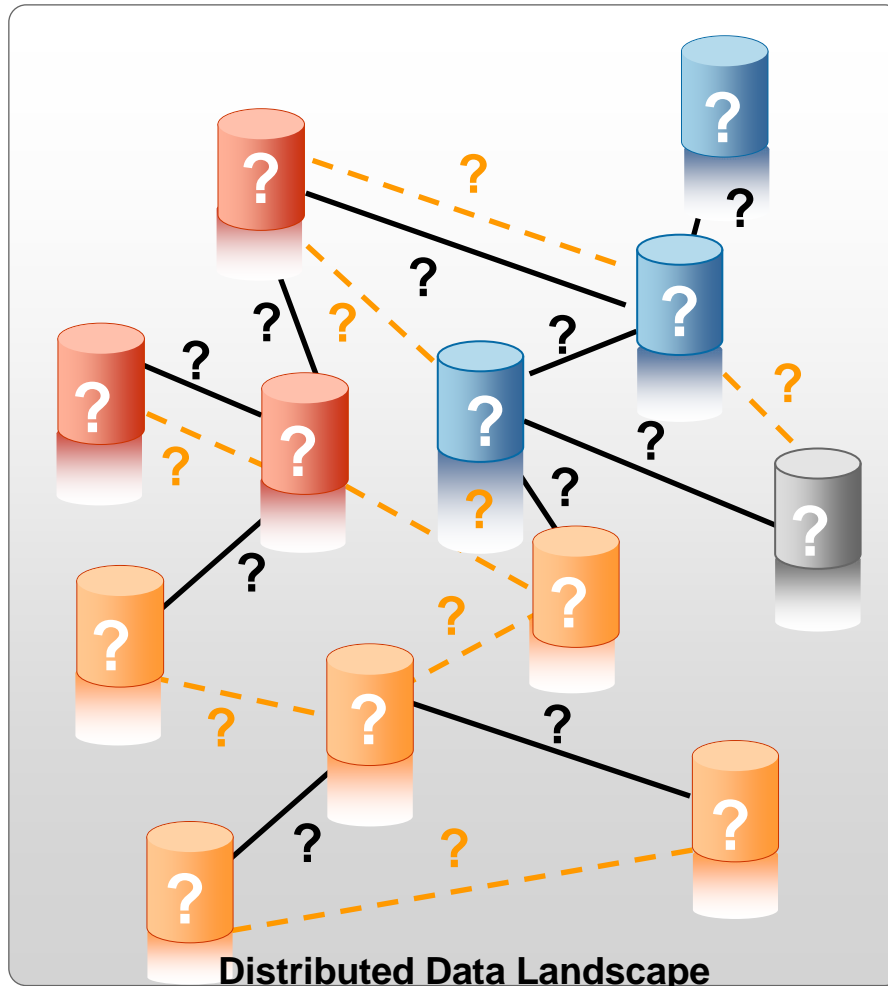


Agenda

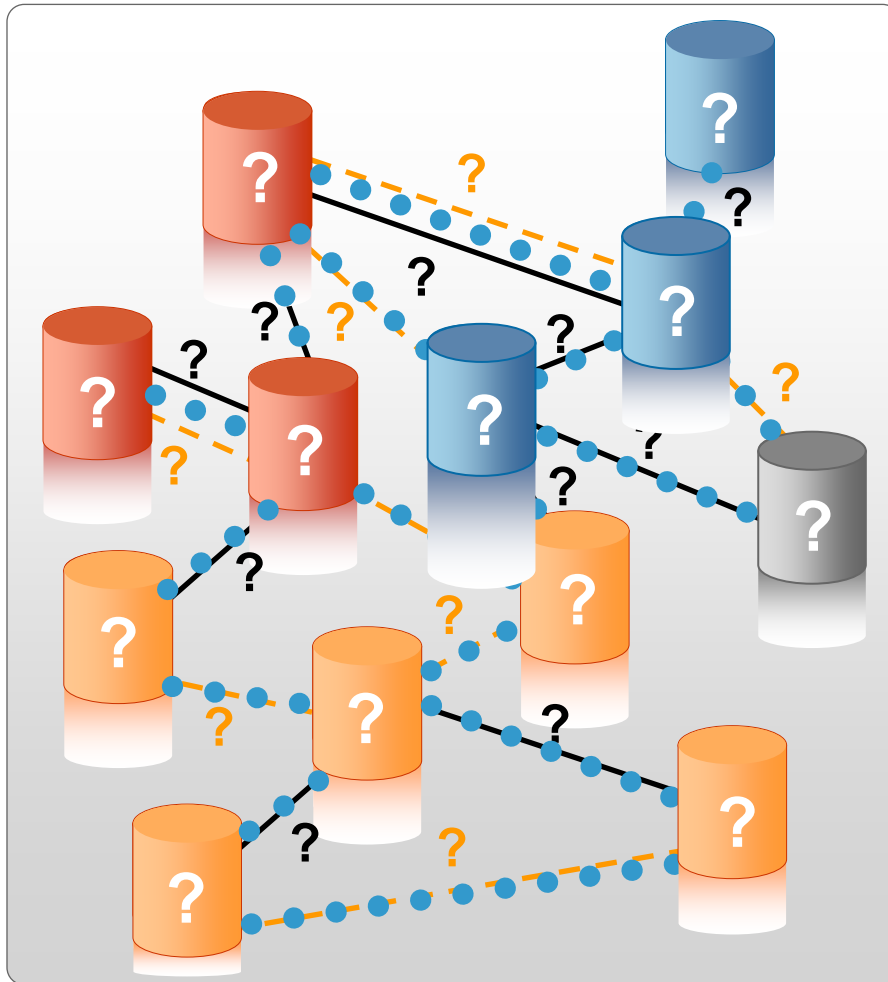
- Data Discovery Challenges
- Data Value Based Discovery
 - Complete Business Object Discovery
 - Archive Use Case
 - Data Privacy Use Case
- Application MetaData Discovery for Oracle Packaged Applications

You can't manage what you don't understand



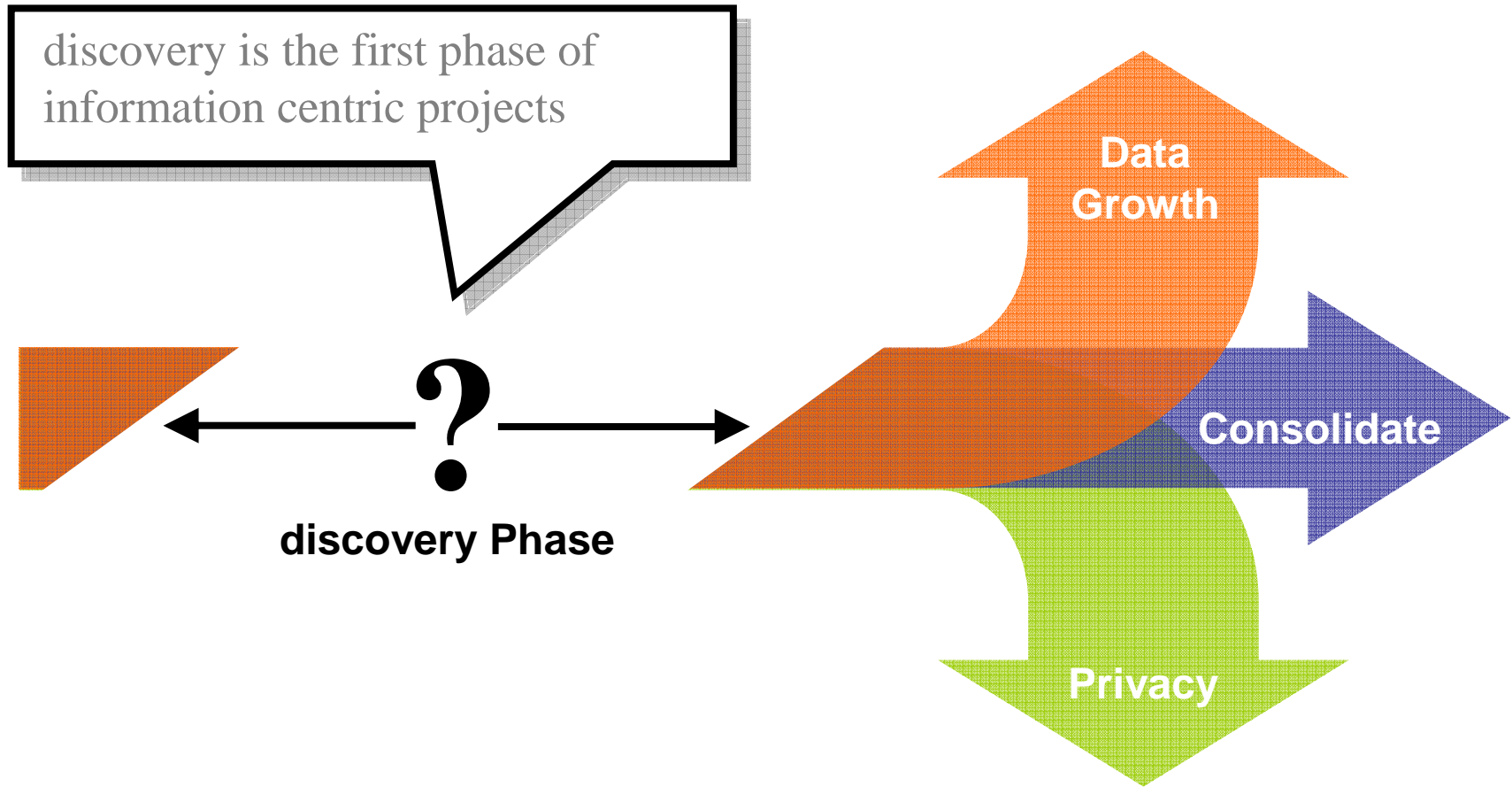
- Increasingly distributed
- Complex, poorly documented data & relationships within & across sources
- Data not understood because:
 - Corporate memory is poor
 - Documentation is poor or nonexistent
 - Logical relationships (enforced through application logic or business rules) are **hidden**

InfoSphere Discovery: An Invaluable Data Analysis Tool



- Automated analysis of data and relationships for **complete understanding** of data assets:
 - **Characterizes** data elements **within** a Source
 - Identifies **relationships** that link data elements into “business entities” **within** a source
 - Customer, counterparty, invoice
 - Identifies **complex logic** that relates **multiple** sources

Poor Understanding = Unpredictable Project Deployment



IBM InfoSphere Discovery: Automation that accelerates time to value:

Data Growth Management:

Automates discovery of referential integrity and business objects

Data Consolidation, Integration & Migration :

Discovers transformation and business logic between data sources

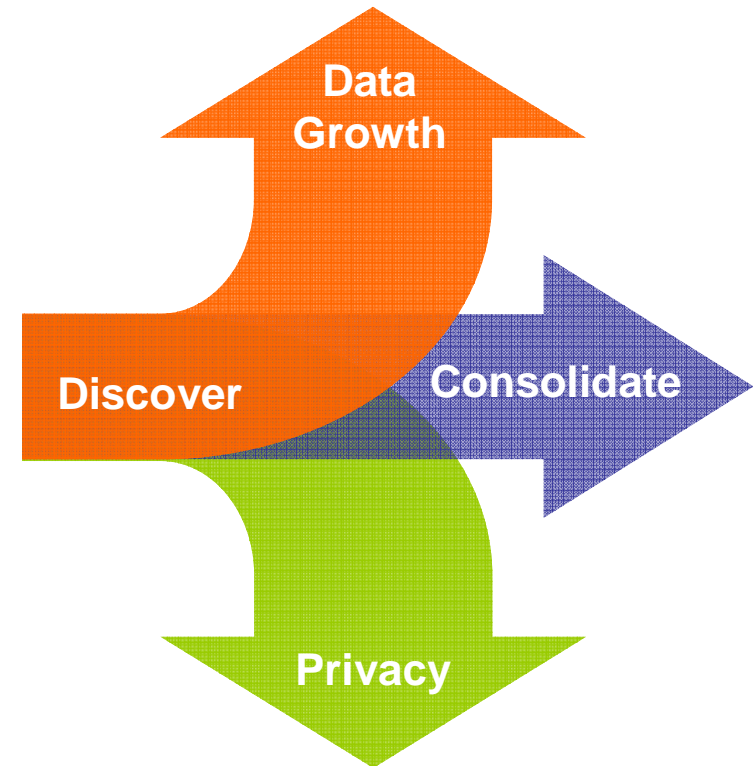
Prototypes empty targets from the combination of many data sources

Data Privacy:

Discovers hidden sensitive data

Data Landscape Rationalization:

Connecting business to unfamiliar assets.



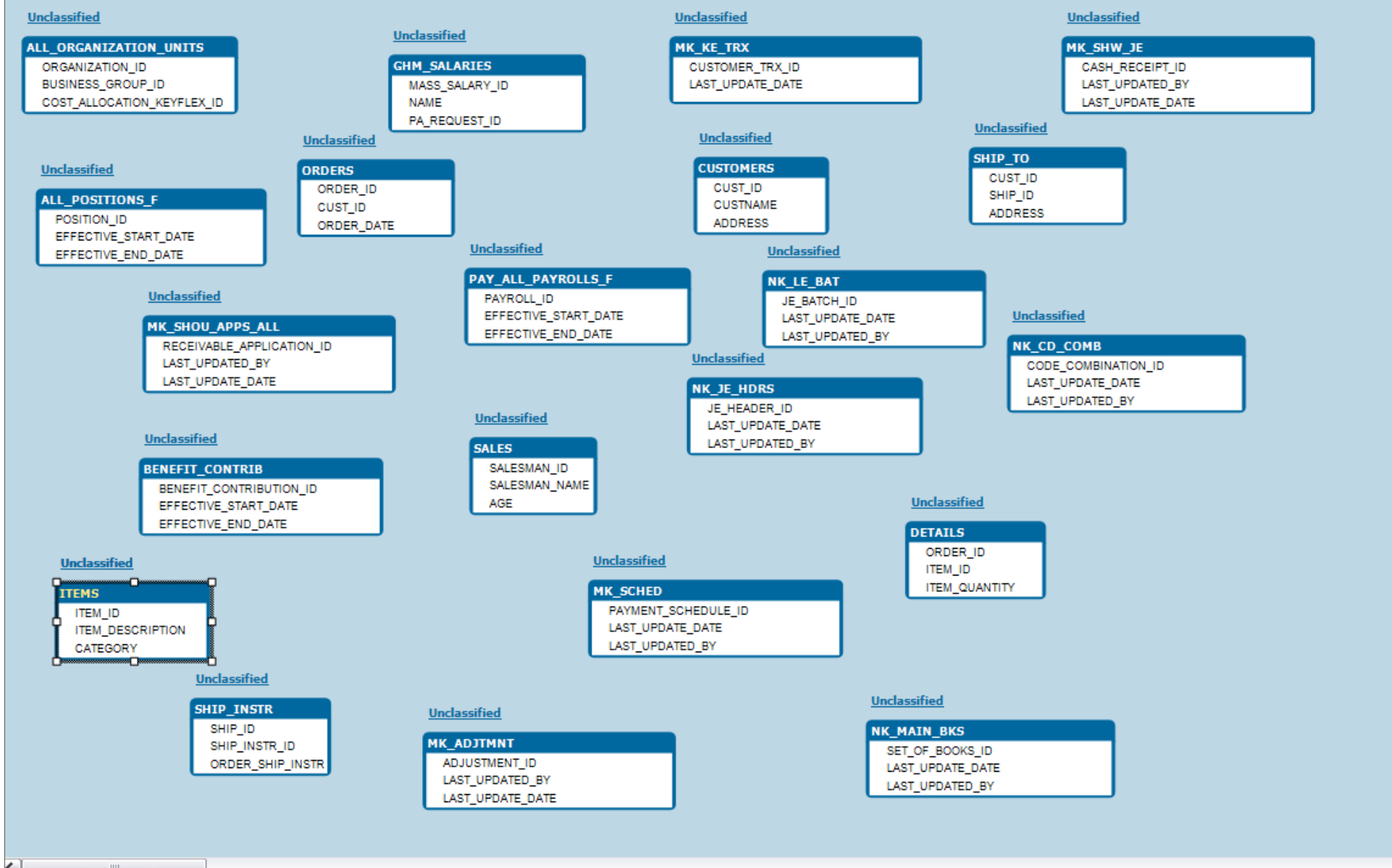
What is unique?

- Analyzes data values and patterns and produces actionable results
- Discovers complex relationships within and between data sources
- Patented approach with greatest level of automation in the industry

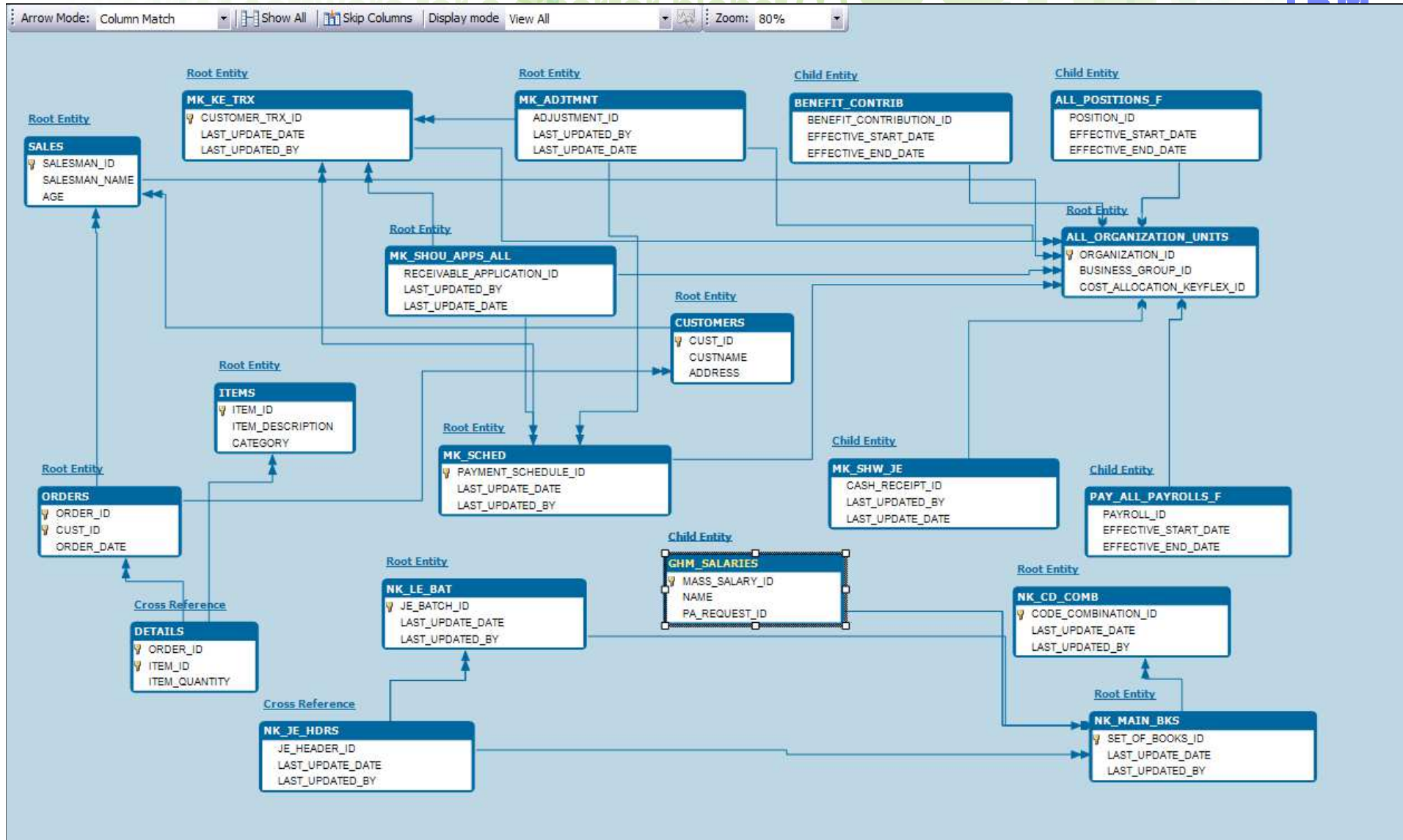
Discovery Deep Dive

Use Cases for InfoSphere Discovery and Optim

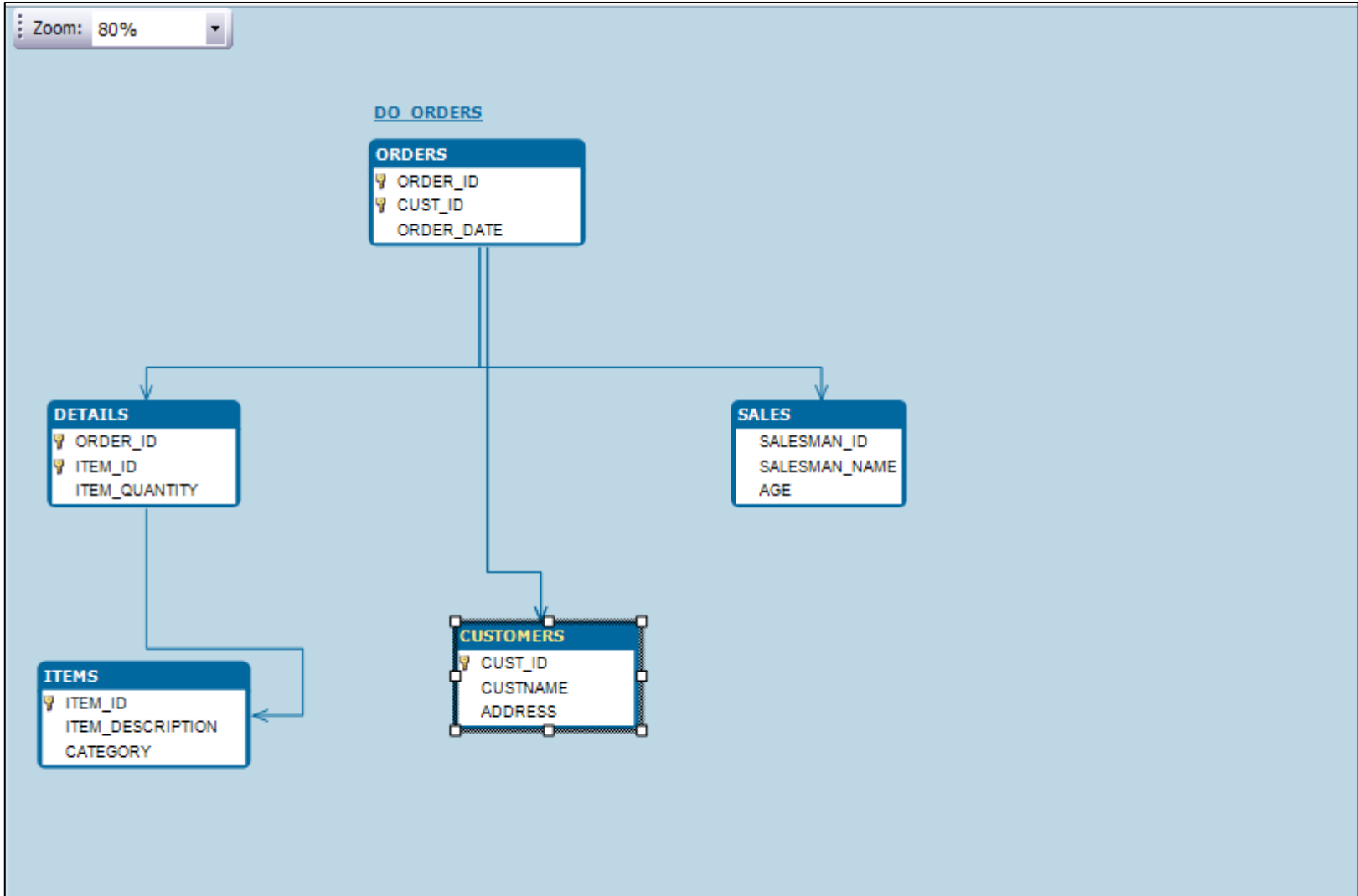
- Perform Complete Business Object Discovery For Archiving
- Discover Sensitive Data for Test Data Management and Data Privacy



How many tables to archive for Orders, where are they?

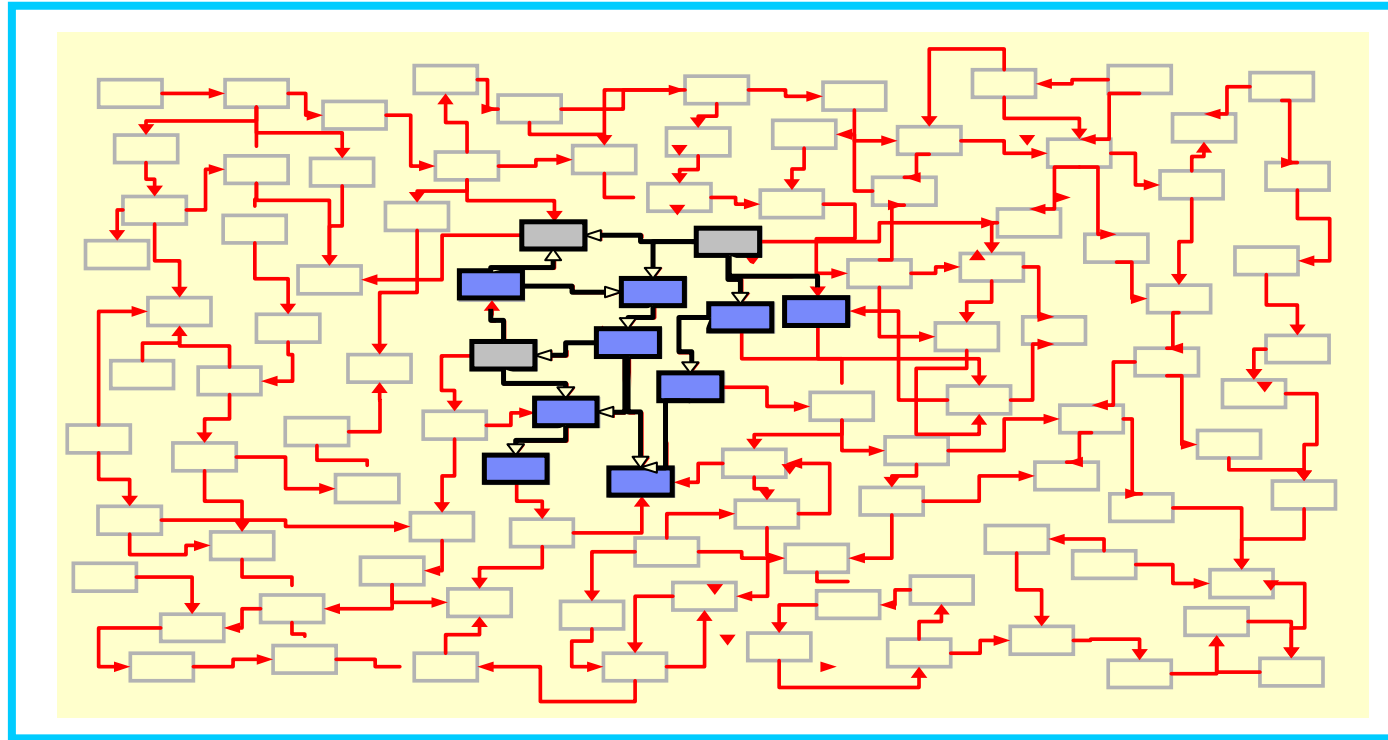


After PFkey discovery...still a lot to work with. Use zoom-and-focus features to review and confirm relationships around Orders table

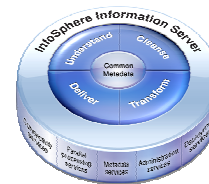


Once confirmed all relevant keys, use Data Object Discovery to produce business object. Export this object to Optim.

Discover Complete Business Object for Optim Projects: Archiving, Test Data, Application Retirement



Optim



InfoSphere projects

IBM InfoSphere Discovery Finds Sensitive Data (Even if Hidden)

Discovery Studio - LOCALHOST - HR MDM 2

Project View Tools Map Help

Home Data Sets Column Analysis PF Keys Data Objects Target Matches Maps

Maps: HQ EMP and EMPERS to WEMPLG HR MDM 2* | No Activity

Maps: <Type keyword>

Summary Joins Bindings Where Clause Transformations Reverse Pivots

Name: HQ EMP and EMPERS to WEMPLG Notes:

Show Map SQL

Query

Join Condition: HQ_EMP JOIN HQ_EMPERS ON (HQ_EMP.EMPLOYEE_ID = HQ_EMPERS.EMPID)
Binding Condition: (HQ_EMP.FNAME = WEMPLG.EFN) AND (HQ_EMP.LNAME = WEMPLG.ELN)
Group By: <Not Applicable>
Where Clause: <Not Specified>

Transformations:

EID	datarule(DR_EID_0, HQ_EMP.EMPLOYEE_ID)
SALUTATION	HQ_EMP.TITLE_OF_COURTESY
EFN	HQ_EMP.FNAME
ELN	HQ_EMP.LNAME
SSN	substr(HQ_EMPERS.SSN, 1, 3) substr(HQ_EMPERS.SSN, 5, 2) substr(HQ_EMPERS.SSN, 8, 4)
BEGIN_DATE	HQ_EMPERS.DOI
END_DATE	CASE WHEN HQ_EMP.STATUS in ('Current', 'Fired', 'Resigned') or HQ_EMP.STATUS is null THEN HQ_EMP.TERMINATION_DATE ELSE HQ_EMP.RETURN_DATE END
DATE_OF_BIRTH	HQ_EMPERS.DOB
CURR_STAT	substr(HQ_EMP.STATUS, 1, 1)

Zoom: Zoom to Fit

HQ_EMP

- EMPLOYEE_ID
- TITLE_OF_COURTESY
- FNAME
- LNAME
- MIDDLEINITIAL
- STREET_ADDR
- CITY
- STATE
- COUNTRY
- ZIP
- COURTILE

HQ_EMPERS

- EMPID
- SSN
- DOB
- DOI
- OEI
- ED
- DANLNO
- PARBARY_CONTACT_NAME
- PARBARY_CONTACT_PA
- PARBARY_CONTACT_REL

WEMPLG

- EID
- SALUTATION
- EFN
- ELN
- SSN
- BEGIN_DATE
- END_DATE
- DATE_OF_BIRTH
- CURR_STAT

Run Next Steps...

Error List

Maps: HQ EMP and EMPERS to WEMPLG

- Automates discovery of sensitive data in both basic as well as complex business rules
 - Finds sensitive data based on basic patterns contained within a column
 - Finds sensitive data hidden within longer fields (e.g. SSN hidden in a 46 digit routing number)
 - Finds sensitive data that has been divided up across multiple columns (e.g. SSN divided into three separate columns)
 - Finds sensitive data that has been transformed (i.e. items converted into codes)

That's not all....

custom “sensitive”, custom “algorithm”

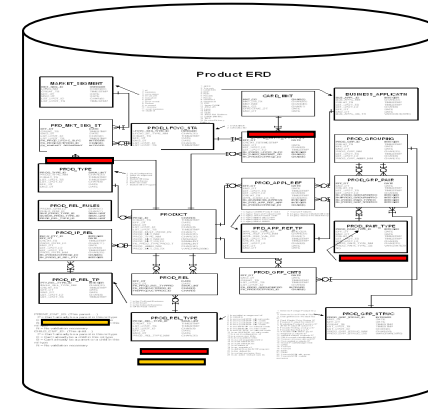
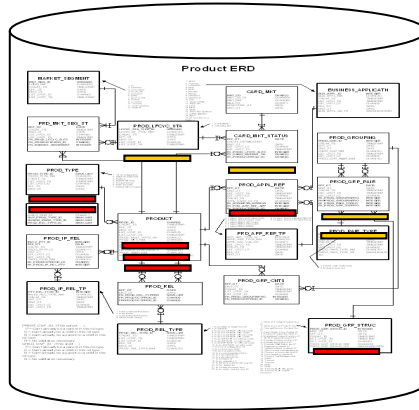
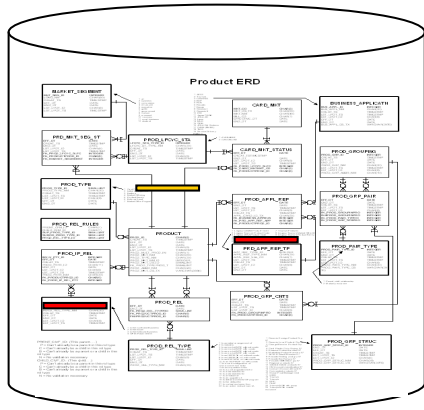
■ How do we find custom sensitive data?

- User defined algorithm -- Once deployed, behave the same as “built-in”
 - Algorithm is run as part of profiling
 - Hit/miss metrics and data view will be available on custom sensitive.
- data similarity – exact value matching
- data similarity -- fuzzy value matching
- metadata similarity including known classification
- data relationships

From “Sensitive Data” Discovery to “Critical Data” Discovery: Using Discovery to connect Business to IT

- Discover “critical data elements”(CDEs) and map them to Business Glossary
 - Data driven term mapping
- Once we find these CDEs, explore its surroundings to identify other data relevant to the business of interest.

Sensitive/Critical Data Discovery Values



Term mapping



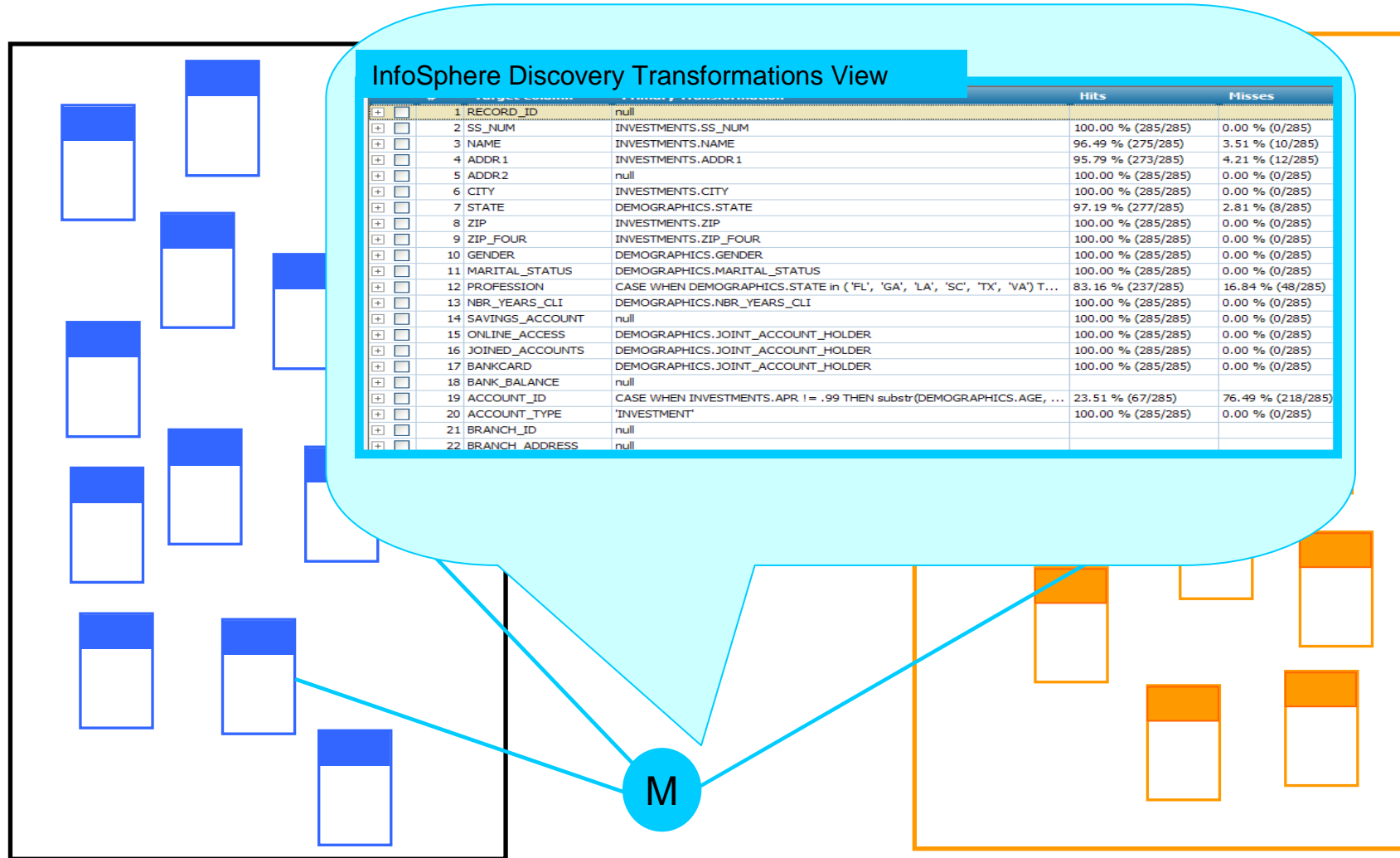
Business Glossary
Connect business to IT

Sensitive data domains with default masking

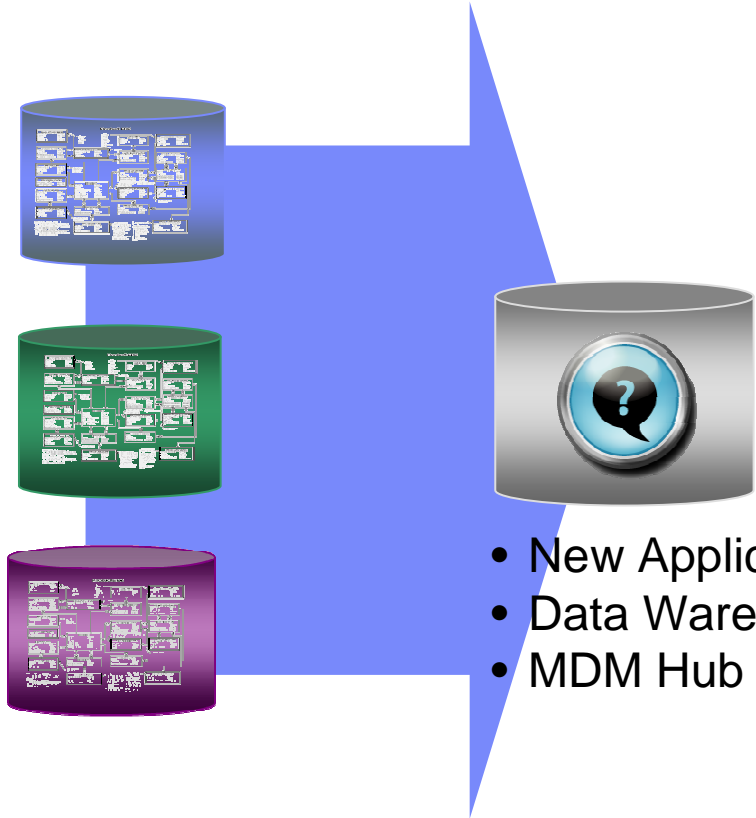


Optim
Masking privacy data in test data extract

And there is still more ... Transformation Discovery



InfoSphere Discovery - Unified Schema Builder



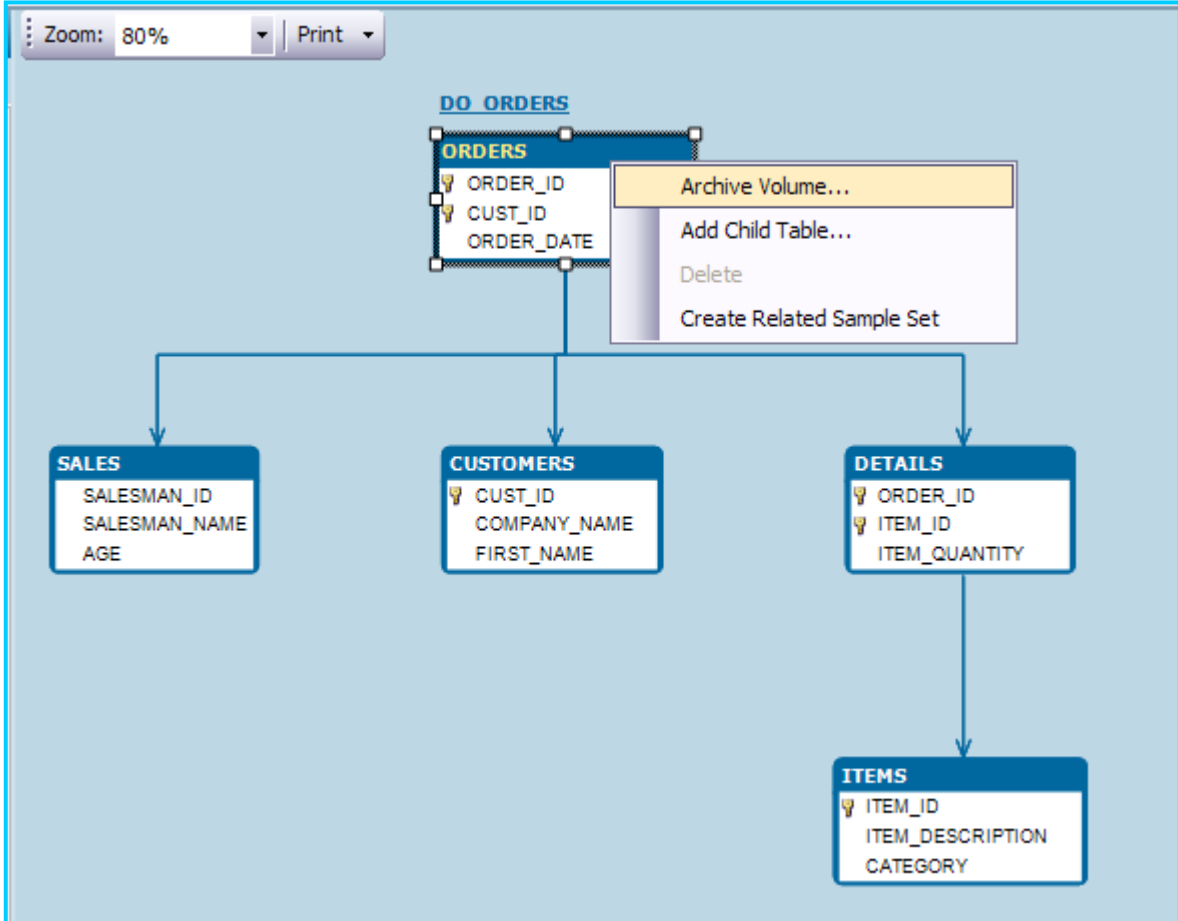
- Unified Schema Builder:
 - Data analyst workbench for data consolidation projects
 - Profile data sources
 - Perform overlap analysis
 - Unified data models
 - Unified data profiles
 - Analyze Matching keys
 - Propose conflict resolution precedence
 - Cross source trouble-shooting workbench

What is unique?

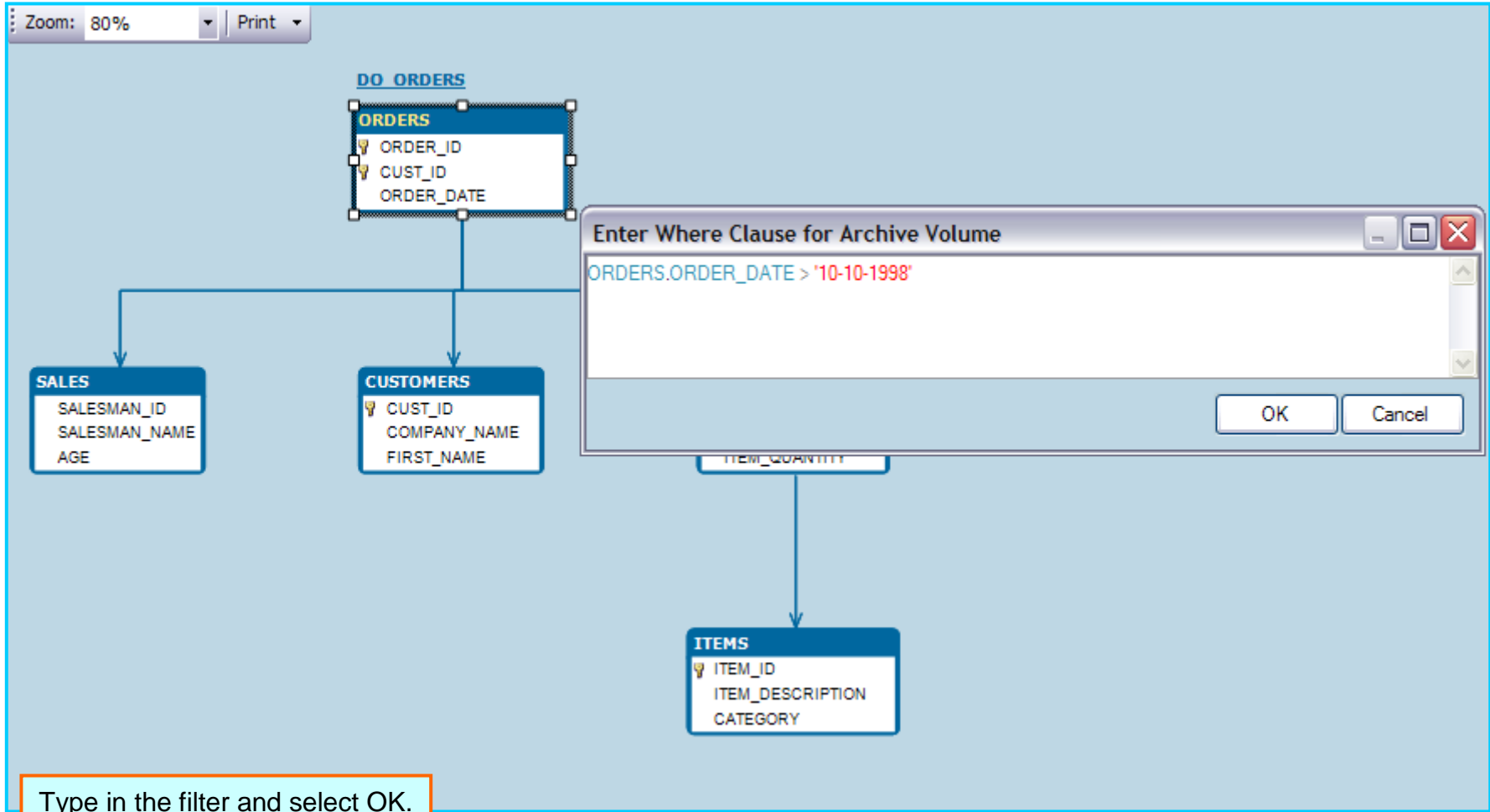
Prototypes empty targets from existing source data (MDM – Master Data Management, EDW – Enterprise Data Warehouse, data migration)

- Applicability
 - Application/Data Consolidation, Migration & Retirement

Archive Volume Projection in Discovery



Volume projections for a filter entered in any table in the Object diagram. Here I will enter a filter on the ORDERS table.



The screenshot shows a database management interface. At the top, there is a zoom control set to 80% and a print button. Below this is a hierarchical diagram of tables. The root table is 'DO ORDERS', which contains 'ORDERS'. 'ORDERS' has three child tables: 'SALES', 'CUSTOMERS', and 'DETAILS'. Each table lists its columns and primary keys.

Below the diagram is a window titled 'Filter: ORDERS.ORDER_DATE > '10-10-1998''. This window contains a table with columns for 'Source', 'Filter', 'Include', 'Table Name', 'Size (MB)', 'Percent', 'Projected Reduction', and 'Realized Reduction'. The table lists the tables from the diagram and their sizes and percentages under both projected and realized conditions. A 'Close' button is located at the bottom right of the window.

Source	Filter	Include	Table Name	Size (MB)	Percent	Projected Reduction	Realized Reduction
						Size (MB)	Percent
		<input checked="" type="checkbox"/>	SALES	0.01	0	0.00	0
		<input checked="" type="checkbox"/>	ITEMS	0.05	2	0.00	0
		<input checked="" type="checkbox"/>	CUSTOMERS	0.85	28	0.27	32
		<input checked="" type="checkbox"/>	DETAILS	1.04	34	0.35	34
	ORDERS.ORDER_DATE > '10-10-1998'	<input checked="" type="checkbox"/>	ORDERS	1.11	36	0.35	32
		<input checked="" type="checkbox"/>	All Tables	3.07	100	0.98	32

A tabular display showing the volumes. You can check/uncheck the “Include” flag to calculate the volume if that table is removed or kept on source.

Application Metadata Discovery for Oracle Packaged Applications

Supporting the Lifecycle for the Oracle ERP/CRM packaged applications

Oracle application owners need to understand the data structures – including customizations – to support these lifecycle activities

- **Manage the data growth of enterprise applications**
 - Identify the complete data structure of a business object (e.g. purchase orders) to intelligently archive historical transaction data
- **Deploy new functionality and manage application upgrades**
 - Understand what needs to be tested, create secure and right-sized test/development environments to ensure smooth deployment
 - Ensure customizations & integrations are not impacted by upgrades and changes
- **Rationalization of application portfolios**
 - Understand and capture the related custom tables and metadata of a retired application to support data retention and retrieval needs
- **Enterprise Application Integration**
 - Understand integration points for Master Data Management, Business Intelligence/Data Warehousing and interfacing with other applications

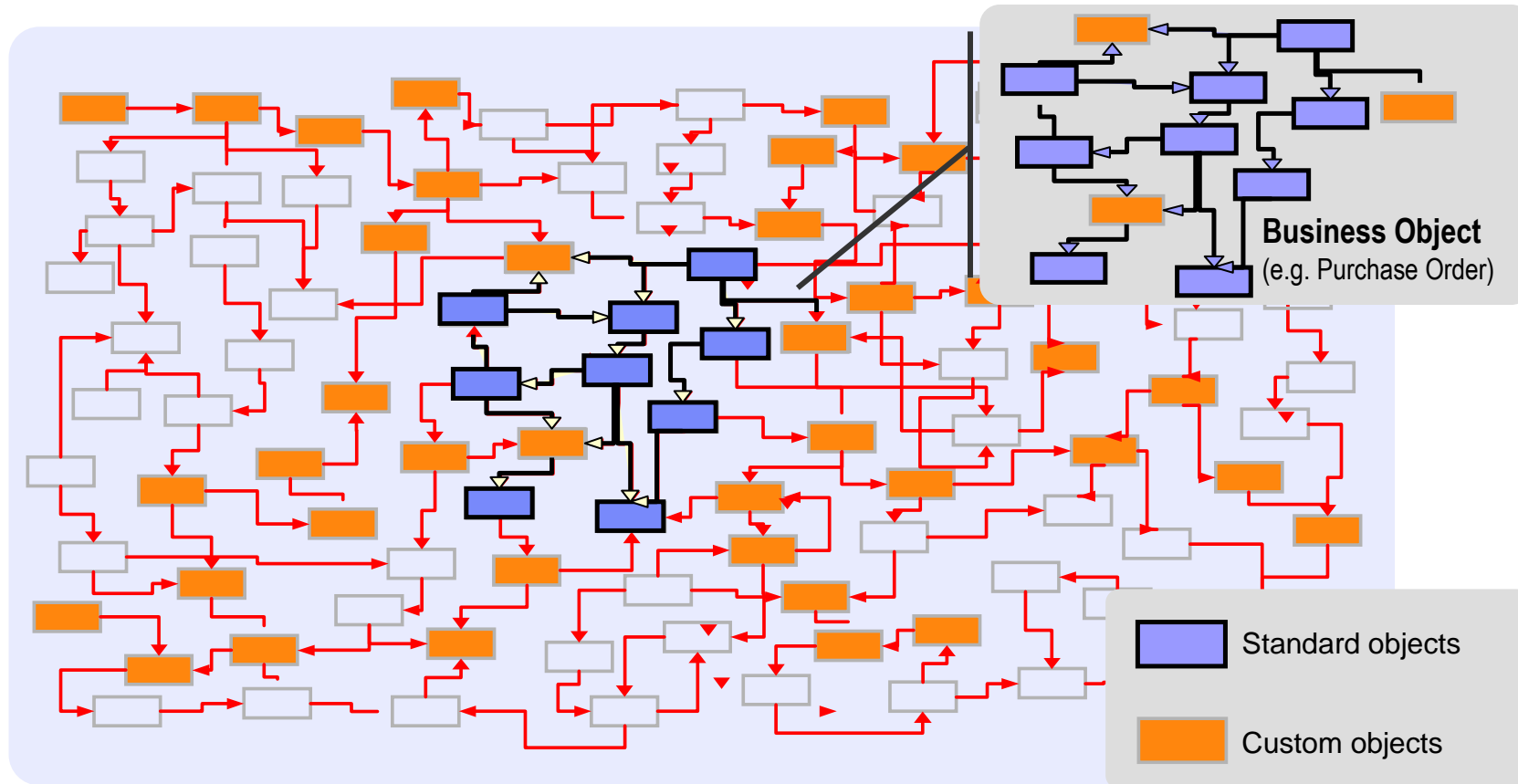


Pain Points for Owners Supporting Oracle packaged applications

- Understanding the customized application data models and complex data relationships
 - Database schemas consist of thousands of tables with complex relationships
 - Proprietary ERP metadata only holds ‘Logical View’ of data
- How to determine the data model for application modules not yet provided “out-of-the-box” by InfoSphere Optim
- How to understand and compare existing data models to new data models for perform impact analysis during an application upgrade
- How to easily define the complete business object to support archiving, subsetting and application retirement
- Dependence on application specific Subject Matter Experts, manual scripts & analysis techniques, reverse engineering ...etc



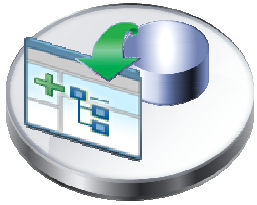
Understanding Oracle Application Business Modules



- Are you implementing a new or heavily customized module?
- Are the custom objects & relationships documented?
- Are you able to define a complete business object?

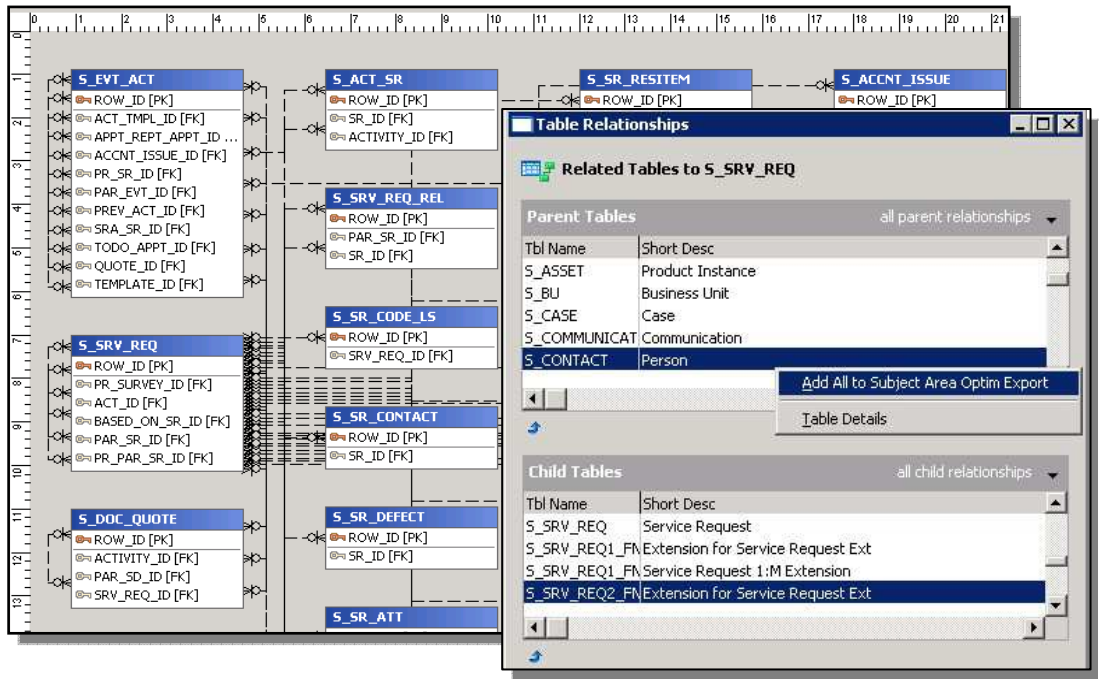


InfoSphere Optim Application Repository Analyzer



ARA

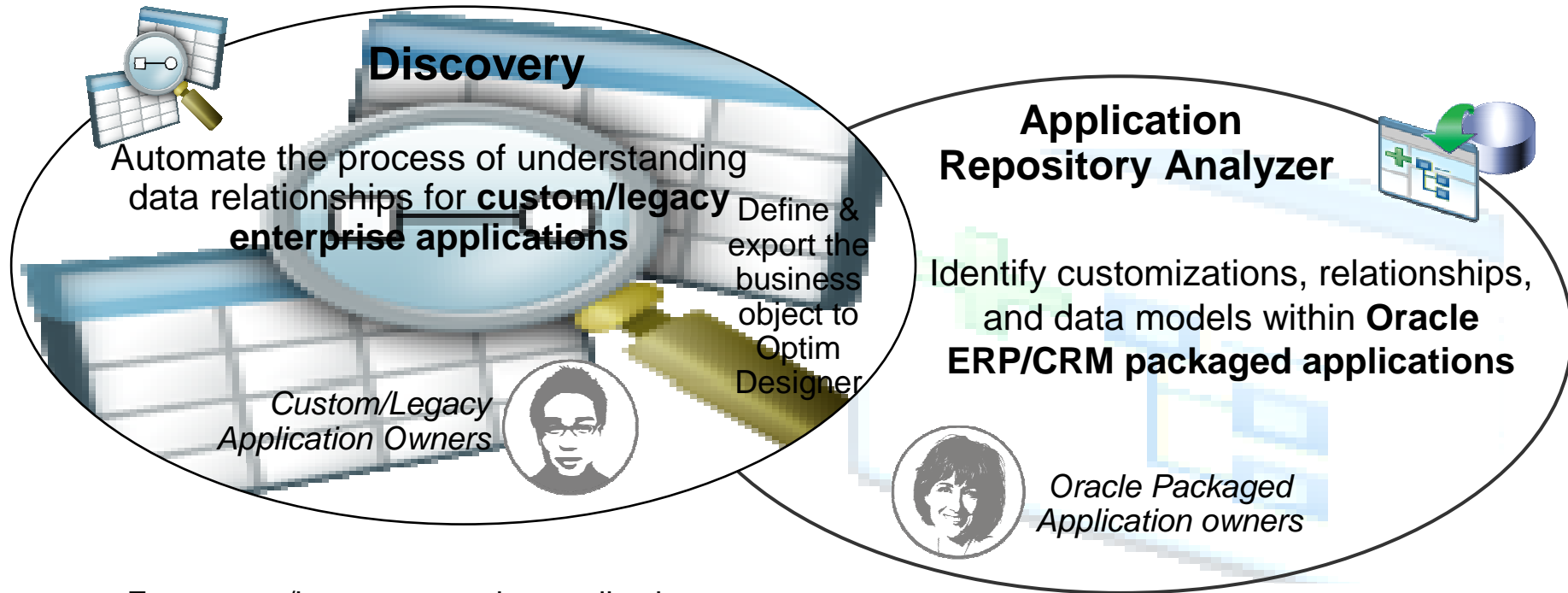
Analyze application metadata to identify relationships & customizations within the Oracle ERP/CRM packaged applications



- ### Requirements
- Analyze application metadata to identify data models, relationships & customizations
 - Compare data model structures across application versions & releases
 - Integrate with InfoSphere Optim solutions

- ### Benefits
- Quickly identify application customizations to speed data lifecycle projects
 - Reduce time and improve quality of application updates
 - Easily export the complete business object for archive, subsetting & masking projects

Strategy for Complete Data Discovery & Analysis across the Enterprise



- For custom/legacy enterprise applications
- Analyzes data samples and data values to infer relationships
- Provides for analysis across multiple applications & databases
- Only way to find relationships for custom and legacy apps that do not have a published metadata repository

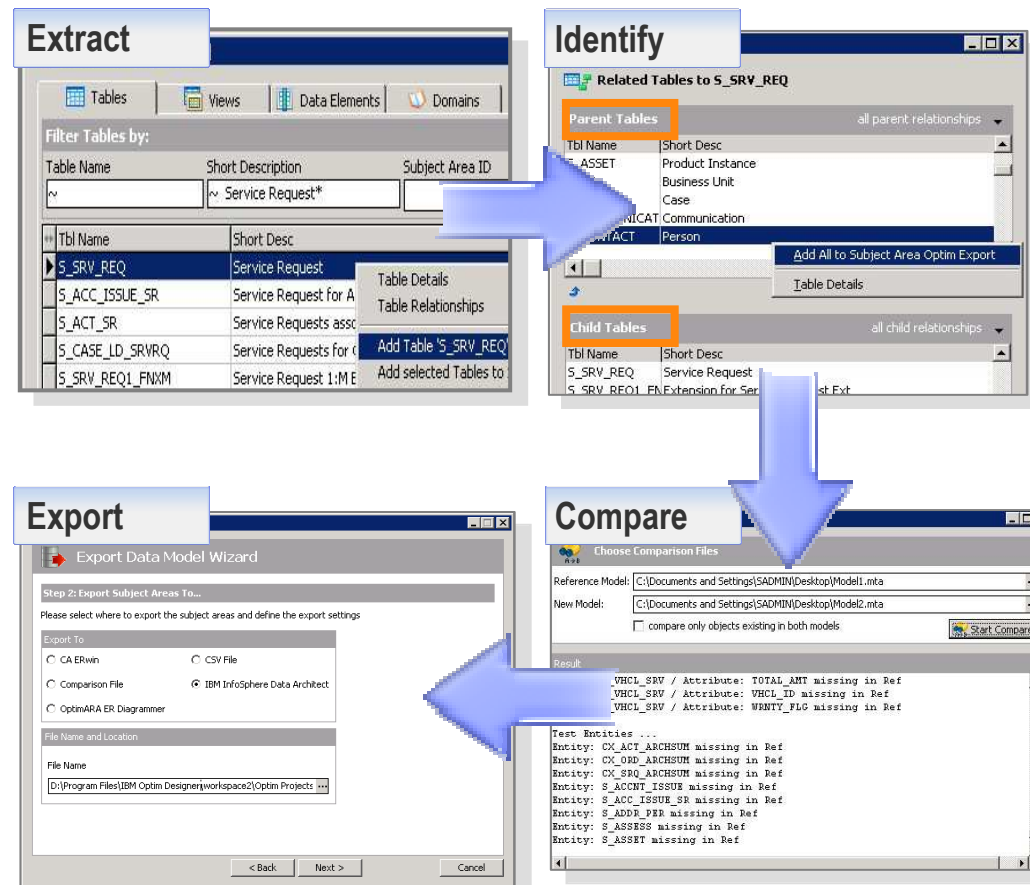
- For Oracle E-Business Suite, Siebel CRM, PeopleSoft Enterprise, JD Edwards applications
- Analyzes the application metadata to extract defined relationships and customizations and understand application business objects
- Compare models between application versions
- Quick, accurate & easiest way for finding relationships in Oracle ERP/CRM applications

InfoSphere Optim Application Repository Analyzer

How does it work



- **Extracts** the relevant metadata from the application, allows metadata browsing & subsetting
- **Identifies** parent-child table relationships to define the complete business object (e.g. Purchase Order) quickly
- **Compares** data models to quickly identify changes associated with application upgrades & consolidation
- **Exports** data model information for business objects in several formats, compatible with:
 - IBM InfoSphere Optim Designer
 - IBM InfoSphere Data Architect
 - CA ERwin Data Modeler
 - ER Diagrams



InfoSphere Optim Application Repository Analyzer

How do you use it

