might need to increase this number.
- Did the pipeline run out of memory resources?
- Is the database causing the problem because of one of the following reasons:
   - Not enough disk space?
   - Lost connectivity to the pipeline?
   - Has the user name and password changed for this database?

3. **Pipelines do not honor configuration changes made in the Configuration Console**
   - Before pipelines apply configuration changes, they must be stopped and restarted. When the pipelines restart, the configuration changes are applied as part of the pipeline initialization process.
   - For data-integrity purposes, stop and restart all running pipelines after a configuration change.
4. **Pipeline does not start on AIX**
   - Did you receive an error message indicating that the "dependent module libicuio.a could not be found"?
      - If so, make sure that the library is found in one of the following directories: `/usr/lib`, `/lib`, `$DB2INSTHOM/sqllib/lib`. Or set the `LIBPATH` environment variable to include the product *`installation_home`*`/lib` directory.
   - Check the version and location of the C++ runtime libraries. The problem might be a result of incorrect settings in the `RunTime Update` and the `LIBPATH` environment. Please refer to the "IBM InfoSphere Identity Insight Installation and Configuration Guide" to the latest support information.
5. **Pipeline processes only part of an incoming record and not the full record**
   - Check the *.BAD log file for invalid UMF messages. This log file indicates the name of the incoming data source file that was being processed.
   - Check the **UMF Exceptions** tab in the Configuration Console.
6. **Pipeline transport does not work**
   - Make sure that the syntax used for the transport is correct. For example, if you are specifying a database transport, did you include quotes in the appropriate place?
   - If the transport is a queue transport, does the message queue exist?
   - If the transport is a file, does the file exist? Is the file located in the directory specified in the transport?
7. **Pipeline does not load a scientific notation or a floating point number**
   - This is a known pipeline limitation. Revise the scientific notations or floating point numbers in the UMF to multiply out the exponent, so the number is in standard numeric notation. For example, `-1.267E-05` multiplied out is `-0.00001267`.
8. **After starting a pipeline, I receive a warning message that indicates no routes are defined**
   - This is an information warning message only. It is safe to ignore it. (The message just informs you that there are no routes defined for this pipeline. Routes are not required to run a pipeline.)

Analyst toolkit web applications troubleshooting checklist

If you experience problems with the web applications, before calling IBM Software Support, review this list of the most common problems encountered:

1. I cannot see the Configuration Console Login screen
2. I cannot log in to the Configuration Console
3. A report opens in the Web browser, but nothing displays on the report
4. I cannot see the status of a pipeline on the Pipeline Status tab
5. Configuration changes made in the Configuration Console are not being honored by the pipelines.

1. **I cannot see the Login screen.**
   - Do you see the message "The page cannot be displayed"?
     - The web application URL is probably incorrect. Retype the URL. If you are unsure of the correct URL, contact your system administrator or internal technical support for assistance.
     - Other potential reasons: The port that connects your machine to the WebSphere Liberty server might be blocked, or the WebSphere Liberty server might not be started. Contact your system administrator or internal technical support for assistance.
   - Is the screen blank?
     - Contact your system administrator or internal technical support. The port that connects your machine to the WebSphere Liberty server might not be started, or the Identity Insight database password might have changed.
   - If none of these solutions resolves the problem, contact your system administrator or internal technical support for assistance.
2. **I cannot log in to the web application.**
   - Make sure that you are entering the correct user name and password. The Analyst Toolkit applications do not lock user accounts, regardless of the number of incorrect login attempts, so try entering your user name and password again.
   - If you have forgotten your user name and password, contact your system administrator or internal technical support for assistance. You might need to have your password reset.
3. **Configuration changes made in the Configuration Console are not being honored by the pipeline.**
   - Before pipelines apply configuration changes, they must be stopped and restarted. When the pipelines restart, the configuration changes are applied as part of the pipeline initialization process.
   - For data-integrity purposes, stop and restart all running pipelines after a configuration change.

Visualizer troubleshooting checklist

If you experience problems with the Visualizer, before calling IBM Support, check this list of the most common problems encountered while using the Visualizer. You might be able to resolve your Visualizer problem or issue by yourself.

1. I cannot start the Visualizer
2. I cannot log in to the Visualizer
3. I generated a Visualizer report. The report opens in my Web browser, but nothing displays on the report
4. I am receiving error messages about the pipeline
5. The Visualizer 'hangs' or 'freezes'
6. Find by Attribute does not return the expected results

7. I am receiving an error message about "insufficient indexes" when using the Find by Attribute window
8. The custom icons for Visualizer graphs do not display or display incorrectly
9. Links (or hyperlinks) do not work in the Visualizer

1. **I cannot start the Visualizer**
   - First, ensure that your workstation client has the required client version of Java installed.
   - If you have multiple versions of Java installed on your client machine, it is likely that the default system version of Java Web Start is not the version required to run the Visualizer. Also keep in mind that the client Java version required to open and run the Visualizer may not be the latest version of Java installed on your machine. There are two ways to resolve this issue: Either associate the required client version of Java Web Start in your Web browser or use a direct-launch approach.
     - Is the Visualizer the only Web Start application that you use on this workstation client? If yes, set your Web browser to associate the *.JNLP file type to use the required client version of Java Web Start.
     - Do you run additional Web Start applications besides the Visualizer on this workstation or do you want to avoid changing system and Java settings ? If yes, directly launch the Visualizer from the Java Web Start file.
   - Are you receiving an error message indicating that the application has requested a version of JRE that is not installed? If yes, configure Java version 1.6 to accept auto-downloads.
   - Can you see the Visualizer Web start page?
     - Yes, I see the Visualizer Web Start page, but I see a message indicating that "Java Web Start is required to launch the Visualizer." I do not see a **"Click here to start the IBM InfoSphere Identity Insight Visualizer"** link.
       - Do you only use this workstation client for the Visualizer? If yes, set your Web browser to associate the JNLP file type to use the required client version of Java Web Start.
       - Do you use this workstation client to open other Web Start applications or do you want to avoid changing system and Java settings ? If yes, directly launch the Visualizer from the Java Web Start file
     - Yes, I saw the Visualizer Web Start Launch page and a Visualizer Splash screen, but I cannot see a Visualizer **Login** window.
       - Did you click the **"Click here to start the IBM InfoSphere Identity Insight Visualizer"** link?
         - If yes, Java might be in the process of opening the Visualizer, which might take a few minutes. If the Visualizer is in the process of opening, you typically see a Java Splash screen or Java Web Start window.
         - If no, click the link to start the Visualizer.
       - The problem is most likely with the embedded WebSphere Application Server. Either the application server is experiencing an error or problem and might need to be restarted, or the application server cannot connect to the correct product database. Contact your system administrator or internal technical support.
     - No, I do not see the Visualizer Web Start page.
       - If you see the message "The page cannot be displayed", check the Visualizer URL. The URL might contain a typo or might be the incorrect

URL for the Visualizer. Retype the URL. If you do not know the Visualizer URL, contact your system administrator or internal technical support.

  - If the URL is correct, here are some other potential reasons why the Visualizer Web Start page does not display:
    - The WebSphere Application Server is experiencing an error or problem and might need to be restarted.
    - The port that connects your workstation client to the WebSphere Application Server might be blocked or already in use by another application.
- If none of these actions resolves the problem, ask your system administrator or internal technical support to contact IBM Software Support.

2. **I cannot log into the Visualizer.**
   - Can you see the Visualizer **Login** screen?
     - No, I cannot see the Visualizer **Login** screen.
       - The problem is most likely with the embedded WebSphere Application Server. Either the application server is experiencing an error or a problem (not connected) or the application server cannot connect to the correct product database. Contact system administrator or internal technical support for assistance.
     - Yes, I can see the Visualizer**Login** screen, but I cannot login.
       - Check to be sure that you are entering the correct user name and password for your Visualizer user account. The Visualizer does not lock user accounts, regardless of the number of incorrect login attempts. So try entering your user name and password again. You cannot lock yourself out of your account.
       - Make that you click **Login**. The **Login** button is not automatically selected, so if you enter your user name and password and press **Enter**, happens. You must use either the mouse to click on **Login** or select **Login** using the keyboard.
   - Have you forgotten your user name and password?
     - Yes. Contact system administrator or internal technical support to look up your user name or reset your Visualizer account password in the Configuration Console.
3. **I generated a Visualizer report. The report opens in my Web browser, but nothing displays on the report.**
   - Wait a minute or two more, because the report may still be generating. As the system is generating a report, it starts by displaying a blank screen in the browser. Once the report is completely generated and ready for display, the system displays the report.
   - Make sure you have Adobe Acrobat Reader version 7.0 or higher installed on your local machine. If not, you can download the latest Adobe Acrobat Reader for free from the Adobe Web site.
   - Is there a firewall on your system? If so, check the firewall to make sure that localhost and the application server are both granted access through the firewall.
4. **I am receiving error messages about the pipeline.**
   - Review the error message carefully for more information about what is causing the problem.
   - Make sure that the Visualizer pipeline is an HTTP pipeline.
   - Does your workstation have Visualizer client logging turned on?

  - No.
    - "Turning Visualizer client logging on" on page 368 on your machine. Set the log level to Debug. Then contact your system administrator or internal technical support, providing the error message text and let that person know that you have turned on Visualizer client logging. Your system administrator or internal technical support may want you to try to connect to the pipeline again, and then examine the log file.
    - When you have resolved the problem, turn off Visualizer client logging.
  - Yes.
    - Review the Visualizer client log files located at `installation_directory/logs/ewas`.
    - Contact system administrator or internal technical support. Your system administrator or internal technical support may want to review the Visualizer client log file.

5. **The Visualizer 'hangs' or 'freezes'.**
   - The port that connects your machine to the embedded WebSphere Application Server may be blocked, or the embedded WebSphere Application Server may not be started. Contact your system administrator or internal technical support
   - Information for DBAs, system administrators, or internal technical support:
     - Consider running stats against entity database tables that affect the Visualizer.
     - If all the Visualizer users are having difficulties with the Visualizer 'hanging', check to be sure that the database table indexes have not been modified. Modifying the indexes on the database tables may cause unpredictable and undesirable results. If you find that the indexes have been modified, contact IBM Software Support.
6. **Find by Attribute does not return the expected results.**
   - Review your search criteria.
     - If you see fewer results than you expected, you may need to broaden the criteria.
     - If you see more results than you expected, you may need to narrow the search criteria.
     - By default, the system only returns up to 1000 records per search. (However, the setting is the configurable. This setting is controlled by the `MAX_ENTITIES_RETURNED` parameter on the **System Parameters** tab of the Configuration Console. You may want to contact your system administrator or internal technical support to verify or modify this setting.)
   - The issue may be related to a case sensitivity database configuration. Contact your system administrator or internal technical support to check the database configuration for case sensitivity settings.
     - For DB2 databases: The DBA, system administrator, or internal technical support may need to apply a script to support case insensitive database searches. Have your system administrator contact IBM Software Support to get the script and the instructions on how to run it.
     - For Microsoft SQL Server databases: The database may be set to be case sensitive. The DBA, system administrator, or internal technical support may need to change the case sensitivity setting for the database.

   - For Oracle databases: The DBA, system administrator, or internal technical support may need to create functional-based indexes with UPPER to support case insensitive database searches.

7. **I am receiving an error message about "insufficient indexes" when using the Find by Attribute window.**
   - You are trying to search on a field that is not indexed.
     - Try narrowing the search by entering additional search criteria.
     - Or contact your system administrator or internal technical support. Depending on the impact on system performance, your system administrator may create a new index on this field. (Your system administrator or technical support might also check the `ENABLE_SEARCH_INDEX_CHECK` parameter on the **System Parameters** tab in the Configuration Console. If this setting is not set to 1, system performance may be affected.)
8. **The custom icons for the Visualizer graphs do not display or display incorrectly.**
   - The icons might not be located in the correct directory on the application server. Contact your system administrator or internal technical support to verify the path location of the custom graph icons.
   - The icon names might be in mixed case rather than all lowercase or might not match their corresponding attribute type. For example, if **`Evidence Photo`** is the name of the attribute type, then the image file name must be in all lowercase characters and include the space between the words `evidence` and `photo`. The file name must look like this:**`evidence photo.gif`**. Contact your system administrator or internal technical support to make sure the icon file name is correct.
   - The icons might not be in the recommended .GIF file format. Or the icons may not be the recommended size, which is 24-by-24 pixels. Contact your system administrator or internal technical support to make sure the icon is in the right file format and uses the recommended image size..
9. **Links (or hyperlinks) in the Visualizer do not work. I see an error message when I click an attribute link.**
   - Configure the hyperlink settings for your workstation. In Visualizer system preferences, you choose the Web browser or program that is used to open files associated with identity record attributes. This setting must be configured on each workstation that runs the Visualizer.
   - After configuring the hyperlink settings, make sure to close the Visualizer and restart it again.

System Health

Here are some tips for DBAs and system administrators on keeping your IBM InfoSphere Identity Insight system healthy.

Performance tips

If you notice a degradation in overall system performance, review this list for ideas on possible causes:

- Database tuning: When was the last time that someone ran database statistics against IBM InfoSphere Identity Insight tables?
- Very large entities: Does the entity database contain very large entities– entities with numerous identities?

While this list is not all-inclusive, it provides a place to start to validate that the system is at peak performance.

Monitoring the entity database tips

Here are some specific items to check to help monitor the health of the entity database:

- Database tuning: What is the schedule for running database statistics against IBM InfoSphere Identity Insight tables?
- Unique numbers: Does the entity database contain multiple entities that share the same unique number?
- Entities: Does the entity database contain entities with many unique numbers?
- Over-resolution: Does the entity database contain very large entities– entities with numerous identities?

While this list is not all-inclusive, it provides some quick tips for monitoring overall system health.

Database tables that affect system performance

If system performance seems slow, Database Administrators can run database statistics against several entity database tables to improve both pipeline performance and the Visualizer user experience.

Pipeline tables

If pipeline performance seems slow, try running database statistics against the following entity database tables:

- DQM_NAME_DICT
- NAME
- ADDRESS
- NUMS
- ATTRIBUTES
- EMAIL_ADDR
- DSRC_ACCT
- SEP_RELATIONS
- SEP_ROLES
- ENTITY
- DISCLOSED_RELATIONS
- UMF_LOG
- UMF_EXCEPT

Visualizer tables

If Visualizer users complain that the Visualizer performance seems slow, try running database statistics against the following entity database tables:

- ER_ENTITY_SCORE
- ER_HISTORY
- ER_RELOCATION
- ER_DETAIL
- ER_ACCT_SCORE

- ER_ENTITY_STATE
- ER_FORCED_LOG
- SEP_CONFLICT
- SEP_CONFLICT_REL
- SEARCH
- APP_ACTIVITY_CODES
- APP_ACTIVITY_HISTORY
- APP_CONFLICT_GROUP
- APP_INBOX
- APP_ROLE
- APP_SEND
- MATCH_MERGE_RULES
- CONFLICT_RULES

In addition, because the Visualizer also uses a background pipeline to perform several Visualizer tasks (such as adding entities, finding entities by entity resolution, and disclosing relationships), Database Administrators should also run database statistics against the database tables listed in the Pipeline tables section.

Large Entities query

This SQL query looks for large entities. The more identity records an entity has, the larger it becomes. Sometimes, the results of processing incoming identity data can cause the system to over-resolve identity records during entity and relationship resolution. Large entities can cause a marked slow down in system performance.

Large Entities SQL Query statement

```
select entity_id
 count(dsrc_acct) as IDENTITY_CNT
from
 DSRC_ACCT
where
 sys_delete_dt is null
group by
 entity_id
count(dsrc_acct) > 100
order by count(dsrc_acct)desc;
```

What next?

In the Identity Insight plug-in for i2 or the Explorer application, use the **Find by Entity ID** screen to look up the Entity IDs returned by the Large Entities query results. Verify that the identities associated with this entity are properly associated. For properly constructed entities, the entity has many different data source accounts, while most of the data for associated names, addresses, and numbers of are very similar. If you have questions about whether the entity is properly constructed, contact your IBM Services or Support for further assistance.

Example results of Large Entities query

Here is an example of what the results of running the Large Entities query might look like:

ENTITY_ID	IDENTITY_CNT
3015	22

ENTITY_ID	IDENTITY_CNT
5241	41
7854	36

Total Unique Numbers by Entity query

This query returns information about how many different unique numbers are associated with a particular entity, by Entity ID. You might find this query useful if each entity typically only has one unique number. Checking to see if entities contain many different types of unique numbers is an excellent way to check for data anomalies and verify that your resolution rules are working as expected.

Total Number of Unique Numbers Associated with a Single Entity SQL Query statement

```
select distinct *
from
 (select entity_id,
  (select count(distinct num_value)
  from
   nums,
   num_type
  Where
   nums.num_type_id=num.type.num_type_id
   and num_type.unique_FLAG='Y'
   and nums.entity_id=dsrc_acct.entity_id
  ) as UNIQUE_NUMBER_CNT
 from dscr_acct
 )as tab1
where
 UNIQUE_NUMBER_CNT>1
order by
 UNIQUE_NUMBER_CNT DESC;
```

What next?

In the Identity Insight plug-in for i2 or the Explorer application, use the **Find by Entity ID** screen to look up the Entity IDs returned from the Total Number of Unique Numbers by Entity ID query results. By reviewing the entity resume for each entity, you can determine if the entity should have more than one unique number. In some cases, this situation might be an indication of fraud. For example, in the United States Social Security Numbers (SSNs) are unique numbers. Typically, each U.S. entity only has one SSN. If this query uncovers an entity that has multiple SSNs, the next step is probably to do further investigation and analysis of why the entity has multiple SSNs.

Example results of Total Number of Unique Numbers by Entity query

Here is an example of what the results of running the Total Number of Unique Number by Entity query might look like:

ENTITY_ID	UNIQUE_NUMBER_CNT
3003	2
3030	2
3039	2

Unique Number Shared by Multiple Entities query

Unique numbers are numbers that, typically, only belong to one entity and are not shared by multiple entities. Checking to see if multiple entities share the same unique numbers is an excellent way to test for data anomalies and verify that your resolution rules are working as expected. You can use the Unique Number Shared by Multiple Entities query to uncover entities that share the same unique number. The query counts a unique number for a single entity only once, regardless of how many identity records for that entity contain the same unique number.

Unique Numbers Shared by Multiple Entities SQL Query statement

```
select num_type,
 num_value,
 count(distinct ENTITY_ID) as cnt
from nums,
 num_type
Where   nums.num_type_id=num_type.num_type_id
 and num_type.unique_FLAG='Y'
Group by
 num_type
 num_value
Having
 count(distinct ENTITY_ID)>1
Order by
 count(distinct ENTITY_ID)desc;
```

What next?

In the Identity Insight plug-in for i2 or the Explorer application, use the **Find by Attribute** screen to look up each number returned by the Unique Numbers Shared by Multiple Entities SQL query. In the **Results** pane, review the entity information for each entity that shares the unique number. You can also review the entity resumes of these entities to help determine why the entities share the same unique number.

You might discover interesting relationships between entities, based on the unique number. For example, you could discover that two different entities are using the same Social Security Number.

Or you might detect an issue with the UMF coding for unique numbers. For example, you could discover that the same passport number is shared between two entities, because the incoming UMF identity record did not use NUM_LOC to indicate the country (location) issuing the passport number. Numbers like passports and drivers licenses are only unique to a particular location, such as country or state. In themselves, these numbers might not be as unique as you think.

Example results of Unique Number Shared by Multiple Entities query

Here is an example of what the results of running the Unique Number Shared by Multiple Entities query might look like:

NUM_TYPE	NUM_VALUE	cnt
SSN	000-00-0000	9
SSN	111-11-1111	9

NUM_TYPE	NUM_VALUE	cnt
SSN	555-55-5555	5
SSN	611-00-6666	2
SSN	999-99-9999	3

Searching knowledge bases

You can often find solutions to problems by searching IBM knowledge bases. This topic describes how to optimize your results by using available resources, support tools, and search methods.

Available technical resources

In addition to this information center, the following technical resources are available to help you answer questions and resolve problems:

IBM InfoSphere Identity Insight technotes at www.ibm.com/software/support/isa/

Searching with support tools

The following desktop tools are available to help you search across IBM knowledge bases:

- **IBM Support Assistant (ISA)** is a free software serviceability workbench that helps you resolve questions and problems with IBM software products. Instructions for downloading and installing the ISA can be found on the ISA Web site at www.ibm.com/software/support/isa/
- **IBM Software Support Toolbar** is a browser plug-in that provides you with a mechanism to easily search IBM support sites. You can download the toolbar at www.ibm.com/software/support/toolbar/.

Search tips

The following resources describe how to optimize your search results:

- Searching the IBM Support Web site
- Using the Google search engine

Receiving automatic updates

- **My support**. To receive weekly e-mail notifications regarding fixes and other support news, follow these steps:
  1. Go to the IBM Software Support Web site at www.ibm.com/software/support/.
  2. Click **My support** in the far upper-right corner of the page under **Personalized support**.
  3. If you have already registered for My support, sign in and skip to the next step. If you have not registered, click **register now**. Complete the registration form using your e-mail address as your IBM ID and click **Submit**.
  4. Click **Edit profile**.
  5. In the **Products list**, select **Software** . A second list is displayed.

6. In the second list, select a product segment, for example, **Systems management**. A third list is displayed.
7. In the third list, select a product sub-segment, for example, **Application Performance & Availability**. A list of applicable products is displayed.
8. Select the products for which you want to receive updates.
9. Click **Add products**.
10. After selecting all products that are of interest to you, click **Subscribe to email** on the **Edit profile** tab.
11. Select **Please send these documents by weekly email**.
12. Update your e-mail address as needed.
13. In the **Documents list**, select **Software**.
14. Select the types of documents that you want to receive information about.
15. Click **Update**.

Messages overview

When you receive a message from a system component, you can often resolve the problem by reading the entire message text and the recovery actions that are associated with the message.

Message identifiers are 10 characters long, and the characters in the message identifier provide more information about the message.

- The first three characters identify the product.
  - **CWU** is the product identifier for IBM InfoSphere Identity Insight.
- The next two characters identify the specific component within the product that is generating the message.
  - **AE** is the component identifier for the pipeline.
  - **AI** is the component identifier for the Configuration Console.
  - **AK** is the component identifier for the Event Manager.
  - **AL** is the component identifier for Web services.
- The next four characters are the message number.
- The last character is the message type code, which describes the severity of the message:
  - **E** indicates an error message. This type of message indicates a problem with a specific product component that requires immediate action. Review the component log files for information to help you troubleshoot and resolve the error.
  - **I** indicates an informational message. This type of message does not require immediate action, but you might want to review the component log files for more information.
  - **W** indicates a warning message. This type of message indicates that a condition has occurred that may need attention. Review the component log files for more information as to what the warning condition is and how you can resolve the condition.

Message examples

If you receive a message with the message identifier of CWUAE0001E, the message indicates an error message from a pipeline that most likely caused the pipeline to shut down and stop processing. You should review the pipeline log files to resolve the problem, so you can restart the pipeline.

If you receive a message with the message identifier of CWUAE325W, the message indicates that a warning message occurred in the pipeline, but the warning did not stop the pipeline from continuing to process incoming records. You can check the pipeline log files for more information about the warning, to see what actions you might need to take to correct the problem or incoming data record. If this particular pipeline is being monitored by the application monitor, you can also check the application monitor windows in the Configuration Console for more information.

UMF parsing errors

UMF parsing errors occur when incoming UMF identity records are incorrectly formatted, such as an end tag is missing, or there are invalid characters in the UMF.

Table 27. UMF Parsing Errors

UMF Error Code	Code Description	Severity
005	Leading spaces are not allowed in the tag name *string*	Major
010	Root level begin tag is missing <*string*>	Major
015	Encountered an unexpected end tag </*string*>	Major
020	Incorrect close tag </*string*> encountered, expecting </*string*>	Major
025	Document is incomplete, not enough end tags...Last segment: <*string*>	Major
030	Document is empty	Warning
035	Segments cannot contain tag data '*string*' when having children	Major

Logs

IBM InfoSphere Identity Insight contains logging mechanisms that write information to a series of log files. Typically, the system begins writing information to the log files when a qualifying condition occurs to a specific system component, such as the component is installed or started, a user logs into the component, or an error occurs during processing.

The following system components create log files:

- Pipelines
- Analyst Toolkit web applications
- Web services
- Event Manager

Pipeline log files

Whenever you start a pipeline, the system automatically starts logging, based on the current pipeline logging configuration in the pipeline configuration file. Logging files are created for each pipeline, by pipeline name, even if you started multiple pipelines using the same configuration file.

Types of pipeline log files

By default, all pipeline log files are written to the directory on the pipeline node where the pipeline was started. There are several different types of pipeline log files. Which message is logged to which file depends upon the mode in which the pipeline was started (debug `-d` mode or daemon/service `-s` mode), the type of message being logged, and the current logging configuration.

Table 28. Pipeline Logging Files by Type of Message, Log File Name, and Logging Modes

Type of message	Log file name	Action	Logging mode(s)
Error messages	*pipeline_name*.err Logs critical errors that occurred in the pipeline.	After reviewing the log files, fix the errors or issues indicated with the pipeline.	Service Debug
SQL error messages	*pipeline_name*.SqlErr.log Logs SQL errors that occurred in the pipeline. This file has a size limit of 1 megabyte. When the file reaches that size limit, the system automatically archives the current log file and creates a new one.	After reviewing this log file, fix the SQL errors or issues indicated.	Service Debug
Queue errors	*pipeline_name*.MQErr.log Logs queue errors.	After reviewing this log file, fix the MQ errors or issues indicated.	
Windows Event Viewer	(Microsoft Windows platforms only) If the pipeline has services installed and was started using the service mode (`-s` pipeline option), the pipeline also sends errors and important messages to the Windows Event Viewer.	Monitor the messages in the Windows Event View and fix any errors or issues indicated..	Service (Microsoft Windows platforms only)
Bad/invalid UMF messages that could not be processed	*pipeline_name*.bad Logs information about records in the incoming data source file that contain malformed or invalid UMF. The pipeline could not process the portion of the record containing this bad or invalid UMF, which sometimes mean that the pipeline processes partial records.	After reviewing this log file, fix the records in the incoming data source file with bad or invalid UMF. Then send the corrected records back through a pipeline for processing.	Service Debug

Table 28. Pipeline Logging Files by Type of Message, Log File Name, and Logging Modes (continued)

Type of message	Log file name	Action	Logging mode(s)
UMF messages that generated exceptions	*pipeline_name*.msg Logs information about records in the incoming data source file that contain generated exceptions during processing. The pipeline did process the record. This type of message may indicate a problem with the data quality for this data source file.	After reviewing this log file, you might still need to fix records in the incoming data source file that generated the UMF exception. Then send the corrected records back through a pipeline for processing. You can also review the Load Summary Report or the Data Source Summary Report for more information.	Service Debug
Debug tracing	Logs debug tracing information when a pipeline was started using the debug mode (`-d` pipeline option). There is no log file. The pipeline runs in the foreground with output messages sent directly to the command shell. You can use the redirection feature to create a file from the pipeline command output: `pipeline -d -f my_umf.xml > my_log_file.log`		Debug
SQL statements and performance statistics	*pipeline_name*.SqlDebug.log Logs SQL statements and performance statistics that can assist with troubleshooting problems and monitoring performance. This file has a size limit of 48 megabytes. When a file reaches the size limit, the system automatically archives the current log file and creates a new log file.		Debug

Table 28. Pipeline Logging Files by Type of Message, Log File Name, and Logging Modes (continued)

Type of message	Log file name	Action	Logging mode(s)
Pipeline shuts down while processing a file	*pipeline_name*.cnt As the pipeline processes incoming records, it logs the name of the data source file being processed, as well as a record count for every 100 records in the file successfully processed. If a pipeline shuts down while processing an incoming data source file, this file can help you determine which of the records in the data source file need to be reloaded into the pipeline for processing.	After reviewing this log file and fixing the problem that shut down the pipeline, reload the unprocessed records into the pipeline for processing.	File

Pipeline logging configurations

IBM InfoSphere Identity Insight provides a default logging configuration that logs pipeline events and errors. This default logging configuration is automatically used, unless a custom pipeline logging configuration is specified in the pipeline configuration file.

There are two primary ways that pipelines are started: debug mode (-d pipeline option) or service/daemon mode (-s pipeline option).

- Debug mode is useful when for testing and troubleshooting the system. It is not usually used in production environments. Logging for the debug mode includes more tracing and pipeline operation information.
- Service/daemon mode is the typical production environment mode. Logging for the service/daemon mode is usually limited to errors and problems that require action.

All pipeline logging configurations (both default and custom) must specify how to log pipeline events in debug mode and in service/daemon mode. If the default logging configuration does not meet your needs, you can create a custom logging configuration by adding a logging section in the pipeline configuration file and using the pipeline configuration components to specify how the system logs pipeline events and errors for both the debug pipeline mode and the service/daemon pipeline mode.

Default debug mode logging configuration

```
console://stdout $NODE_NAME.*;*.CRIT;*.ERR;*.NOTE
 cmeadmin:/// *.CRIT;*.ERR file://./$NODE_NAME.err *.CRIT
 file://./$NODE_NAME.SqlDebug.log?rotateSize=49152 sql.DBUG;sql.PERF
 file://./$NODE_NAME.SqlErr.log?rotateSize=1024 sql.ERR;sql.CRIT
 file://./$NODE_NAME.MQErr.log mq.!DBUG
 file://./$NODE_NAME.bad?style=bare bad_xml.*
 file://./$NODE_NAME.msg?style=bare msg.*
```

Default Microsoft Windows service mode logging configuration

```
eventlog://./ *.NOTE;*.CRIT;*.ERR
 cmeadmin:/// *.CRIT;*.ERR file://./$NODE_NAME.err *.CRIT
 file://./$NODE_NAME.SqlDebug.log?rotateSize=49152 sql.DBUG;sql.PERF
 file://./$NODE_NAME.SqlErr.log?rotateSize=1024 sql.ERR;sql.CRIT
 file://./$NODE_NAME.MQErr.log mq.!DBUG
 file://./$NODE_NAME.bad?style=bare bad_xml.*
 file://./$NODE_NAME.msg?style=bare msg.*
```

Default UNIX daemon mode logging configuration

```
file://./$NODE_NAME.log *.CRIT;*.ERR;*.NOTE;*.INFO;logger.!DBUG
 cmeadmin:/// *.CRIT;*.ERR file://./$NODE_NAME.err *.CRIT
 file://./$NODE_NAME.SqlDebug.log?rotateSize=49152 sql.DBUG;sql.PERF
 file://./$NODE_NAME.SqlErr.log?rotateSize=1024 sql.ERR;sql.CRIT
 file://./$NODE_NAME.MQErr.log mq.!DBUG
 file://./$NODE_NAME.bad?style=bare bad_xml.*
 file://./$NODE_NAME.msg?style=bare msg.*
```

Pipeline logging components

Pipeline logging components help you create custom pipeline logging configurations. They provide the system with instructions on how to log pipeline events and messages.

Log writer

Specifies which log writer to use to write or display the log file:

file
Writes the log events and messages to a specified file name.

The file log writer uses the Path, Parameter, White Space, and Filter logging components. For example:

```
file://absolute path?parameters [white space] filter
```

cmeadmin
Writes the log events and messages to the cmeadmin log.

The cmeadmin log writer uses the White Space and Filter logging components. For example:

```
cmeadmin://[white space] filter
```

console
Writes the log events and messages to the command line console.

The console log writer uses the Location, Parameters, and Filter logging components. For example:

```
console://file location?parameters filter
```

eventlog
(Microsoft Windows platforms only) Writes the log events and messages to the Microsoft Windows Event Viewer.

The eventlog log writer uses the Filter logging components. For example:

```
eventlog://./filter
```

Path Specifies the file location and file name to write the log information to:

File location
Valid values include:

- stdout - used with the console log writer
- stderr - used with the console log writer
- *absolute path* - used with the file log writer

file name

Indicates which standard product log file to write the information to. The file name extension determines the type of log file. Valid log file extension names include:

- .err
- .bad
- .msg
- .SqlDebug.log
- .SqlErr.log
- .MQErr.log

Parameter

Specifies optional logging parameters. Valid values include:

style=bare

Indicates that the logging does not include timestamps or other header information. This parameter is typically includes in files that log UMF messages.

rotateSize=*maximum file size number*

Indicates the maximum file size in kilobytes for the log file. When the file exceeds the maximum file size, the system automatically archives the log file and creates a new file to use for logging. The system appends a 0 to the archive file name, and the new file takes on the original file name. This process continues until the system reaches the maximum number of archive files indicated in the `keep` parameter.

keep=*maximum number of archive files*

Indicates the maximum number of archive files to keep during the automatic file rotation, based on the `rotateSize` parameter. When the maximum number of files is exceeded, the system writes over the oldest archive log file with the new log information.

White Space

Indicates what type of white space to place in the log file. Valid values include:

- Space
- Tab

Filter Indicates the log information to be recorded. Valid values include:

Module

Indicates the type of messages to log.Valid values include:

- $NODE_NAME - generic messages
- sql - SQL messages
- mq - message queue messages
- bad_xml - invalid or malformed UMF messages
- msg - UMF exceptions
- logger - logger messages

If you want to include all module types, use an asterisk wild card character. For example:

```
console://stdout *.ERR
```

Severity

Indicates the severity level of the log message. Valid values include:

- CRIT - critical messages
- ERR - error messages
- WARN - warning messages
- NOTE - notices
- INFO - informational messages
- PERF - performance messages
- DBUG - debug messages

If you want to include all severity types, use an asterisk wild card character. For example:

```
console://stdout *.*
```

If you want to exclude one severity from being reported, use the exclamation point. For example:

```
console://stdout mq.!DBUG
```

Configuring custom pipeline logging

IBM InfoSphere Identity Insight provides default pipeline logging configurations that determine how pipelines log errors and messages in both debug mode and service/daemon mode. But you may want to modify the default pipeline logging configuration or create a custom logging configuration to meet your organization's needs. To do so, you must create two logging files that specify the custom logging configuration and then modify the pipeline configuration file to use those custom logging files.

About this task

Pipeline logging is by pipeline node, so you will need to make these changes on each pipeline node. After you create them, you can copy the debug and standard configuration files to each pipeline node. You can also either copy-and-paste the text from the [logging] section from one pipeline configuration file to another, or you can copy the entire pipeline configuration file from one pipeline node to another. Just remember to adjust the connection settings, as appropriate.

Procedure

1. Using any text editor, create two files:
   a. A debug configuration file, used to specify the logging for pipelines operating in debug mode
   b. A standard configuration file, used to specify the logging for pipelines operating in service/daemon mode
2. In each file, use the appropriate pipeline logging components to instruct the system how to log in that mode.
3. Save each file. It is a good idea to save these files to the same directory where the pipeline configuration file is located.
4. In the pipeline configuration file, add a new section named `[logging]`. This is the section where you will specify the names of the two logging configuration files that you created.
5. Under the `[logging]` section heading, add the following two settings:
   a. `DebugConfigFile=`*`debug logging configuration filename`*

b. `ConfigFile=`*`service/daemon logging configuration filename`*

   Note: If you saved the logging configuration files in a directory other than the directory where the pipeline configuration file is located, be sure to indicate the full path to the file.

6. Save the changes to the pipeline configuration file.

What to do next

Before these logging changes take effect, you will need to stop and restart all running pipelines on each affected pipeline node.

Analyst Toolkit web application log files

The web applications rely on IBM WebSphere Liberty to communicate with and connect to IBM InfoSphere Identity Insight. WebSphere Liberty log files include information about the Web services and the Analyst Toolkit applications, as well as WebSphere Liberty errors. If your system is enabled for event processing (using Event Manager), event errors are also logged in the web error log files.

The application server contains two primary log files that can be used to troubleshoot problems:

- Standard output and error streams, which are logged in the file named `console.log`
- Messages captured by the logging components, which are logged in the file named `messages.log`. Messages written to this file contain additional information such as message time stamp and ID of the thread that wrote the message.

These log files are located in the following directory:

```
installation_directory/wlp/usr/servers/iiServer/logs
```

WebSphere Liberty log files are configured by a system administrator on the application server.

Visualizer log files

The Visualizer has two types of log files available to assist users in troubleshooting Visualizer problems or messages: a local log file for each Visualizer client, and log files for the WebSphere Application Server that hosts the Visualizer.

Visualizer client logging

You can configure the Visualizer to log errors, warnings, and informational messages that occur on your local Visualizer client. Each workstation contains one Visualizer client, so you can determine whether or not to log Visualizer messages by workstation.

By default, Visualizer client logging is turned off. You turn Visualizer logging on or off and select logging settings in the **Configure Screen Preferences** window on the **Log Settings** tab.

You determine the directory location of the Visualizer client log file when you turn Visualizer client logging on, by entering the name of the directory or by browsing to an existing directory. The default name of Visualizer client log files is visualizer.log. This file is a text file, that can be viewed using any text editor.

Messages append to the existing log file, until the maximum file size is reached. The maximum size for a Visualizer client log is 1 megabyte.

- If the log file reaches the maximum file size, the system creates another Visualizer client log file in the configured directory location and begins logging messages to that log file.
- Once the second log file reaches the maximum size limit, the system then automatically rotates message logging to the first log file, until it is full.

This automatic log file rotation continues each time the current log file reaches its file size limit. As the system rotates log files, it writes over the previous messages in that log file.

WebSphere Application Service logging

The Visualizer relies on the WebSphere Application Server to communicate with and connect to IBM InfoSphere Identity Insight. Web services events are logged in the application server log files, along with the Configuration Console and Web services events, which also rely on the WebSphere Application Server.

The application server contains two primary log files that can be used to troubleshoot problems:

- System messages, which are logged in the file named `SystemOut.log`
- System error messages, which are logged in the file named `SystemErr.log`

These log files are located in the following directory:

installation_directory/logs/ewas

WebSphere Application Server log files are configured by a system administrator on the application server or through the IBM InfoSphere Identity Insight configuration utility.

Turning Visualizer client logging on

Use these instructions to turn Visualizer client logging on and to configure the settings for Visualizer client logging. If you make changes to Visualizer client logging or settings, you must restart the Visualizer before the changes take effect.

About this task

Visualizer client logging settings are configured for each local Visualizer client. By using these instructions to turn logging on, you only affect the settings for the Visualizer client on this local machine.

Procedure

1. From the **File** menu, select **Preferences**.
2. Select the **Log Settings** tab.
3. Under **Log Settings** in the **Turn on Logging** check box, select the check box so that a check mark displays in the box. (The check box should contain a check mark when logging is turned on.)
4. Select the level of log detail from the **Log detail level** selection box:

a. Select **Errors** to log Visualizer client events that caused error messages. This log level is the default log level when logging it turned on. This log level provides a good balance of performance and logging information.

b. Select **Warnings** to log Visualizer client events that caused warning or error messages.

c. Select **Informational** to log Visualizer client events that caused informational, warning, or error messages.

d. Select **Debugging** to log tracing messages for all Visualizer events. This log level is typically set only when you are troubleshooting a specific Visualizer error, often with IBM Support assistance. The debugging log level may generate a large volume of tracing messages, which are useful for troubleshooting but may adversely affect Visualizer performance for normal operations.

5. In the **Log file directory path** field, enter the full directory path and file name for the Visualizer client log file or browse to an existing directory.
   - Enter the full directory path for the Visualizer client log file
   - Or browse to an existing directory on your local machine to select it as the Visualizer client log directory.
6. Click the **Submit** button to save your changes.
7. Restart your Visualizer by logging out of the Visualizer and then logging back in. Changes to logging settings for your Visualizer client do not take effect until you restart the Visualizer.

Turning Visualizer client logging off

Use these instructions to turn off Visualizer client logging, especially if you turned on debug level logging to troubleshoot a specific Visualizer problem. While log files can help you troubleshoot problems, some levels of logging, like the debug log level, may affect Visualizer performance. If you make changes to Visualizer client logging or settings, you must restart the Visualizer before the changes take effect.

Before you begin

Make sure that you are logged into an active Visualizer session.

About this task

Visualizer client logging settings are configured for each local Visualizer client. By using these instructions to turn logging off, you only affect the settings for the Visualizer client on this local machine.

Procedure

1. From the **File** menu, select **Preferences**.
2. Select the **Log Settings** tab.
3. Under **Log Settings** in the **Turn on Logging** check box, select the check box so that no check mark displays in the box. (The check box should be empty when logging is turned off.) When logging is turned off, logging configuration settings are disabled.
4. Click the **Submit** button to save your changes.
5. Restart your Visualizer by logging out of the Visualizer and then logging back in. Changes to logging settings for your Visualizer client do not take effect until you restart the Visualizer.

Event Manager log files

If your system is enabled to process events using Event Manager, the system creates a log file, which contains program information about events. Error messages from the external event processor are logged to the WebSphere Liberty error log files. Standard pipeline errors encountered during pipeline processing are logged to the pipeline log files, based on the current pipeline logging configuration.

The application server contains the primary log files that can be used to troubleshoot Event Manager messages and problems:

- Event Manager program information, which is logged in the file named `gem_prog_`*`date`*`.log`
- Event Manager error messages, which are logged in the *`installation_directory`*`/logs` directory.

Messages append to the program and data logs by event date. These log files should be reviewed periodically, and then either archived or deleted, according to your organization's policies.

These log files are located in the following directory:

`installation_directory``/logs`

Tracing

Traces are records of component or transaction processing. The information collected from a trace can be used to assess problems and performance. In IBM InfoSphere Identity Insight, traces are part of the debug component logging.

Getting fixes

A product fix might be available to resolve your problem. You can download product fixes by following these steps.

Procedure

1. Determine which fix you need. Go to the *Fixes by version for IBM InfoSphere Identity Insight* document located at http://www-1.ibm.com/support/docview.wss?rs=2216&uid=swg27008307 and click on one of the fixes listed to view more information about all fixes in that particular version. (Fixes are listed in version, release, modification format.)
2. Download the fix. From the fix list, click on the Download information link. In the “Download package” section, click on the “Download Options” link for your environment.
   - If the IBM license agreement screen displays, read the information and click **I Accept** if you accept the agreement and wish to continue downloading the fix.
   - If you click **I Do Not Accept**, the fix will not download.

   At the **File Download** windows, click **Save** and save the fix file locally.
3. Apply the fix. Go to the location where the fix file was saved. Extract or unzip the files from the zipped fix file and follow the instructions in the "readme" document to install the fix.

Learning about fixes and service updates

If you encounter a problem with IBM InfoSphere Identity Insight, first check the list of recommended updates to confirm that your software is at the latest maintenance level. Next, check the list of problems fixed to see if IBM has already published an individual fix to resolve your problem.

Individual fixes are published as often as necessary to resolve defects in the product. In addition, two kinds of cumulative collections of fixes, called fix packs and refresh packs, are published periodically, to bring users up to the latest maintenance level. You should install these update packages as early as possible in order to prevent problems.

To receive weekly notification of fixes and updates, subscribe to My Support e-mail updates.

The following table describes the characteristics of each maintenance delivery vehicle.

Table 29. Characteristics of a fix, a fix pack, and a refresh pack

Name	Characteristics
Fix	• A single fix that is published between updates to resolve a specific problem, for example, PQ79582. • After you install a fix, test any functions that the fixed component has an impact on.
Fix pack	• A cumulative fix package that contains all fixes that have been published since the previous fix pack or refresh pack; a fix pack might also contain new fixes. • Fix packs increment the modification level of the product and are named accordingly, for example, 4.0.2. • A fix pack can update specific components, or it can update the entire product image. • During fix pack installation, all previously applied fixes are automatically uninstalled. • After you install a refresh pack, you should regression-test all critical functions. • The most recent two fix packs are available for download (for example, 4.0.2 and 4.0.1). Earlier fix packs are not available.
Refresh pack	• A cumulative fix package that contains all fixes that have been published since the previous fix pack or refresh pack, as well as new fixes. • A refresh pack typically contains new function, in addition to fixes, and it updates the entire product image. • Refresh packs increment the modification level of the product and are named accordingly, for example, 4.0.2. • During fix pack installation, all previously applied fixes are automatically uninstalled. • After you install a refresh pack, you should regression-test all critical functions.

Service updates

Service updates help you keep your system at the latest software maintenance level.

You can access the latest service updates from the IBM InfoSphere Identity Insight product support page. The URL is

```
https://www-947.ibm.com/support/entry/portal/Overview/Software/
Information_Management/InfoSphere_Identity_Insight
```

To determine the pipeline service level on your system:

1. At a command line on the pipeline node, enter the following command:
   `pipeline`
2. The pipeline version is located on the first line. The number determines the service level.

To determine the Configuration Console service level on your system:

1. Start the Configuration Console.
2. Log in to the Configuration Console.
3. Select **About** from the top menu.
4. Look at the version number listed in the About window. The number determines the service level.

Contacting IBM Software Support

IBM Software Support provides assistance with product defects.

Before you begin

Before contacting IBM Software Support, your company must have an active IBM software maintenance contract, and you must be authorized to submit problems to IBM. For information about the types of maintenance contracts available, see "Enhanced Support" in the *Software Support Handbook* at techsupport.services.ibm.com/guides/services.html

About this task

Complete the following steps to contact IBM Software Support with a problem:

Procedure

1. Define the problem, gather background information, and determine the severity of the problem. For help, see the "Contacting IBM" in the *Software Support Handbook* at techsupport.services.ibm.com/guides/beforecontacting.html
2. Gather diagnostic information.
3. Be prepared to provide the following information in the problem report to assist IBM Software Support:
   - Product name and version
   - Database type and version
   - Operating system name and version
4. Submit your problem to IBM Software Support in one of the following ways:
   - Online: Click **Submit and track problems** on the IBM Software Support site at http://www.ibm.com/software/support/probsub.html

- By phone: For the phone number to call in your country, go to the Contacts page of the IBM Software Support Handbook at techsupport.services.ibm.com/guides/contacts.html

What to do next

If the problem you submit is for a software defect or for missing or inaccurate documentation, IBM Software Support creates an Authorized Program Analysis Report (APAR). The APAR describes the problem in detail. Whenever possible, IBM Software Support provides a workaround that you can implement until the APAR is resolved and a fix is delivered. IBM publishes resolved APARs on the Software Support Web site daily, so that other users who experience the same problem can benefit from the same resolution.

Notices

This information was developed for products and services offered in the U.S.A. IBM InfoSphere Identity Insight Version 9.0.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

IBM World Trade Asia Corporation
Licensing 2-31 Roppongi 3-chome, Minato-ku
Tokyo 106-0032, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement, or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only. This information is for planning purposes only. The information herein is subject to change before the products described become available.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Trademarks

IBM trademarks and certain non-IBM trademarks are marked on their first occurrence in this information with the appropriate symbol.

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

The following terms are trademarks or registered trademarks of other companies:

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks are trademarks of Oracle Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

Index

A

accessing
- Configuration Console 60
- Visualizer 83

accounts (identities) 10

acquisition programs
- description 4, 202
- routing rules 185

activity codes
- configuring 85
- creating for event alerts 87
- creating for role alerts 86
- creating for searches 85
- deleting for event alerts 88
- deleting for role alerts 86
- deleting for searches 86
- editing for event alerts 87
- pre-defined codes for event alerts 87

adding
- comments to alerts 239
- Configuration Console users 61
- criteria to candidate builder configurations 153
- database tables to dictionary 211
- fields to entity database tables 210
- groups of Visualizer users 84
- new data sources 201
- single entity using Visualizer 262
- tables to entity database 209
- Visualizer users 83

address hygiene and standardization
- description 12
- recognize phase 11

address precision
- description 139
- examples 140

addresses
- address precision 139
- hygiene and standardization 12
- kinds of attributes 10

administering 55
- Console 57
- Visualizer 79

administrative tasks for the Configuration Console 55

alert graph
- description 302

Alert Summary window
- configuring default alert display filter options 222
- displaying alerts 236
- filtering the alerts that display 237

alerts 116, 117
- adding comments to alerts 239
- alert graph description, graphing tool 302
- alert indicators in the graphing tool 315
- analyzing in the Visualizer 234
- assigning alerts to other analyst groups 238

alerts *(continued)*
- assigning alerts to yourself 237
- Attribute Alert Generator History report 272
- Attribute Alert Generator report 273
- Attribute Alert report 273
- attribute alerts 20, 235
- changing status of alerts 239
- configuring activity codes for Visualizer dispositioning 85
- configuring Alert Summary display filter default settings 222
- configuring graph options in Visualizer 224
- configuring role alert rules 115
- configuring role alerts system parameter 160
- creating attribute alert generators 248
- criteria for selecting which alerts to analyze 234
- description 19
- Disclosures report 276
- displaying in the Visualizer 236
- editing attribute alert generators 248
- enabling Visualizer users to see all alerts 162
- Event Alert Detail report 277
- event alerts 25, 236
- filtering the display on the Alert Summary window 237
- Role Alert Detail report 285
- role alert invalidation 22
- role alert rules 21
- Role Alert Status report 288
- role alerts 20, 236
- viewing role alert graphs 256
- WS_ALERT format code 336

algorithms
- Name Manager name scoring 104, 145

All Events report
- description 279

alternate name parses
- configuring name data 99
- description 98

Analyst toolkit
- cannot log in 349
- troubleshooting 349

analyzing 254
- alerts in the Visualizer 234
- data 219
- data sources 208
- entity data in the Visualizer, description 219

application monitor
- checking pipeline status 190
- deleting pipeline registrations 182
- description 5
- editing pipeline registrations 181
- registering pipelines 179, 180

application monitor *(continued)*
- routing rules 185
- viewing events 192

application server
- Visualizer log files 367

architecture
- description 2

assigning
- alerts to yourself 237
- event alerts to other analyst groups 238
- role alerts to other analyst groups 238

ATTR_LARGE_DATA 165

ATTR_VALUE 165

Attribute Alert Generator History report
- description 272

Attribute Alert Generator report
- description 273

attribute alert generator system parameter
- configuring 160

attribute alert generators 247
- changing expiration dates 248
- configuring minimum score values 222
- creating 248
- editing 248
- history report 272
- report 273
- updating 248

attribute alerts
- adding comments 239
- assigning to yourself 237
- Attribute Alert Generator History report 272
- Attribute Alert Generator report 273
- Attribute Alert report 273
- changing status 239
- description 20, 235

attribute data
- configuring for UMF 167

Attribute data
- configuring for UMF 167
- description 164
- developing custom scoring plug-ins 170
- overview 165

ATTRIBUTE data segment definitions 165, 168, 169

attribute explorer
- description 310
- description (graphing tool component) 307

Attribute Result report
- description 273

attributes 89, 92
- attribute explorer in the graphing tool, description 307
- attribute icons in the graphing tool 315

attributes *(continued)*
candidate lists 14, 152
customization 164, 165
data to UMF 164
data to UMF 164
description 10
developing custom scoring plug-ins 170
displaying selected properties in the graphing tool 310
finding entities by attribute 244
finding entities with similar attributes 338
identities 10
large data
storing 164, 165
automatic updates
receiving 358

B

BIRT Report Viewer
exporting a Configuration Console report into other applications 77
exporting Configuration Console reports 77
exporting report data from Configuration Console reports 78
business names
categorizing by type, description 102

C

callouts
data to UMF 174
candidate builder
customizing 151
description 151
candidate builder configurations
adding criteria 153
creating 152
deleting 153
description 151
candidate builders
configuring data sources to use a specific candidate builder 97, 128
candidate list building
benefits of enhanced name hashing 95
candidate lists
candidate thresholds 139
description 14, 152
resolve phase 14
candidate thresholds 139
resolution rules 15, 138
categorizing
names by type personal or business, description 102
Centrifuge
setting the default path in Visualizer 163, 221
CEP
creating a basic COUNT event rule 39
creating a basic SUM event rule 37
creating a new project 33
CEP *(continued)*
defining complex event rules 34
description 27
exporting a new cep.xml file 34
importing the cep.xml file 33
installing the event rule authoring tool 28
starting the Rule Author tool 29
terms 29
cep.xml file
importing to define event rules 33
changing
alert display filter settings on the Alert Summary window 237
characteristic confirmations and denials
viewing 155
characteristic types 89, 90
configuring 89
creating 90
deleting 91
characteristics 89, 90
creating characteristic confirmations and denials 155
creating characteristic types 90
deleting characteristic confirmations and denials 156
deleting characteristic types 91
kinds of attributes 10
client authentication 52
cloning
resolution configurations 137
codes
lookup codes, description 108
Cognos
deploying reports 299
installing 298
modifying database configuration 300
verifying report deployment 300
Cognos Reports 8
command line interfaces
checking pipeline status 191
description 8
pwdmgr command 62
queue utility 204
UMF formatting utility 207
commands
starting pipelines 177
starting Web services pipelines 325
stopping pipelines 178
wsutil.jar 329
comments
adding to alerts 239
sending vii
components
pipeline logging components 364
concepts
core product 9
concurrency system parameter
configuring 161
configuration
customizing the graphing tool graph icons 319
optimal Visualizer browser settings 81
requirements for customizing graph icons 320
configuration *(continued)*
viewing Configuration Console settings 73
viewing system configuration settings 73
Configuration Console 7, 57
adding users 61
application monitor 5
creating groups of Visualizer users 84
creating Visualizer users 83
deactivating a Visualizer user 84
deleting users 62
log files 367
logging in 59
logging out 59
managing access 60, 62
managing access using database login information 60
managing access using password manager utility 60
pipeline status and statistics 188
registering pipelines 179
resetting passwords 62
resetting Visualizer user passwords 84
resolution configurations 136
running reports 66
viewing application monitor events 192
viewing users and their status 61
Web browser settings 58
configuration files
queue utility 203
Configuration report
description 73
running 73
configuration settings
deploying fixes 64
updating 64, 89
configuration tasks 89
configuring 55, 89, 116
activity codes 85
advanced pipeline logging 366
alert display filter options 222
attribute alert generator system parameter 160
categorizing names using Name Manager 102
characteristic types 89
concurrency system parameter 161
configuring Java v1.6 for Windows workstations 232
confirmation and denial system parameter 159
confirmations and denials 154
creating a composite name hash 97
data mappings 211
data quality management system parameter 161
data sources 126, 127
data sources to use enhanced name hashing 97, 128
data sources, Name Manager match level 128
database system parameter 158

configuring *(continued)*
  default path for Centrifuge in Visualizer 163, 221
  default path for UMF files in Visualizer 163, 221
  direct launch approach to open the Visualizer 231
  display options in Visualizer 220
  DQM function 255 for IBM Global Name Recognition Name Hasher 96
  DQM rules 105
  entity database 89
  entity model 208
  entity resolution 136
  event business rules 31
  Event Manager 25
  Event Manager system parameters 162
  event rules, importing the cep.xml file 33
  generic data values 112
  generic thresholds 112
  graph options in Visualizer 224
  hyperlink browser options in the Visualizer 224
  IBM Global Name Recognition Name Hasher, disabling DQM 252 rule 96
  Internet Explorer to open the Visualizer 230
  Java v1.6 for Windows workstations 232
  Java Web Start 230, 231
  location of Name Manager support libraries 103, 158
  log options in Visualizer 223
  logs system parameter 159
  minimum score values for search entities 222
  Mozilla Firefox to open the Visualizer 231
  name cultures for Name Manager 105
  name data to create alternate name parses 99
  name data, description 95
  Name Manager system parameter 103, 158
  name scoring system parameter 157
  names to assign culture 102
  number types 92
  output documents 126
  personal and organizational name categorization 103, 158
  personal and organizational names 102
  pipeline logging configurations 363
  pipelines 178
  pipelines configuration check 175
  product options system parameter 161
  resolution configurations 136
  resolution rules 138
  resolution rules, Name Manager name score confirmation and denial thresholds 104
  role alert rules 115

configuring *(continued)*
  role alerts system parameter 160
  roles 113
  routing rules 183
  system parameters 157
  system parameters for enhanced name hashing 96
  system parameters for the Name Hasher 96
  the Visualizer 219
  UMF documents 125
  URL syntax and parameters for the graphing component 318
  Visualizer 220
  Visualizer logging 369
  Visualizer logging settings 368
  Visualizer system parameter 162
configuring role alert rules 116
confirm/deny thresholds
  resolution rules 15, 138
confirmation and denial system parameter
  configuring 159
confirmations and denials
  configuring 154
  creating characteristic confirmations and denials 155
  deleting characteristic confirmations and denials 156
  description 154
  resolution rules 15, 138
  viewing characteristic confirmations and denials 155
conflicts
  role alert invalidation 22
contacting
  IBM software support viii, 372
converting
  data to UMF 202
  format of UMF files 207
creating 106, 109
  activity codes for event alerts 87
  activity codes for role alerts 86
  activity codes for searches 85
  attribute alert generators 248
  candidate builder configurations 152
  characteristic confirmations and denials 155
  characteristic types 90
  Configuration Console users 61
  data mappings 212
  data source locations 128
  entity types 121
  event types 134
  groups of Visualizer users 84
  number types 93
  resolution rules 147
  roles 114
  Visualizer users 83
culture
  categorizing personal names to assign culture 102
customizing
  resolution configurations 137

D

daemons
  default UNIX daemon mode logging 363
data
  loading from UMF files in the Visualizer 263
data mappings
  creating 212
  defining 211
  deleting 213
  description 211
  mapping data to UMF 209
  viewing 212
data quality 13
  recognize phase 11
data quality management
  description 11
data quality management system parameter
  configuring 161
Data Source Summary report
  description 67, 275
data sources
  adding 201
  adding tables to entity database 209
  analyzing 208
  configuring 126, 127
  configuring default path in Visualizer 163, 221
  configuring Name Manager match level 128
  configuring to use enhanced name hashing 97, 128
  converting to UMF 202
  creating data source locations 128
  deleting 128
  description 6, 126
  determining the quality of data within data sources 67, 275
  finding entities by data source account 245
  viewing 127
  viewing the Data Source Summary report 67, 275
  viewing the Load Summary report 68, 280
database login information
  managing access to Configuration Console 60
database system parameter
  configuring 158
databases
  configuring 52
  creating 52
  setting up 49
date of birth precision
  description 146
  examples 146
dates of birth
  date of birth precision 146
DB2
  client authentication, configuring 52
deactivating
  Visualizer users 84
debugging
  default debug logging 363

debugging *(continued)*
- log files 361

Default w/ Name Only
- setting the required candidate builder by data source for the Name Hasher 97, 128

degrees of separation
- creating a new separation configuration 124
- editing separation configurations 125
- example 123
- impersonal awareness 18, 120
- viewing separation configurations 124

Degrees of Separation
- overview 17, 122

deleting 106, 109, 117
- activity codes for event alerts 88
- activity codes for role alerts 86
- activity codes for searches 86
- candidate builder configurations 153
- characteristic confirmations and denials 156
- characteristic types 91
- Configuration Console users 62
- data mappings 213
- data sources 128
- entity types 121
- event types 135
- generic thresholds 113
- number types 93
- pipeline registrations 182
- resolution configurations 137
- resolution rules 148
- roles 115
- routing rules 188

deleting role alert rules 117

description 89, 93, 119

developing
- Web queries 323
- Web services 323

dictionary
- adding database tables 211

disclosing
- relationships between entities 265

Disclosures report
- description 276

display
- configuring alert display filter options for the Visualizer 222
- configuring Visualizer display options 220
- filtering alerts that display on the Alert Summary window 237

downloading
- fixes and service updates 370

DQM functions
- 258, dynamically assigning gender on names 101
- configuring DQM function 255 for IBM Global Name Recognition Name Hasher 96
- configuring NAME segment to assign culture using DQM function 260 102

DQM functions *(continued)*
- disabling DQM 252 rule for IBM Global Name Recognition Name Hasher 96
- enabling DQM function 610 for IBM Global Name Recognition Name Hasher 97

DQM rules 106
- configuring 105
- data quality 13
- data quality management 11
- description 105
- turning off 107
- validating 107
- viewing 106

E

e-mails
- kinds of attributes 10

editing
- activity codes for event alerts 87
- attribute alert generators 248
- event types 134
- pipeline registrations 181

enabling
- categorizing names by type 103, 158
- IBM Global Name Recognition Name Hasher 95

Enabling
- Event Manager 27

entities 254
- adding data in the Visualizer, description 262
- adding through the Visualizer 262
- alert graph description, graphing tool 302
- attribute alerts 20, 235
- description 10, 254
- disclosed relationships 265
- displaying selected properties in the graphing tool 310
- entity database 7
- entity graph description, graphing tool 303
- entity icons in the graphing tool 315
- entity resumes 254
- event alerts 25, 236
- finding by attribute 244
- finding by data source account 245
- finding by Entity ID 246
- finding by resolution 246
- finding in the Visualizer 244
- finding multiple entities that share the same unique number 357
- identities 10
- Large Entities query 355
- linking from the graphing tool to the entity resume 321
- pipelines process only part of incoming record 347
- printing 255
- related entities indicator in the graphing tool 315
- role alert rules 21
- role alerts 20, 236
- roles 19, 113

entities *(continued)*
- social network graph description, graphing tool 305
- SQL query to find the total number of unique number by Entity ID 356
- SQL query to find very large entities 355
- Total Unique Number by Entity query 356
- Unique Number Shared by Multiple Entities query 357
- using the Visualizer to analyze entity data description 219
- validating UMF files in the Visualizer 264
- viewing entity resumes 254
- WS_DETAIL format code 335

entity database
- adding a new data source 201
- adding fields to tables 210
- adding tables 209
- adding tables to dictionary 211
- building queries 330
- configuring data sources 127
- configuring for the system 89
- creating 52
- creating data mappings 212
- data mappings 211
- deleting data sources 128
- description 7
- finding entities 244
- generic data values 112
- other databases 7
- querying 332, 338
- risks in modifying 209
- searching for entities by attribute 244
- searching for entities by data source account 245
- searching for entities by Entity ID 246
- searching for entities byresolution 246

entity graph 17, 122
- description 303

entity graphs
- viewing in Visualizer 255

Entity ID
- finding entities by Entity ID 246
- finding entities by resolution 246

entity model
- extending 208

entity resolution 4, 175
- adding criteria to candidate builder configurations 153
- address precision 139
- address precision examples 140
- candidate lists 14, 152
- configuring 136
- configuring Name Manager name score confirmation and denial thresholds 104
- confirmations and denials 154
- date of birth precision 146
- date of birth precision examples 146
- description 11, 136
- developing custom scoring plug-ins 170

entity resolution *(continued)*
name precision 143
re-resolve process 15
recognize phase 11
relate phase 16
relationship scores 23
relationships 17
resolution configurations 136
resolution rules 15, 138
resolution scores 23
resolve phase 14
scoring 23
turning off relationship detection 132
unresolve process 16
entity resumes 254
copying to another application 255
printing 255
viewing 254
entity types 119, 121
creating 121
deleting 121
EntitySearcher 8
environment variables 49, 50
setting 49
errors
Configuration Console log files 367
Event Manager log files 370
pipeline log files 361
queue log files 361
SQL log files 361
UMF log files 361
UMF parsing errors 360
ETL tools
vs. acquisition programs 4, 202
Event Alert Detail report
description 277
event alerts
adding comments 239
assigning to other analyst groups 238
assigning to yourself 237
changing status 239
creating activity codes 87
deleting activity codes 88
description 25, 236
editing activity codes 87
pre-defined activity codes 87
transferring to other analyst groups 238
Event Manager
configuring 25
configuring event business rules 31
configuring event types 133
configuring the CEP URI connection in the Configuration Console 27
creating a basic COUNT event rule 39
creating a basic SUM event rule 37
creating a CEP project 33
creating event types 134
defining complex event rules 34
deleting event types 135
description 24
editing event types 134
enabling in the Configuration Console 27
event alerts 25, 236
Event Manager *(continued)*
event business rules description 25
exporting a new cep.xml file 34
importing the cep.xml file 33
installing the Rules Author tool 28
integrating CEP with Event Manager 27
log files 370
starting the Rule Author tool 29
event types
configuring 133
creating 134
deleting 135
editing 134
events
All Events report 279
configuring event business rules 31
configuring event types 133
configuring system parameters 162
configuring the CEP URI connection 27
creating a basic COUNT event rule 39
creating a CEP project 33
creating a SUM event rule 37
creating event types 134
defining complex event rules 34
deleting event types 135
description 25
description of event processing 24
editing event types 134
enabling Event Manager 27
Event Alert Detail report 277
event alerts 25, 236
event business rules description 25
event type definition 25, 133
exporting a new cep.xml file 34
importing the cep.xml file 33
installing the Rule Author tool 28
integrating CEP with Event Manager 27
starting the Rule Author tool 29
examples 89, 93
address precision 140
alerts 19
building a UMF_QUERY query 332
building a UMF_SEARCH query 338
confirmations and denials 154
data mappings 211
data quality 13
date of birth precision 146
degrees of separation 123
DQM rules 11
generic data values 112
impersonal awareness 18, 120
relationships 17
roles 19, 113
SNMP agents 188
UMF_QUERY input document 333
UMF_SEARCH input documents 340
unresolve 16
Web services alert queries, WS_ALERT 336
Web services alert queries, WS_RELATION 337
Web services detail entity queries, WS_DETAIL 335
examples *(continued)*
WS_SUMMARY_TOP10 342
wsutil.jar commands 329
exceptions
viewing UMF exceptions 193
expiration dates
changing for attribute alert generators 248
Explorer 8
exporting
cep.xml file 34
Configuration Console reports into other applications 77
data from Configuration Console reports into spreadsheet applications 78

F

file types
.bad file 361
.cnt file 361
.log file 361
.MQErr.log file 361
.msg file 361
.SqlDebug.log file 361
.SqlErr.log file 361
Visualizer.log 367
files
adding data in the Visualizer, description 262
configuration file for queue utility 203
configuring default path for UMF files in Visualizer 163, 221
console.log files 367
Event Manager log files 370
formatting UMF 207
gem_prog_date.log files 370
messages.log files 367
validating UMF files in the Visualizer 264
filtering
alerts that display on the Alert Summary window 237
filters
configuring Alert Summary display default settings 222
routing rules 185
find
entities by attribute 244
entities by data source account 245
entities by Entity ID 246
entities by resolution 246
Find by Resolution
configuring minimum score values 222
finding
multiple entities that share the same unique number 357
the total number of unique numbers associated with a single entity 356
very large entities 355
fixes
description 371
downloading 370

- format codes
  - WS_ALERT 336
  - WS_DETAIL 335
  - WS_RELATION 337
  - WS_SUMMARY 342
  - WS_SUMMARY_TOP10 342
  - WS_SUMMARY_TOP100 342

G

- gender
  - assigning gender dynamically for names 101
  - assigning to names, description 100
- generating 271
- generic data values
  - configuring 112
  - configuring generic thresholds 112
  - deleting generic thresholds 113
  - description 112
  - viewing 112
- generic thresholds
  - configuring 112
  - deleting 113
- Graph 8
- graphing component
  - URL syntax and parameters 318
- graphing tool
  - alert graph, description 302
  - alert indicators 315
  - attribute explorer, description 307
  - common graph elements 315
  - description 301
  - entity graph, description 303
  - icons 315
  - line indicators 315
  - linking to the entity resume 321
  - navigating the graphs 310
  - related entity indicators 315
  - selected properties, description 310
  - social network graph, description 305
- graphs
  - alert graph description, graphing tool 302
  - alert indicators in the graphing tool 315
  - attribute explorer in the graphing tool, description 307
  - common elements in the graphing tool 315
  - configuring graph options in Visualizer 224
  - customizing the graphing tool graph icons 319
  - customizing Visualizer graph icons 256
  - description of the graphing tool 301
  - entity graph description, graphing tool 303
  - icons in the graphing tool 315
  - lines in the graphing tool 315
  - linking from the graphing tool to the entity resume 321
  - navigating the graphing tool graphs 310

- graphs *(continued)*
  - related entity indicators in the graphing tool 315
  - requirements for customizing graph icons 320
  - selected properties in the graphing tool, description 310
  - social network graph description, graphing tool 305
  - URL syntax and parameters for the graphing component 318
  - viewing entity graphs in Visualizer 255
  - viewing role alert graphs 256
- groups of users 46, 57, 80

H

- hashes
  - creating a composite name hash 97
- hashing
  - benefits of enhanced name hashing 95
- hyperlinks
  - selecting browser to open 224

I

- i2 Notebook plug-in 8
- IBM Degrees of Separation
  - impersonal awareness 18, 120
- IBM Global Name Recognition Name Hasher 96
  - configuring DQM function 255 UMF Exclude 96
  - creating a composite name hash 97
  - description 95
  - disabling DQM 252 rule 96
  - enabling 95
- IBM Global Recognition Name Hasher
  - migrating to V8 FP2 from a previous version 97
- IBM InfoSphere Identity Insight
  - description 1
- IBM software support
  - contacting viii, 372
- icons
  - attribute icons in the graphing tool 315
  - customizing the graphing tool graph icons 319
  - customizing Visualizer graph icons 256
  - entity icons in the graphing tool 315
  - requirements for customizing graph icons 320
- identities
  - description 10
  - entities 10, 254
  - entity database 7
  - pipelines process only part of incoming record 347
  - roles 19, 113
  - viewing new identities 194
- impersonal awareness 119
  - description 18, 120

- installation
  - Event Manager, Rules Author tool 28
  - event rule authoring tool 28
- interfaces
  - command line 8
  - user interfaces 7
- Internet Explorer
  - setting to use the required client Java Web Start version 230

J

- Java
  - configuring Java v1.6 for Windows workstations 232
  - configuring Java Web Start 230, 231
  - direct launch approach to open the Visualizer 231
- Java Web Start
  - direct launch approach to open the Visualizer 231
  - setting Web browser to use the required client version Java Web Start 230

K

- knowledge bases
  - finding known product problems and work arounds 358
  - optimizing search results 358
  - searching 358

L

- Launching
  - Visualizer, setting Web browser to use required Java Web Start 230
- lines
  - dashed lines in the graphing tool 315
  - thick lines in the graphing tool 315
- load
  - SRDWebService method 327
- Load Summary report
  - description 68, 280
- loading 201
- loading data
  - data mappings 211
  - from UMF files in the Visualizer 263
- locations
  - creating data source locations 128
- locking
  - Visualizer 233
- log files
  - .bad file 361
  - .cnt file 361
  - .log file 361
  - .MQErr.log file 361
  - .msg file 361
  - .SqlDebug.log file 361
  - .SqlErr.log file 361
  - Configuration Console 367
  - Event Manager 370
  - pipelines 361
  - Visualizer 367
  - Visualizer.log 367

log writers 364
logging
  Configuration Console 367
  configuring custom pipeline 366
  default debug logging 363
  default pipeline logging configurations 363
  default service/daemon logging 363
  Event Manager 370
  pipeline logging components 364
  turning off Visualizer logging 369
  Visualizer log files 367
logging in
  Configuration Console 59
  Visualizer 82, 229
Logging in
  Visualizer, setting Web browser to use required Java Web Start 230
logging out
  Configuration Console 59
  Visualizer 83, 232
logs
  configuring logging options in Visualizer 223
  definition 360
  turning on Visualizer logging 368
logs system parameter
  configuring 159
lookup codes 109
  description 108
  turning off 109
  viewing 108

M

managing access
  Configuration Console using database login information 60
  Configuration Console using password manager utility 60
mapping data
  using data mappings 209
message identifiers
  description 359
messages
  description 359
Microsoft Message Queuing queues
  log files 361
Microsoft Windows
  default service logging 363
migrating
  NameHasher to V8 FP2 97
minimum score thresholds
  configuring for Visualizer search entities 222
Mozilla Firefox
  setting to use the required client Java Web Start version 231

N

Name Comparator versions
  comparing 143
  Name Comparator 1.0 143
  Name Comparator 2.0 144
Name Hasher
  configuring candidate builders for enhanced name hashing 97
  configuring DQM function 255 UMF Exclude 96
  configuring system parameters for enhanced name hashing 96
  creating a composite name hash 97
  description 95
  disabling DQM 252 rule 96
  migrating to V8 FP2 from a previous version 97
name hygiene and standardization
  description 12
  recognize phase 11
Name Manager
  categorizing names by type, description 102
  configuring match level 128
  configuring name scoring confirmation and denial thresholds 104
  configuring system parameters 103, 158
  configuring to categorize names 102
  description 102
  name scoring description 104, 145
name matching
  enabling name cultures for Name Manager 105
name matching algorithms
  Name Comparator 1.0 143, 144
name scoring
  configuring Name Manager confirmation and denial thresholds 104
  Name Manager algorithm 104, 145
name scoring algorithms
  configuring NC1 or NC2 157
name scoring system parameters
  configuring 157
Name Sifter
  categorizing names by type, description 102
names
  alternate name parses, description 98
  assigning gender, description 100
  benefits of enhanced name hashing 95
  categorizing by type, description 102
  categorizing personal names to assign culture 102
  comparing with Name Comparator 1.0 143
  comparing with Name Comparator 2.0 144
  configuring name data, description 95
  configuring Name Manager to categorize names 102
  configuring system parameters for enhanced name hashing 96
  configuring system parameters for the Name Hasher 96
  configuring to create alternate name parses 99
  creating a composite name hash 97
names *(continued)*
  hygiene and standardization 12
  kinds of attributes 10
  migrating to NameHasher V8 FP2 97
  name precision 143
  name scoring, Name Manager algorithm 104, 145
  selecting name cultures for Name Manager 105
  turning on gender 101
nodes
  icons that represent nodes in the graphing tool 315
number types 93
  configuring 92
  creating 93
  deleting 93
  viewing 93
numbers 92, 93
  configuring number types 92
  creating number types 93
  deleting number types 93
  description 92
  finding multiple entities that share the same unique number 357
  finding the total number of unique numbers associated with a single entity 356
  kinds of attributes 10
  pipelines do not load scientific or floating decimal point numbers 347
  viewing number types 93

O

opening
  Visualizer 229
Oracle
  cache statement, sizing 53
  client authentication, configuring 52
  CREATE VIEW privileges 51
output documents
  configuring 126

P

parallel pipeline processing 4, 175
parameter groups
  configuring system parameters 157
parameters
  URL syntax and parameters for the graphing component 318
password authentication
  locking the Visualizer 233
password manager
  command syntax 62
passwords
  changing for Visualizer 233
  resetting Configuration Console passwords 62
  resetting Visualizer user passwords 84
performance
  system performance slow 355
  tables that affect pipeline performance 354

performance *(continued)*
- tables that affect Visualizer performance 354
- tips for good system health 353

performance considerations
- candidate builder configurations 151

persistent searches
- creating 248
- editing 248

persistent searches (attribute alert generators) 247

personal names
- categorizing by type, description 102

pipeline
- deployments 46
- parallel processing threads 46

pipeline nodes 4, 5, 175, 176

pipeline searches
- description 330
- finding a specific entity 332
- finding entities with similar attributes 338
- UMF_QUERY 333
- UMF_SEARCH 340
- WS_ALERT queries 336
- WS_DETAIL queries 335
- WS_RELATION queries 337
- WS_SUMMARY 342
- WS_SUMMARY_TOP10 342
- WS_SUMMARY_TOP100 342

pipelines 4, 5, 175, 176
- address hygiene and standardization 12
- application monitor 5
- cannot see pipeline status 347
- cannot start on AIX 347
- checking status 190
- checking status using the pipeline command 191
- configuration check 175
- configuring 178
- configuring advanced logging 366
- configuring concurrency parameters 161
- configuring routing rules 183
- data quality 13
- default logging configurations 363
- deleting registrations 182
- deleting routing rules 188
- do not load scientific or floating decimal point numbers 347
- do not reflect configuration changes 347, 349
- editing registrations 181
- entity resolution 11, 136
- log files 361
- managing 175
- name hygiene and standardization 12
- pipeline logging components 364
- process only part of an incoming record 347
- recognize phase 11
- registering 179, 180
- relate phase 16
- resolution scores 23
- resolve phase 14

pipelines *(continued)*
- routing rules 185
- shut down 347
- SNMP agents 188
- starting 177
- starting Web services pipelines 325
- status and statistics 188
- status of Down 347
- stopping 178
- tables that affect pipeline performancestatistics 354
- transports 6
- troubleshooting checklist 347
- viewing application monitor events 192
- viewing data source statistics 67, 275
- viewing quality characteristics of data by data load 68, 280
- viewing registration details 180
- viewing UMF exceptions 193
- warning message indicates "no routes defined" 347
- Web services queries 330

precision
- addresses 139
- date of birth 146
- Name Comparator 1.0 143
- Name Comparator 2.0 144
- names 143

prerequisite information vii

printing 271
- current window in Visualizer 255
- entity resumes 255

problems
- Visualizer, checklist 349

problems and work arounds
- searching knowledge bases 358

problems and workarounds
- describing problems 345

process
- SRDWebService method 327

product architecture 5, 176
- description 2

product options system parameter
- configuring 161

protected user
- creating 46

pwdmgr command
- adding users to the Configuration Console 61
- command syntax 62
- deleting users to the Configuration Console 62
- managing access to Configuration Console 60
- resetting passwords 62
- viewing users of the Configuration Console 61

Q

QS-AVI
- overview 215
- requirements 215
- task overview 215
- troubleshooting 216

quality
- determining the quality of data within data sources 67, 275
- viewing quality characteristics of data by data loads 68, 280

QualityStage Address Verification interface
- overview 215
- requirements 215
- task overview 215
- troubleshooting 216

queries
- developing for your Web services environment 323
- finding a specific entity 332
- finding entities with similar attributes 338
- methods for building, description 330
- UMF_QUERY input document 333
- UMF_SEARCH input documents 340
- Web services 330

queue utility
- command syntax 204
- configuration file 203
- description 202
- transferring files 202

QUtil (queue utility) 202

R

re-resolve
- description 15

recognize phase 11

registering
- pipelines 180

relate phase 16

related information vii

relationship chain 17, 122

relationship detection
- description 17
- relate phase 16
- relationship scores 23
- turning off 132

relationship resolution
- creating a new separation configuration 124
- editing separation configurations 125
- role alert invalidation 22
- viewing separation configurations 124

relationship scores
- description 23

relationships 116, 117
- creating a new separation configuration 124
- description 17
- disclosing between entities 265
- Disclosures report 276
- editing separation configurations 125
- entity graph description, graphing tool 303
- impersonal awareness 18, 120
- related entities indicator in the graphing tool 315
- role alert rules 21
- role alerts 20, 236

relationships *(continued)*
- social network graph description, graphing tool 305
- turning off relationship detection 132
- viewing separation configurations 124
- WS_RELATION format code 337

removing
- Visualizer users 84

reports 271
- All Events report 279
- Attribute Alert Generator History report 272
- Attribute Alert Generator report 273
- Attribute Alert report 273
- Configuration Console 66
- Configuration report, ATTRIBUTE data segment definitions 169
- Disclosures report 276
- Event Alert Detail report 277
- exporting a Configuration Console report 77
- exporting Configuration Console reports 77
- exporting data from a Configuration Console report 78
- Role Alert Detail report 285
- Role Alert Status report 288
- running the Configuration report 73
- viewing statistical reports 66
- viewing the Configuration report 73
- viewing the Data Source Summary report 67, 275
- viewing the Load Summary report 68, 280
- Visualizer 271

requirements
- Web services 324

resetting
- Configuration Console user passwords 62
- Visualizer user passwords 84

resolution configurations
- cloning and customizing 137
- configuring 136
- deleting 137
- description 136
- viewing 137

resolution rules
- candidate thresholds 139
- configuring 138
- configuring Name Manager name score confirmation and denial thresholds 104
- creating 147
- deleting 148
- description 15, 138
- viewing 147

resolution scores
- configuring confirmations and denials 154
- description 23
- resolution rules 15, 138

resolve phase 14

reviewing
- default UMF specification 209

role alert chain 17, 122

Role Alert Detail report
- description 285

role alert rules 116, 117
- configuring 115
- description 21
- role alerts 115
- viewing 116

Role Alert Status report
- description 288

role alerts
- adding comments 239
- assigning to other analyst groups 238
- assigning to yourself 237
- changing status 239
- configuring system parameter 160
- creating activity codes 86
- deleting activity codes 86
- description 20, 115, 236
- Role Alert Detail report 285
- role alert invalidation 22
- Role Alert Status report 288
- transferring to other analyst groups 238
- viewing role alert graphs 256

role alerts system parameter
- configuring 160

roles
- configuring 113
- creating 114
- deleting 115
- description 19, 113
- role alert rules 21
- viewing 114

roles and responsibilities 46, 57, 80

routing
- deleting pipeline registrations 182
- editing pipeline registrations 181
- registering pipelines 179, 180

routing process 185

routing rules
- configuring 183
- deleting 188
- description 185

Rule Author tool
- starting 29

rules
- configuring event business rules 31
- creating a basic COUNT event rule in CEP 39
- creating a basic SUM event rule in CEP 37
- event business rules description 25

S

Sample Cognos report
- Entity Resume 297
- role alert 296

score
- SRDWebService method 327

scoring
- customization 164
- description 23
- plug-in
  - user-created 164

scoring *(continued)*
- plugin
  - user-created 164
- relationship scores 23
- resolution scores 23

scoring plug-ins
- developing 170

Scoring plug-ins
- configuring 169

scoring processes
- address precision 139
- date of birth precision 146
- name precision 143

search
- EntitySearcher 294
- SRDWebService method 327
- thin client 294

searches
- configuring minimum score values for search entities 222
- creating activity codes 85
- deleting activity codes 86
- finding a specific entity 332
- Web services 330

searching
- entity database 244
- resources and tools 358

security
- changing your Visualizer password 233

sending comments vii

separation configurations
- creating a new separation configuration 124
- editing separation configurations 125
- viewing settings for degrees of separation 124

service updates
- description 371
- downloading 370
- overview 372

services
- default Microsoft Windows service mode logging 363

settings 55, 89
- configuring alert filter display options 222
- configuring candidate builders by data source 97, 128
- configuring Internet Explorer to open the Visualizer 230
- configuring Java v1.6 for Windows workstations 232
- configuring Java Web Start 230, 231
- configuring Mozilla Firefox to open the Visualizer 231
- configuring name data to create alternate name parses 99
- configuring name data, description 95
- configuring Name Manager match level 128
- customizing the graphing tool graph icons 319
- direct launch approach to open the Visualizer 231

settings *(continued)*
hyperlink browser options in the Visualizer 224
optimal Visualizer browser settings 81
requirements for customizing graph icons 320
selecting name cultures for Name Manager 105
turning off Visualizer logging 369
turning on Visualizer logging 368
URL syntax and parameters for the graphing component 318
viewing Configuration Console settings 73
viewing system configuration settings 73
SNMP agents
description 188
starting 189
stopping 190
social network graph
description 305
software requirements
Web services 324
source locations
data sources 6, 126
source systems
data sources 6, 126
SQL
.SqlDebug.log 361
.SqlErr.log 361
log files 361
SQL queries
Total Unique Number by Entity query 355, 356
Unique Number Shared by Multiple Entities query 357
srd.wsdl file
description 9, 323
SRDWebService
load method 327
process method 327
score method 327
search method 327
starting
pipelines 177
SNMP agents 189
Visualizer 229
Web services pipelines 325
statistics
tables that affect pipeline performance 354
tables that affect Visualizer performance 354
viewing data source statistics 67, 275
viewing quality characteristics of data by data loads 68, 280
viewing statistical Configuration Console reports 66
viewing the Load Summary report 68, 280
status and statistics
deleting pipeline registrations 182
editing pipeline registrations 181
registering pipelines 179, 180
stopping
pipelines 178
SNMP agents 190
support
contacting viii, 372
searching knowledge bases 358
system architecture
defining 46
description 2
system health
Large Entities query 355
tips 353
Total Unique Number by Entity query 356
Unique Number Shared by Multiple Entities query 357
system parameters
attribute alert generator 160
concurrency default 161
configuring 157
confirmation and denial 159
data quality management 161
database 158
Event Manager 162
logs 159
Name Manager 103, 158
name scoring 157
product options 161
role alerts 160
Visualizer 162
system requirements
64-bit Linux, System z 44
details 41
IBM AIX 41
Linux Power Systems 43
Linux System x 42
Microsoft Windows Server (64-bit) 45
system requirements and planning
details 41
system-created 89
system-created characteristic types 89

T

technical resources
finding 358
testing
Web services 326
tooks
entity graph description, graphing tool 303
tools
Centrifuge default path 163, 221
queue utility 202
searching knowledge bases 358
social network graph description, graphing tool 305
support tools 358
UMF formatting utility 207
tracing
description 370
log files 361
transferring
event alerts to other analyst groups 238
role alerts to other analyst groups 238
transports
description 6
troubleshooting 347
troubleshooting
cannot log in to the Analyst toolkit 349
cannot see pipeline status 347
cannot start AIX pipelines 347
description 345
downloading fixes 370
finding multiple entities that share the same unique numbery 357
fixes and service updates 371
general checklist 347
Large Entities query 355
logging 360
pipeline processes only part of an incoming record 347
pipeline shuts down 347
pipeline transport does not work 347
pipelines do not honor configuration changes 347
searching knowledge bases 358
service updates 372
system performance slow 355
tips for good system health 353
Total Unique Numbers by Entity query 356
tracing 370
Unique Number Shared by Multiple Entities query 357
Visualizer, checklist 349
Troubleshooting
Visualizer, direct launch approach 231
Visualizer, error message when starting on Windows workstations 232
Visualizer, setting Internet Explorer use the required client Java Web Start version 230
Visualizer, setting Mozilla Firefox use the required client Java Web Start version 231
Visualizer, setting Web browser to use the required Java Web Start 230
troubleshooting checklist
Analyst toolkit 349
pipelines 347

U

UMF
configuring default path for UMF files in Visualizer 163, 221
converting data 164, 167, 174, 202
converting the format of UMF files 207
creating data mappings 212
description 4, 208
loading data in the Visualizer 263
parsing errors 360
reviewing the default specification 209
tall format 207
transforming to using acquisition programs 4, 202

UMF *(continued)*
- validating files in the Visualizer 264
- wide format 207

UMF data
- transfer to queues 202

UMF document types
- routing rules 185

UMF documents
- configuring 125
- description 4, 208

UMF exceptions
- viewing 193

UMF files
- adding data in the Visualizer, description 262

UMF formatting utility
- command syntax 207
- description 207

UMF input documents
- UMF_QUERY 333
- UMF_SEARCH 340
- viewing 125

UMF records
- description 4, 208

UMF segments
- ATTRIBUTE data segment definitions 165, 168, 169
- data mappings 211
- defining data mappings 211
- description 4, 208
- mapping to entity database 209

UMF_QUERY input document
- building Web services pipeline searches 332

UMF_SEARCH input document
- building Web services pipeline searches 338

Universal Message Format (UMF) 4, 208

UNIX
- default daemon mode logging 363

unresolve
- description 16

updates
- receiving automatically 358

updating
- attribute alert generators 248
- pipeline configuration 178

user accounts
- Configuration Console 59

user interfaces 7, 8, 57
- description 7
- Visualizer 79, 219

user roles 46, 57, 80

users
- adding Configuration Console users 61
- changing your Visualizer password 233
- creating groups of Visualizer users 84
- creating Visualizer users 83
- deactivating a Visualizer user 84
- deleting Configuration Console users 62
- resetting Configuration Console passwords 62

users *(continued)*
- resetting passwords for Visualizer user 84
- viewing Configuration Console users 61

V

validating
- DQM rules 107
- UMF files in the Visualizer 264

version 8.1
- Sample Cognos Entity Resume report 297
- Sample Cognos role alert report 296

viewing 90, 121, 271
- application monitor events 192
- characteristic confirmations and denials 155
- Configuration Console users and their status 61
- data mappings 212
- data sources 127
- DQM rules 106
- entity graphs in Visualizer 255
- entity resumes 254
- generic data values 112
- identities 194
- lookup codes 108
- number types 93
- pipeline registrations 180
- pipeline status 190
- pipeline status using the pipeline command 191
- resolution configurations 137
- resolution rules 147
- role alert graphs 256
- role alert rules 116
- roles 114
- UMF exceptions 193
- UMF input documents 125

viewing characteristic types 90

Visualizer 271
- adding entity data, description 262
- All Events report 279
- analyzing entity data description 219
- Attribute Alert Generator History report 272
- Attribute Alert Generator report 273
- Attribute Result report 273
- cannot log in 349
- cannot start 349
- changing your password 233
- client log files 367
- configuring 219, 220
- configuring alert display filter options 222
- configuring display options 220
- configuring graph options 224
- configuring log options 223
- configuring minimum score values for search entities 222
- configuring the default path for Centrifuge 163, 221
- configuring the default path for UMF files 163, 221

Visualizer *(continued)*
- configuring Visualizer logging settings 368
- description 79, 219
- Disclosures report 276
- Event Alert Detail report 277
- exiting 83, 232
- finding entities 244
- finding entities by attribute 244
- finding entities by data source account 245
- finding entities by Entity ID 246
- finding entities by resolution 246
- loading data from UMF files 263
- locking 233
- logging in 82, 229
- logging out 83, 232
- managing access 83
- opening 229
- reports 271
- reports, nothing displays on report 349
- Role Alert Detail report 285
- Role Alert Status report 288
- starting 229
- tables that affect Visualizer performance 354
- troubleshooting 349
- troubleshooting, direct launch approach 231
- troubleshooting, error message when starting on Windows workstations 232
- troubleshooting, unable to start with Internet Explorer 230
- troubleshooting, unable to start with Mozilla Firefox 231
- turning off Visualizer logging 369
- turning on logging 368
- validating UMF files 264
- Web browser settings 81

Visualizer system parameter
- configuring 162

W

Web browser
- setting Internet Explorer to use the required client Java Web Start version 230
- setting Mozilla Firefox to use the required client Java Web Start version 231

Web browser settings
- Configuration Console 58
- Visualizer 81

Web services
- alert queries 336
- building a UMF_QUERY search 332
- building a UMF_SEARCH query 338
- description 9, 323
- detail entity queries 335
- developing for your environment 323
- pipeline searches, UMF_SEARCH documents 340
- queries 330

Web services *(continued)*
relationship queries 337
software requirements 324
srd.wsdl 327
SRDWebService methods 327
starting pipelines 325
summary 342
test client 328
testing 326
UMF_QUERY documents 333
using the wsutil.jar to test 326
wsutil.jar 328
wsutil.jar command syntax 329
WebSphere application server
Visualizer log files 367
WebSphere Liberty
log files 367
Windows Event Viewer
log files 361
WS_ALERT
Web services alert queries 336
WS_DETAIL
Web services detail entity queries 335
WS_RELATION
Web services relationship queries 337
WS_SUMMARY
Web services pipeline searches 342
WS_SUMMARY_TOP10
Web services pipeline searches 342
WS_SUMMARY_TOP100
Web services pipeline searches 342
wsdl files
srd.wsdl description 327
wsutil.jar
command syntax 329
description 328
using to test Web services 326
wsutil.jar file
description 9, 323

X

XUtil (UMF file conversion utility) 207

Printed in USA