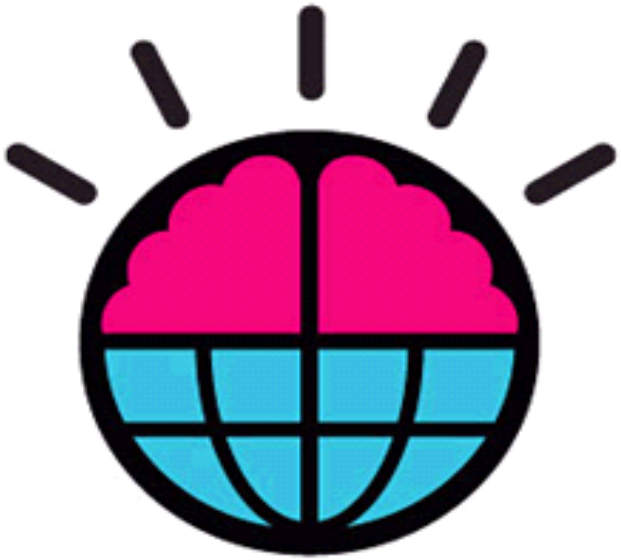
Information Server 8.5를 통한 정보통합솔루션의 진화

김장원 , 실장, IM(SWG) CP : 010-4995-8761 e-mail : kjustin@kr.ibm.com

Agenda

- Evolution form ETL to Data Govern
- What's New with Information Server v8.5
- How to migrate to Information Server v8
- InfoSphere CDC (Change Data Capture)
- Enterprise data architecture and Value of Information Server
- Demo



- **Evolution form ETL to Data Govern**
- What's New with Information Server v8.5
- How to migrate to Information Server v8
- InfoSphere CDC (Change Data Capture)
- Enterprise data architecture and Value of Information Server
- Demo

History of IBM Information Integration Solutions

Compare product lists from Ascential to InfoSphere

	Ascential Suite (~2007)	Information Server v8.0 (2008.01)	Information Server v8.5 (2010.11)
Understand	▪ Profile Stage ▪ Audit Stage	▪ Data Architect ▪ Information Analyzer ▪ Business Glossary	▪ Data Architect ▪ Information Analyzer ▪ Business Glossary ▪ Business Glossary and Anywhere ▪ Blueprint Director
Cleans	▪ Data Stage ▪ Quality Stage	▪ DataStage ▪ QualityStage	▪ DataStage & QualityStage
Transform	▪ Data Stage	▪ Data Stage	▪ Data Stage ▪ Fast Track
Deliver	▪ Data Stage	▪ Data Stage ▪ Federation Server	▪ Data Stage ▪ Federation Server
SOA	▪ RTI	▪ Information Service Director	▪ Information Service Director
Metadata	▪ Meta Stage	▪ Shared Metadata Server	▪ Shared Metadata Server ▪ Metadata Workbench
CDC	▪ Ascential CDC	▪ Information Integration ▪ Data Mirror	▪ InfoSphere CDC

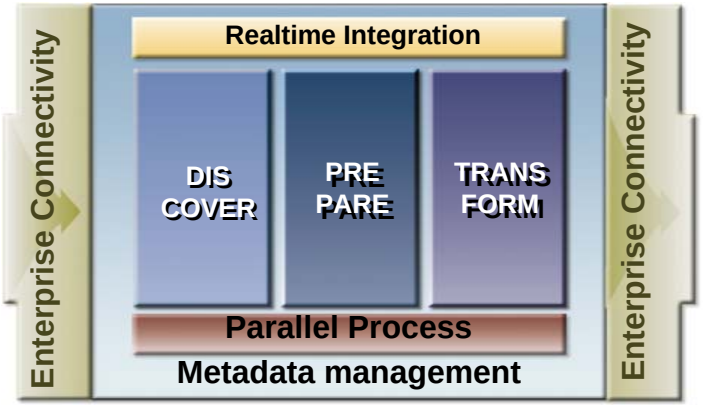
Evolution form ETL to Data Govern

Data Governance Support

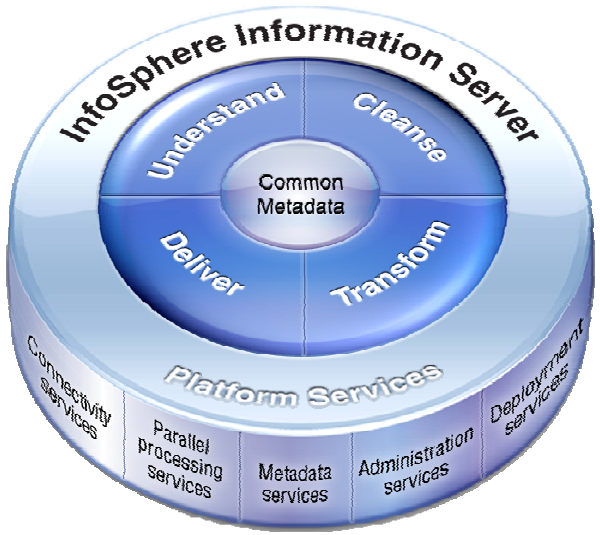
ETL(DataStage) Tool

- Parallel Processing
- Common Repository
- Functional Integration
- Right time data delivery

Information Integration Suite

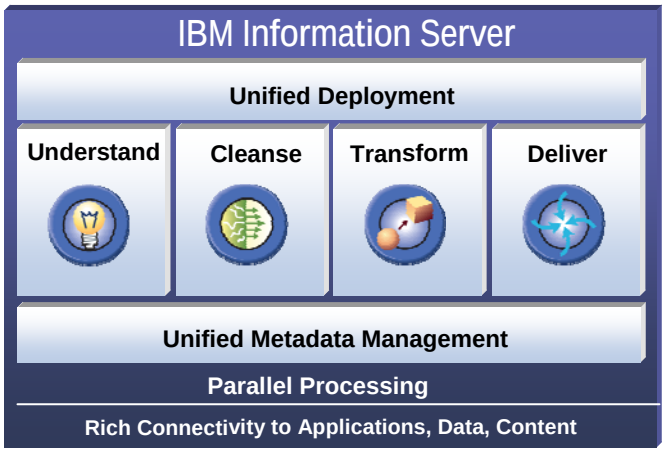


- Simplified Scalability at low cost
- Easily and Rightly deployment
- Delivering end-to-end Metadata



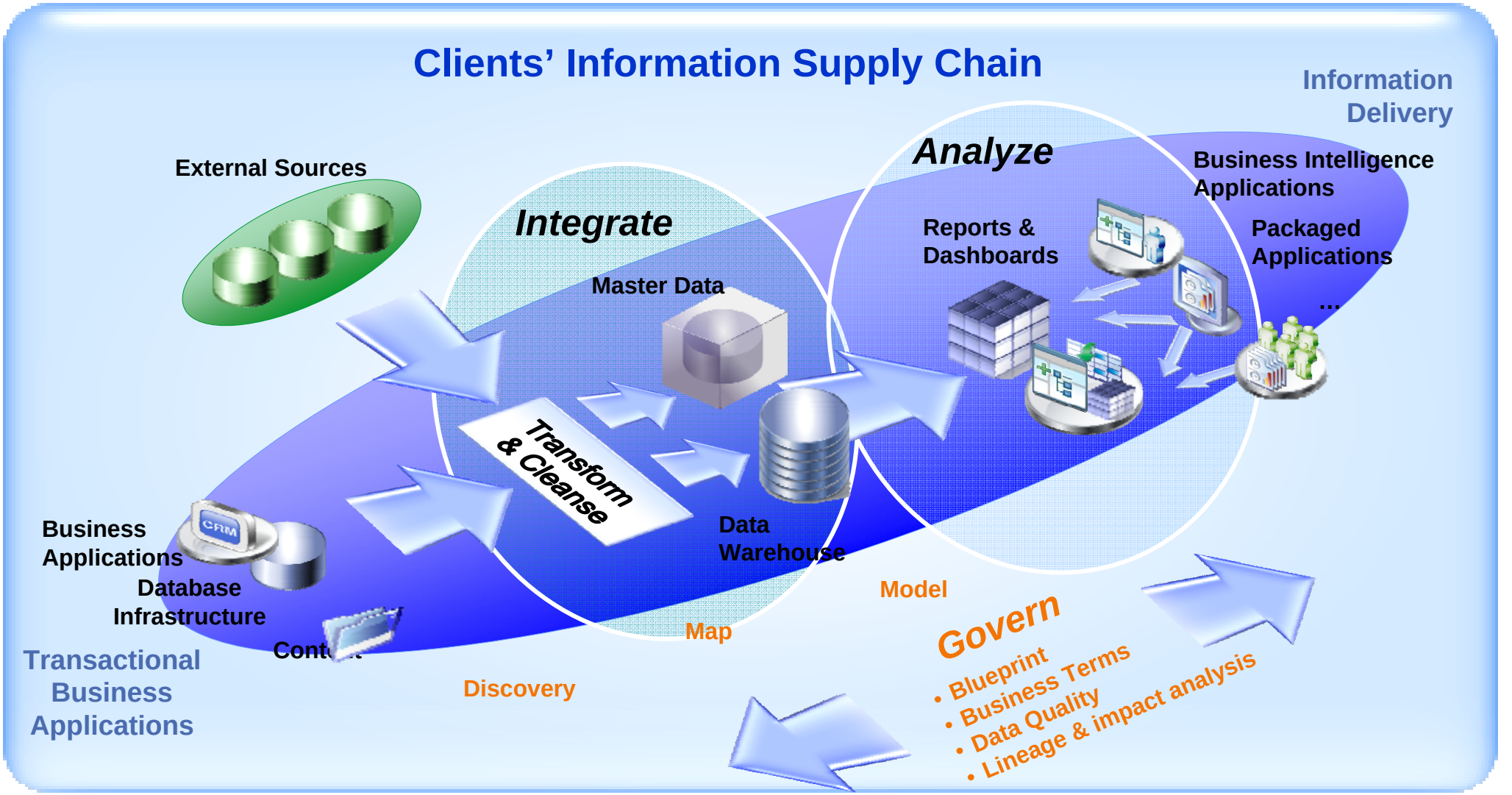
- Deep integration across components
- Shared Understanding of existing assets
- Data Govern
  - Blue Print
  - Business Terms
  - Data Quality
  - Lineage & Impact Analysis

IS v8.5



Delivering business insight requires mastering your information

Going beyond simply managing... to Optimizing Business Results

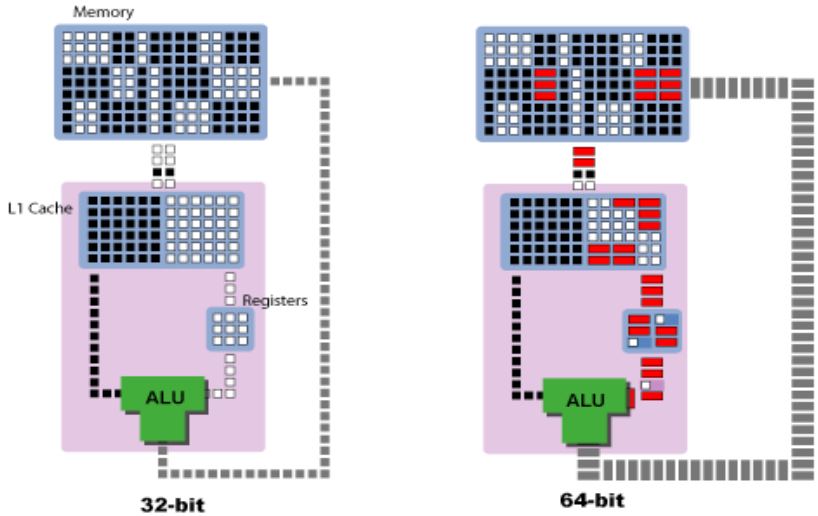


- Evolution form ETL to Data Govern
- **What's New with Information Server v8.5**
- How to migrate to Information Server v8
- InfoSphere CDC (Change Data Capture)
- Enterprise data architecture and Value of Information Server
- Demo

IBM Information Server– What's new in 8.5

Platform support and General Enhancement

- Red Hat Enterprise Linux 5,6 (64 bit)
- SUSE Linux Enterprise Server 9,10 (64 bit)
- Windows Server 2008 64bit (32-bit app)
- AIX 5.3, 6.1 (64 bit)
- Solaris 9,10 (64 bit)
- HP-UX Itanium (64 bit)
- Red Hat Enterprise Linux for System Z (64 bit)
- SUSE Linux Enterprise Server for System Z (64 bit)
- Windows Server 2003 (32-bit)
- Red Hat Enterprise Linux 5, 6 (as 32-bit app)
- SUSE Linux Enterprise Server 9, 10 (as 32-bit app)
- Clients – Windows XP, Vista and 7 (32 & 64 bit)
- Repository – DB2 9.5, 9.7, Oracle 10g, 11g, SQL Server 2005,



- Design Time Performance compare with v8.0
  - 40% Performance improvement in Job Open, Save, Compile etc.
- Improvements in startup time for jobs
- Runtime Performance
- New Utility, String and Date/Time functions
  - eg: IsValidTime, NthWeekdayFromDate, DecimalToTimestamp

IBM Information Server – What's new in 8.5

Certifications delivered (DataStage & QualityStage)

New database connectivity enhancements include support for -

- IBM DB2 LUW 9.7, 9.8 (Pure Scale)
- IBM DB2/Z v10
- IBM WebSphere MQ Series 7.0
- Informix IDS v11.7
- Teradata 13
- Oracle 11gR2
- Sybase ASE v15.5
- Sybase IQ v15.2
- Netezza 5
- DataDirect ODBC v6.0 SP2 (including new Bulk Loading for SQL Server from DataStage on Linux/Unix)
- DB2/Z Bulk Load/UnLoad Stage – enables FTP data movement and invocation of native DB2/Z utilities

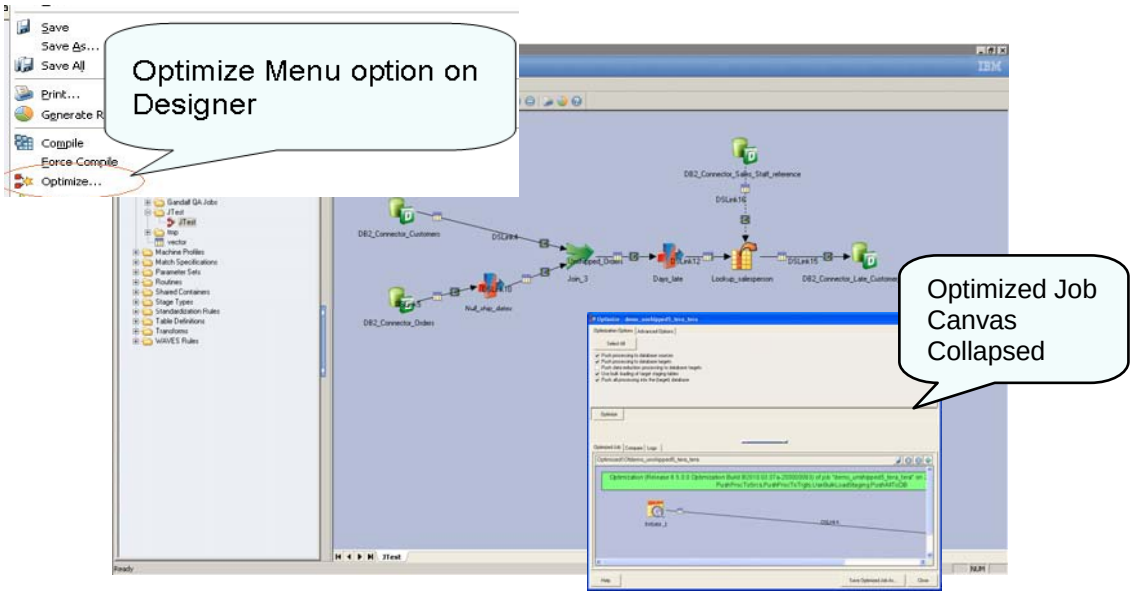
Add-ons:

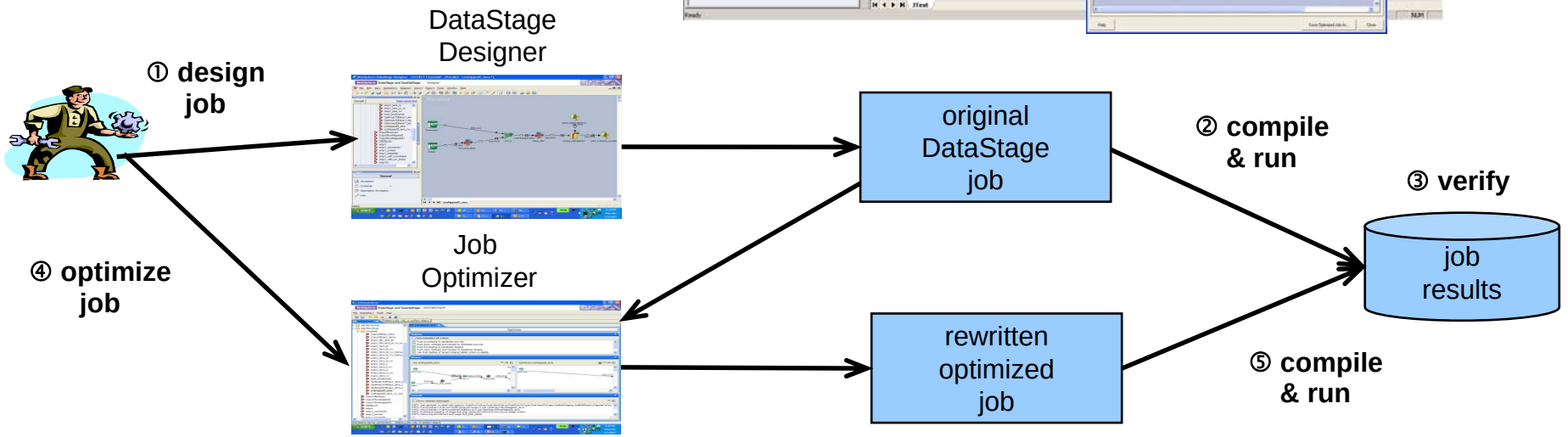
- z/OS File Stage – new component that enables more direct/easy access to native mainframe files.
- Optimized Integration with Changed Data Capture (CDC) and Changed Data Delivery (CDD)
- Packs all (for SAP R/3 and BW, Oracle eBusiness/Siebel/PeopleSoft/Hyperion Essbase/JD Edwards), updated for use with Information Server/DataStage v8.5 with new functionality for salesforce.com bulk loading.

DataStage– What's new in 8.5

Balanced optimization

- InfoSphere DataStage 메뉴에서 Balanced Optimization 실행
- Optimization Log 처리 개선
- Optimized Job 개발 기능 개선
  - create, delete, edit, rename, move folder, compile, deploy, export/import

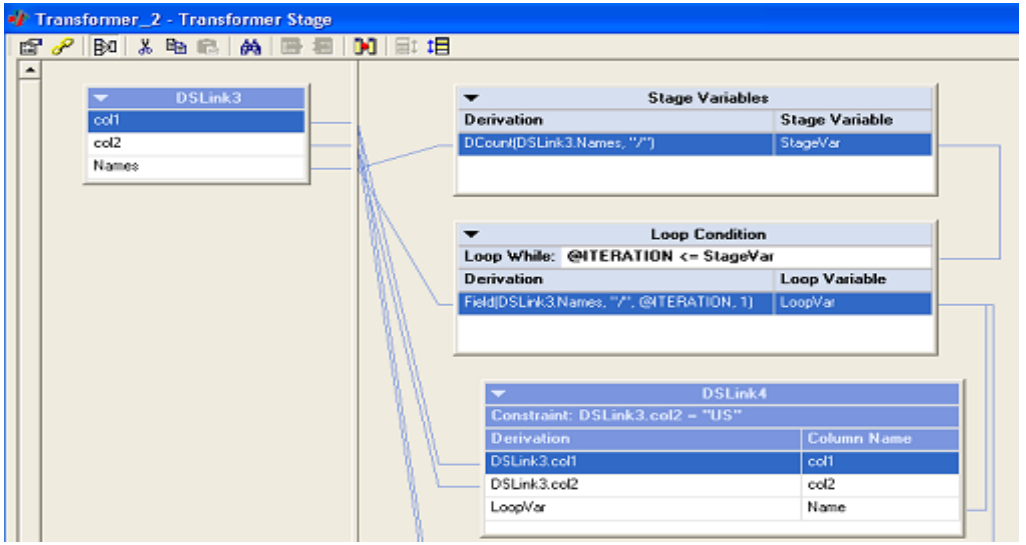




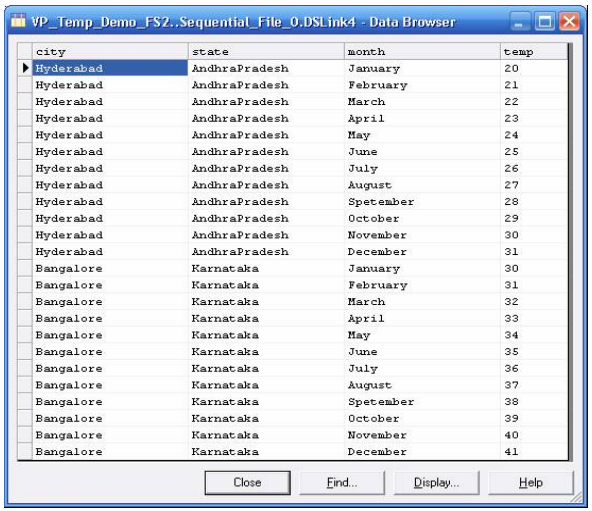
DataStage– What's new in 8.5

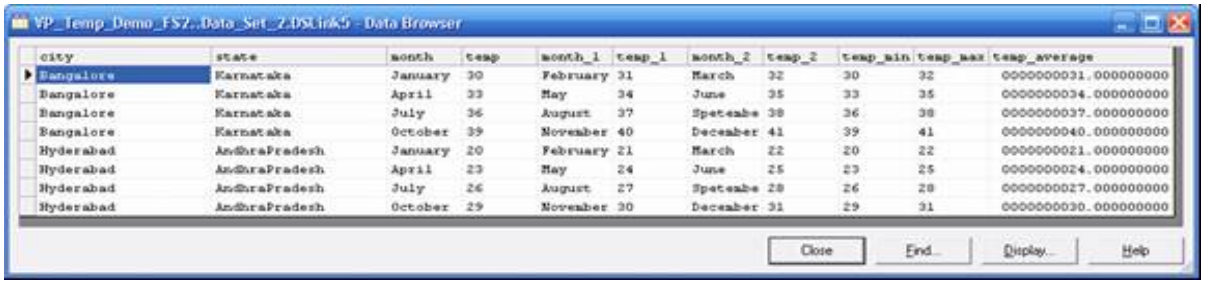
Transformer enhancements

- 조건에 따른 Looping 처리
- End of Data Flag 처리 로직 제공
- System variables, function 추가
  - @ITERATION, Loop Count
  - @EOD, End of data flag for last row
  - LastRowInGroup(InputColumn), automate change detection
- Null 처리 로직 추가



Vertical pivot





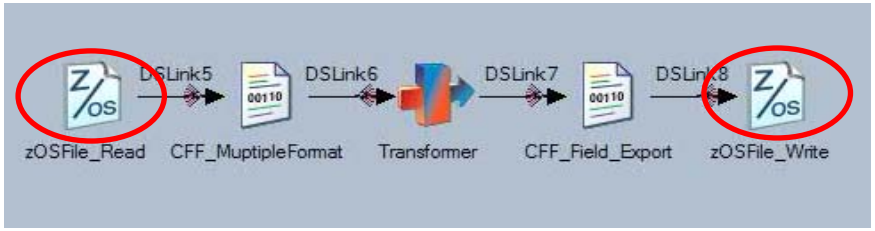
- *Pivot* stage 기능 개선 - Vertical Pivoting

  - mapping multiple input rows with a common key

DataStage– What's new in 8.5

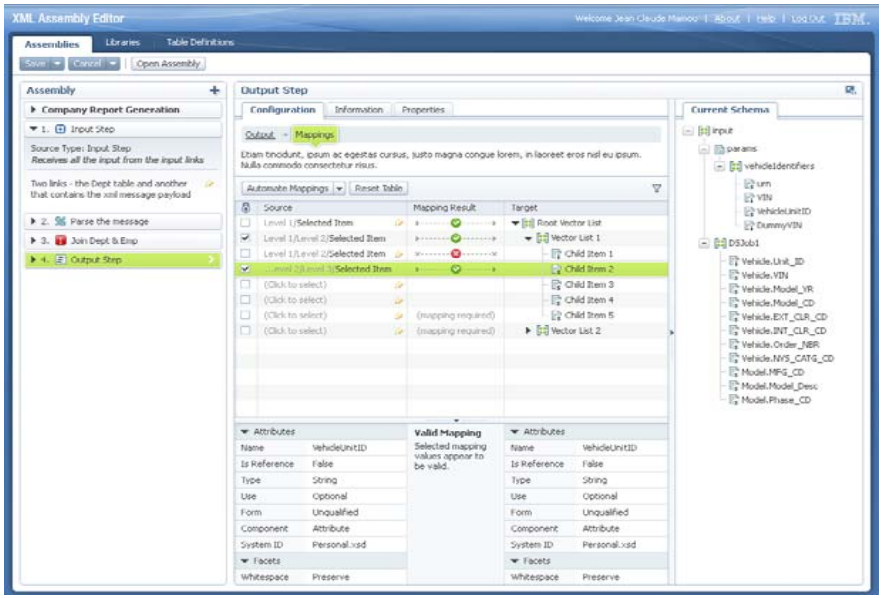
z/OS file stage

- Mainframe z/OS 파일 처리 (Open system)
  - VSAM files - KSDS, ESDS, RRDS
  - Sequential files - QSAM, Sequential read of BDAM/BSAM, PDS members, GDG files.
- Mainframe z/OS 파일 직접 연결
  - InfoSphere Classic Federation



XML pack (post vNext)

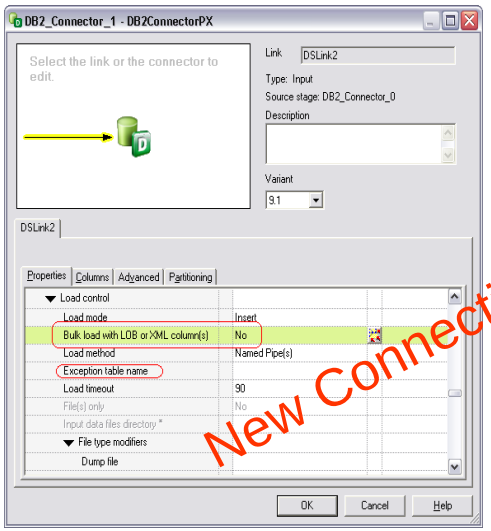
- 복잡한 XML schema 설정 가능
- XSD XML Schema 1.0, WSDL 1.1 지원
- XML 변환 로직 구현
  - 계층 구조, Join, Filter, Switch, Sort, Union, Regroup, Row To Columns, Columns To Rows, Aggregate, Distinct
- 변환 단계에서 Web Services 호출
- 병렬처리 지원
- 제한 없는 XML 문서 Size

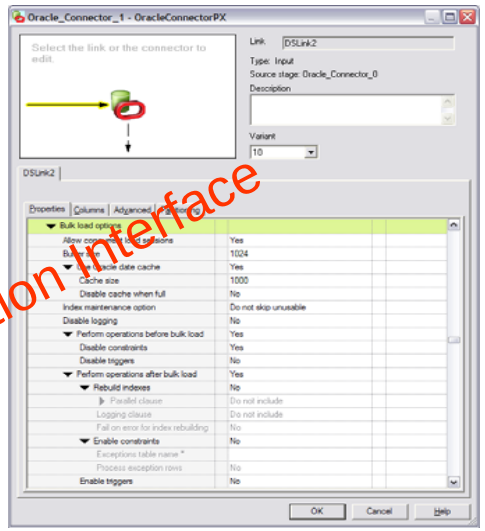


DataStage– What's new in 8.5

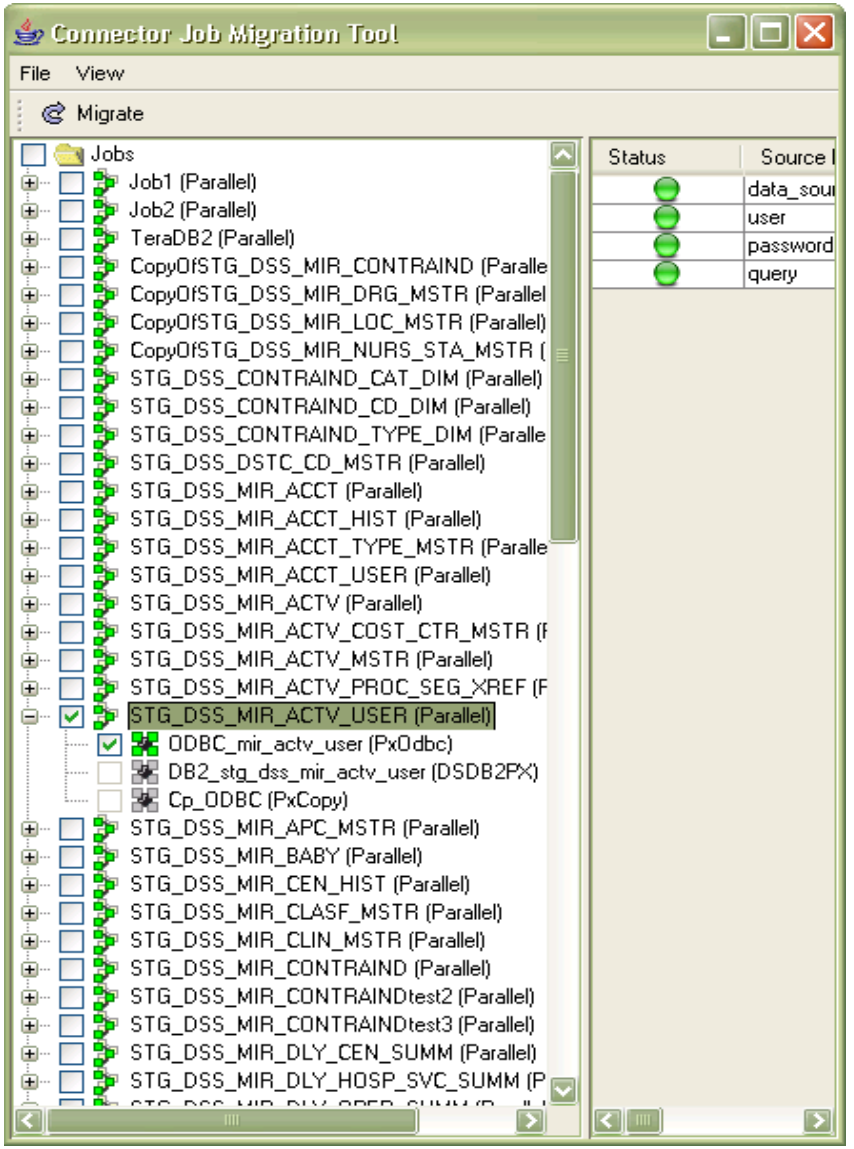
Connection migration tool

- 기존 Job 에 대하여 새로운 Connector Stage 적용
- 전환 가능한 Connection stage 에 대한 자동 변환
- GUI (우측 화면) and command line 제공
- Server and Parallel job 동시 적용 가능
- 기존 Job 에 대한 Backup, copy, replace
- Job 단위의 Annotated(설명) 전환



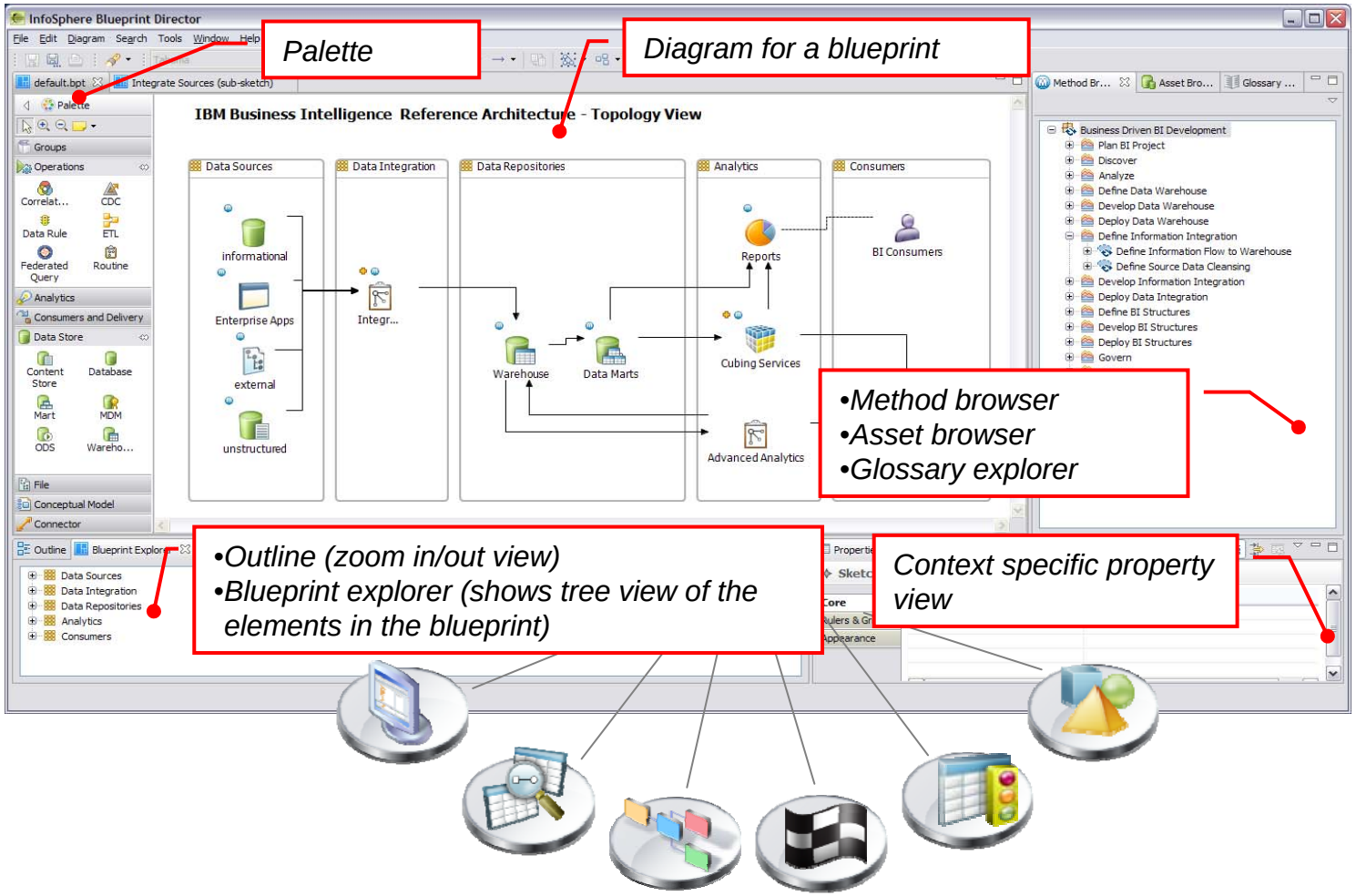


New Connection Interface



Blueprint Director

The GPS for your information project



- Information 프로젝트에 대한 Blue Print 작성
- Blue Print 템플릿 제공
- 비즈니스 & IT View 제공
- Navigate information project by information roadmap & overtime 관리

Information project leadership team

(stakeholders, bus. Analysts, stewards, specialty architects, etc.)

Business Glossary

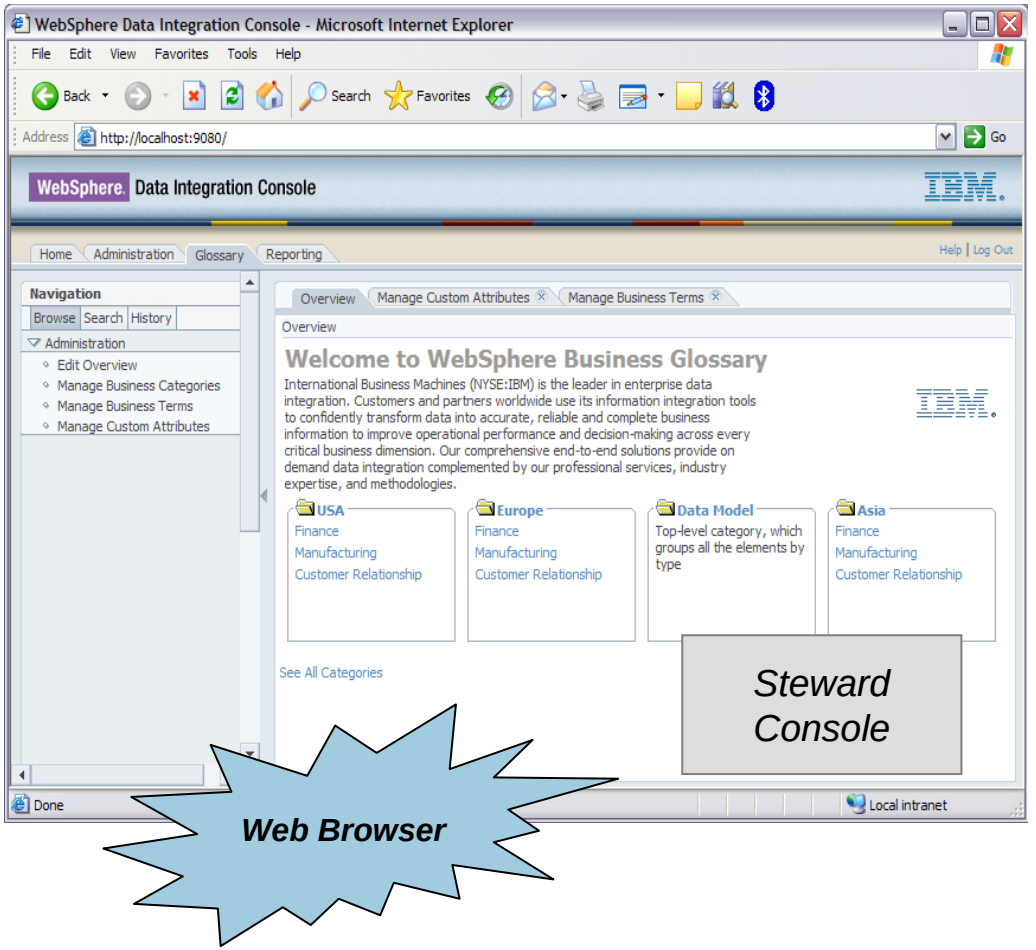
Create and manage business vocabulary and relationships

Features

- 비즈니스 사용자 & IT 사용자의 의사소통을 위한 common business vocabulary.
- Web 환경에서 비즈니스 용어를 공유
- 비즈니스 용어 와 설명에 대한 데이터 오너쉽 (stewards) 정의.
- 비니지스 용어 vs IT 용어 연결 정의

Benefits

- 비즈니스 목적 과 일치 하는 IT 목표 설정.
- 비즈니스 용어 와 IT 자산의 연계 설정
- Data governance 정책 수립을 위한 역할 정의 및 관리 기반 마련.



Business Glossary Anywhere

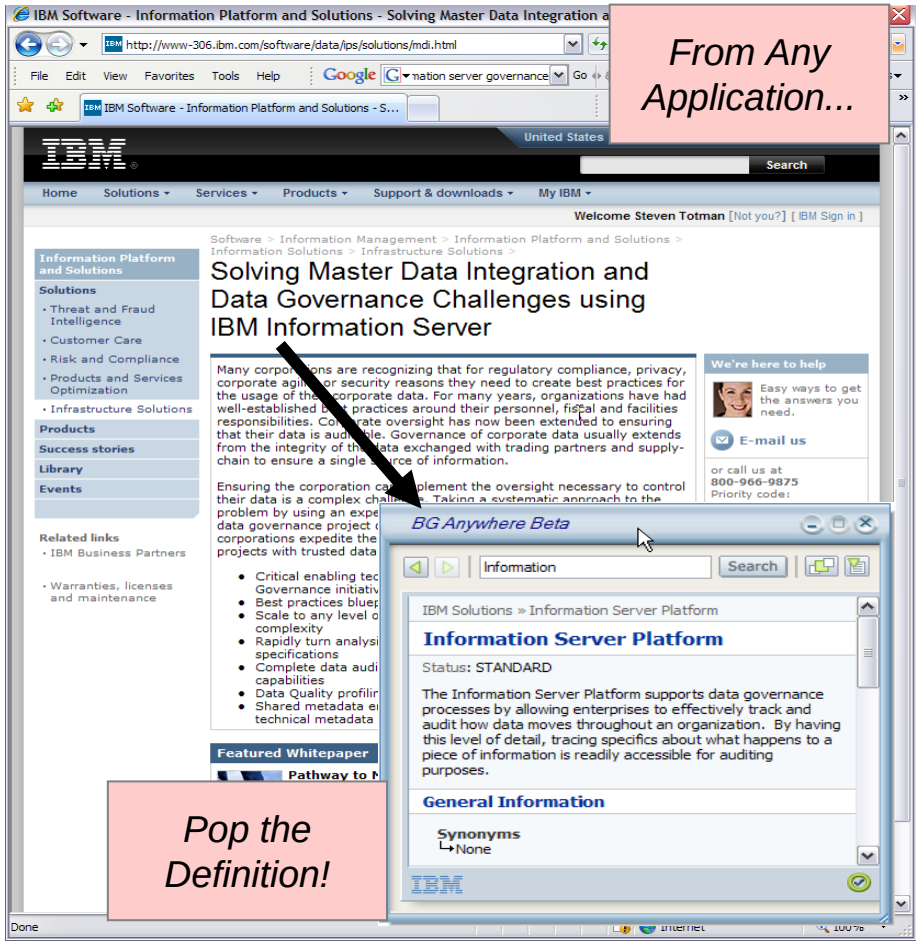
Real-time access to Business Glossary from any desktop application

Features

- Windows 에 Client 모듈이 설치 되는 Desktop 기반의 application.
- Desktop 에 있는 어떤 단어에 대하여 Click 하게 되면, 해당 비즈니스 용어 정의에 대한 내용이 Pop-up window 형태로 조회됨.
- 조회되는 내용은 Business Glossary 에 정의된 내용에 한하며, 추가적인 비즈니스 용어 정의 및 변경은 Business Glossary 에서 관리함.
- 조회를 위한 Hot-key 는 유저가 정의.

Benefits

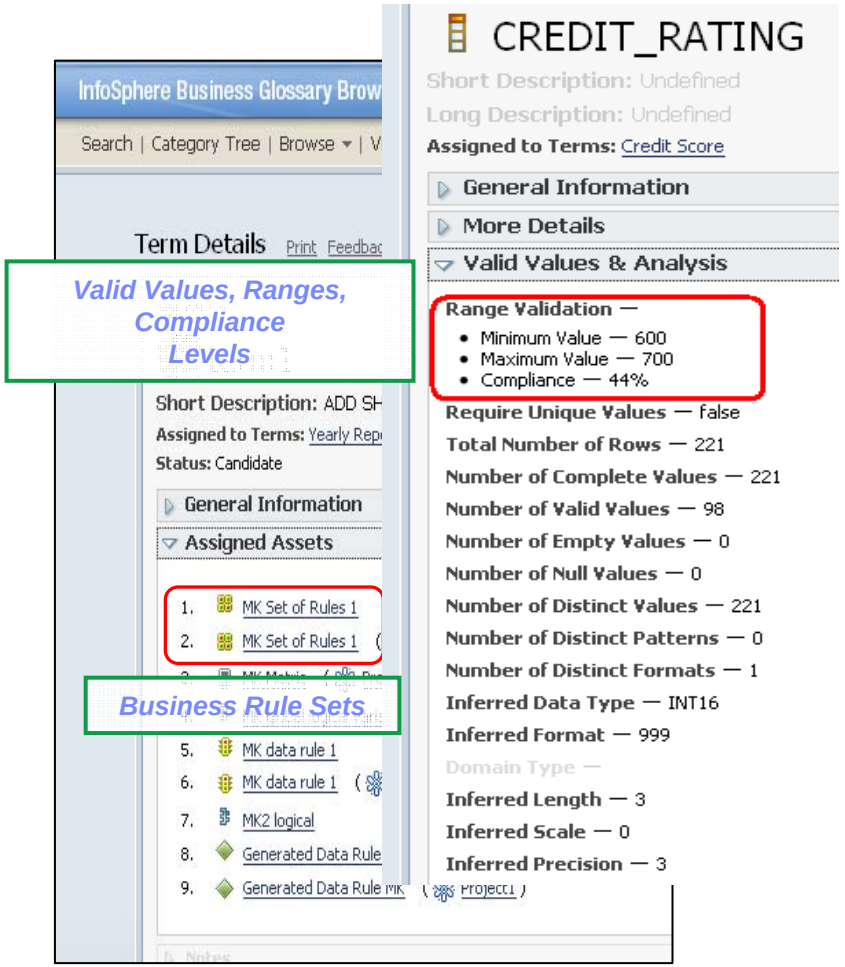
- 비즈니스 용어에 대한 신뢰감 증진 및 객관적인 이고 일반적인 용어 및 문장을 사용 하게 됨.
- 기업 차원의 비즈니스 용어 정의로 확장 가증한 Information Platform technology 지원
- Desktop 과 인터넷 연결이 가능하면, 사내 위치의 제한 없이 비즈니스 용어 조회 및 공유가 가능함.



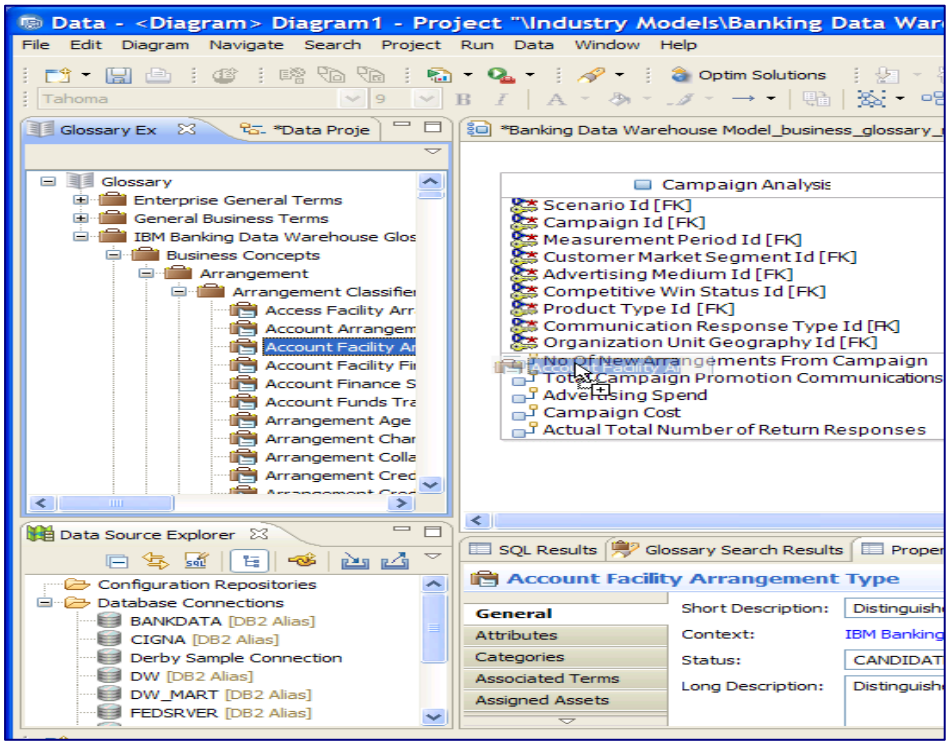
Business Glossary – What's new in 8.5?

Extend the collaboration between business and IT

- 비즈니스 용어에 대한 Rule 설정 및 Validation



- 비즈니스 용어 와 데이터 모델(IT) 연계 정의



InfoSphere Data Architect 와 Business Glossary 연관된 용어 정의

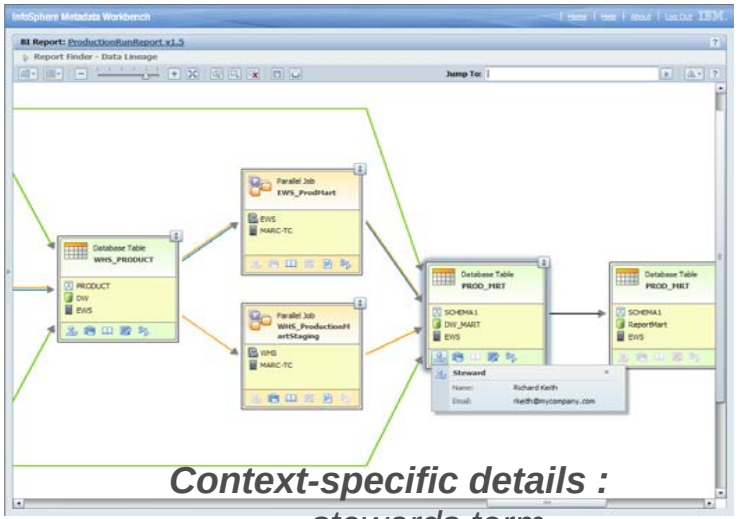
Metadata Workbench

Metadata lineage and Impact Anaysis

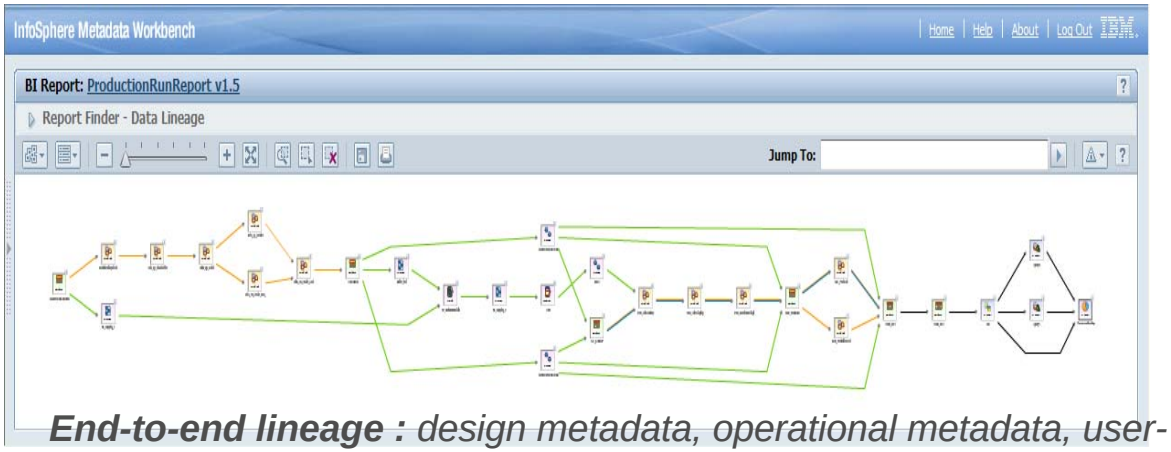
- Web 기반의 Information Assets 탐색
- Tool 간의 data movement, data lineage, business meaning, impact of changes and dependencies 보고서 작성
- Sarbanes-Oxley and Basel II 와 같이 컴플라이언스 관점의 보고서 작성

IBM Metadata Workbench®

Provides IT professionals with a tool for exploring and understanding the assets generated and used by the Information Server suite.

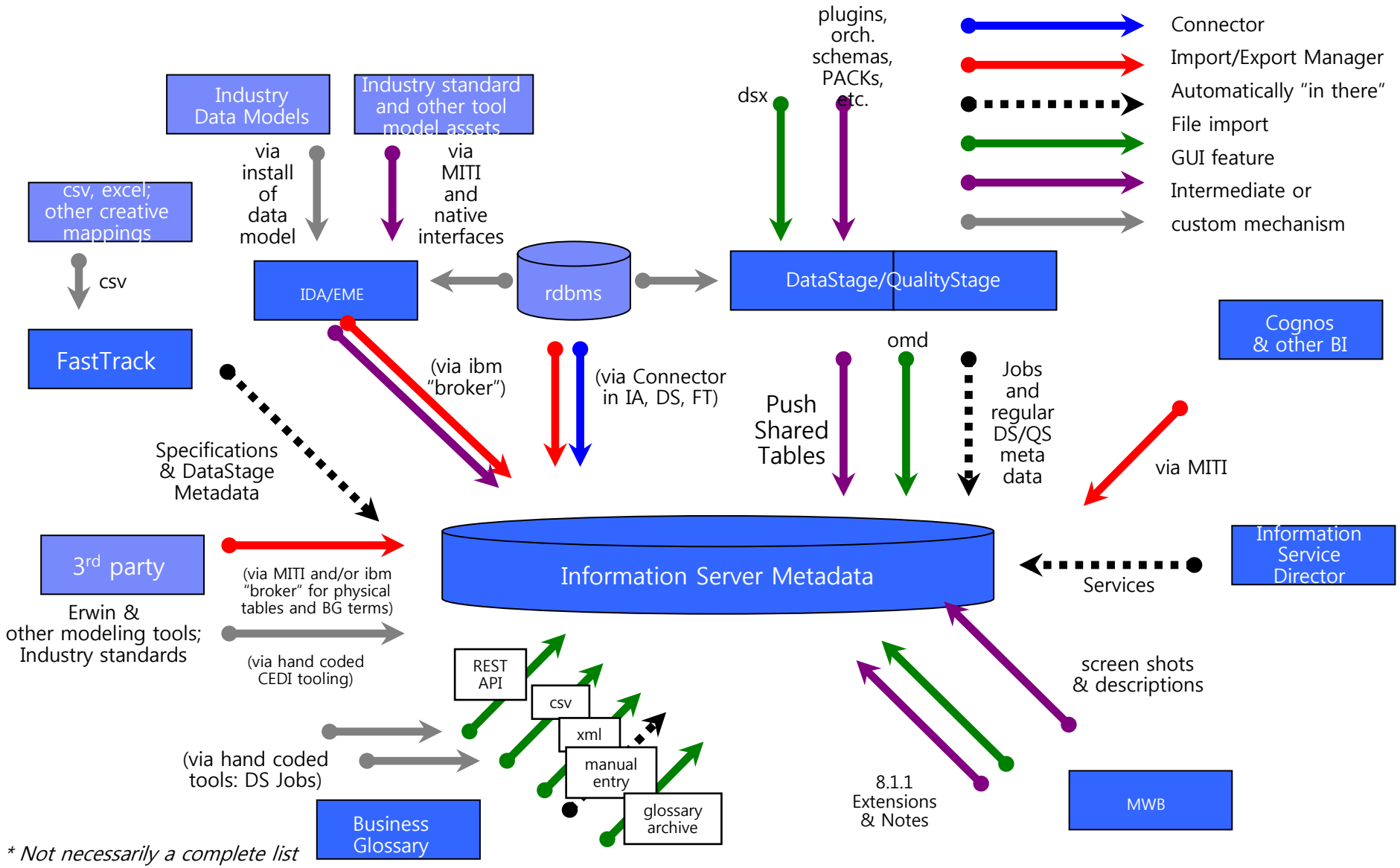


Context-specific details : *stewards,term, description,Job image, Job operational metadata details, etc.*



End-to-end lineage : *design metadata, operational metadata, user-defined metadata*

Metadata Collection



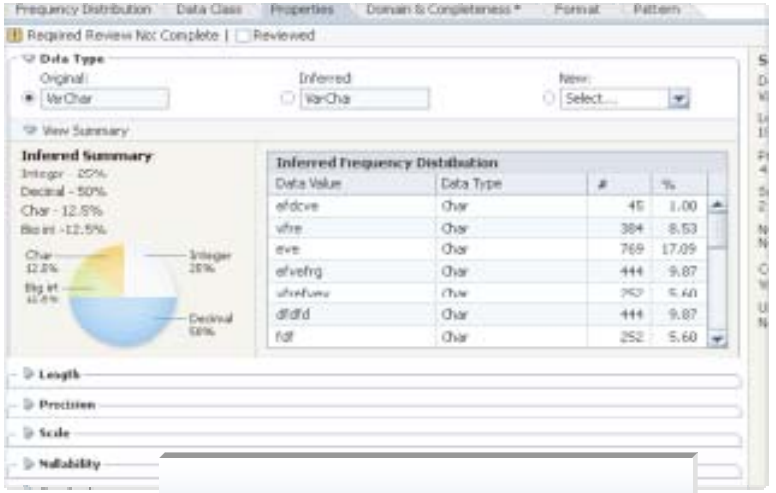
Information Analyzer

Analyze the Technical Metadata

- 애플리케이션, 데이터베이스, 파일기반 소스에 대한 데이터 및 프로세스 중심적 분석
- 컬럼, 컬럼 간, 테이블 간에 신뢰할 수 있는 상세한 프로파일링 정보 제공
- 프로파일링 결과로부터 메타데이터 생성
- 프로파일링 결과를 IBM Information Server 전체에 걸쳐 활용되도록 공유 가능
- 30여 가지의 다양한 Report Template 제공

IBM Information Analyzer

Analyze source data structures, and monitor adherence to integration and quality rules



Physical View

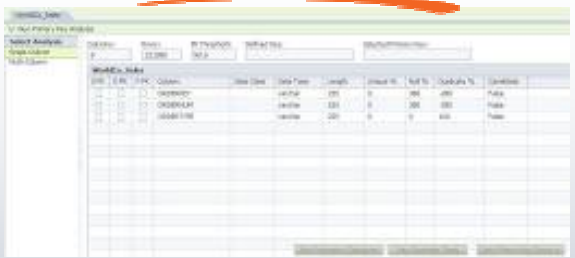
Information Analyzer – 주요 기능

Column Analysis

- *Frequency Distribution Analysis*
- *Data Classification*
- *Data Properties*
- *Formats*

도메인 값의 완전성, 정확성, 유효성

Table Analysis



- *Primary Key Analysis (single or multi columns)*
- *Key Duplicates*

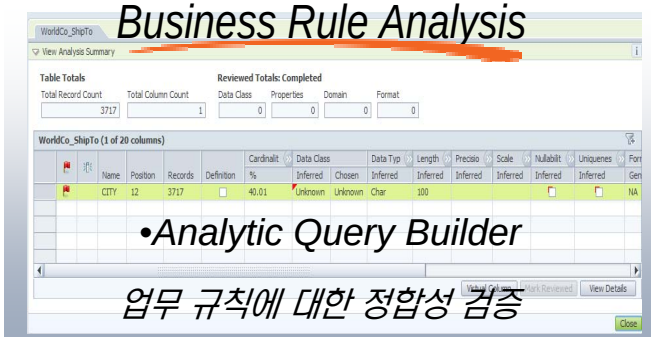
정규화 구조의 최적성

Cross-Table Analysis

- *Foreign Key & Similarity Analysis*
- *Referential Integrity*
- *Cross-Domain Relationships*
- *Data Redundancy*

참조 무결성, 데이터 중복도

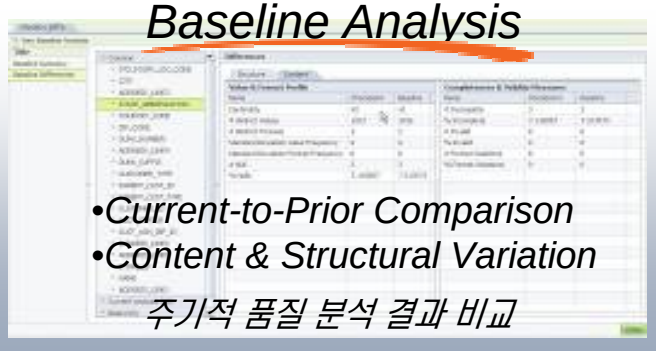
Business Rule Analysis



- *Analytic Query Builder*

업무 규칙에 대한 정합성 검증

Baseline Analysis



- *Current-to-Prior Comparison*
- *Content & Structural Variation*

주기적 품질 분석 결과 비교

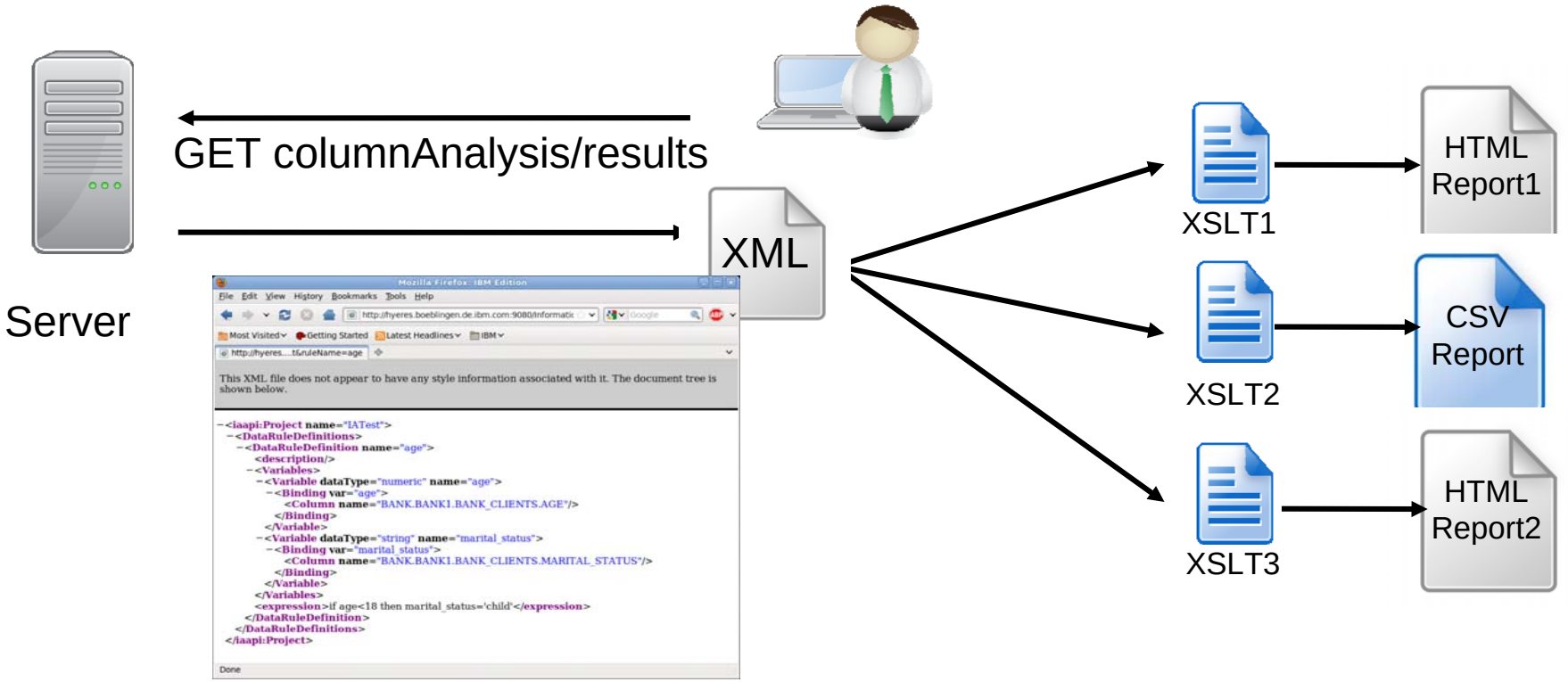
Report Templates

30여가지 보고서 기능

Information Analyzer – What's new in v8.5?

Flexible Processing and Result Publication - REST API / CLI

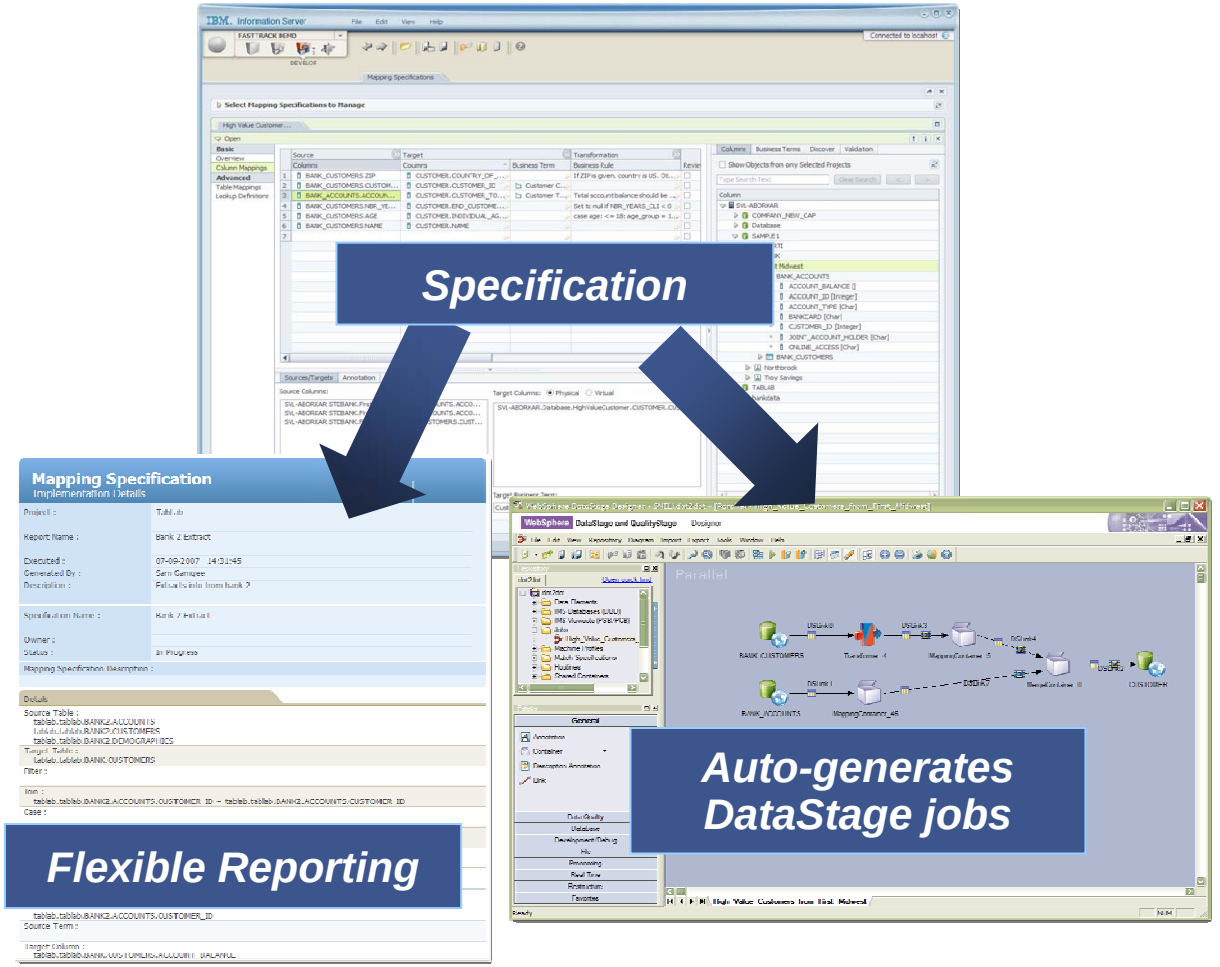
- InfoSphere Information Analyzer 결과 보고서, 규칙, project 정보를 외부데이터 형태로 제공
- 고객의 요구사항에 맞는 형태의 external or downstream (such as custom reports)
  - 제공되는 Scripts를 사용하여 Customer dashboards , or applications
  - 분석결과 내용을 XML file 로 제공하여 다양하게 사용 할수 있도록 함.



FastTrack

To reduce Costs of Integration Projects through Automation

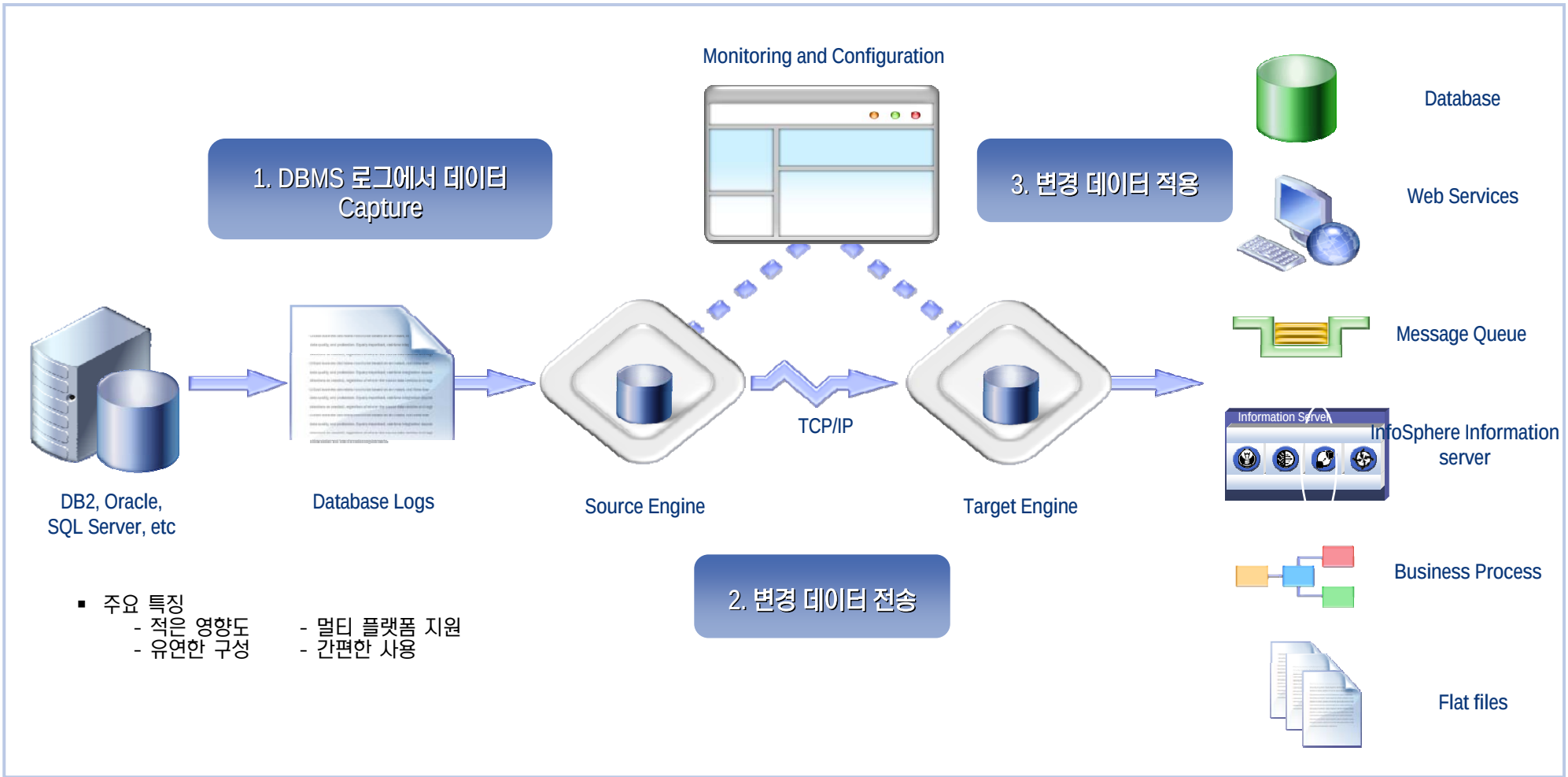
- 변환 규칙을 설계하여 비즈니스 정보를 최적화
  - -IA의 프로파일링 결과를 활용하여 업무 요건에 가장 최적화된 설계
  - -유연한 설계 규칙 제공 – 다양한 function logic 적용
- 설계된 내용을 중앙에서 검증하고 통제
- 협업 기능 강화
  - -강력한 사용자 편의성
  - -Excel로 작성된 변환 설계 import/export
  - -BG에 정의된 표준 업무 용어와 물리적인 테이블 간의 정의 및 연결
- 정보 통합 프로세스 개발을 가속화할 수 있는 솔루션



- Evolution form ETL to Data Govern
- What's New with Information Server v8.5
- How to migrate to Information Server v8
- **InfoSphere CDC (Change Data Capture)**
- Enterprise data architecture and Value of Information Server
- Demo

What is InfoSphere CDC (Change Data Capture)

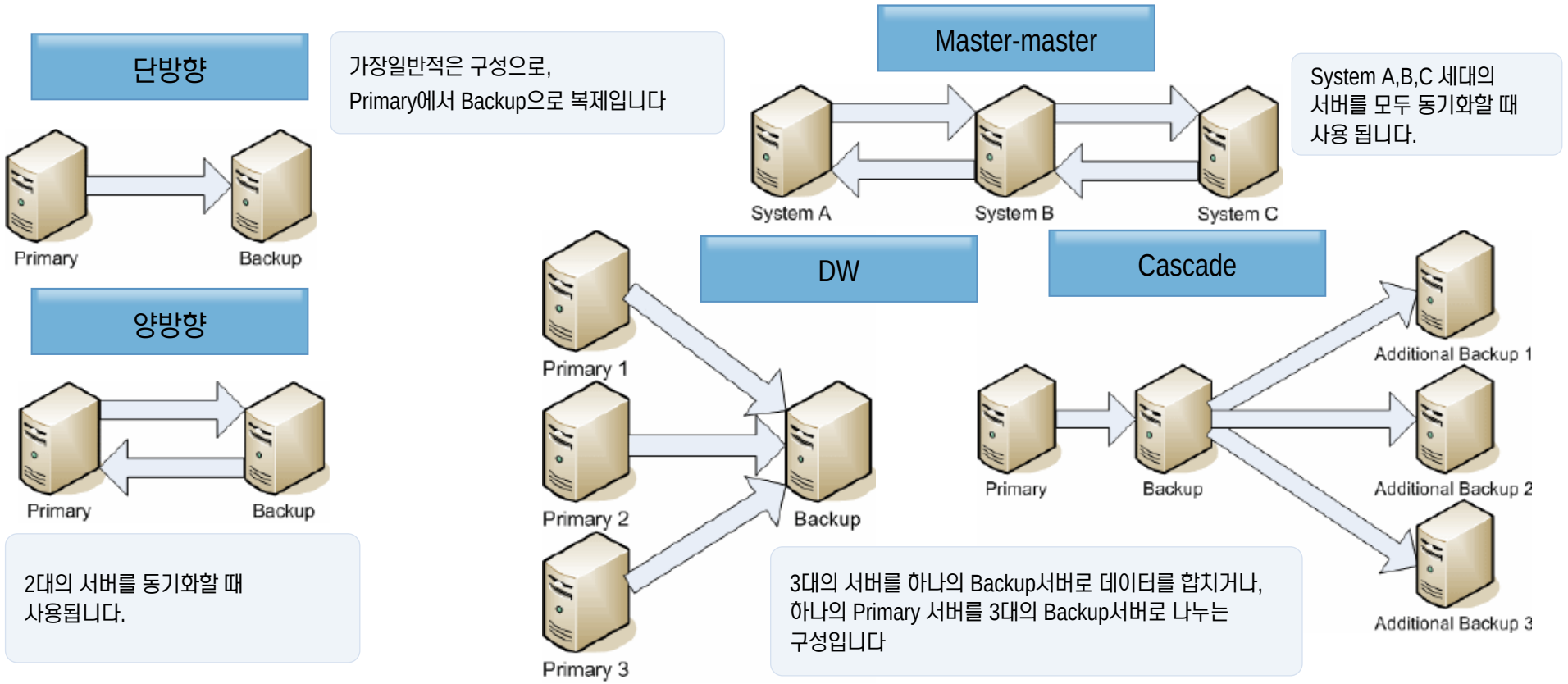
InfoSphere CDC는 다양한 DBMS에서 발생하는 데이터 변경내용을 실시간으로 Capture하여, 해당 데이터가 필요한 시스템으로 데이터를 변환 및 복제해주는 데이터통합 도구입니다.



- 주요 특징
  - 적은 영향도
  - 유연한 구성
  - 멀티 플랫폼 지원
  - 간편한 사용

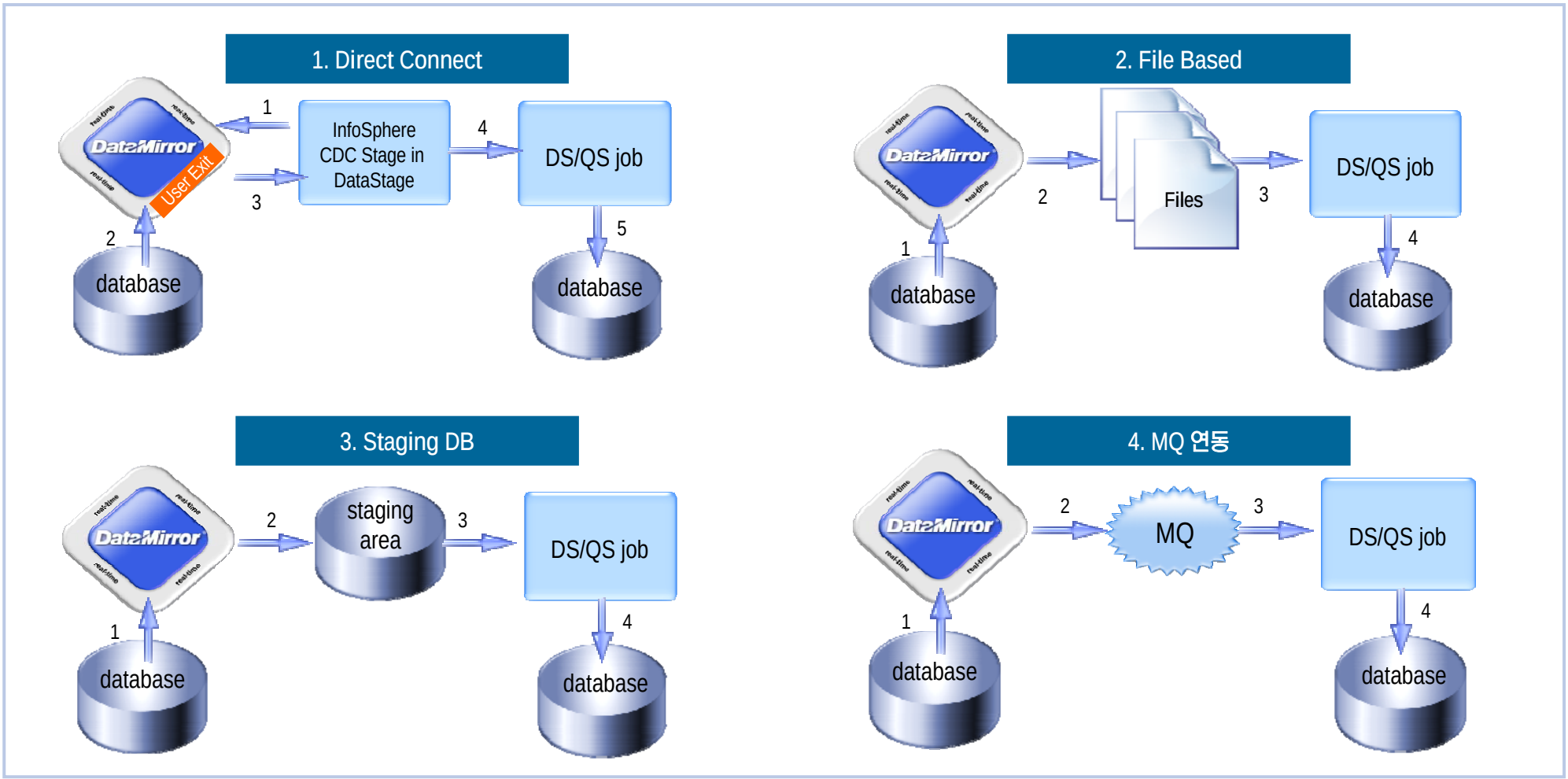
What is InfoSphere CDC (Change Data Capture)

InfoSphere CDC는 1:1, 1:N, N:1, N:N 등 다양한 형태의 복제 구성을 기본으로 하고 관리 및 운영자가 생각하시는 대부분의 복제 구성이 가능합니다. 또한, CDC는 복제 구성을 위해 별도의 전용 스크립트가 필요하지 않으며 제공되는 자바 기반의 GUI화면에서 사용방법만 습득하시면 손쉽게 복제 구성을 할 수 있습니다.



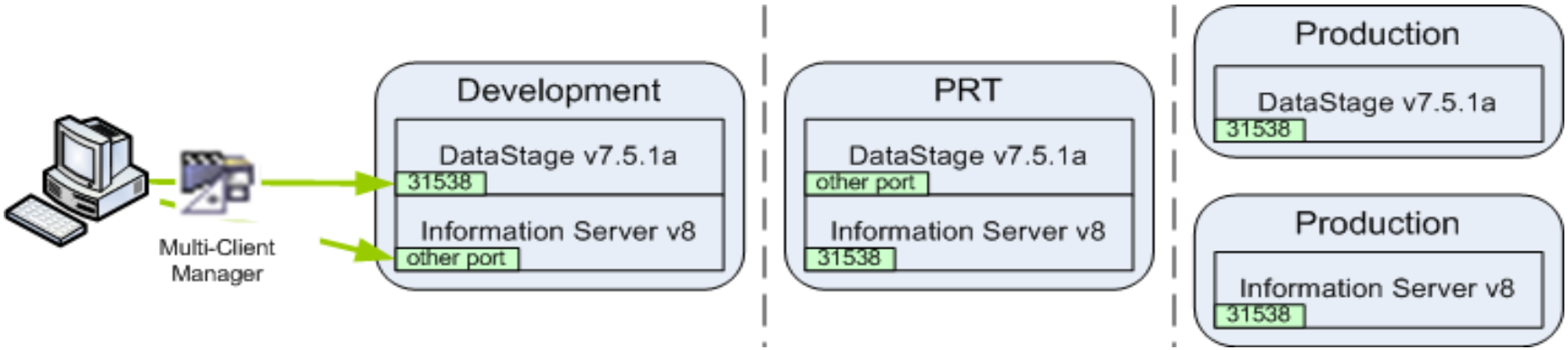
What is InfoSphere CDD (Change Data Delivery)

CDC툴인 InfoSphere CDC와 ETL 툴인 DataStage가 완벽하게 연동되어 구성의 유연성을 증대시킵니다. .



- Evolution form ETL to Data Govern
- What's New with Information Server v8.5
- InfoSphere CDC (Change Data Capture)
- **How to migrate to Information Server v8**
- Enterprise data architecture and Value of Information Server
- Demo

Upgrading to Information Server from v7 to v8



Approach

- I-Tag 설치 옵션으로 업그레이드 대상 Version을 dev 또는 test 환경으로 구축
  - 다른 version 의 DataStage Engine 이 다른 port 로 동일 서버에 공존.
  - 다른 OS user 를 통하여 각각의 DataStage 엔진을 Start / Stop 하도록 정의함.
  - Domain switch 를 통하여 동일한 "dsjob" 을 다른 버전의 엔진에서 실행 하도록 함
- 여러 개의 DataStage Client 를 관리할 수 있는 Multi-client manager 사용.
- 단계별(Phased) migration of projects.
- DataStage Designer import / export command 를 사용한 Repository Migration.
- IBM 에서 제공하는 technical guidance / support 조직 활용.

Upgrading to Information Server from v7 to v8

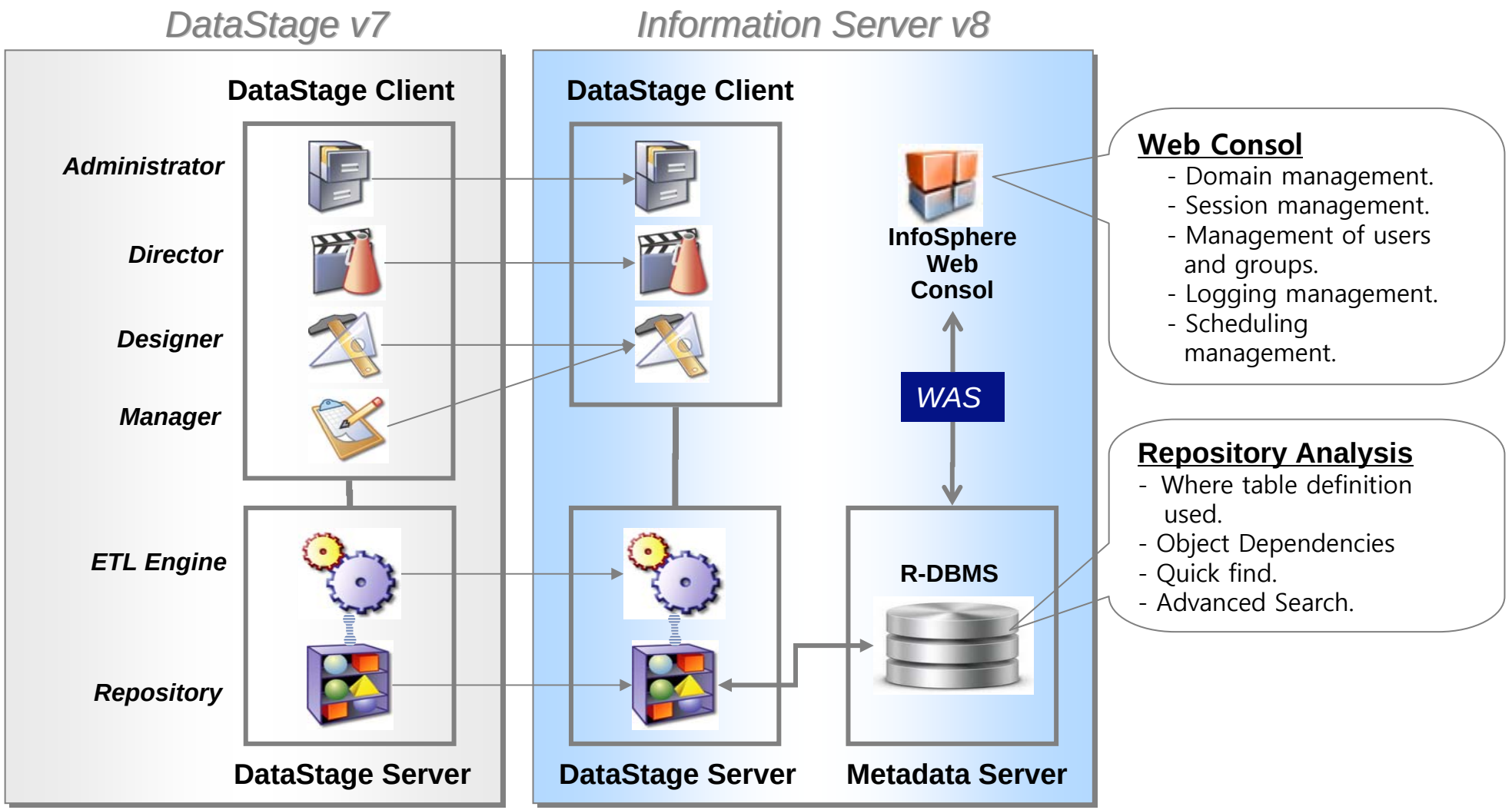
Lessons and Learned

- Planning is key including
  - Topology (적용 위치)
  - Capacity (수용 용량)
  - DR requirements (Recovery).
- 테스트 계획 작성이 중요하며, 회귀 테스트(regression test) 정의
  - 실제 데이터에 근간한 테스트 수행
  - 기능 항목별 테스트 수행.
- 복잡한 아키텍처 - DataStage v8 is more complex than v7 : WAS and R-DBMS repository.
- Install 과정 중에 Upgrade 하는 것 보다는 (I-Tag) 을 사용한 Test or dev 환경을 이용한 upgrade to v8.1.
- Upgrade 전에 복구를 위한 소스 코드 Back-up 및 Export 등의 준비.
- "big bang" 접근방식 보다는 단계별(Phased approach) 를 권장함.

- Evolution form ETL to Data Govern
- What's New with Information Server v8.5
- InfoSphere CDC (Change Data Capture)
- How to migrate to Information Server v8
- **Enterprise data architecture and Value of Information Server**
- Demo

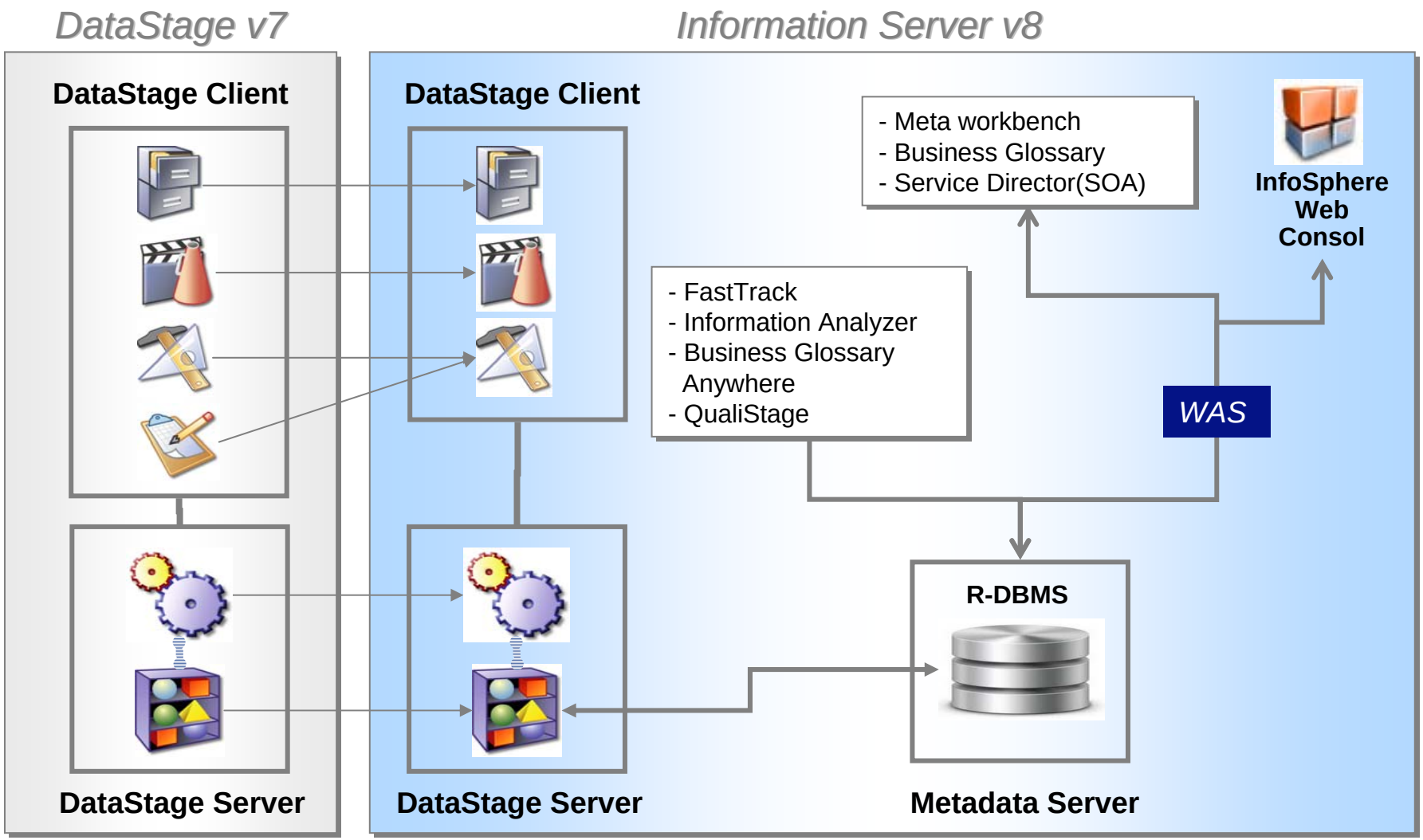
Enterprise Architecture and Value of Information Server

ETL (DataStage) 업그레이드



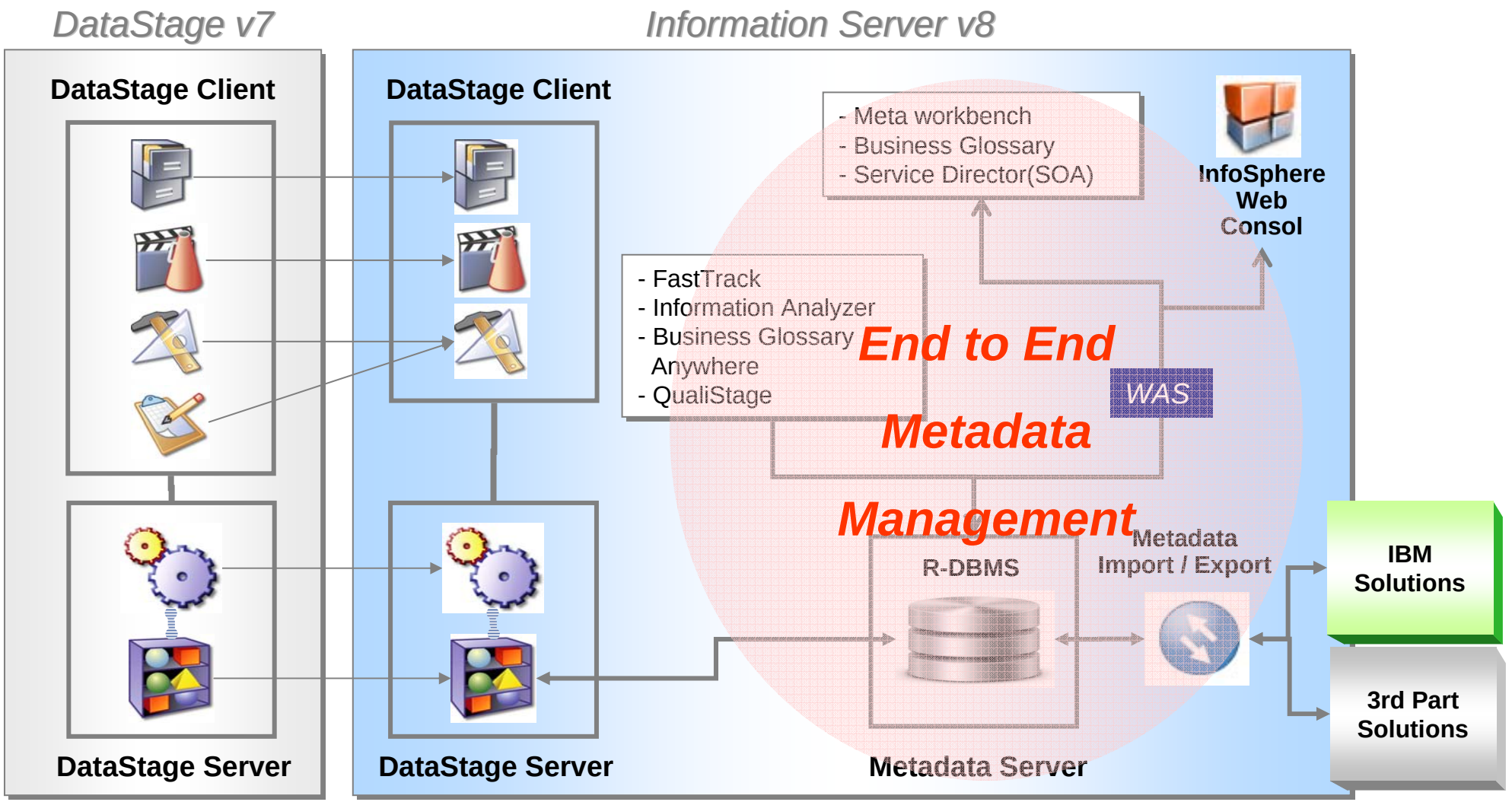
Enterprise Architecture and Value of Information Server

InfoSphere Information Server 도입



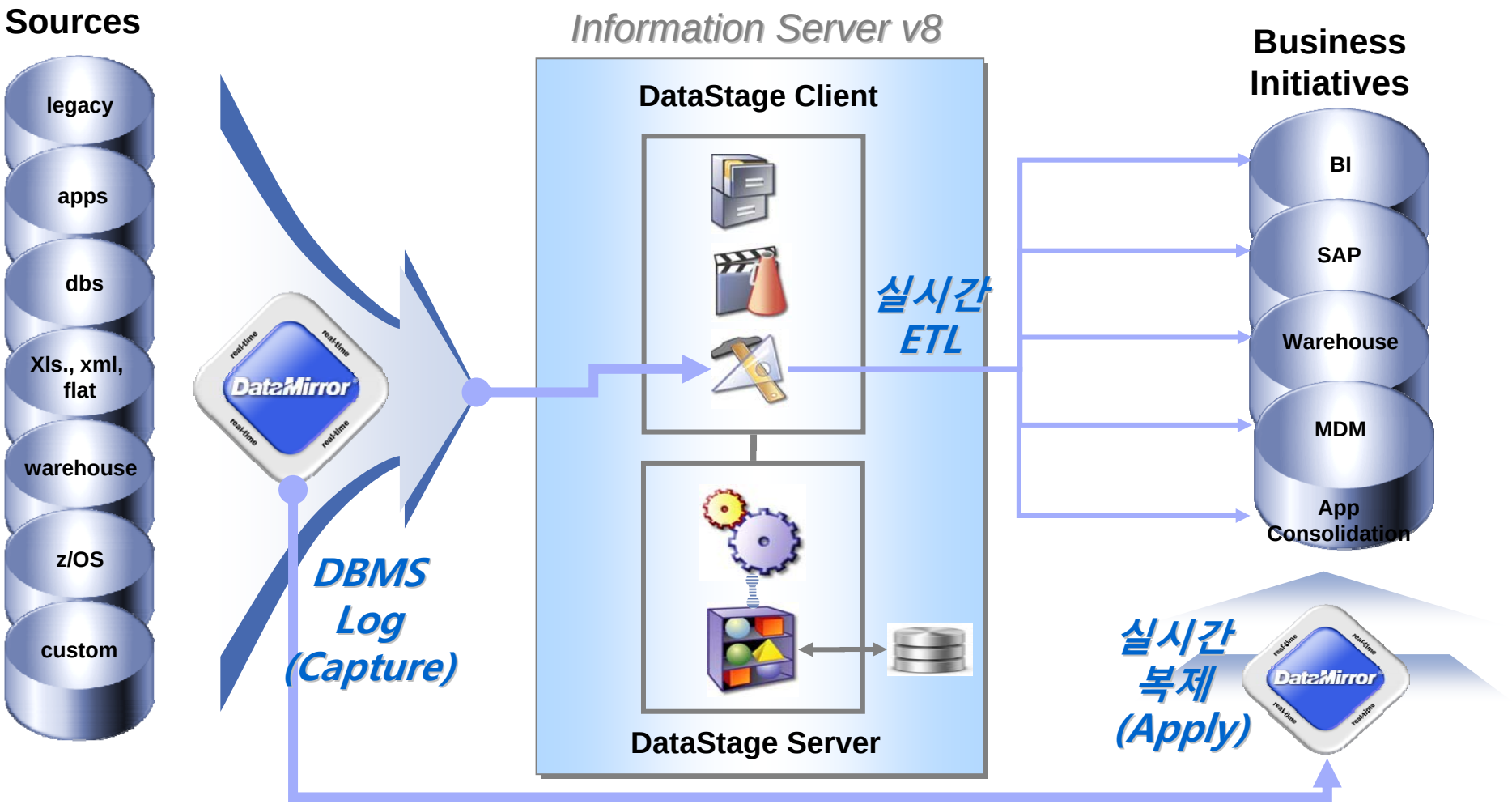
Enterprise Architecture and Value of Information Server

InfoSphere Information Server 도입



Enterprise Architecture and Value of Information Server

실시간 데이터 통합 아키텍처 구현



- Evolution form ETL to Data Govern
- What's New with Information Server v8.5
- How to migrate to Information Server v8
- InfoSphere CDC (Change Data Capture)
- Enterprise data architecture and Value of Information Server
- **Demo**