



# **Data Warehouse**

Reto Cavegn Margrit Obrecht



# Agenda

- Datawarehouse New Intelligence
  - Was aus Daten entstehen kann
- Datenintegrations Lifecycle mit Infoserver
  - Präsentation und Demo

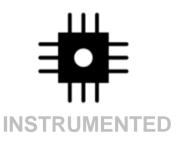


# Agenda

- Datawarehouse New Intelligence
  - Was aus Daten entstehen kann
- Datenintegrations Lifecycle mit Infoserver
  - Präsentation und Demo



# **Smarter Planet Vision and Implementation**







INTERCONNECTED

### **NEW INTELLIGENCE**

**SMART WORK** 



# GREEN AND BEYOND DYNAMIC INFRASTRUCTURE

#### **Strategic / Industry Initiatives**

**Smarter Cities** 

Smarter Transportation

Smarter Energy

Smarter Grids

Smarter Governments

Smarter Healthcare

Smarter Education

Smarter Retail ...

#### **Cross Brand Sales Plays**

Information Agenda

Information Infrastructure

Virtualization / Consolidation

Service Management

Smart SOA Foundation

Dynamic Business Awareness

Smart Collaboration

Green IT

Aligned Brand sales initiatives

Aligned

Industry, MM and

GB

Initiatives

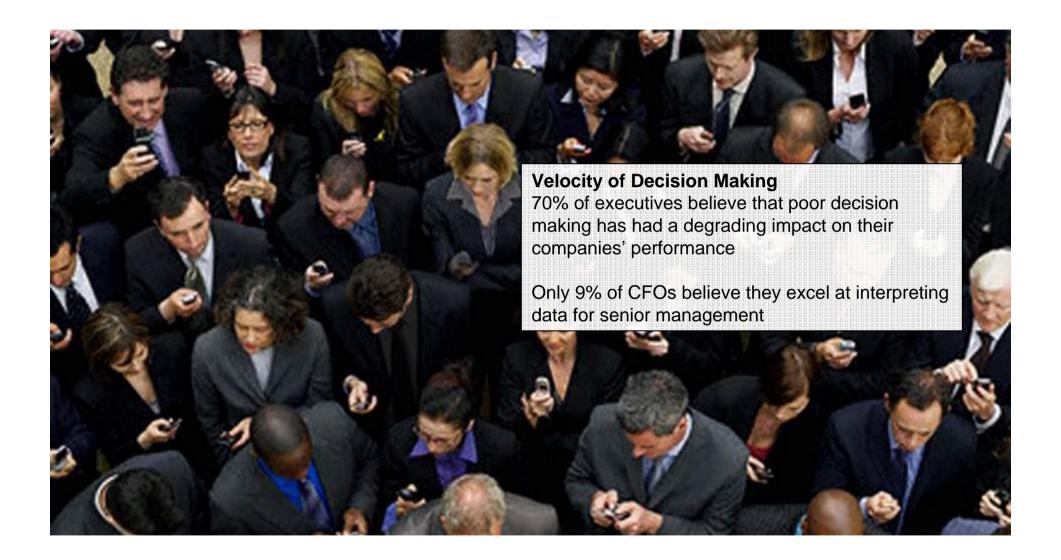














# We are seeing a mandate for change. We are seeing a mandate for smart.



We now have the ability to measure, sense and see the exact condition of everything.



People, systems and objects can communicate and interact with each other in entirely new ways.



We can respond to changes quickly and accurately, and get better results by predicting and optimizing for future events.



# We must overcome the waste, inaccuracy and missed opportunities that stem from information raging out of control

Information is not being effectively captured, managed, analyzed and made broadly available to everyone who needs it

Businesses are not optimizing the use of information to understand past and current performance and plan for the future



85% of CIO's do not believe that their information is currently well managed

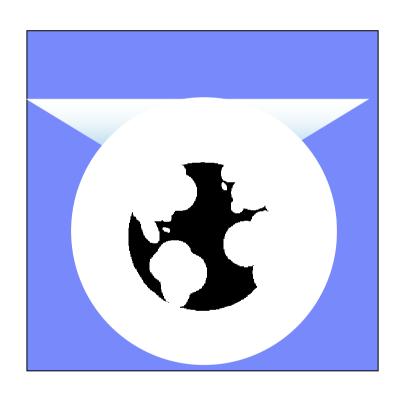


70% do not get predictions on future opportunities and problems



59% do not have access to information across the value chain that would be most useful to them





#### **SMART IS**

Effective Management of the explosive volume and variety of historical and real-time information

#### **SMART IS**

Identifying and reacting to opportunities and threats with predictive analytics in real-time

#### **SMART IS**

Pervasive information, collected from a multitude of sources across the value chain and made useful to all who need it

#### **SMART IS**

Speeding time to value by delivering accurate and timely information to the right decision makers



#### **SMART IS**

Using Real Time Information to control power production sources.

#### **Smart Power**

A German power company produces power from two sources: wind mills in the North Sea and a coal-burning power plant. The goal is to minimize the coal-burning power plant. They installed wind sensor in the sea to forecast the amount of power that will be available from the wind mills and reduce the production in the coal burning plant accordingly.



#### **SMART IS**

Tapping the collective intelligence of the entire value chain seizing the benefits of collaboration with partners, suppliers and customers

### **Smart Supply Chain**

The METRO Group is taking a new approach to the way it tracks stock and collaborates with suppliers. By incorporating RFID and the latest traceability software in their retail operations, they are able to monitor frenshness throughout the value chain, reduce food waste, and respond dynamically to shifts in customer demand.



#### **SMART IS**

Effective Information Management designed for the explosive volume and vast variety of historical and real-time information available today

### **Smart Trading**

TD Bank uses real-time market analysis for trade process transformation. Using new stream computing capabilities TD bank is able to interconnect with a variety of text, voice, image, video, data feeds, and application data at volumes of millions of messages per second... and identify and execute trades in millisecond speeds... faster than a hummingbird can flap its wing.



#### **SMART IS**

Identifying and reacting to opportunities and threats with predictive analytics in real-time

### **Smart Policing**

The New York City Police Department is combining business intelligence software with GIS mapping and visualization tools to detect crime patterns as they are forming, enabling precinct commanders to take proactive measures to keep ahead of these trends—and head off spikes in criminal activity.



#### **SMART IS**

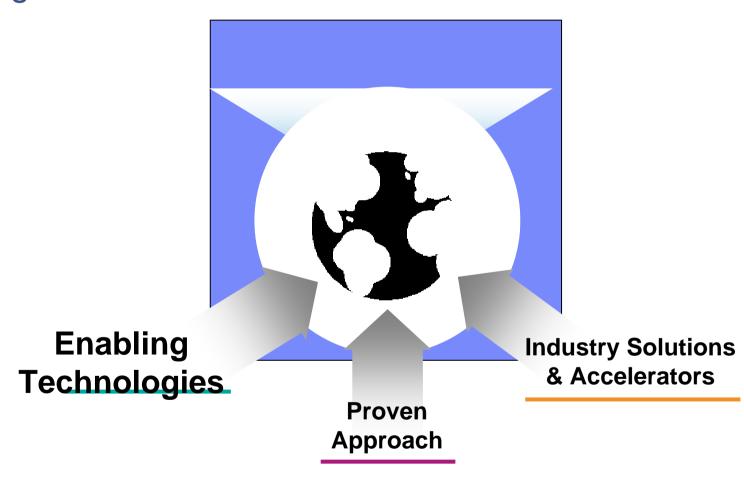
Speeding time to value by delivering accurate and timely information to the right decision makers so they can optimize business results

#### **Smart Retail**

To stay ahead of the competition, international retailer Elie Tahari deployed a new solution to integrate order, sales, inventory and financial data. By delivering to managers the real-time information they need to make crucial inventory decisions, Elie Tahari is best able to anticipate and meet demand for the latest fashions.

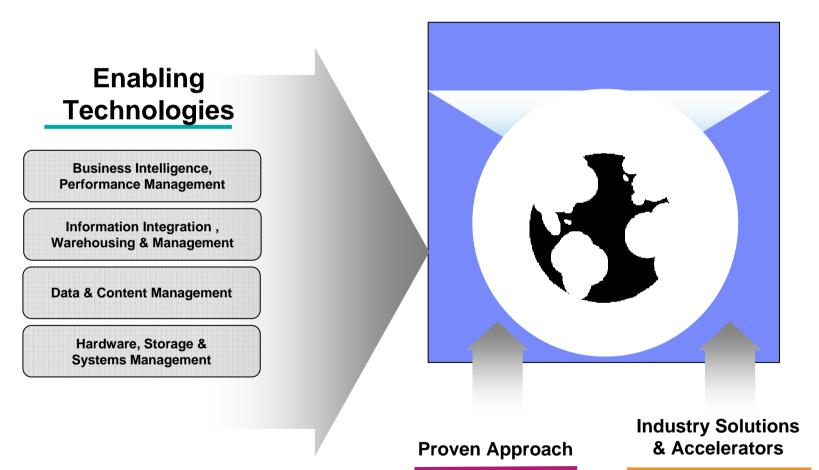


# Helping a smarter planet discover a new kind of intelligence





# Enabling Technologies: Establishing a platform to support intelligent decision making





# Information Management solutions provide key enabling technologies

# **Enabling Technologies**

Business Intelligence, Performance Management

Information Integration, Warehousing & Management

**Data & Content Management** 

Hardware, Storage & Systems Management





# Enabling Technologies: Establishing a platform that supports intelligent decision making



Information Integration , Warehousing & Management

**Data & Content Management** 

Hardware, Storage & Systems Management

# Manage information over its lifecycle

Optimize content-based processes



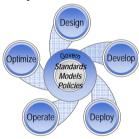
# **Database Software**

Helping smarter businesses manage your data smarter, faster, and at lower cost



#### Enterprise Content Management

An integrated content, process, compliance and discovery portfolio that delivers agile ECM capabilities for better decisions faster.



# Integrated Data Management

An integrated, modular environment to design, develop, deploy, operate, optimize and govern enterprise data.



### DB2 9.7 Feature Overview

### Application Freedom

- SQL Procedure Language (SQL PL) enhancements
- Oracle Application Enablement Made Easy
- Application development enhancements and changes

### Service Level Confidence

- Workload management enhancements
- High availability and resiliency enhancements
- Manageability enhancements
- Performance enhancements
- Database administration changes
- Security enhancements and changes

## XML Insight

## Storage TCO Reduction

Enhanced Compression



# Enabling Technologies: Establishing a platform that supports intelligent decision making



Information Integration , Warehousing & Management

Data & Content Management

Hardware, Storage & Systems Management

#### Trusted information

#### **InfoSphere Streams**

Deep analytics on huge volumes of heterogeneous data in-motion with ultra low latencies

#### **InfoSphere Foundation Tools**

Profile, model, define, map and govern information that is spread across your enterprise

#### **Enterprise Content Management**

An integrated portfolio that delivers agile ECM capabilities for better decisions faster,

#### **Information Integration**

Integrates and transforms data improving productivity, flexibility and performance

#### **Master Data Management**

Manages master data domains that impacts the most important business processes

#### **Data Warehousing**

Warehousing and analytics capabilities for structured and unstructured information

#### **IBM Mashup Center**

Integrate structured, unstructured, desktop and cloud data into contextual applications



# Enabling Technologies: Establishing a platform to support intelligent decision making



Information Integration, Warehousing & Management

Data & Content Management

Hardware, Storage & Systems Management

# Plan, understand and manage financial and operational performance

#### **Financial Performance Management**

Supports the finance specifics of managing revenue and cost drivers, closing the books, consolidating results and reporting out with confidence.

#### **Business Intelligence**

Reports, analysis and dashboards provide the right information to guide decision making and optimize performance.

#### **Analytic Applications**

Prepackaged best practices content for executives, managers, analysts and specialists.

#### **Mid Market Solutions**

Solutions to provide flexible deployments that fit user need and budgets and grow as the business grows.

#### **Performance Management Framework**

An integrated approach to getting "Information Sweet Spots" to decision makers across the organization to answer, "How are we doing?" "Why?" and "What should we be doing?"



# Agenda

- Datawarehouse New Intelligence
  - Was aus Daten entstehen kann
- Datenintegrations Lifecycle mit Infoserver
  - Präsentation und Demo



## Unlocking the Business Value of Information

Financial Risk Insight Workforce Optimization

Dynamic Supply Chain

Customer & Product Profitability

**Business Optimization** 

Multi-channel Marketing

Better Business Outcomes

Cognos. software





InfoSphere software

Information Integration, Warehousing, & Management

Maragk

FileNet. software





Information Assets

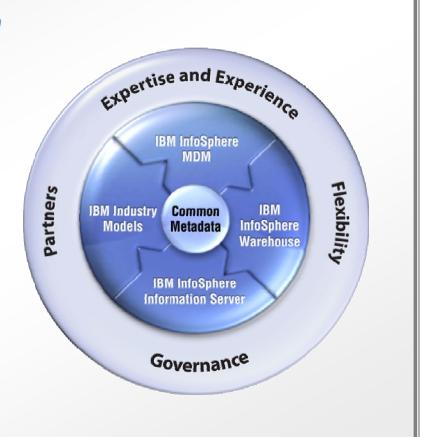
**DB2.** Informix. software



## The IBM InfoSphere Vision

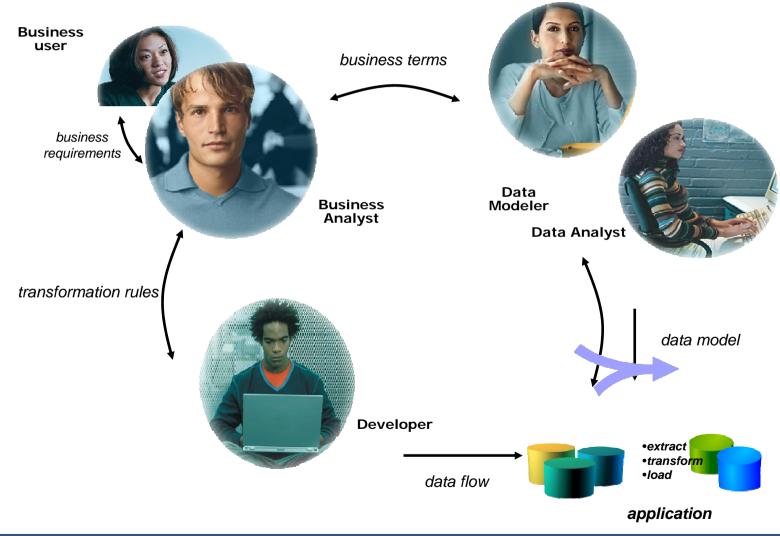
### An Industry Unique Information Platform

- Simplify delivery of Trusted Information
- Accelerate client value
- Promote collaboration
- Mitigate risk
- Modular, yet integrated
- Scalable project to enterprise



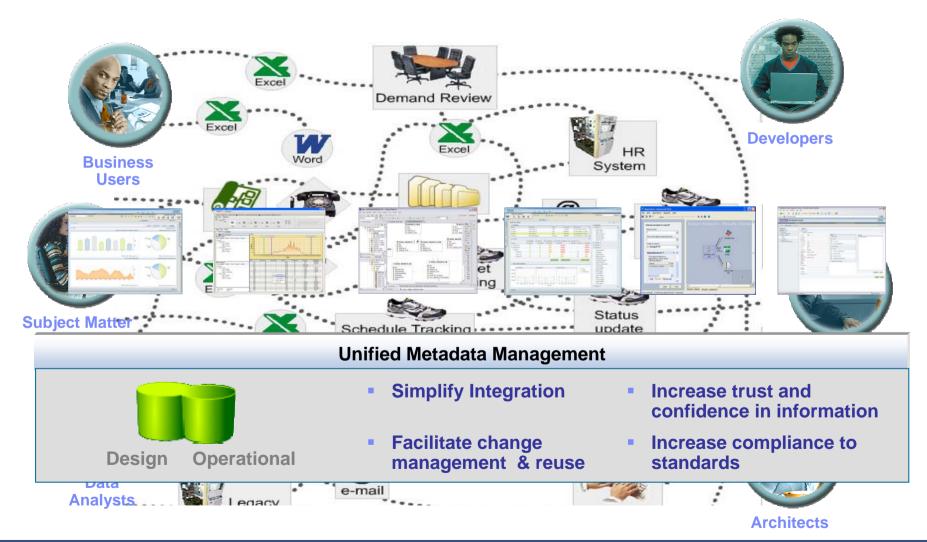


# Data Integration Projects require a Collaborative Effort



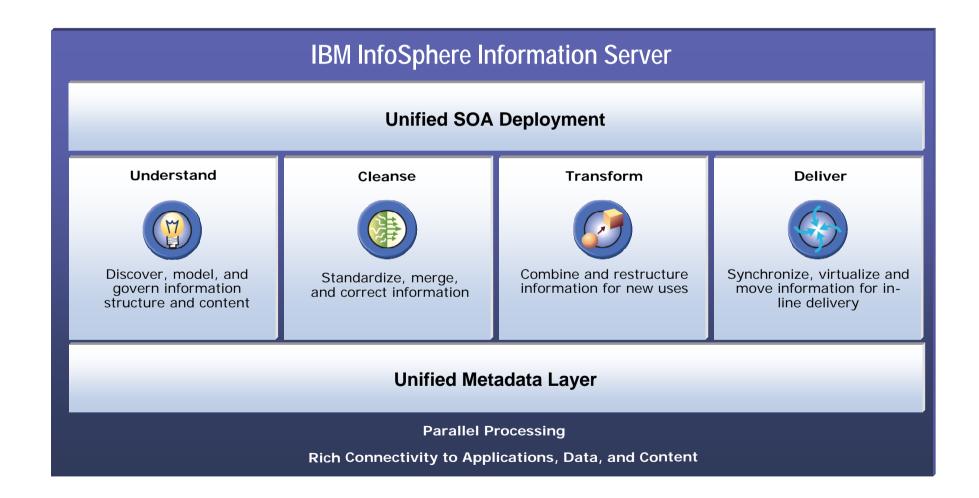


## Role Based Tools with Shared Metadata



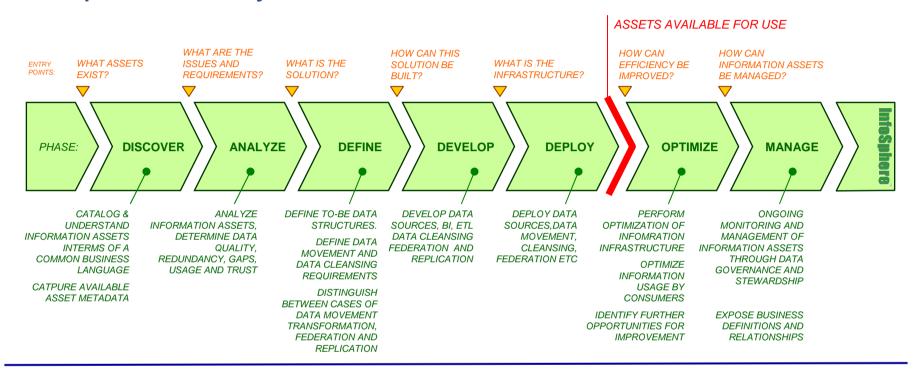


# IBM Information Server Delivering information you can trust

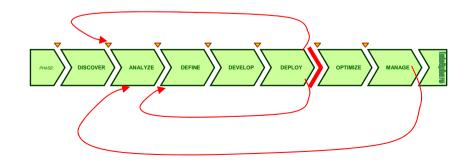




# InfoSphere Lifecycle Phases

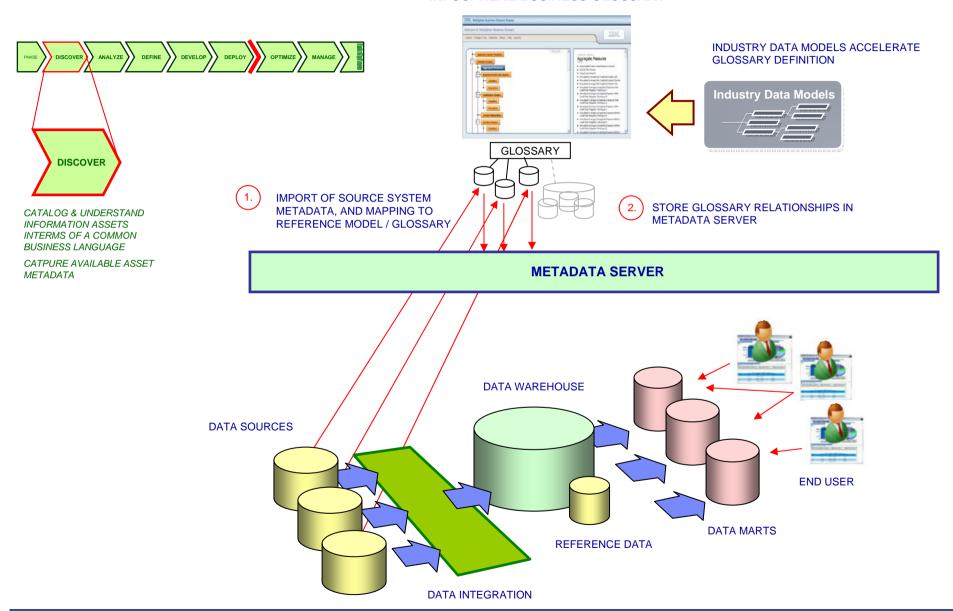


CYCLE IS PROGRESSIVE / ITERATIVE - PHASES MAY OCCUR IN PARALLEL, AND MAY EXIT AND RE-ENTER AT EARLIER PHASES...





#### **INFOSPHERE BUSINESS GLOSSARY**





## InfoSphere Business Glossary

- Starting point for designing information governance
- Web-based authoring, sharing & management of business metadata
- Enhanced collaboration
  - Aligns the goals of the business with the efforts of IT
  - Accelerate project delivery with information sharing
- Define relationship between business definitions and IT assets
- Drive information trust across enterprise applications
  - From any application, click and automatically search glossary
- Establish accountability
  - Creation of stewards and assignment of responsibilities





IBM InfoSphere Information Server & IBM Cognos 8 BI: Information Understanding







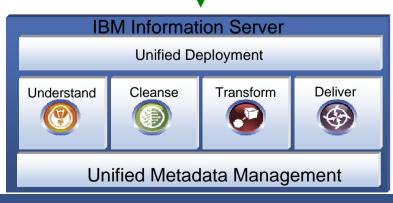
Users

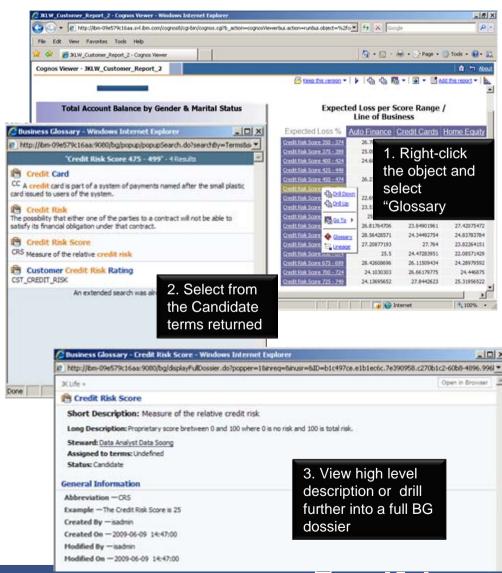
Subject Matter **Experts** 

**Analysts** 

Reports in Cognos Viewer and Metadata Tree in Query Studio. Report Studio, or Analysis Studio

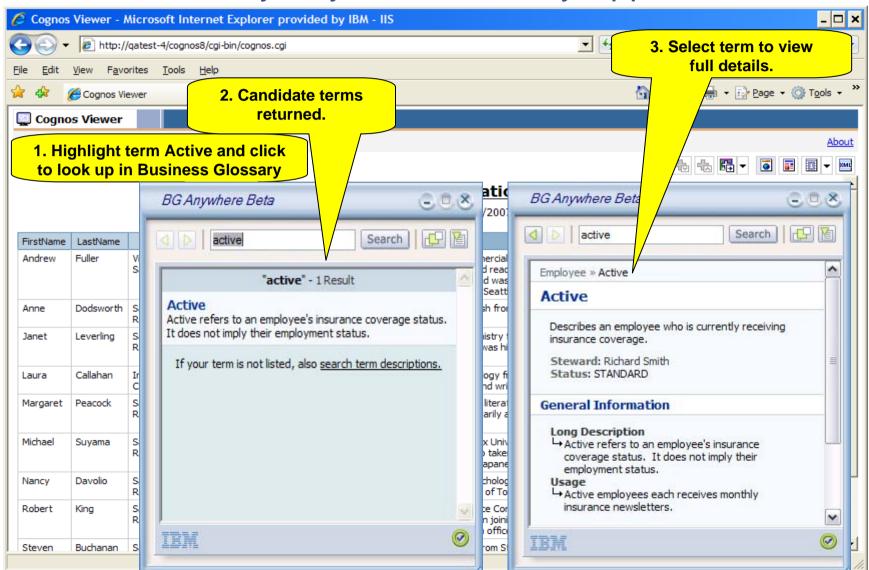
Information Understanding



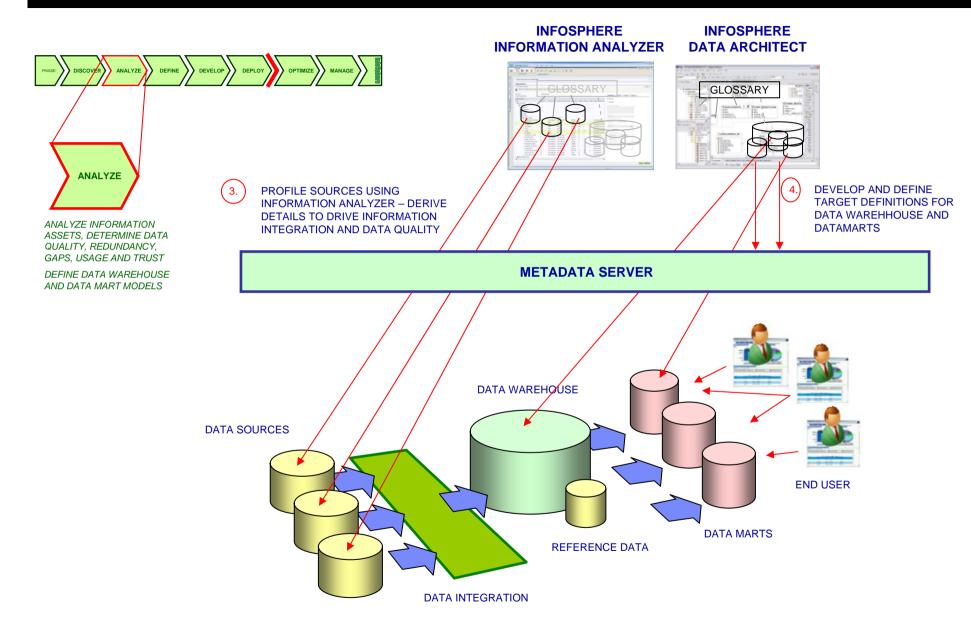




# Business Glossary Anywhere with any Application









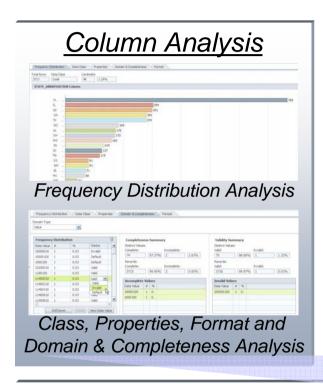
# InfoSphere Information Analyzer

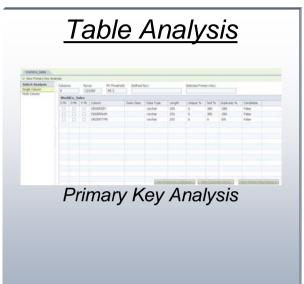
- First step in defining an information infrastructure across enterprise ecosystem
- In-depth analysis of heterogeneous information – IBM or non-IBM sources
  - Data-centric analysis of databases, files and enterprise applications for content, quality, and structure
  - Secure, detailed profiling of fields, and relationship analysis across fields and across sources
- On-going measurement and baseline reporting of information quality
- Know where information is managed across systems
  - Capture fitness of sources and reengineering requirements for downstream use

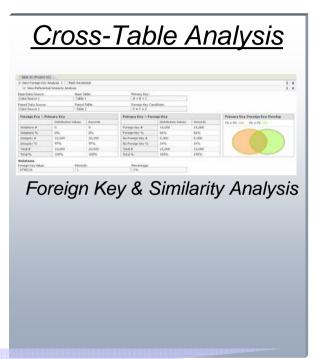




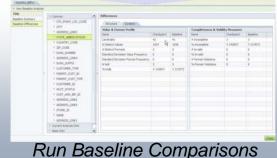
# How Does InfoSphere Information Analyzer Work?

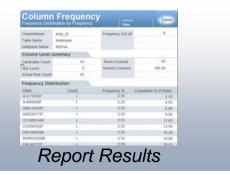








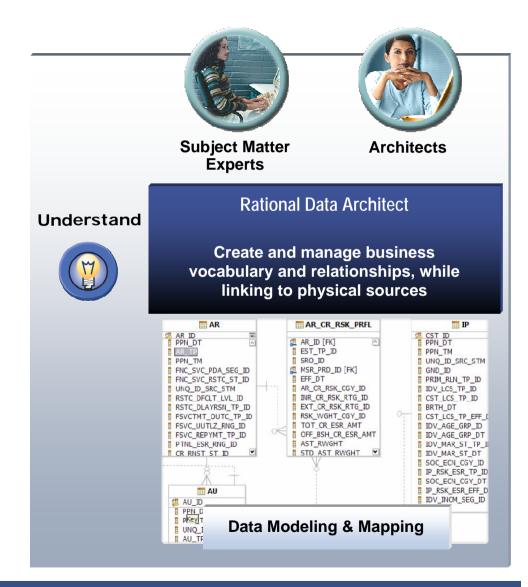




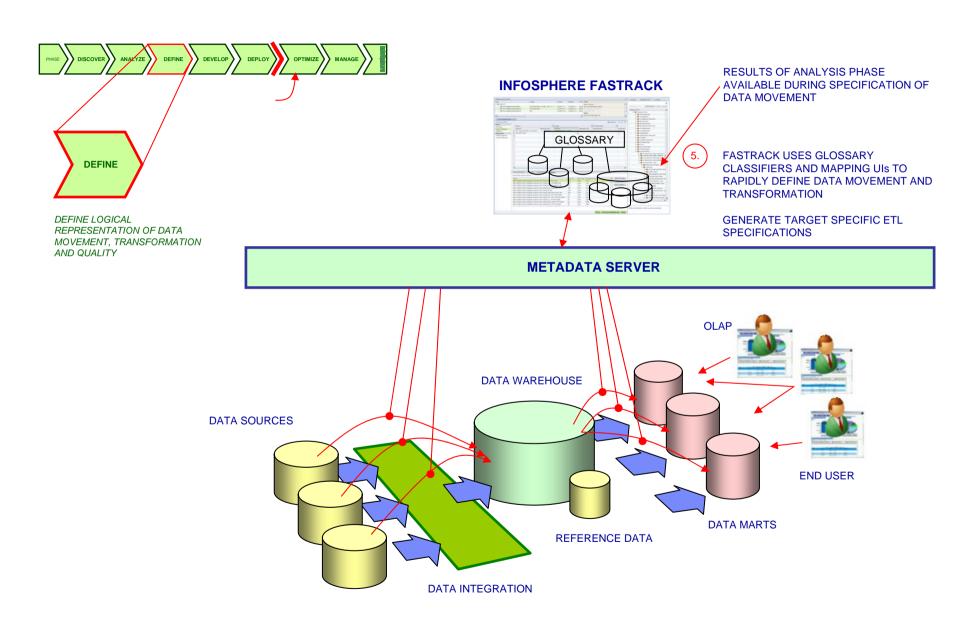


### InfoSphere Data Architect

- Data modeling for data structures and federations
- Federated data discovery
- Metadata relationship discovery & mapping
- Impact analysis, and synchronization across models
- SQL & XML generation capabilities







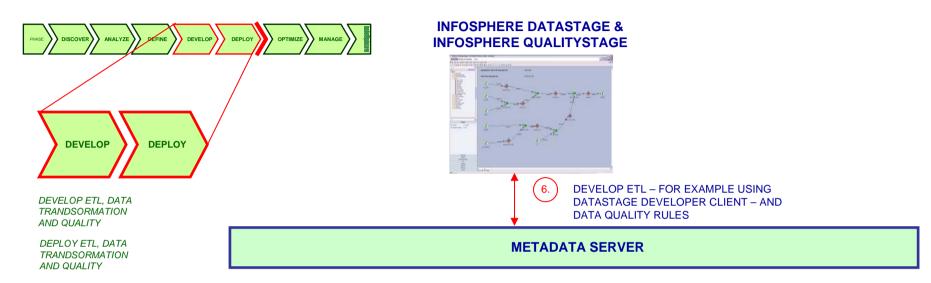


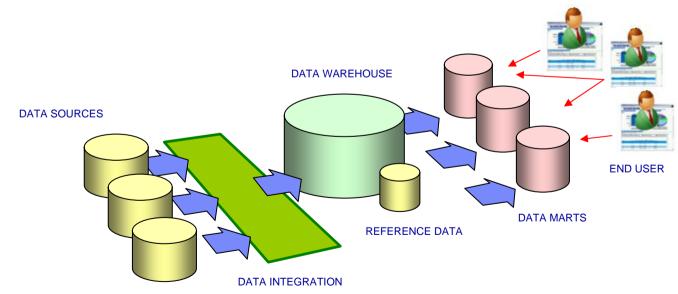
### InfoSphere FastTrack

- Create a solid information infrastructure
- 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





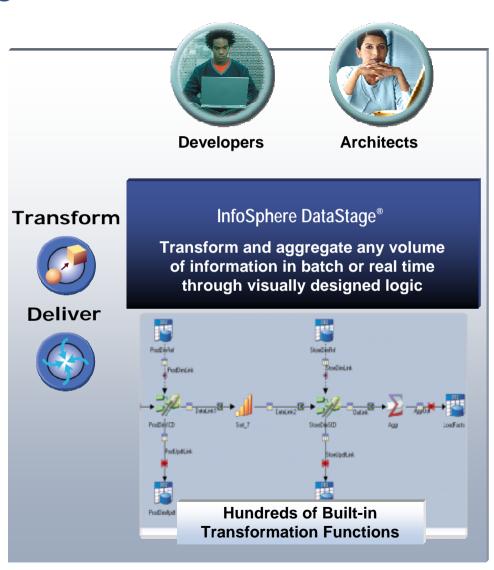






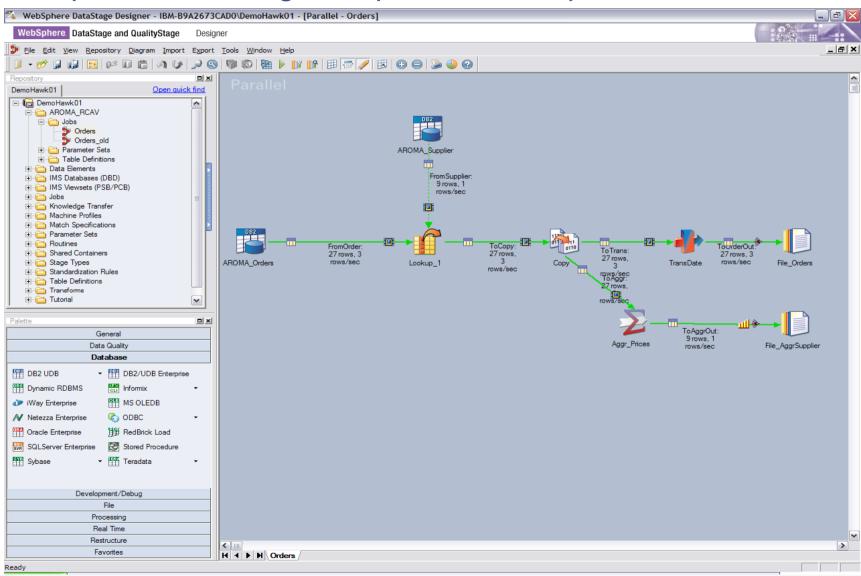
## What is InfoSphere DataStage?

- Provides codeless visual design of data flows with hundreds of built-in transformation functions
  - Optimized reuse of integration objects
  - Supports batch & real-time operations
  - Produces reusable components that can be shared across projects
- Complete ETL functionality with metadata-driven productivity
- Supports team-based development and collaboration
- Provides integration from across the broadest range of sources





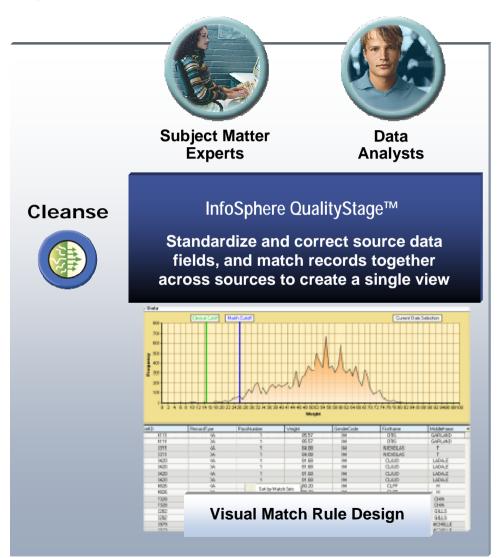
### InfoSphere DataStage Graphical Development Interface





## What is InfoSphere QualityStage?

- Provides specialized data quality processing
  - Ensures clean, standardized, deduplicated information
  - Enables a single version of the truth
  - Supports global postal verification
- Provides visual tools for designing quality rules and matching logic
  - Seamlessly integrated with DataStage (one engine, one metamodel, one UI)
  - Precisely calibrates matching rules
- Allows quality logic to be deployed seamlessly within ETL, or as shared services





## Typical Data Quality Challenges

- Lack of information standards
  - Different formats & structures across different systems

Kate A. Roberts 416 Columbus Ave #2, Boston, Mass 02116

Catherine Roberts Four sixteen Columbus APT2, Boston, MA 02116

Mrs. K. Roberts 416 Columbus Suite #2, Suffolk County 02116

- Data surprises in individual fields
  - Data misplaced in the database

 Name
 Tax ID
 Telephone

 J Smith DBA Lime Cons.
 228-02-1975
 6173380300

 Williams & Co. C/O Bill
 025-37-1888
 415-392-2000

 1st Natl Provident
 34-2671434
 3380321

 HP 15 State St.
 508-466-1200
 Orlando

Information buried in free-form fields

WING ASSY DRILL 4 HOLE USE 5J868A HEXBOLT 1/4 INCH WING ASSEMBY, USE 5J868-A HEX BOLT .25" - DRILL FOUR HOLES USE 4 5J868A BOLTS (HEX .25) - DRILL HOLES FOR EA ON WING ASSEM RUDER, TAP 6 WHOLES, SECURE W/KL2301 RIVETS (10 CM)

- Data myopia
  - Lack of consistent identifiers inhibit a single view

19-84-103	RS232 Cable 6' M-F CandS
CS-89641	6 ft. Cable Male-F, RS232 #87951
C&SUCH6	Male/Female 25 PIN 6 Foot Cable

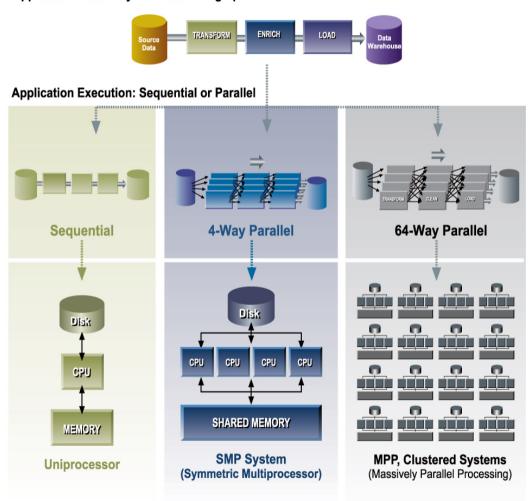
- The redundancy nightmare
  - Duplicate records with a lack of standards

90328574	IBM	187 N.Pk. Str. Salem NH 01456
90328575	I.B.M. Inc.	187 N.Pk. St. Salem NH 01456
90238495	Int. Bus. Machines	187 No. Park St Salem NH 04156
90233479	International Bus. M.	187 Park Ave Salem NH 04156
90233489	Inter-Nation Consults	15 Main Street Andover MA 02341
90345672	I.B. Manufacturing	Park Blvd. Bostno MA 04106



## Parallel Engine - Building Scalable Software

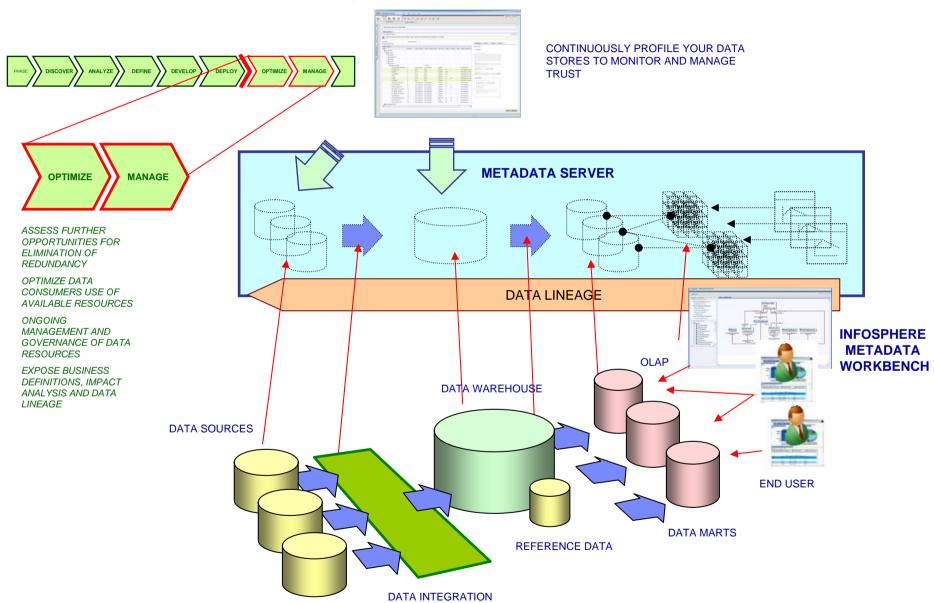
Application Assembly: One dataflow graph



- Designs integration process without concern about data volumes or time constraints
- Leverages database partitioning schemes for optimal load performance
- Simplifies steps to define partitions within each process if needed
- Utilizes a single configuration file to add processors, hardware
- Requires no hand coding of programs to enable more processors
- Supports SMP, Clustered, GRID, and MPP platforms



#### **INFOSPHERE INFORMATION ANALYZER**





## InfoSphere Information Analyzer - Data Rules

#### **Rules Analysis**

- Extend profiling content with:

Validation of data rules individually or within a broader set

Establishment of benchmark thresholds

Tracking trends against initial baselines

#### **Data Source Analysis**

- Enrich profiling content with:

**Definitions & Terms of Metadata** 

Annotations for information or action

Validation of structural properties

Validation of domains, formats and keys

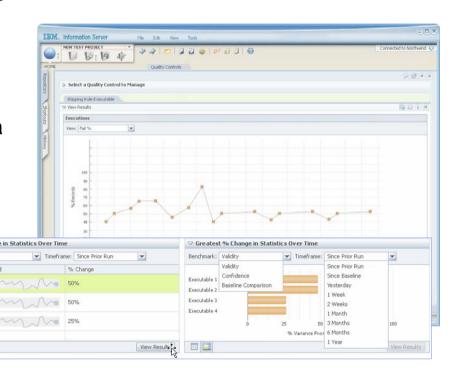
Identification of redundancies

- Deliver information through:

Reports

**Shared Metadata** 

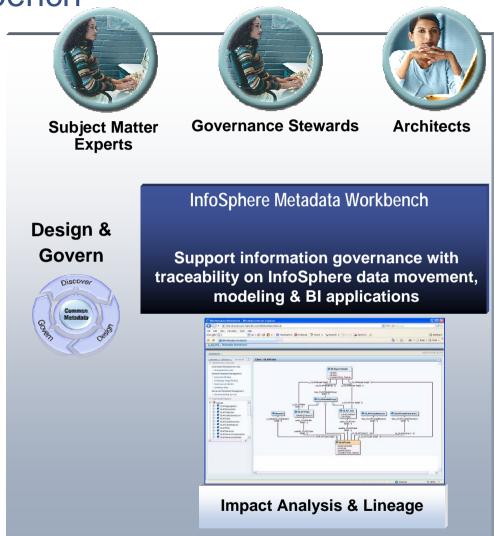
**Published Analytical Results** 





InfoSphere Metadata Workbench

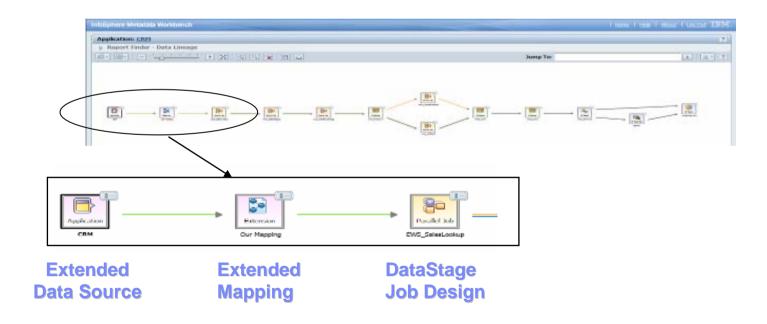
- Understand your end-to-end information infrastructure
- Proactively manage and administer your information infrastructure
  - Web-based exploration of metadata relationships – InfoSphere, modeling and BI applications
  - Explore and analyze both graphically and textually
  - Perform searches and customized queries
- Assess and mitigate change management risk
  - Assess dependencies across InfoSphere and 3rd party tools
- Support compliance and governance initiatives (eg Sarbanes-Oxley, Basel II)
  - Trace data lineage of modeling, BI reports and InfoSphere objects





## Metadata Workbench Where does a field of data come from?

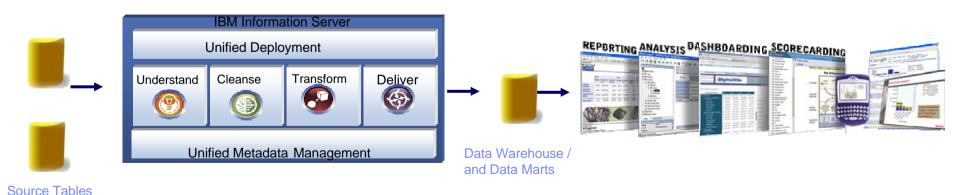
- Support governance requirements for business provenance
- Extended visibility to enterprise data integration flows outside of Information Server
- Comprehensive understanding of data lineage for trusted information
- New graphical control for simplified navigation





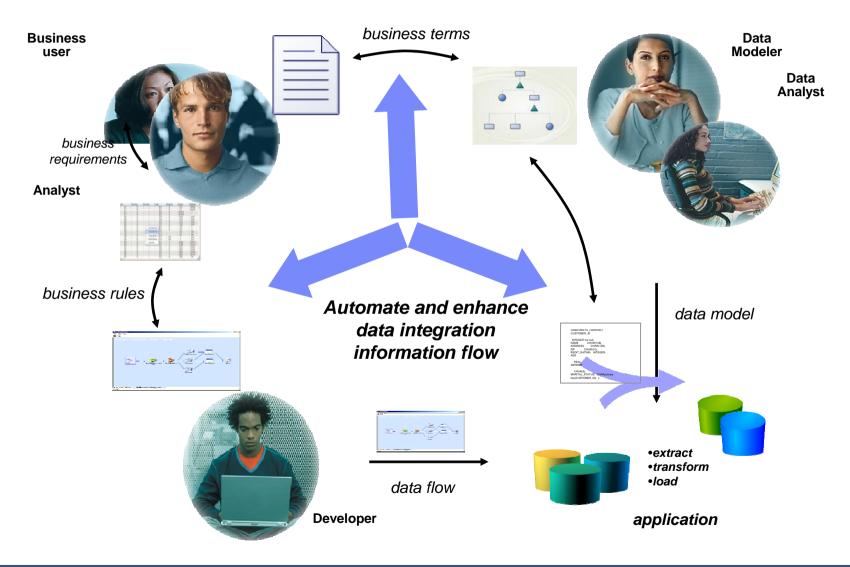
## IBM InfoSphere Information Server & IBM Cognos 8 BI: End-to-End Data Lineage

- Capability: End-to-end data lineage navigation across tools
  - Import & Browse Full Cognos Report Metadata
  - Navigate through report attributes all the way back to the source
  - Combines operational, design, and business viewpoint
  - Allows third party lineage data for integrated view
- Benefit: Answers the question "Where did this data come from & what happens if I change this?"





### Connect The Dots between Collaborators





## Customers Achieve Significant Productivity Benefits<sup>1</sup>

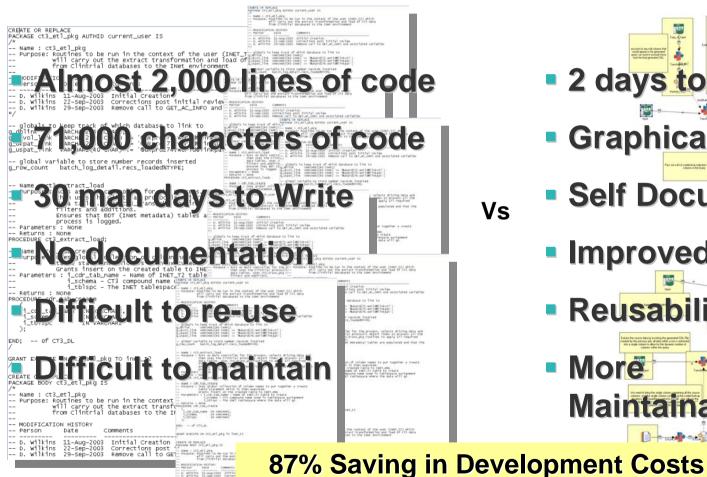
Approx. Project Effort	Example Integration Projects	
30%	Source System Analysis	• 50+% gain
20%	Data Cleansing	• 50+% gain
20%	Transformation Logic Construction	• 40+% gain
20%	Data Management Services	• 20+% gain
10%	Application System Connectivity	● 30+% gain
100%		



## Improved Productivity Example

Real-life example: Pharmaceutical data processing

#### **Legacy Development (Handcoding)**



D. within 13-Aug-2009 Infets D. within 13-Aug-2009 Correct

#### **DataStage**



Self Documenting

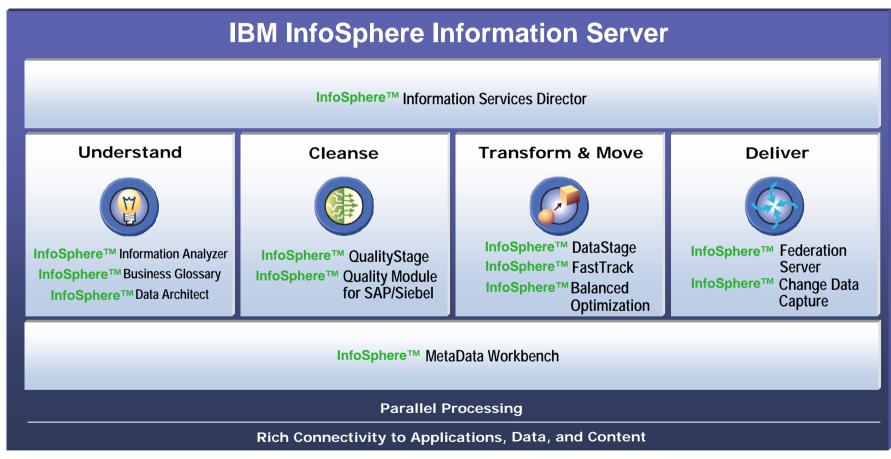
**Improved Performance** 



Trusted Information



# IBM Information Server Delivering information you can trust



InfoSphere™ Information Server Pack for SAP / Siebel / Essbase / ...









## Thank you!