



InfoSphere™
software

Trusted Information

IBM InfoSphere technológiák

Az információ integrációs eszközök teljes skálája

IBM Üzleti Intelligencia Fórum
2009. Június 3..

Temesi Gergely
gergely.temesi@hu.ibm.com

Delivering Trusted Business Information

Financial Risk Insight *Workforce Optimization* *Dynamic Supply Chain*

Customer & Product Profitability *Multi-channel Marketing*



BUSINESS VALUE

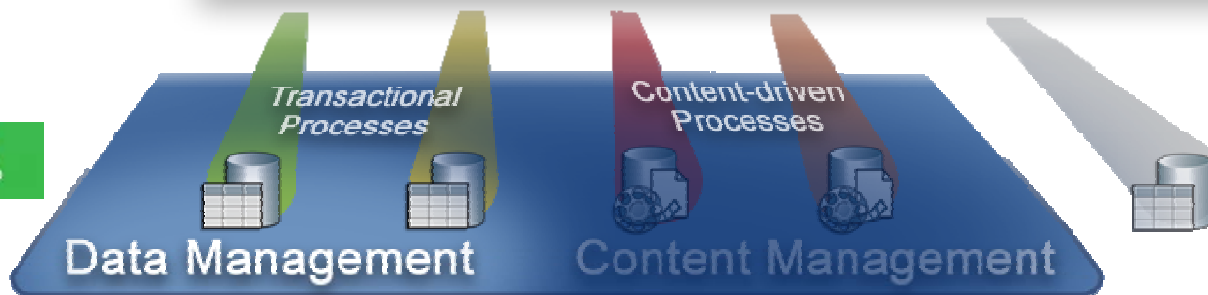
InfoSphere™ software



Accurate & Trusted View,
Managed Over Time

UNLOCK

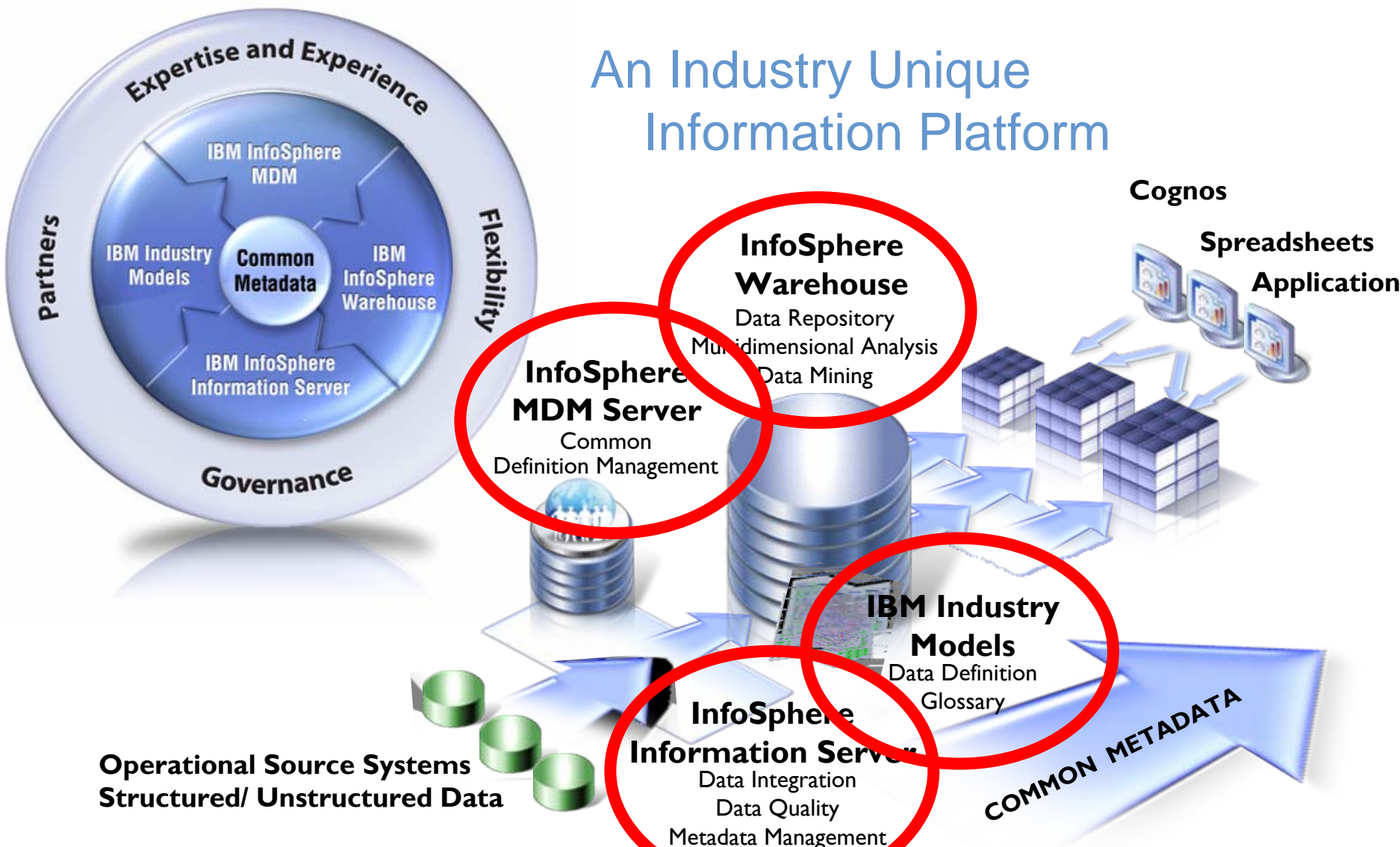
DB2.



INFORMATION ASSETS

The InfoSphere Vision

An Industry Unique Information Platform

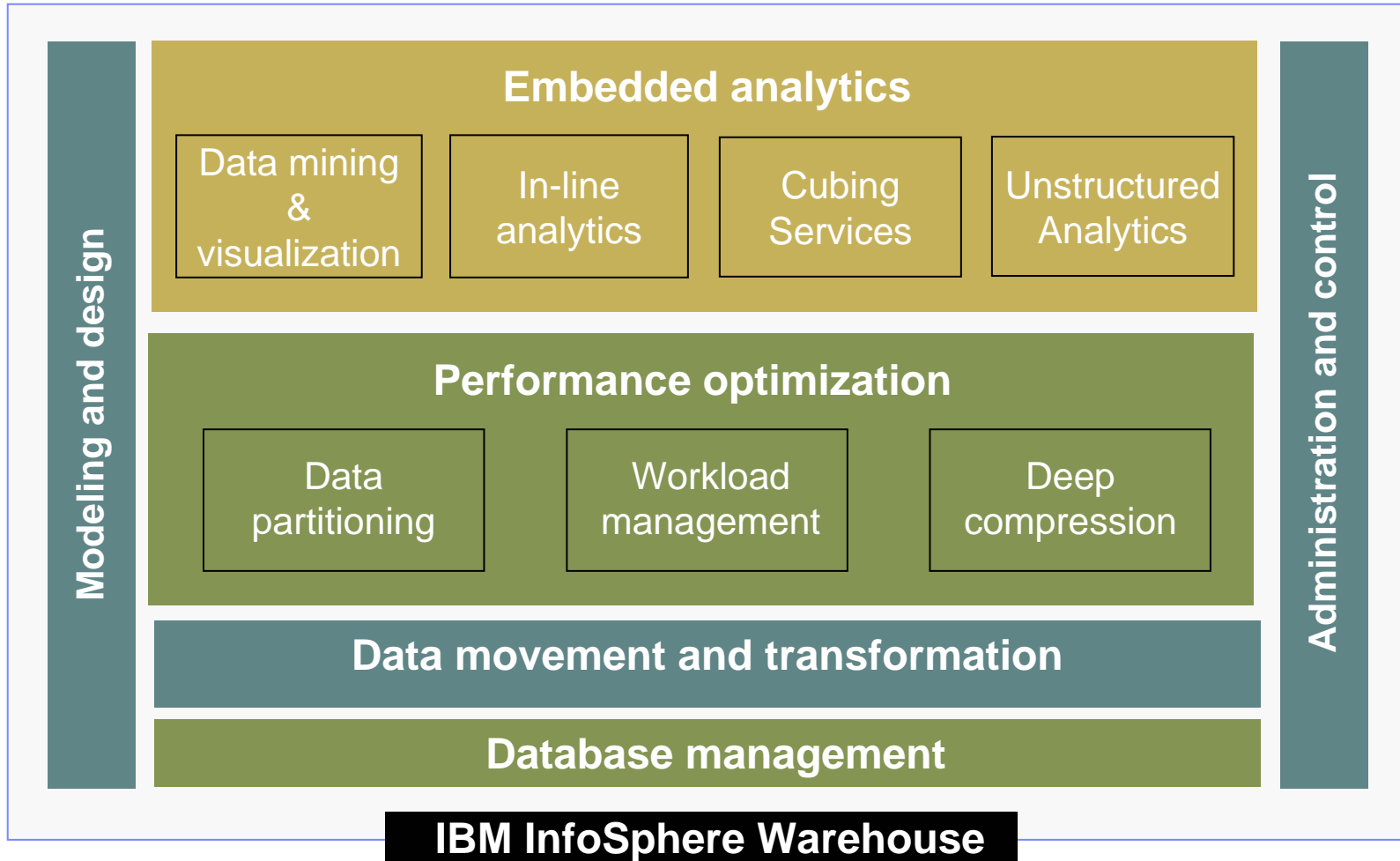


InfoSphere Warehouse



IBM InfoSphere Warehouse (former DB2 Warehouse Edition)

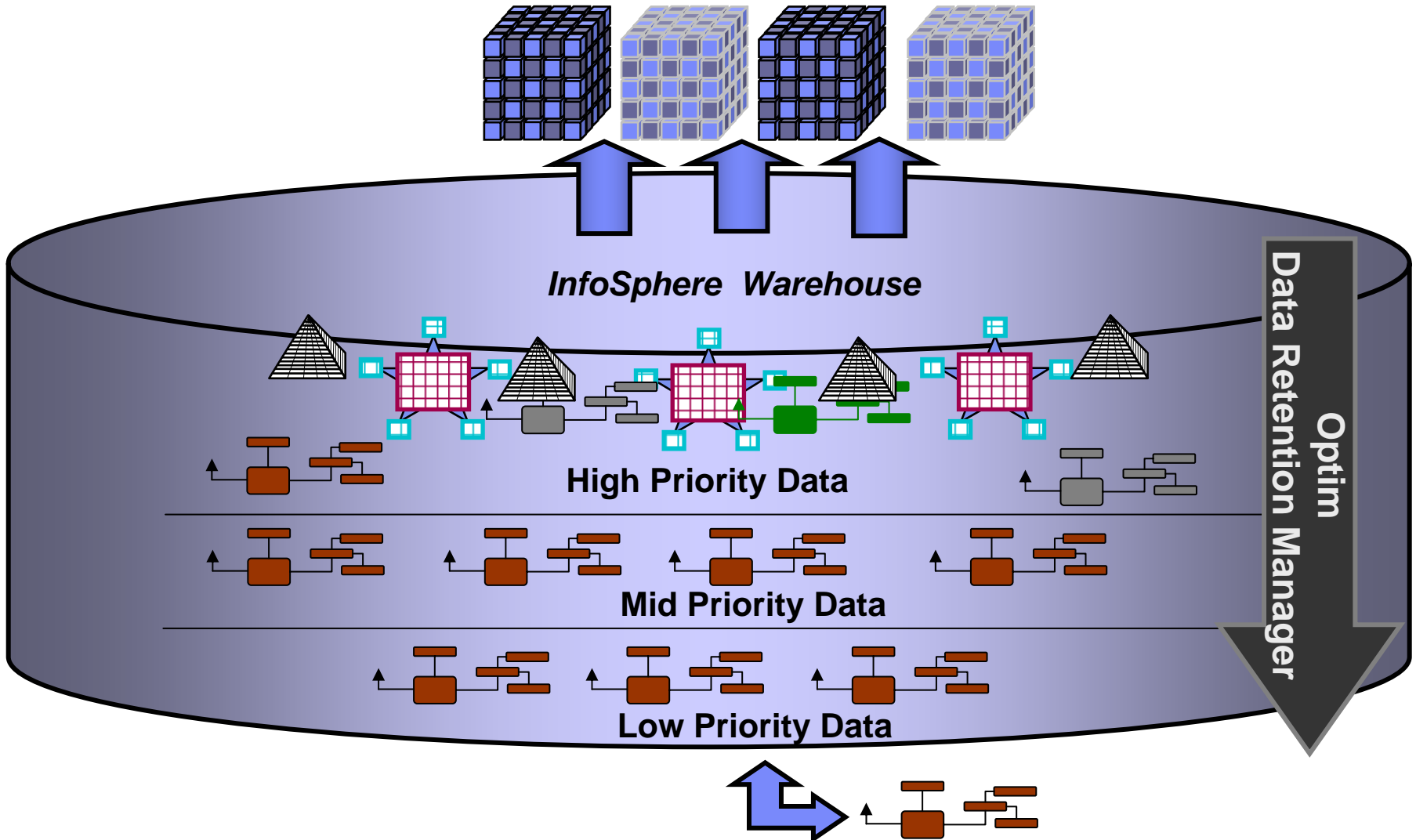
A complete, integrated platform based on DB2 UDB



(+ Cognos 8 BI Starter Edition available in all InfoSphere Warehouse editions)

Information Lifecycle Management

The multi-temperature approach



IBM Industry Models



IBM Industry Models



Banking

(Data, Process and Services Models)

- Profitability, Relationship Marketing
- Risk Management
- Asset and Liability Mgmt
- Compliance
- Business Process re-engineering



Insurance

(Data, Process and Services Models)

- Customer centricity
- Claims, Policy, Underwriting
- Intermediary Performance
- Compliance
- Risk Management
- Business Process Re-engineering



Financial Markets

(Data, Process and Services Models)

- Risk Management
- Asset and Liability Mgmt
- Compliance
- KYC and Account Opening
- Middle/Back Office Transformation



Health Plan

(Health Plan Data Warehouse)

- Claims
- Medical Management
- Provider and Network
- Sales, Marketing and Membership
- Financials



Retail

(Retail Data Warehouse)

- Customer centricity
- Merchandising Management
- Store Operations & Product Mgmt
- Supply Chain Management
- Compliance



Telco

(Telecommunications Data Warehouse)

- Churn Management
- Relationship Mgmt and Segmentation
- Sales and Marketing
- Service Quality and Product Lifecycle
- Usage Profile

The "Understanding" of the Organization varies



Business Users



Data Architect



DBA



Developer

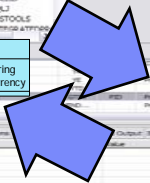
Business Language



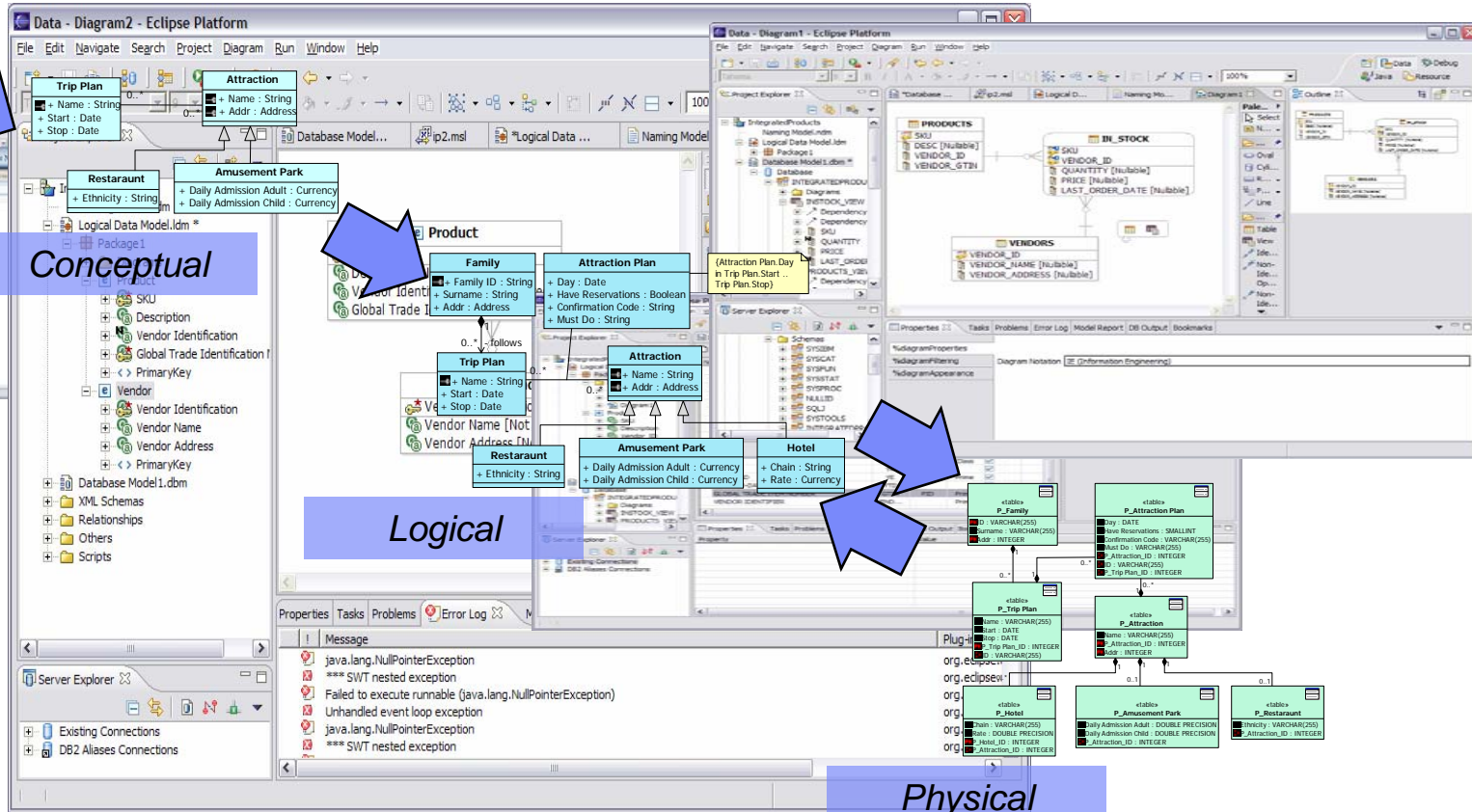
Conceptual



Logical



Physical





Data Modelling – InfoSphere Data Architect

- Data modeling & discovery for data structures and federations
- Metadata relationship discovery & mapping
- Impact analysis, and synchronization across models
- SQL & XML generation capabilities
- Glossary model, Domain model, Logical and Physical data model...

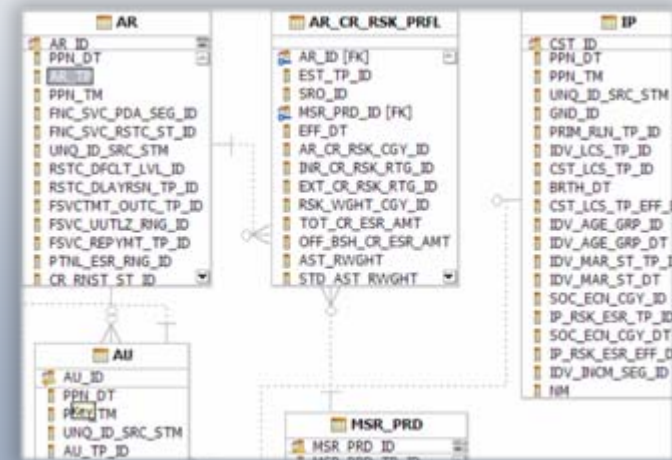


Subject Matter Experts



Architects

Rational Data Architect



InfoSphere Data Architect

The screenshot displays the IBM InfoSphere Data Architect interface. The main workspace shows a data model for an 'Invoice' project. The 'Data Project Explorer' on the left shows a tree view with 'Invoice' as the selected project, containing 'Mappings', 'XML Schemas', 'Data Diagrams', and 'Data Models'. The 'Database Explorer' at the bottom left shows various database connections. The central workspace displays a diagram with entities: 'Invoice Term' (with attributes 'Invoice ID [FK]', 'Term Type Code [FK]', and 'Attribute2'), 'Term Type' (with attributes 'Term Type Code' and 'Term Type Description'), and 'Invoice ID' (with attributes 'Invoice ID', 'Invoice Date', and 'Invoice Description'). Relationships are shown between 'Invoice Term' and 'Term Type', and 'Invoice Term' and 'Invoice ID'. A 'Part of' relationship is also visible. The 'Properties' window at the bottom right shows the details for the '<Entity> Term Type' entity, including a table of attributes.

Name	Primary ...	Surrogate Key	Type	Length/Precision	Scale
Term Type Code	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Code [CHAR(5)]		
Term Type Description	<input type="checkbox"/>	<input type="checkbox"/>	Description [VARCH...		

InfoSphere Information Server



Information Integration layer: IBM InfoSphere Information Server

Delivering information you can trust

IBM InfoSphere Information Server

Understand



Discover, model, and govern information structure and content

Cleanse



Standardize, merge, and correct information

Transform



Combine and restructure information for new uses

Deliver



Capture, virtualize and move information for in-line delivery

Platform Services

Parallel Processing Services



Connectivity Services



Metadata Services



Administration Services

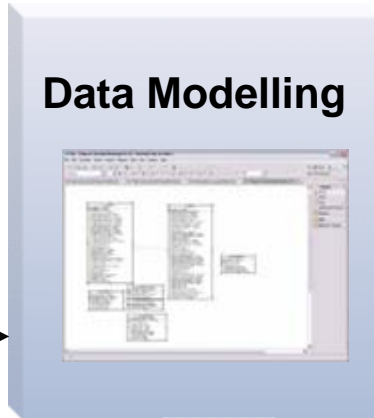


Deployment Services



- **Unified security, user administration, logging, reporting, monitoring, metadata, connectivity, scheduling etc. across modules**

Information Server for Building the Data Warehouse



*Import/
Export*

MetaData Repository

Active Cross-Platform Administration, Management and Reporting



Business Glossary and Business Glossary Anywhere

• Features

- Create and manage a common business vocabulary along with relationships to different IT assets
- Web based interface shared across enterprise business teams
- Real-time and easy access to Business Glossary from any desktop applications with Business Glossary Anywhere
- Allows creation of stewards & assignment of their responsibilities for terms & assets.
- Link business terms to technical assets

• Benefits

- Aligns IT efforts with business goals
- Provides business context to information technology assets

Business Analyst

Data Analyst

IBM InfoSphere Business Glossary

Welcome to JKLW Business Glossary

Term Details

Line of Business

Short Description: X Life & Wealth division

Long Description: BANK and insurance are called divisions of X Life & Wealth. A division is responsible for the creation of the organization's products and/or services. It is not IT, HR and Accounting.

Metadata: Data Analyst (S. Commander) Date Stamp

General Information

Related IT Assets

IBM Information Server

Information

IBM Solutions - Information Server Platform

Status: STANDARD

The Information Server Platform supports data governance processes by allowing enterprises to effectively track and audit how data moves throughout an organization. By having this level of detail, tracing specifics about what happens to a piece of information is readily accessible for auditing purposes.

General Information

Synonyms: None

© 2009 IBM Corporation

InfoSphere Business Glossary

The screenshot shows the IBM Information Server WebSphere Business Glossary interface. At the top, it says "Welcome to WebSphere Business Glossary" and includes navigation links for Search, Category Tree, Stewards, About, Help, and Log Out. The main content area is divided into two panels. The left panel is a tree view of categories, with "Customer Product Summary" selected and highlighted in blue. The right panel displays the details for the selected category, including a description and a list of terms.

Category Tree:

- Classification
 - Customer
 - Customer Product Summary**
 - Customer Summary
 - Customer Attriton Analysis
 - Customer Interaction Analysis
 - Customer Lifetime Value Analysis
 - Customer Trading Profile
 - Enterprise
 - Finance
 - General Ledger
 - Japan
 - United Kingdom
 - United States
 - Transaction
 - Group
 - Manufacturing
 - Procurement

CATEGORY DETAILS:

Customer Product Summary

Customer / Product Summary is the summary of Products sold to or used by Customers of the Financial Institution.

TERMS:

- Activity Fees
- Communications And IT
- Exchange Rate To Original Currency
- Interest Expense
- Interest Income
- Management Fees
- Number Of Arrangements
- Number Of New Arrangements
- Number Of New Arrangements from Cross Sell
- Origination And Renewal Fees
- Other Non-Interest Expense
- Other Non-Interest Income
- Risk Adjustment
- Total Arrangement Closing Balance
- Total New Funds From Cross Sell
- Waived Fees
- Waived Interest Expense
- Waived Interest Income
- Waived Non-Interest Expense

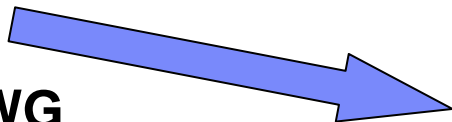
InfoSphere Business Glossary Anywhere

Name: Gergely Temesi

Status: Active

Department: SWG

...





The screenshot shows a window titled "BG Anywhere" with a search bar containing the word "active". The search results display the following information:


- Employee » Active
- Active**
- Describes an employee who is currently receiving insurance coverage.
- Steward: Richard Smith
- Status: STANDARD
- General Information**
- Long Description**
 - ↳ Active refers to an employee's insurance coverage status. It does not imply their employment status.
- Usage**
 - ↳ Active employees each receives monthly insurance newsletters.

The IBM logo is visible in the bottom left corner of the window, and a green checkmark icon is in the bottom right corner.

InfoSphere Business Glossary

Term Details [Print](#)

 [Customer](#) »  [Customer Identifier](#)











 **Customer Identifier**

The unique identifier of the Customer, identical to that of the equivalent Involved standardized database table.

▶ **More Information**

▼ **Related IT Resources**

Toggle List / Paragraph Display

- ▣ [ccb_customer1](#) ( [GMP2EDW](#) »  [VTAS](#) »  [DQA_POC](#))
- ▣ [ccb_customer1 cleansed](#) ( [GMP2EDW](#) »  [VTAS](#) »  [DQA_POC](#))
- ▣ [customer_id](#) ( [GMP2EDW](#) »  [VTAS](#) »  [DQA_POC](#) »  [ccb_customer1](#))
- ▣ [Database VTAS\DQA_POC\ccb_customer1](#)
- ▣ [Database VTAS\DQA_POC\ccb_customer1 cleansed](#)

▶ **Additional Comments**



Data Profiling - InfoSphere Information Analyzer

- **Provides in-depth analysis of existing systems**
 - Identify the best data sources
 - Analyze fields and relationships across fields and across sources
 - Highlight data that may need downstream attention
- **Enables ongoing measurement and baseline reporting of information quality**
- **Creates metadata that describes where information is managed across systems**

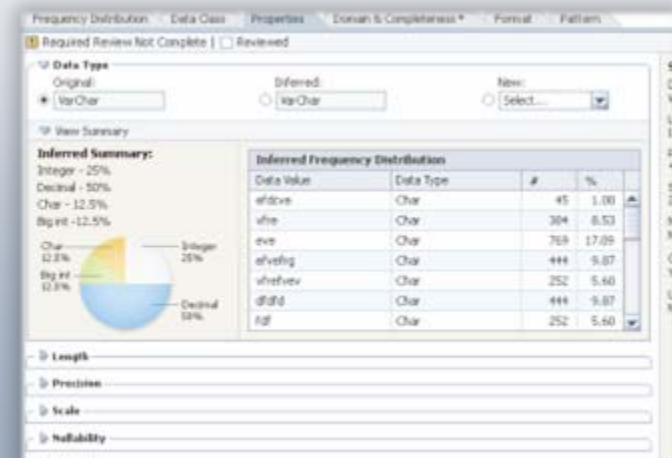


Subject Matter Experts



Data Analysts

Information Analyzer



InfoSphere Information Analyzer

Select Data Source to Work With

WorldCo_BillTo WorldCo_ShipTo

Open Foreign Key Analysis

View Referential Integrity Analysis

1 You can use this pane to view the results of a referential integrity analysis job. You can accept the foreign key inference or remove the foreign key inference from the repository.

Base Data Source: GlobalCo_WorldCo
 Base Table: WorldCo_BillTo
 Primary Key: BILLTO_CUSTOMER

Paired Data Source: GlobalCo_WorldCo
 Paired Table: WorldCo_ShipTo
 Foreign Key Candidate: BILLTO_CUSTOMER

	Distinct Values	Total Records
Violations	2	2
Violations %	0.05380683	0.05380683
Integrity	3715	3715
Integrity %	99.9462	99.9462
Total Values	3717	3717
Total %	100%	100%

	Distinct Values	Total Records
Values that	3715	3715
Match %	100	100
Values that	0	0
% of Values	0	0
Total Values	3715	3715
Total %	100%	100%

Category	Count
FK > PK	2
PK > FK	3715

Foreign Key Value	Records	Percentage
C1013	1	0.027
C7651	1	0.027

Close Foreign Key Status



Source/target mapping - InfoSphere FastTrack

- **Create required transformation rules for source-to-target historical mapping documentation**
- **Design transformation rules to optimize business information**
 - Leverage Information Analyzer profiling results for most complete requirements
 - Flexible rules support textual descriptions or function logic
- **Centrally track design decisions for auditing**
- **Enhance collaboration**
 - Easy to use desktop interface or import work from Excel spreadsheets
 - Define and link InfoSphere Business Glossary terms to physical structures
- **Accelerate development of InfoSphere integration jobs**

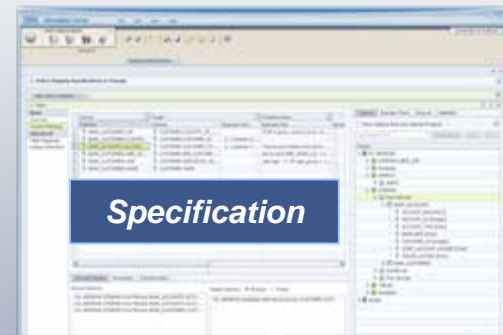


Subject Matter Experts

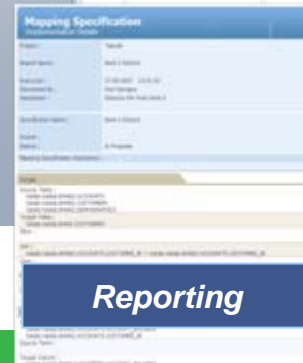


Data & Business Analysts

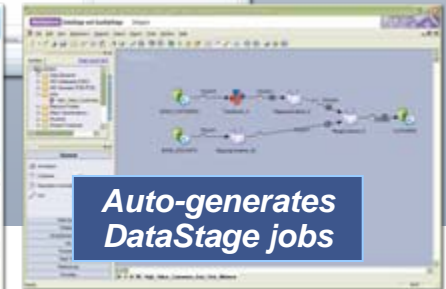
InfoSphere FastTrack



Specification



Reporting



Auto-generates DataStage jobs

InfoSphere Fasttrack

The screenshot displays the IBM Information Server FastTrack Mapping Editor. The main window is titled 'Extract from Bank1' and contains a 'Mapping Editor' pane. This pane features a table with columns for Source, Target, Business Term, Rule, and Reviewed. Below the table are tabs for 'Sources and Targets', 'Annotation', 'Transformation', and 'Discover'. The 'Sources and Targets' tab shows source and target column selections with radio buttons for 'Physical' and 'Candidate'. At the bottom, there are input fields for 'Source Business Term' and 'Target Business Term', each with a 'Create...' button. To the right of the main editor is a 'Database' pane showing a tree view of the 'banking_host' database structure, including tables like 'BANKING', 'BANK', and 'CHECKING', with columns like 'ACCOUNT_BALANCE' and 'CUSTOMER_ID'. Below the database pane is a 'Properties' pane showing metadata for the selected column, such as 'Inferred Length: 10', 'Inferred Precision: 8', and 'Inferred Nullability: true'. The bottom status bar shows 'Mapping Specifications' and the local host 'localhost/demohawk'.

Source	Target	Business Term	Rule	Reviewed
CHECKING.ACCOUNT_BALAN	CUSTOMERS.ACCOUNT_BALA	Account Balan		<input type="checkbox"/>
CHECKING.ADDR1	CUSTOMERS.ADDR1	Address [WB		<input type="checkbox"/>
CHECKING.ADDR2	CUSTOMERS.ADDR2	Address [WB		<input type="checkbox"/>
CHECKING.CITY	CUSTOMERS.CITY	City [WBG]		<input type="checkbox"/>
CHECKING.CUSTOMER_ID	CUSTOMERS.CUSTOMER_ID	Customer ID [<input type="checkbox"/>
	CUSTOMERS.GENDER	Gender [WBG]	Set to null	<input type="checkbox"/>
	CUSTOMERS.LEVEL	Level [WBG]	Set to null	<input type="checkbox"/>
CHECKING.NAME	CUSTOMERS.NAME	Name [WBG]		<input type="checkbox"/>
	CUSTOMERS.ONLINE_ACCESS	Has Online Acc	Set to null	<input type="checkbox"/>
CHECKING.STATE	CUSTOMERS.STATE	State [WBG]		<input type="checkbox"/>
CHECKING.SS_NUM	CUSTOMERS.TAX_ID	Tax ID [WBG]		<input type="checkbox"/>
	CUSTOMERS.YEARS_CLIENT	Number of Yea	Set to null	<input type="checkbox"/>
CHECKING.ZIP	CUSTOMERS.ZIP	ZIP [WBG]		<input type="checkbox"/>

Auto Generated DataStage job

Parallel

Todo's

Source Column(s)	Target Column(s)	Business Rule(s)
DSLink2.C0_ADDR1 DSLink3.C0_ADDR2	M0_STREET_NAME M0_STREET_SUFFIX HOUSE_NUMBER	Standardize with USADDR.SET ruleset. Use first character of house number at blocking column in match. Street name very important.
	STATE	Standardize with USAREA.SET.
	MIDDLE_NAME FIRST_NAME LAST_NAME	Standardize with USNAME.SET ruleset. Use first character of first name as blocking column for match. First name and last name are very important for match.
	TAX_ID	standardize with USTAXID.SET. Very important for match.
	CITY	Standardize with USAREA.SET.
DSLink2.C0_ZIP	M0_ZIP	Standardize with USAREA.SET. Add first three digits as blocking column for match.

Parallel

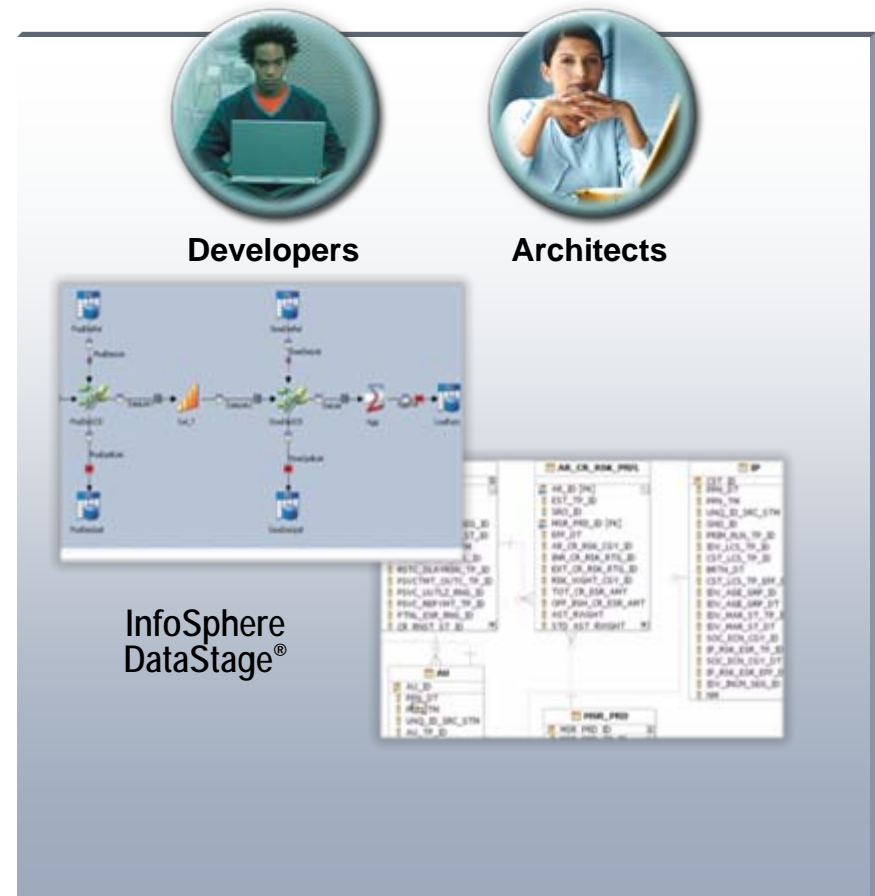
Specification Name: Identify_High_Value_Customer_tab

Generation Date: Sun Jan 06 11:35:09 PST 2008

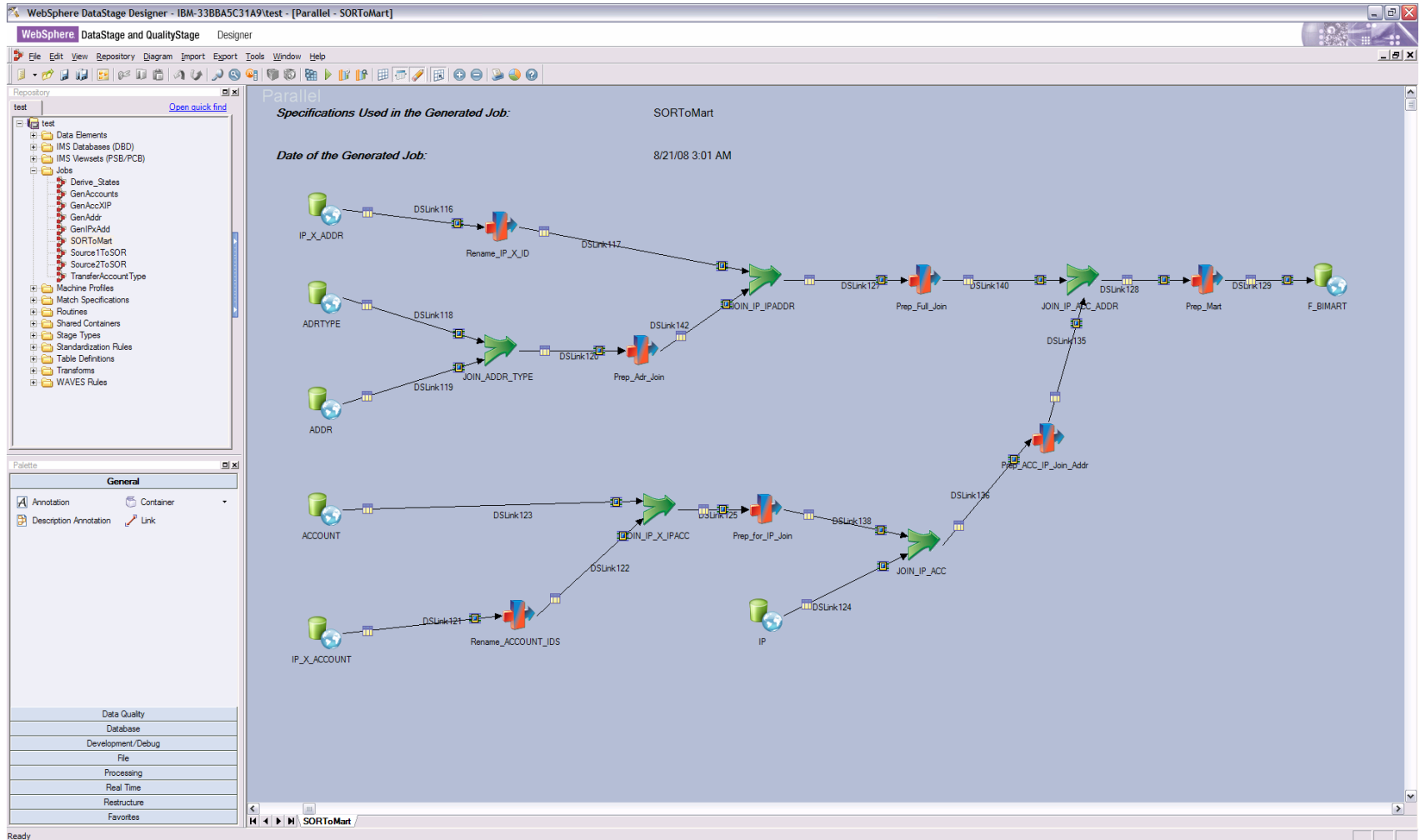


ETL – InfoSphere DataStage

- Transform and aggregate any volume of information
- Deploy in batch, real time, or on demand through visually designed logic
- Complete ETL functionality with metadata-driven productivity
- Supports team-based development and collaboration
- Balanced Optimization pack supports ETL, ELT, ETLT, TETLT, ...



InfoSphere DataStage





Cleanse – InfoSphere QualityStage

- Standardize and correct source data fields
- Match records together across sources to create a single view
- Provides visual tools for designing quality rules and matching logic
- Allows quality logic to be deployed seamlessly within ETL, or as shared services



Subject Matter Experts



Data Analysts

WebSphere QualityStage™





MetaData – InfoSphere Metadata Workbench

- Web-based exploration of metadata
- Out of the box reporting on data movement, data lineage, business meaning, impact of changes and dependencies
- Data lineage tracing from Business Intelligence reports back to sources

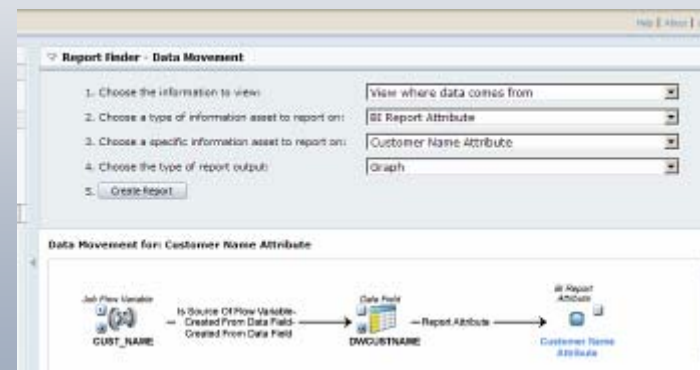


Developers



Data
Integration
Managers

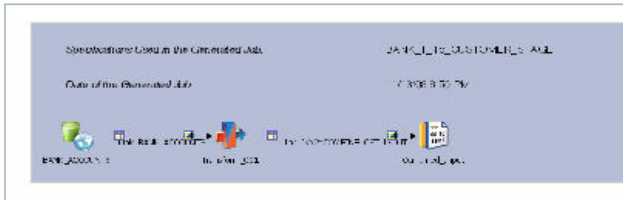
Metadata Workbench



InfoSphere Metadata Workbench

DataStage Parallel Job: BANK_1_To_CUSTOMER_STAGE

Image



DataStage Job

DataStage Job Design Information

DataStage Job Operational Information

Next DataStage Job (Operational) Job2_Q5_Standardize

Previous DataStage Job (Operational) None

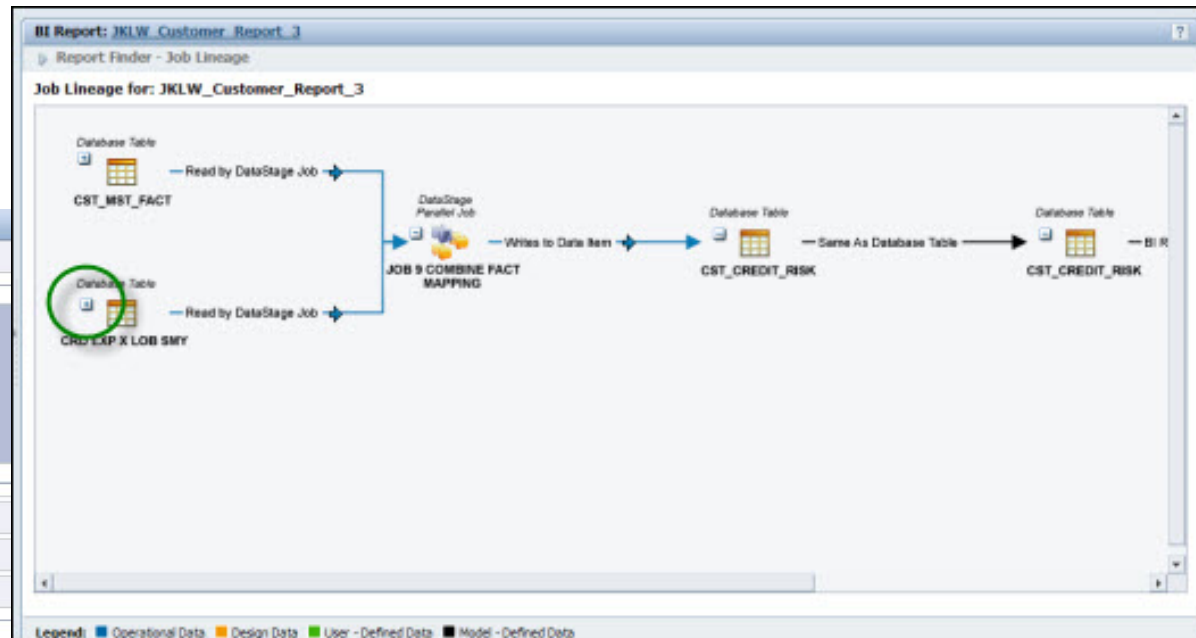
Writes to Data Item (Operational) None

Reads from Data Item (Operational) BANK_ACCOUNTS

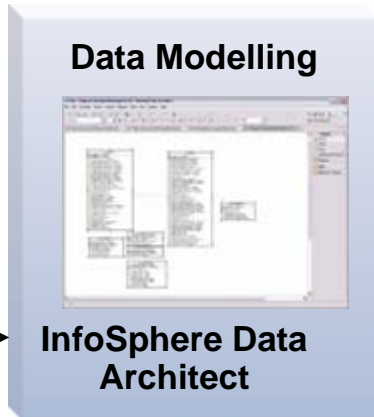
Job Runs

Finished - Dec 4, 2008 7:33 PM

Starting Date	Dec 4, 2008 7:33 PM							
Ending Date	Dec 4, 2008 7:34 PM							
Final Status	Finished Job BANK_1_To_CUSTOMER_STAGE.							
Includes Activities	<ul style="list-style-type: none"> Read 207 Rows from BANK_ACCOUNTS at Dec 4, 2008 7:34 PM Wrote 207 Rows, to Combined_Input.ds at Dec 4, 2008 7:34 PM 							
Includes Parameters	<table border="1"> <thead> <tr> <th>Name</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>InfoServerDemo_Common_Parameters</td> <td>(As pre-defined)</td> </tr> <tr> <td>InfoServerDemo_Common_Parameters.DATADIR</td> <td>C:\IOD_Demo\DSQ5JobOutputData</td> </tr> </tbody> </table>		Name	Value	InfoServerDemo_Common_Parameters	(As pre-defined)	InfoServerDemo_Common_Parameters.DATADIR	C:\IOD_Demo\DSQ5JobOutputData
Name	Value							
InfoServerDemo_Common_Parameters	(As pre-defined)							
InfoServerDemo_Common_Parameters.DATADIR	C:\IOD_Demo\DSQ5JobOutputData							



Information Server for Building the Data Warehouse



*Import/
Export*

InfoSphere Metadata Server and InfoSphere Metadata Workbench
Active Cross-Platform Administration, Management and Reporting

InfoSphere Software and Cognos

InfoSphere provides the foundation of trusted information for Cognos business intelligence & performance management

InfoSphereTM
software

*Trusted
Information*



- ***End-to-end data lineage***
- ***InfoSphere Cubing Services as data source***
- ***Business metadata sharing***
- ***Data Profiling and Quality integration***
- ***InfoSphere DataStage with Cognos Data Manager***
- ***InfoSphere Change Data Capture with Cognos Now!***

InfoSphere Master Data Management Server

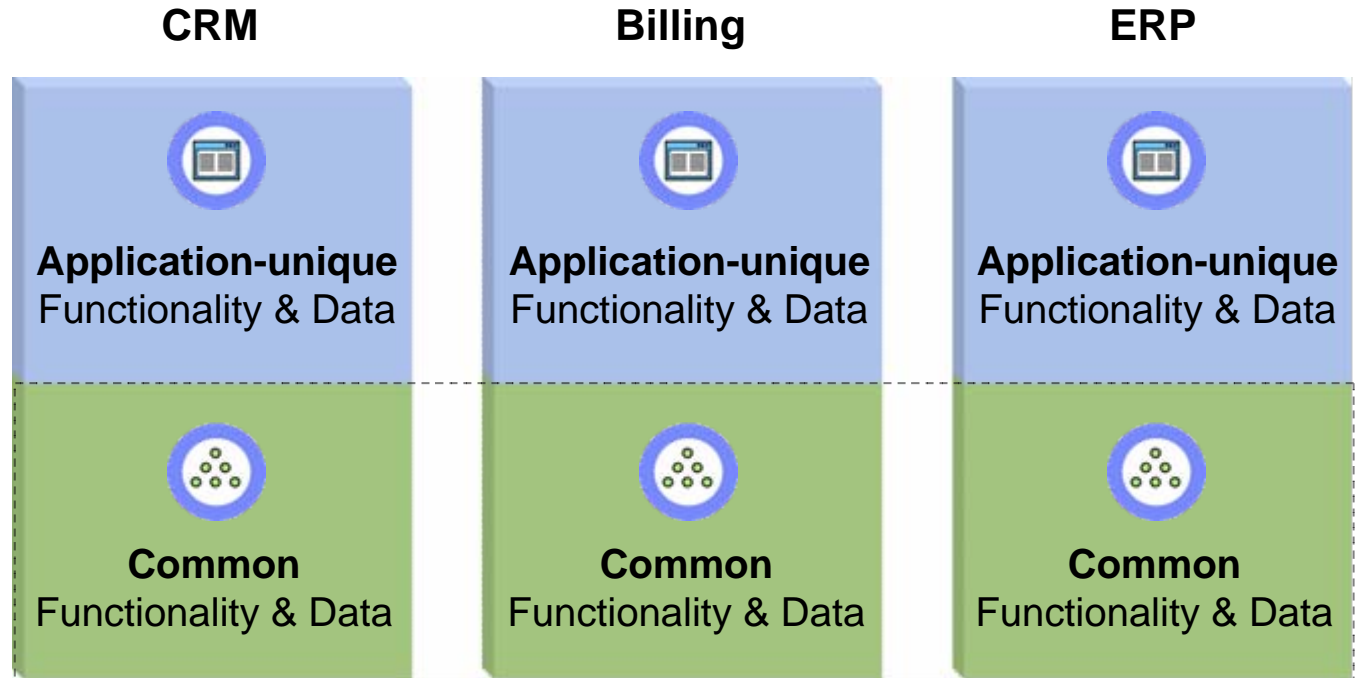


Multiform Master Data Management

The Application Process and Function Problem

Data and Functionality
unique to each application

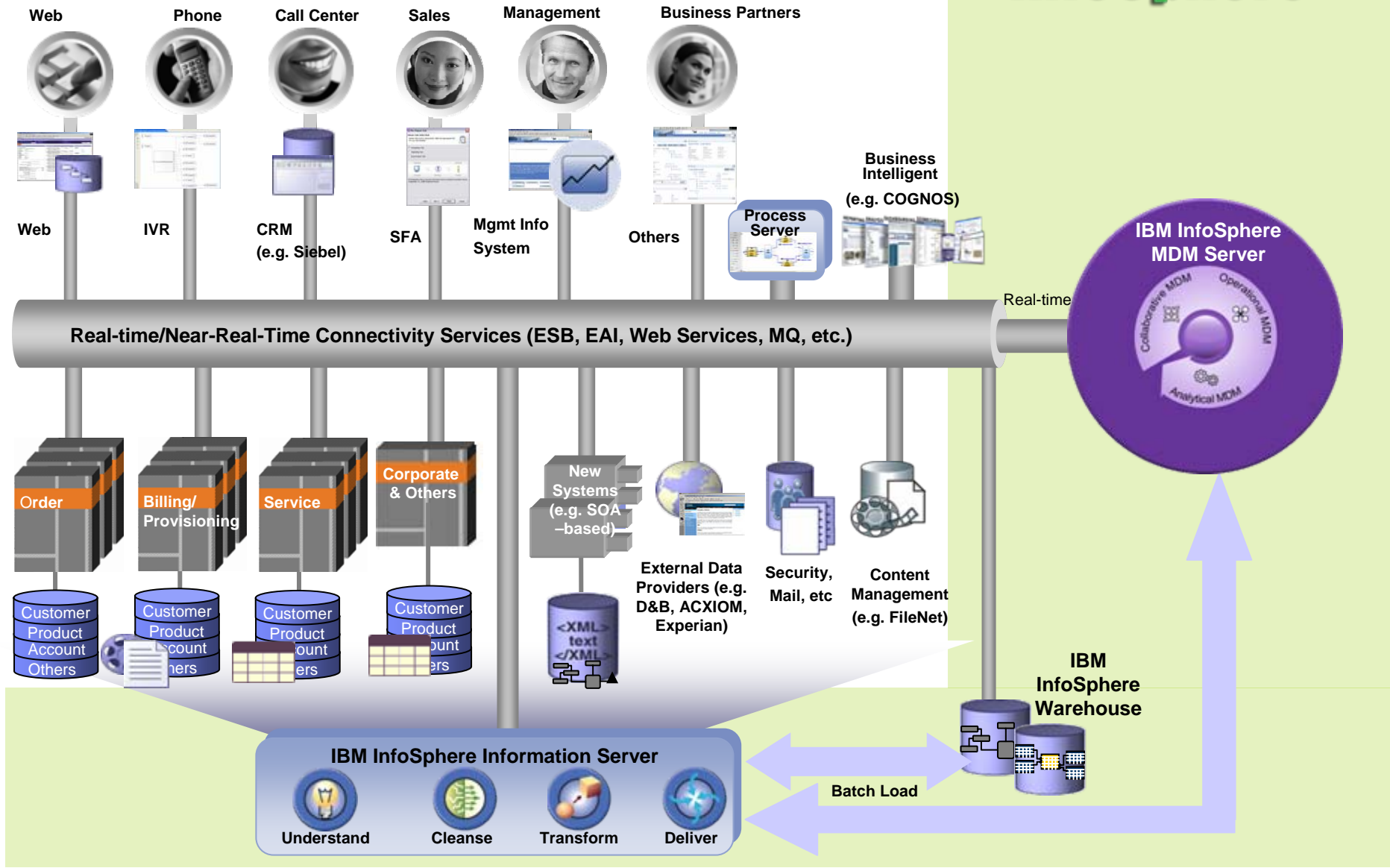
Data and data functionality
common to applications:
Customer, Account, Product



Symptom
Data is inaccurate, incomplete and conflicts with other systems

Root Cause
Common data and data function is siloed

IBM InfoSphere MDM Server Positioning



Questions & Answers

