Introduction to Data Warehousing on System z
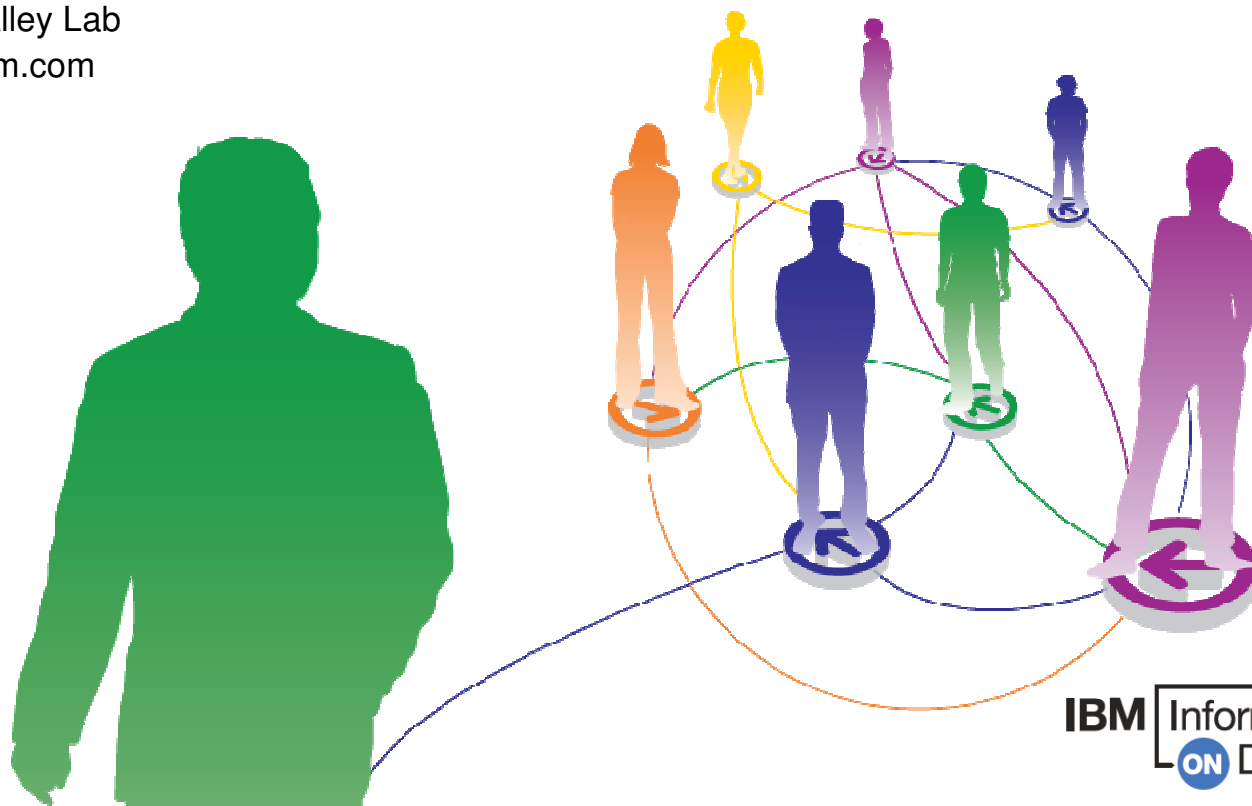
Presenter: Mike Biere
IBM WW Marketing Mgr.
IBM Silicon Valley Lab
mbiere@us.ibm.com
Session: 005

IBM Information on Demand 2010

INFORMATION-LED TRANSFORMATION

LEAD THE WAY

Son … someday you will make a girl very happy for a short period of time. Then she'll leave you and be with new men who are ten times better than you could ever hope to be … these men are called musicians
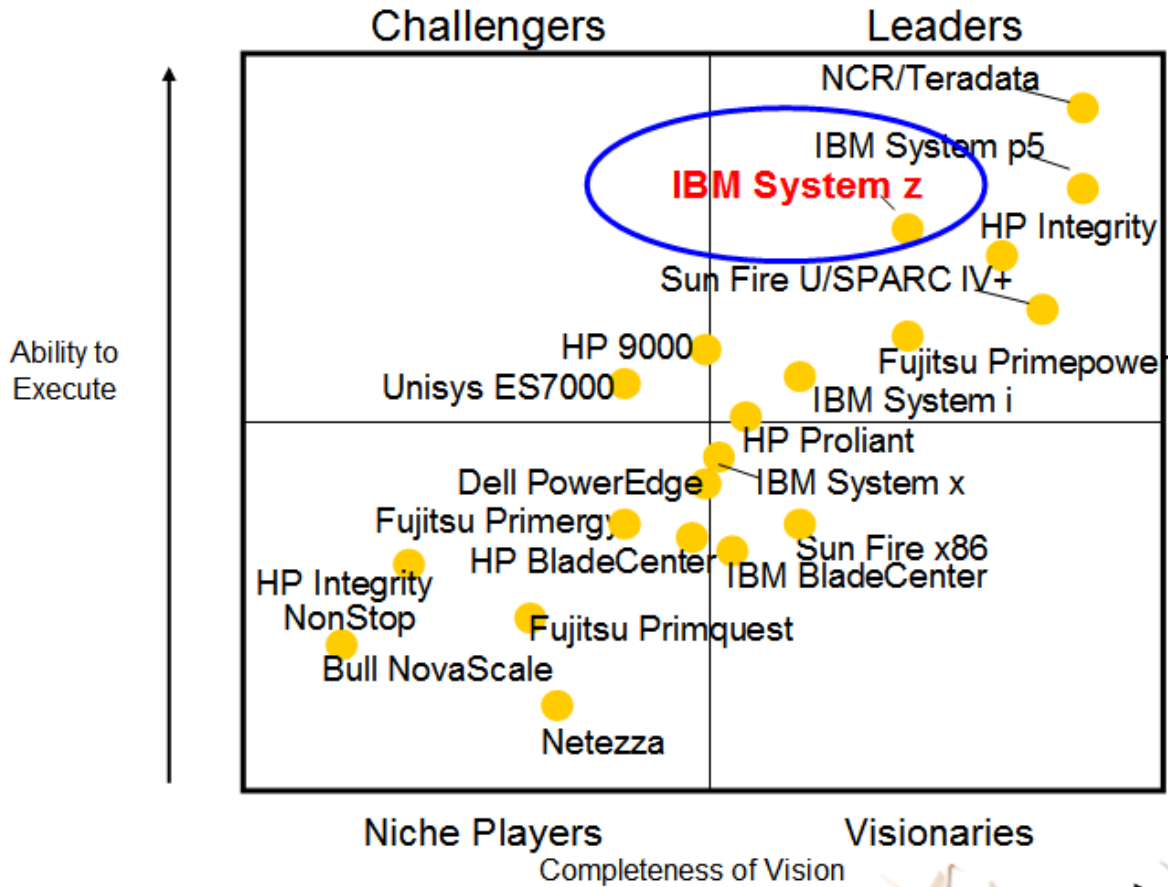
Disclaimer

Gartner's Data Warehouse DBMS Server Magic Quadrant

August 2006

Challengers | Leaders

NCR/Teradata
IBM System p5
IBM System z
HP Integrity
Sun Fire U/SPARC IV+
HP 9000
Fujitsu Primepower
Unisys ES7000
IBM System i
HP Proliant
Dell PowerEdge
IBM System x
Fujitsu Primergy
HP BladeCenter
Sun Fire x86
IBM BladeCenter
HP Integrity NonStop
Fujitsu Primquest
Bull NovaScale
Netezza

Ability to Execute

Niche Players | Visionaries

Completeness of Vision

Mission-Critical Data Warehousing

IBM Information On Demand 2010

Analysts Predict Move to Right-Time BI

By 2012, BI Platform capabilities will be embedded as a service within 75% of new business applications.

By 2012, emerging technologies will significantly drive adoption of BI to 50% of business users

(e.g. interactive visualization, in-memory analytics, search, SaaS and SOA)

Embedded Analytics

Dashboards

Mash ups

Financial Planning

Scorecards

Source : Gartner BI Summit 2008, 4/2008

IBM Information ON Demand 2010

Cost Savings

Your IT Cost may vary:

→ Up to 80% Saving in IT Cost
→ Up to 96% Less Hardware
760 x86 Processor Cores vs 26 IFLs
→ Potential for dramatic reductions in software expense for processor based licenses
→ Potential reductions in power and cooling
Up to 93% Savings in KWatts and Energy Costs in this scenario
→ Up to 46% Less Space
→ Up to 89% People savings
→ Increased processor utilization
→ Industry leading Security

Energize your IT savings with z10 EC.

**Consolidating 760 Linux servers
z/VM Virtualization versus x86**
Oracle DB Workload
3-Year Total IT Cost

**$56 M Savings versus
x86 without Virtualization**

Legend:
- Space
- Power
- People Cost
- Software Maint
- Software
- Maint
- Hardware

5.1 X — 760 x86 CPs — **x86 w/o Virtualization** — **Sun X2100 Single Core servers**

2.0 X — 248 x86 CPs — **x86 Virtualization** — **Dell Power Edge Quad Core servers**

1.0 X — 26 z10 EC IFLs — **z/VM Linux on System z10 EC**

80% Savings

IBM is investing to accelerate your information-led transformation

Over $12B in software investments since 2005

Over 4,000 Dedicated Consultants

Analytics in a Box to Accelerate Time to Value

Largest Math Department in Private Industry

ovum
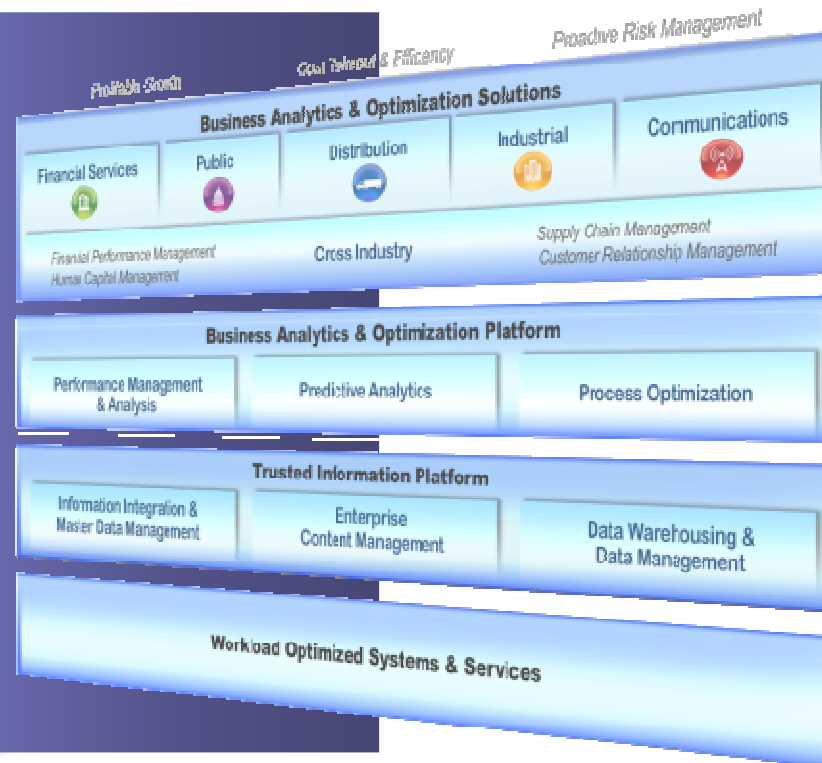
"Since 2006, IBM has deliberately & doggedly constructed an unparalleled portfolio… it's difficult to see how any competitors will be able to compete anytime soon…"

FORRESTER

"IBM, not SAP or Oracle, is now the industry's premo analytics solution/platform vendor…"

Getting started with an information-led transformation

Create Value

Plan an information agenda

Establish a flexible information platform

Apply business analytics to optimize decisions

Costs

Profitable Growth — Cost Takeout & Efficiency — Proactive Risk Management

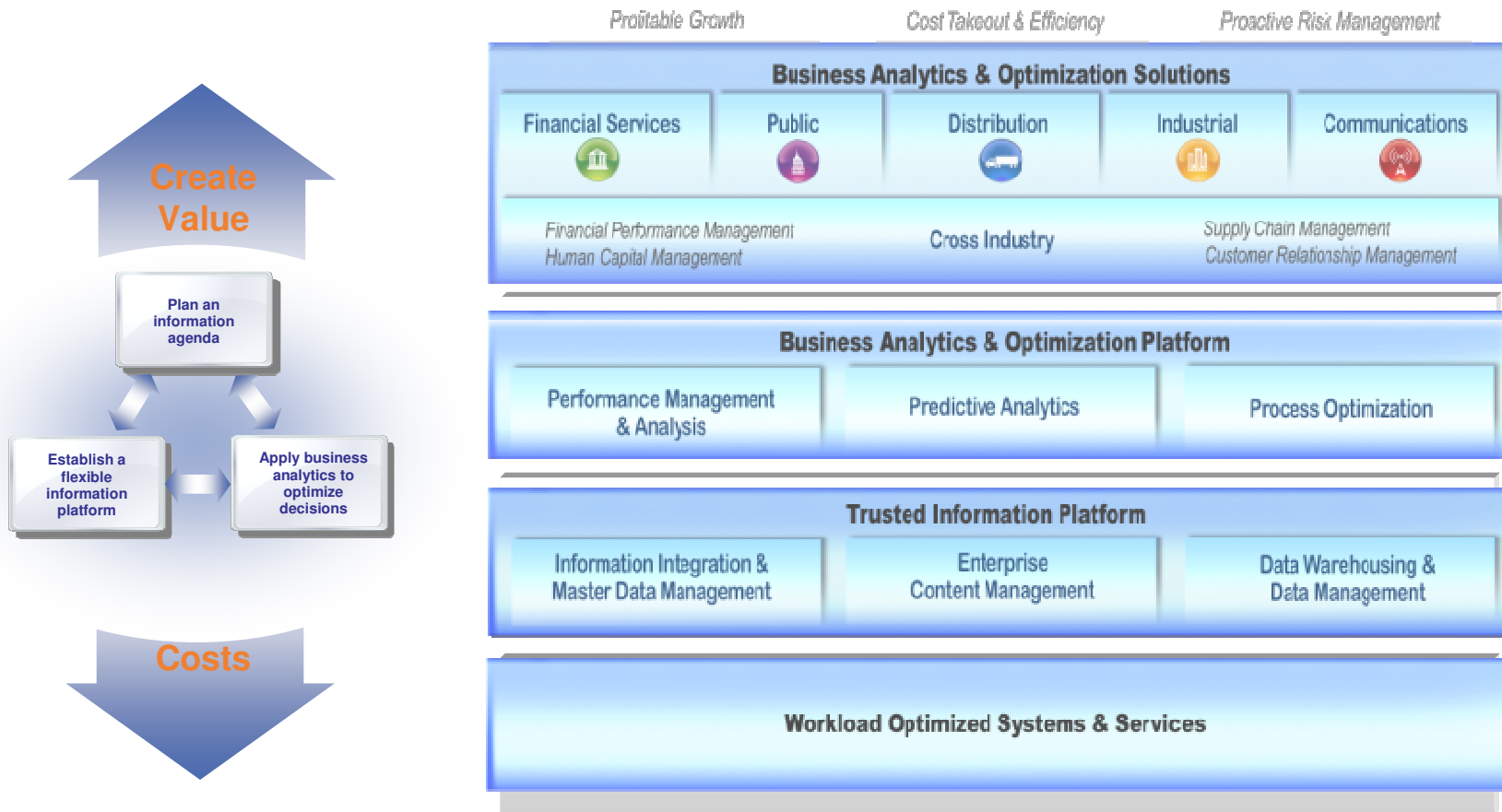**Business Analytics & Optimization Solutions**

| Financial Services | Public | Distribution | Industrial | Communications |

Financial Performance Management
Human Capital Management — Cross Industry — Supply Chain Management
Customer Relationship Management

Business Analytics & Optimization Platform

| Performance Management & Analysis | Predictive Analytics | Process Optimization |

Trusted Information Platform

| Information Integration & Master Data Management | Enterprise Content Management | Data Warehousing & Data Management |

Workload Optimized Systems & Services

60%+ of CEOs: Need to do a better job leveraging information

30%+ of Execs: time spent searching for information

70% of employee time: Spent searching for relevant information

Data and Dynamic Warehousing- Some Definitions

➔ <u>Data Warehousing</u>
  – Provides the underlying data storage facilities required to support any BI Solution.

➔ <u>An Operational Data Store (ODS)</u>
  – A subject oriented database organized by business area. It is up to date (vs. historical) and detailed (vs. summarized).

➔ <u>A Data Warehouse (DW or sometimes WH)</u>
  – A multi-subject oriented database populated from operational systems and/or ODS's. It is historical (vs. point-in-time) in nature and typically contains detailed data. It is often looked upon as the single source of corporate "truth".

➔ <u>A Data Mart (DM)</u>
  – A database designed to support the analysis of a particular business subject area. Data has usually been transformed and aggregated from the source DW or operational system. Data Marts can be relational, multidimensional (OLAP) or statistical in nature.

Data and Dynamic Warehousing- Some Definitions

➔ <u>Star-schema or "snowflake' design</u>
- – Structures where date is organized for BI queries in the form of a centralized 'fact' table that houses values and surrounding 'dimension' tables that contain levels within the enterprise (Product, Department, Time, etc.)

➔ <u>Dynamic Warehousing</u>
- – The next generation of data warehousing to improve customer service, optimize business processes and identify trends and insights for competitive advantage. Emphasis upon unilateral access to data in all formats (structured as well as unstructured, XML, etc.)

➔ <u>Operational Business Intelligence (aka Operational Intelligence)</u>
- – Delivering BI analytics to operations-based individuals … the 85% not serviced today … a VERY sweet spot for System z

Challenges for Traditional Warehousing

Not just for traditional query and reporting purposes anymore

➔ Must address expanding needs for business insight

➔ Must serve increasing number and types of applications

➔ Must support varying service level demands

Increasingly **mixed workload** environments and **constantly changing needs** of different business constituents require more **dynamic** and **balanced** warehousing capabilities

Dynamic Warehousing with System z
Mission-critical analysis of operational data

Rapid and secure user-access to data analysis

➔ *Interactive executive dashboards and information portals*

Improved query and reporting optimizations

➔ *SQL Procedures may be run on zIIP*

Up to 50% reduction of storage for indexes

➔ *Improved index compression*

Up to 50% reduction of CPU utilization

➔ *Across most utilities*

Over 1 Billion per Hour
UPS runs DB2 for z/OS to support the
world's largest known peak database workload
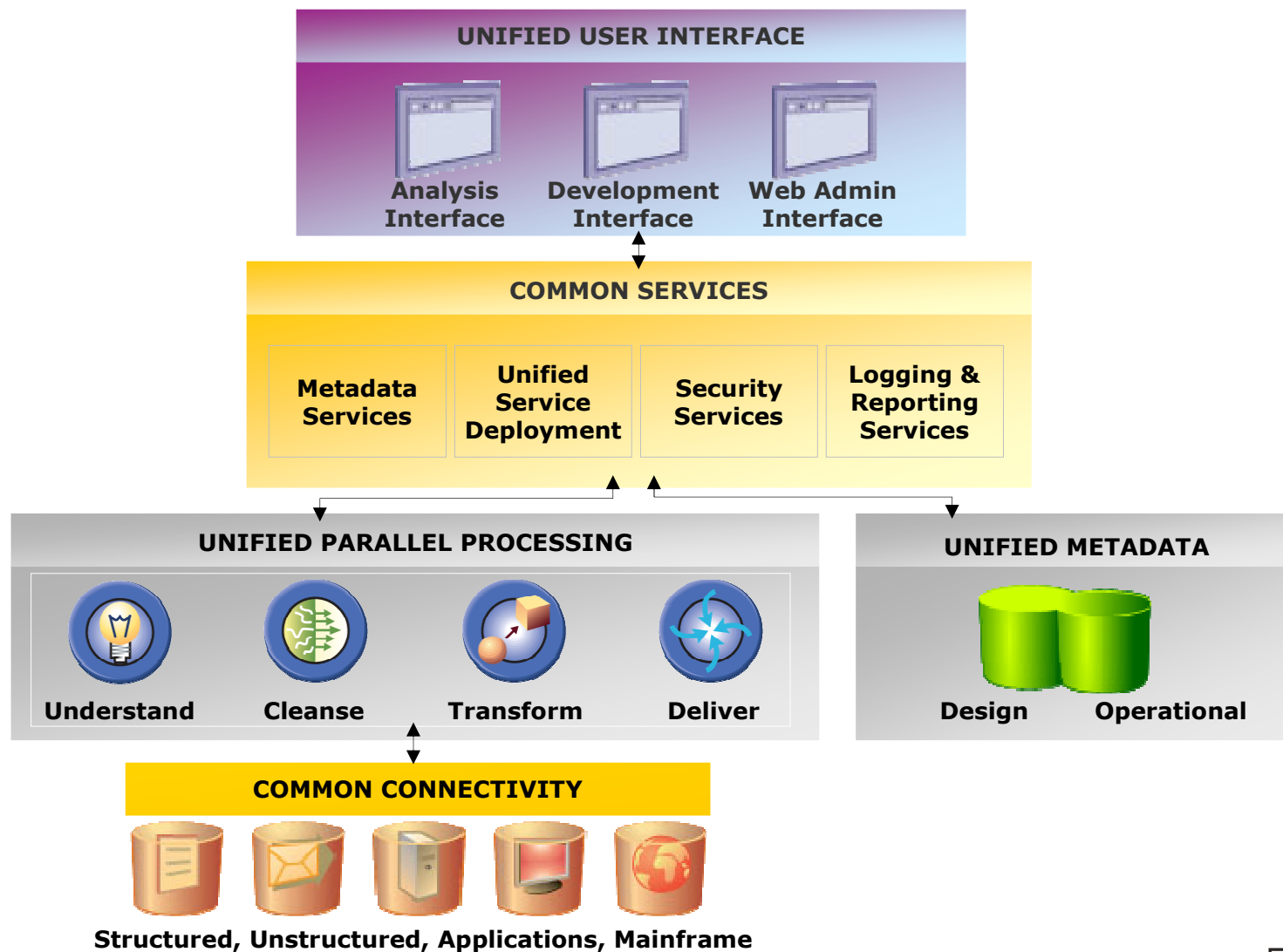- 1.1 Billion SQL statements per hour! [1]
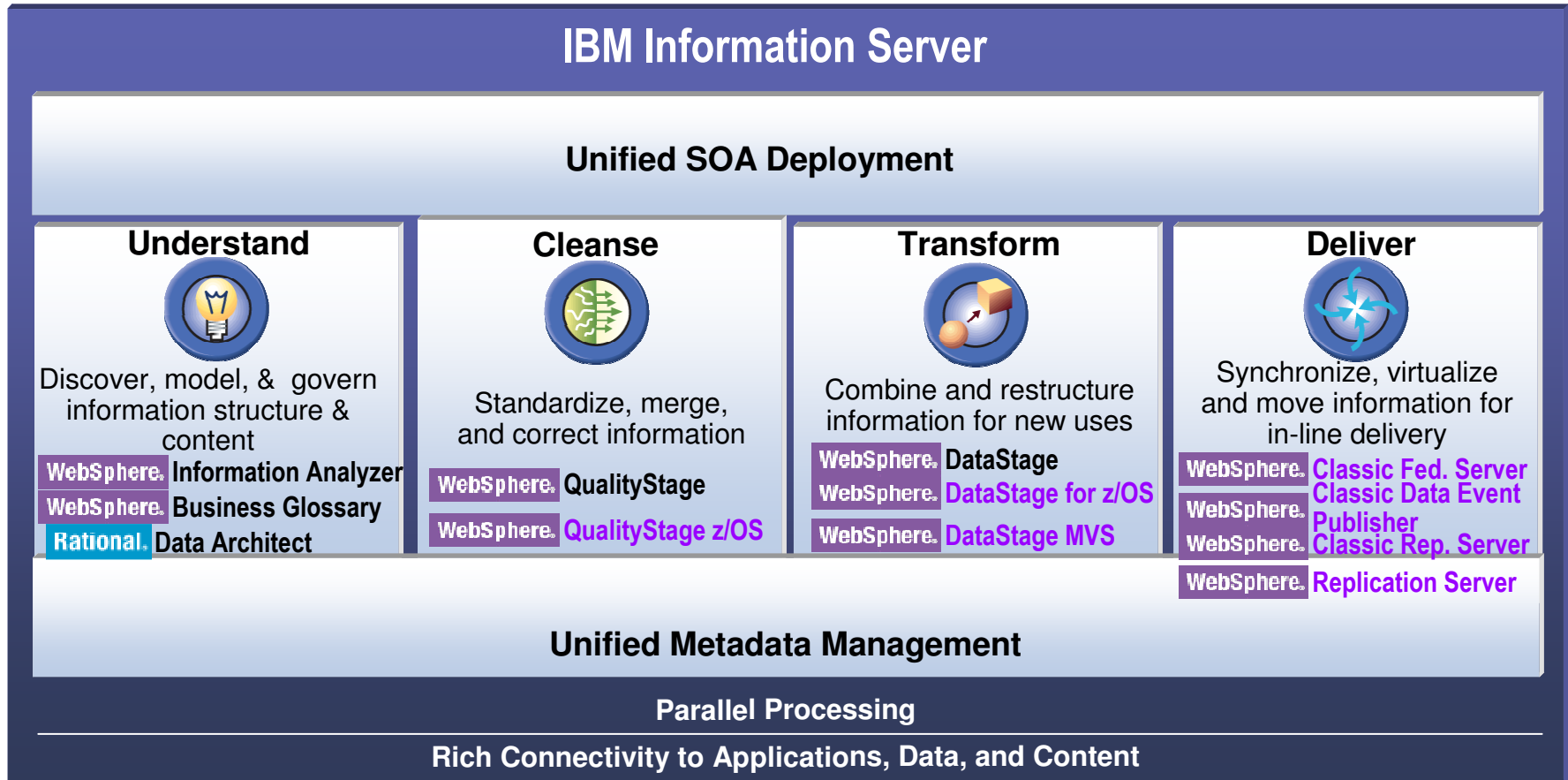
The InfoSphere Information Server
Building the Data Warehouse

IBM Information Server Architecture

First … provide an Infrastructure to enable DW
IBM Information Server - for Linux on System z

IBM Information Server

Unified SOA Deployment

Understand	**Cleanse**	**Transform**	**Deliver**
Discover, model, & govern information structure & content	Standardize, merge, and correct information	Combine and restructure information for new uses	Synchronize, virtualize and move information for in-line delivery
WebSphere. Information Analyzer	WebSphere. QualityStage	WebSphere. DataStage	WebSphere. Classic Fed. Server
WebSphere. Business Glossary	WebSphere. QualityStage z/OS	WebSphere. DataStage for z/OS	WebSphere. Classic Data Event Publisher
Rational. Data Architect		WebSphere. DataStage MVS	WebSphere. Classic Rep. Server
			WebSphere. Replication Server

Unified Metadata Management

Parallel Processing

Rich Connectivity to Applications, Data, and Content

Physical Metadata: IBM WebSphere Information Analyzer

→ Data-centric analysis of
application, database and
file-based sources

→ Secure, detailed profiling of fields,
across fields, and across sources

→ Creation of metadata from
profiling results

→ Results instantly promotable
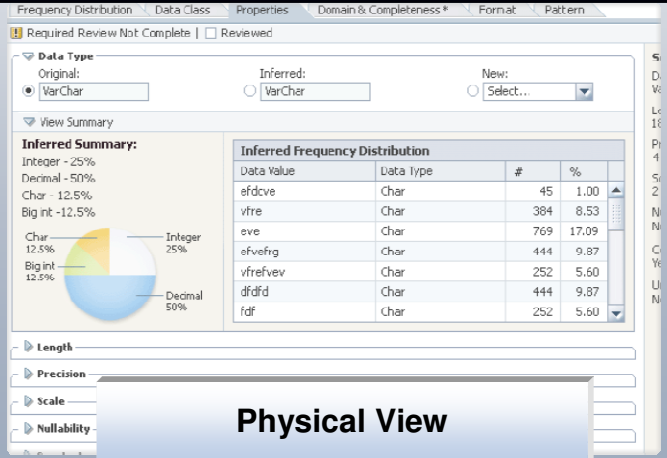across IBM Information Server

Subject Matter Experts

Data Analysts

Understand

IBM WebSphere Information Analyzer

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

| Frequency Distribution | Data Class | Properties | Domain & Completeness * | Format | Pattern |

⚠ Required Review Not Complete | ☐ Reviewed

▽ **Data Type**

Original: ● VarChar   Inferred: ○ VarChar   New: ○ Select...

▽ View Summary

Inferred Summary:
Integer - 25%
Decimal - 50%
Char - 12.5%
Big int -12.5%

Char 12.5% Integer 25%
Big int 12.5% Decimal 50%

Inferred Frequency Distribution			
Data Value	Data Type	#	%
efdcve	Char	45	1.00
vfre	Char	384	8.53
eve	Char	769	17.09
efvefrg	Char	444	9.87
vfrefvev	Char	252	5.60
dfdfd	Char	444	9.87
fdf	Char	252	5.60

▷ Length
▷ Precision
▷ Scale
▷ Nullability

Physical View

IBM Information
ON Demand **2010**

Business Metadata: IBM WebSphere Business Glossary

→ Web-based authoring, managing & sharing of business metadata

→ Aligns the efforts of IT with the goals of the business

→ Provides business context to information technology assets

→ Establishes responsibility and accountability

Subject Matter Experts

Business Users

Understand

IBM WebSphere Business Glossary

Create and manage business vocabulary and relationships, while linking to physical sources

eporting   Help | Log O

Overview | Manage Custom Attributes × | Manage Business Terms ×

Overview

Welcome to WebSphere Business Glossary
International Business Machines (NYSE:IBM) is the leader in enterprise data integration. Customers and partners worldwide use its information integration tools to confidently transform data into accurate, reliable and complete business information to improve operational performance and decision-making across every critical business dimension. Our comprehensive end-to-end solutions provide on demand data integration complemented by our professional services, industry expertise, and methodologies.

IBM.

USA	Europe	Data Model	Asia
Finance	Finance	Top-level category, which groups all the elements by type	Finance
Manufacturing	Manufacturing		Manufacturing
Customer Relationship	Customer Relationship		Customer Relationship

See All Categories

Business View

Database = DB2

Schema = NAACCT

Table = DLYTRANS

Column = ACCT_NO

data type = char(11)

Technical   Business

GL Account Number

The ten digit account number. Sometimes referred to as the account ID. This value is of the form L-FIIIIVVVV.

Logical Metadata: Rational Data Architect

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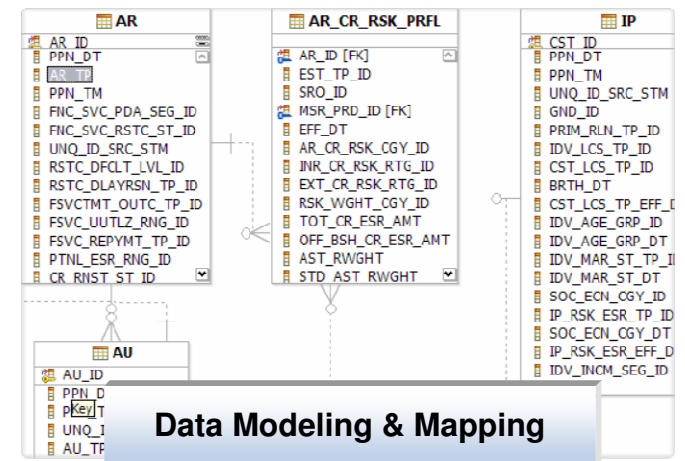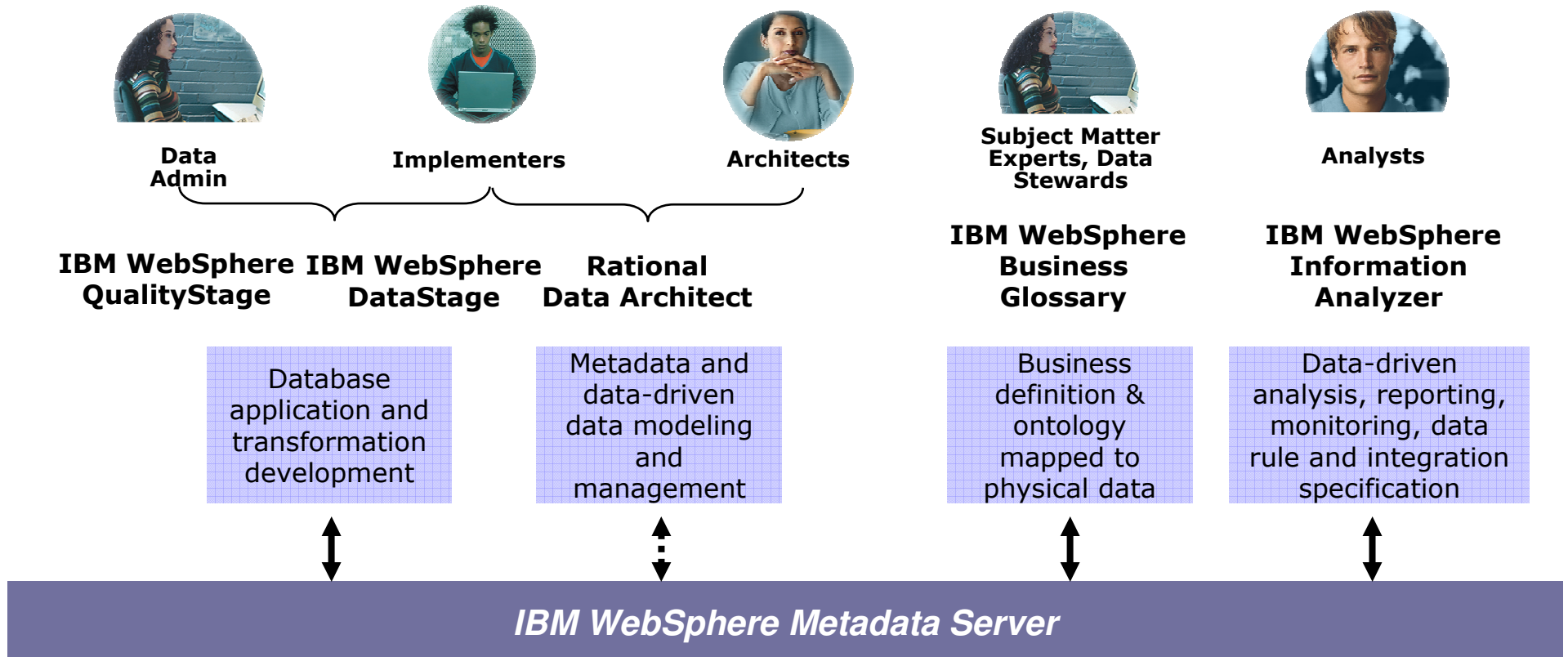- SQL & XML generation capabilities

Subject Matter Experts

Architects

Rational Data Architect

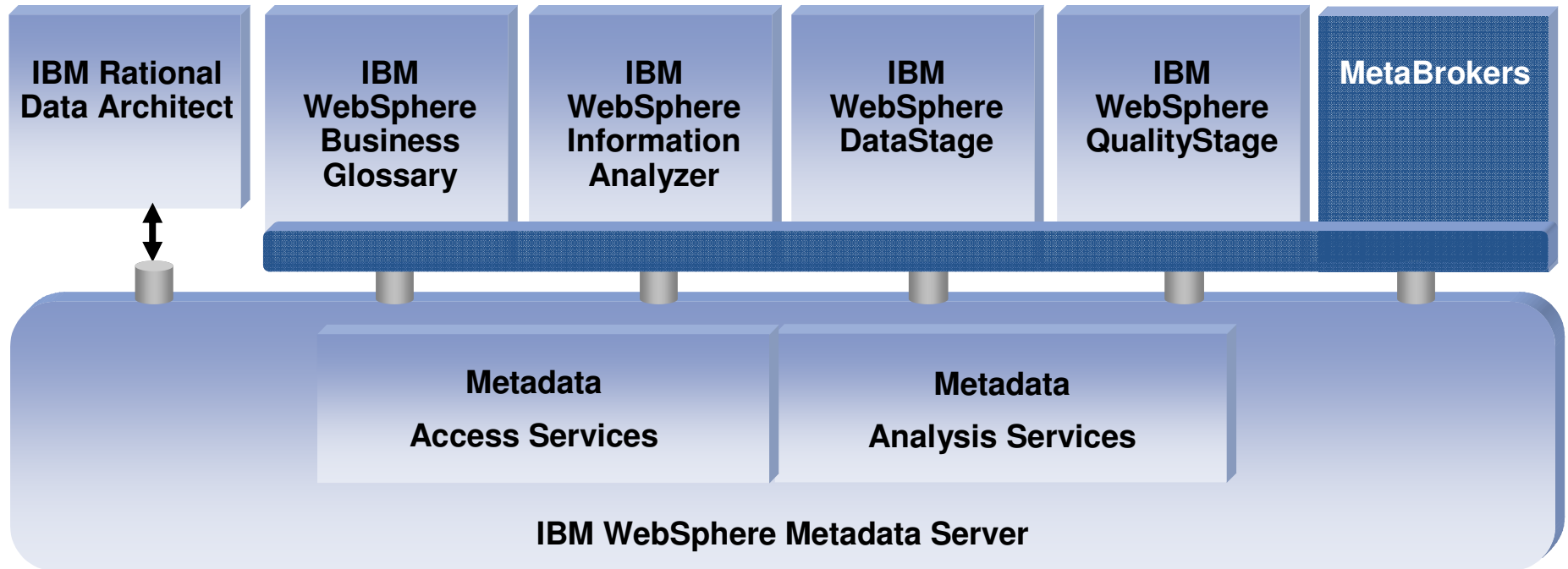**Create and manage business vocabulary and relationships, while linking to physical sources**

Data Modeling & Mapping

Role-Based Tools with Integrated Metadata

Data Admin	Implementers	Architects	Subject Matter Experts, Data Stewards	Analysts
IBM WebSphere QualityStage	**IBM WebSphere DataStage**	**Rational Data Architect**	**IBM WebSphere Business Glossary**	**IBM WebSphere Information Analyzer**
Database application and transformation development		Metadata and data-driven data modeling and management	Business definition & ontology mapped to physical data	Data-driven analysis, reporting, monitoring, data rule and integration specification

IBM WebSphere Metadata Server

- **Simplify integration**
- **Facilitate change management & reuse**
- **Increase compliance to standards**
- **Increase trust and confidence in information**

IBM WebSphere Metadata Server – at the Core of IBM Information Server

| IBM Rational Data Architect | IBM WebSphere Business Glossary | IBM WebSphere Information Analyzer | IBM WebSphere DataStage | IBM WebSphere QualityStage | MetaBrokers |

Metadata Access Services

Metadata Analysis Services

IBM WebSphere Metadata Server

Why Should I Care About Cleansing Information?

→ **Lack of information standards**
  – Different formats & structures across different systems

→ **Data surprises in individual fields**
  – Data misplaced in the database

→ **Information buried in free-form fields**

→ **Data myopia**
  – Lack of consistent identifiers inhibit a single view

→ **The redundancy nightmare**
  – Duplicate records with a lack of standards

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

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

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)

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

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

Data Cleansing: IBM WebSphere QualityStage

→ Specialized data quality functions seamlessly integrated with DataStage

→ Visual tools for defining complex matching and survivorship logic

→ Ensures clean, standardized, de-duplicated information

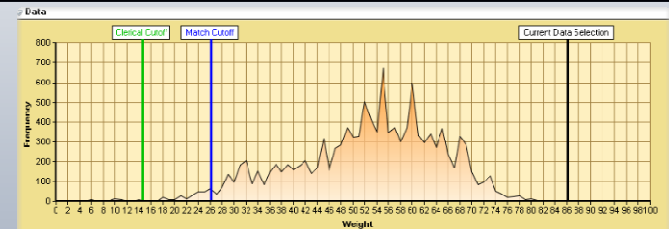→ Enables a single version of the truth

Subject Matter Experts

Data Analysts

Cleanse

IBM WebSphere QualityStage™

**Standardize and correct source data fields, and match records together across sources to create a single view**

Visual Match Rule Design

What Is Important About Transformation & Delivery?

→ Transformation is key to enabling information to be used in new business contexts – it needs to be metadata-driven

→ Designed for use by information experts using the understanding imparted by the metadata

→ Transformation and Delivery can be reused across multiple mechanisms
  - Large volume batch movement
  - Real-time event-driven response
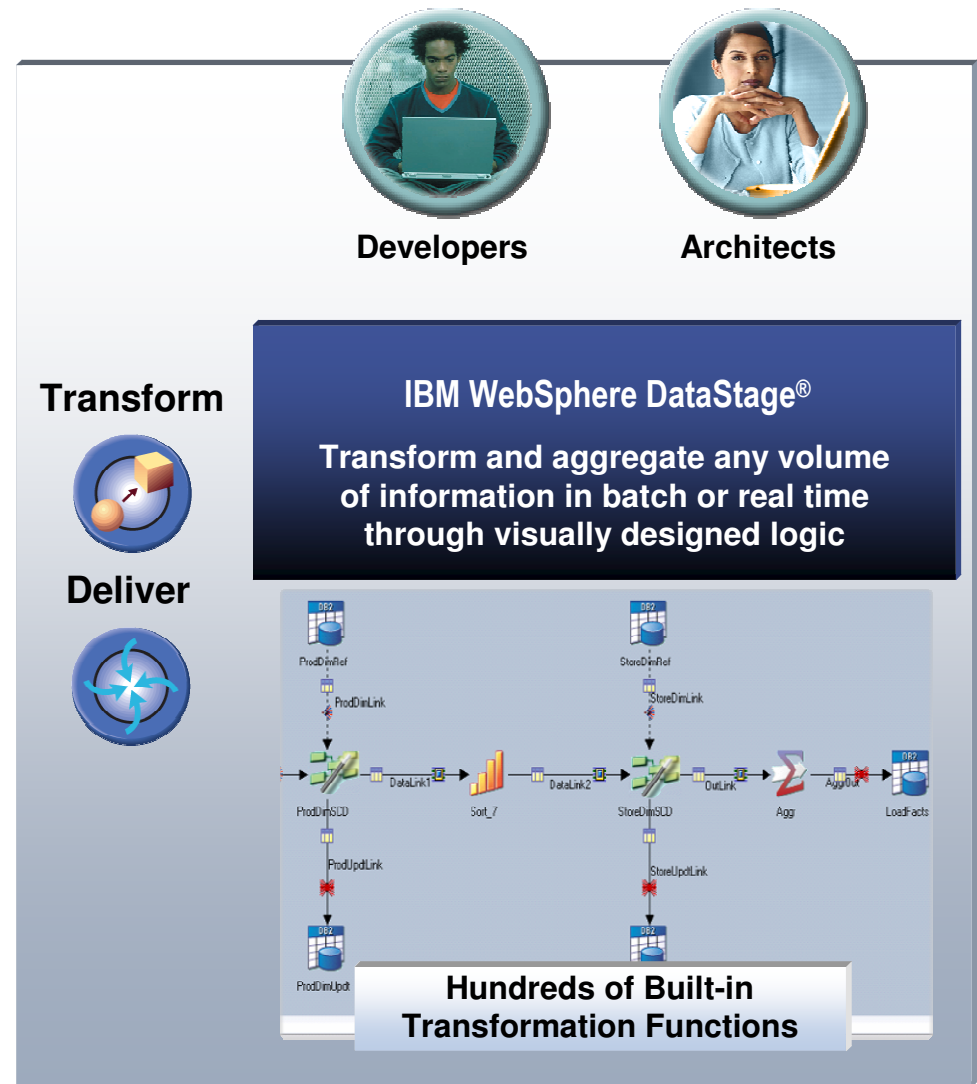 - Service-oriented architecture
  - Federated query
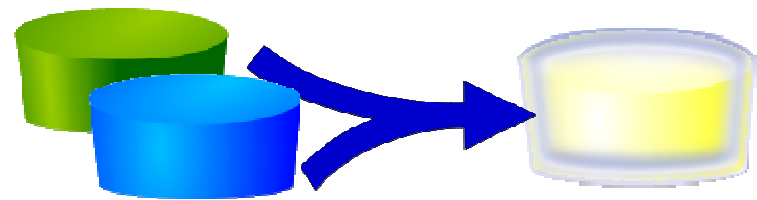
Data Analysts **Data Architects** **DBAs** **Subject Matter Experts**

Logic Reuse

Request Response

Query

IBM Information ON Demand 2010
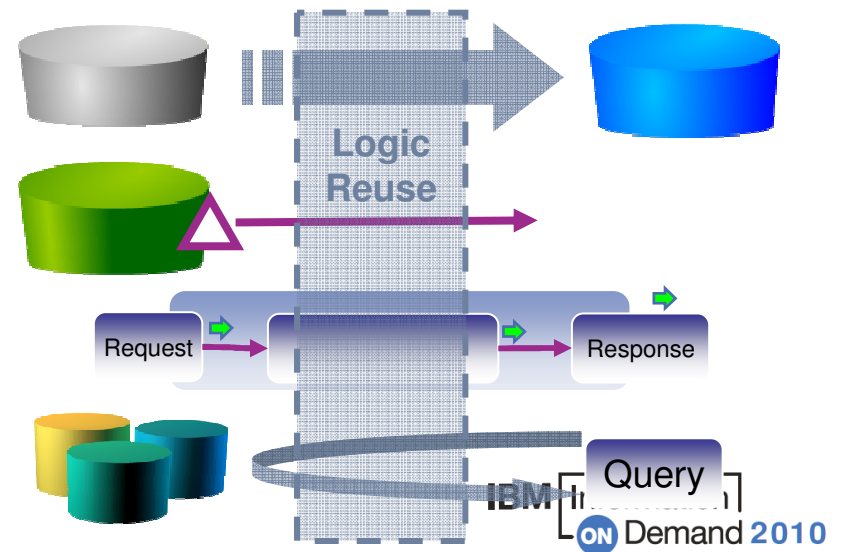
Data Transformation & Movement: IBM WebSphere DataStage

→ Codeless visual design of data flows with hundreds of built-in transformation functions

→ Optimized reuse of data integration objects

→ Leverages parallel processing without requiring design changes

→ Capable of supporting batch and real-time operations

Developers **Architects**

Transform

Deliver

IBM WebSphere DataStage®

Transform and aggregate any volume of information in batch or real time through visually designed logic

Hundreds of Built-in Transformation Functions

Data Federation: IBM WebSphere Federation Server and Classic Federation

- Access diverse & distributed information as if it were in one system

- Industry leading query optimization with single sign-on, unified views, and function compensation

- Transactional write capabilities across heterogeneous sources

- Visual tools for federated data discovery & data modeling

Deliver

IBM WebSphere Federation Server

Access and integrate heterogeneous information across multiple sources as if they were a single source

Extend value of existing analytical applications by providing real-time access to integrated information
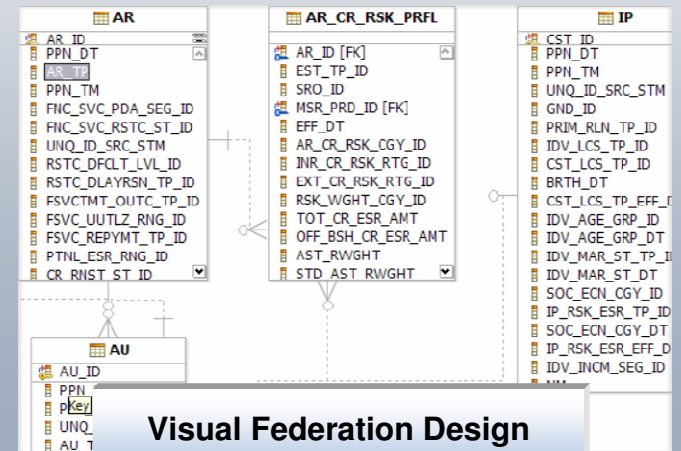
Visual Federation Design

The IBM Information Server Advantage

A Complete Information Infrastructure

➜ A *comprehensive, unified foundation* for enterprise information architectures, scalable to any volume and processing requirement

➜ *Auditable data quality* as a foundation for trusted information across the enterprise

➜ *Metadata-driven integration*, providing breakthrough productivity and flexibility for integrating and enriching information

➜ *Consistent, reusable information services*—along with application services and process services, an enterprise essential

➜ Accelerated time to value with *proven, industry-aligned solutions* and expertise

➜ *Broadest and deepest connectivity* to information across diverse sources: structured, unstructured, mainframe, and applications

The InfoSphere Warehouse for System z
Enhancing the Data Warehouse

Key Component Overview - Warehouse for tooling DB2 for z/OS

➡ **Data Server Component – Prerequisite**

- DB2 for z/OS -- Not included in the offering, prerequisite
  Application Server Components – Linux on System z Partition
- WebSphere Application Server
  - Manages/executes the SQW runtime processes
  - Server for the browser-based Admin Console
- SQW Runtime
  - Execution engine for SQW jobs (control flows, data flows)
  - Runs in the WAS environment
- Cubing Services engine
  - Cube server providing cube access to MDX clients through ODBO and XMLA
  - Cubing engine and cube optimizer ported to Linux on System z.

➡ **Design and Administration Client Components – Windows, Linux**

- Design Studio, Eclipse-based design tool
  - Physical modeling based on Rational Data Architect (RDA)
  - SQW Data flow, control flow editor – extension of RDA
  - Cubing Services cube model editor, optimization tool – extension of RDA
- Admin Console, browser-based administration tool
  - Common admin tasks, data source, role-based security assignments
  - Cube metadata management (import/export)
  - Cube Server management (configure / start / stop, etc.)
  - SQW runtime job administration

IBM Information On Demand 2010

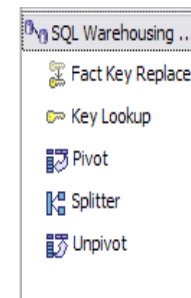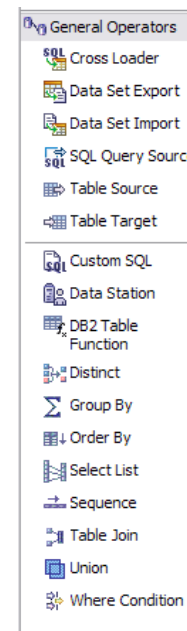# SQW- Data Flows for <u>in-database</u> data movement and transformation

➡ **Eclipse tool that generates SQL instructions, optimized for DB2, to move and transform data within the database**

– Ease of use, drag and drop UI for fast warehouse building

– Integrated with physical data model for source and target tables and other database objects

– Data flow for doing transformations: 30+ SQL operators optimized for DB2

• Leverages DB2 functions for transformations 100+

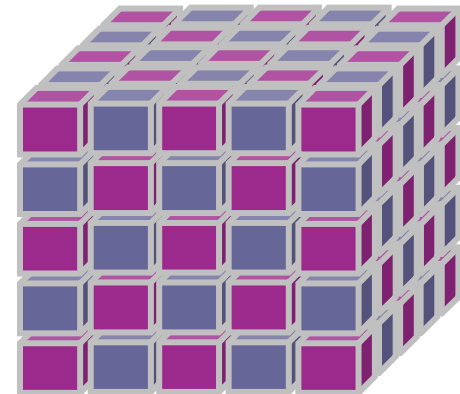• Customize with User-Defined Functions or custom SQL

– Test and debug dataflows
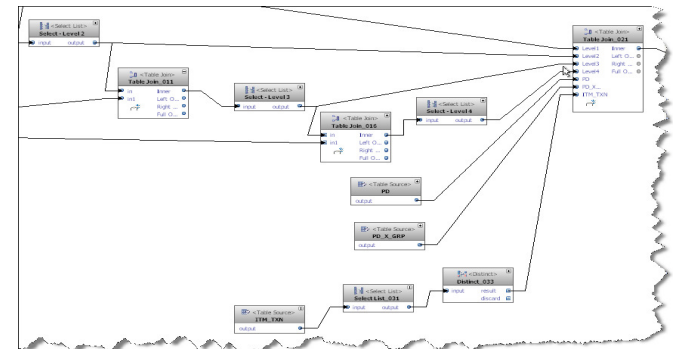
➡ **Perform warehouse building in homogeneous "database" environment**

– Leverage existing database infrastructure

– Leverage existing DBA skills, reuse custom SQL code

– Native integration for fast BI application deployment

• Database modeling and administration

• Data Prep for in-database Report, Mining & OLAP

• Analyze impact from tables to BI applications

• Allow a team of BI designers to work together, and deal with changes quickly

General Operators

SQL Cross Loader
Data Set Export
Data Set Import
SQL Query Source
Table Source
Table Target
Custom SQL
Data Station
DB2 Table Function
Distinct
Group By
Order By
Select List
Sequence
Table Join
Union
Where Condition

SQL Warehousing ...
Fact Key Replace
Key Lookup
Pivot
Splitter
Unpivot

OLAP - Cubing Services

Cubing Services (OLAP)

Cubing Services is a Multidimensional Analysis Server that enables OLAP applications access to Terabyte data volumes via industry standard OLAP connectivity

Benefits

➜ Uses industry standards OLAP Interfaces for access by a wide variety of tools for presentation and reporting

➜ Empowers users with ad hoc access to business information.

– What is the profitability for Product A across the Branches X,Y,Z?

➜ Speed of thought access to OLAP data managed in a DB2 Data Warehouse. Thus for your DB2 Warehouse OLAP data there is:

– OLAP and SQL shared access to the same information

– Single point of Management

– Single point of Maintenance

– Single point of Performance Tuning

➜ Enables Access to up to 1 TB of base OLAP Data

IBM DB2 Warehouse

IBM Information
ON Demand 2010
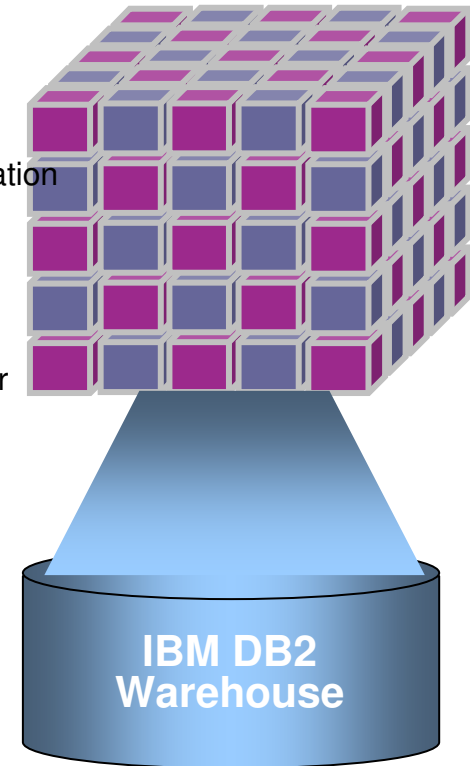
Cubing Services - OLAP Analytics and Open Access
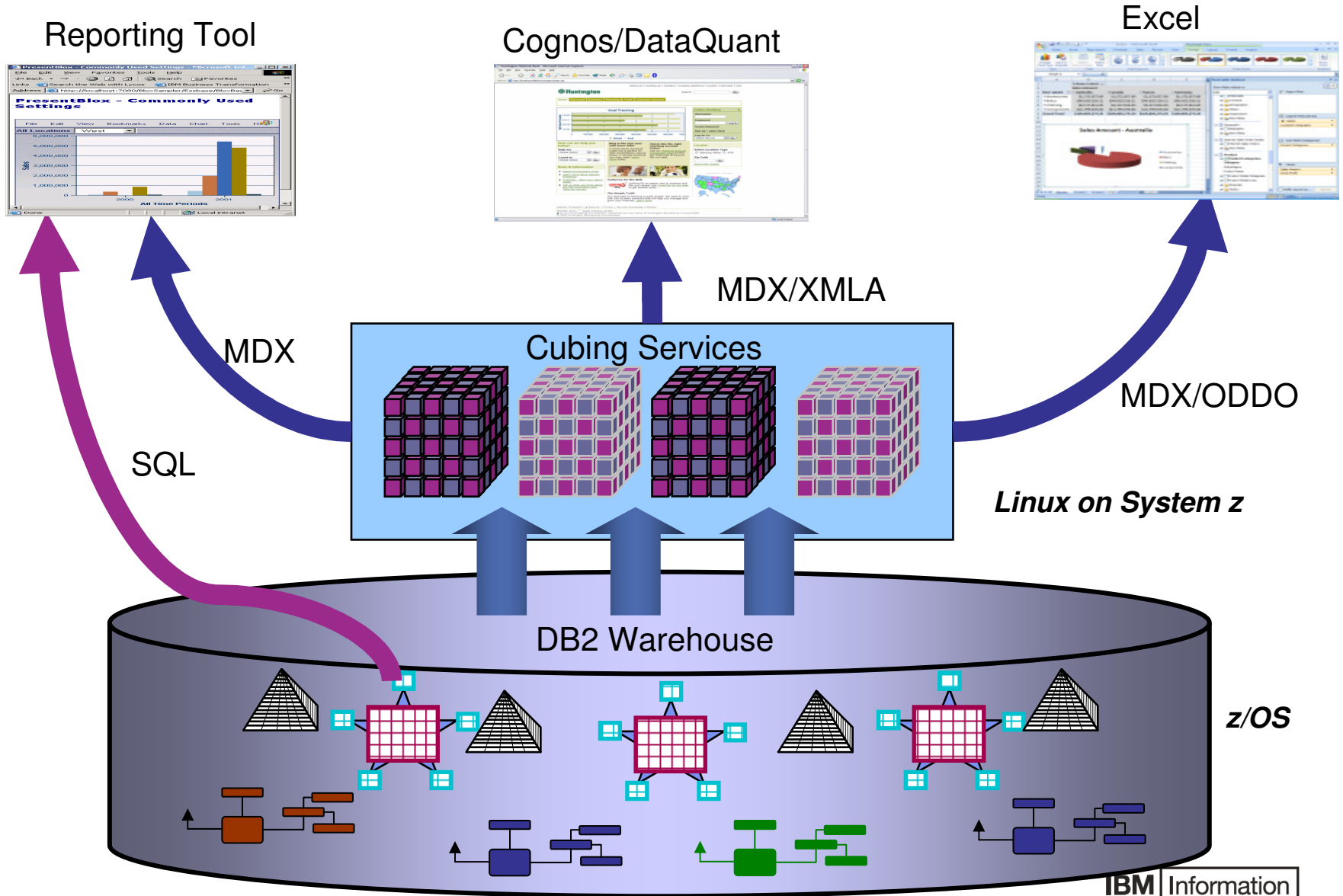
IBM Smart Analytics Optimizer
Technology Preview for System z

What is it?

✓ A high performance extension that easily integrates with IBM data systems, delivering predictable, order-of-magnitude faster, analytic query response times, while lowering operating costs


How is it different

✓ **Deep integration with IBM data management systems**

✓ **High performance query software, based on advanced data in-memory technologies**

✓ **Leveraging existing data system investment and values without any changes to applications**

✓ **For System z, extends gold-standard manageability, security, and availability to high-performance analytic applications**

Pre-announced now deferred to 2010

IBM Information On Demand 2010

Data Warehouse and BI Architecture for System z
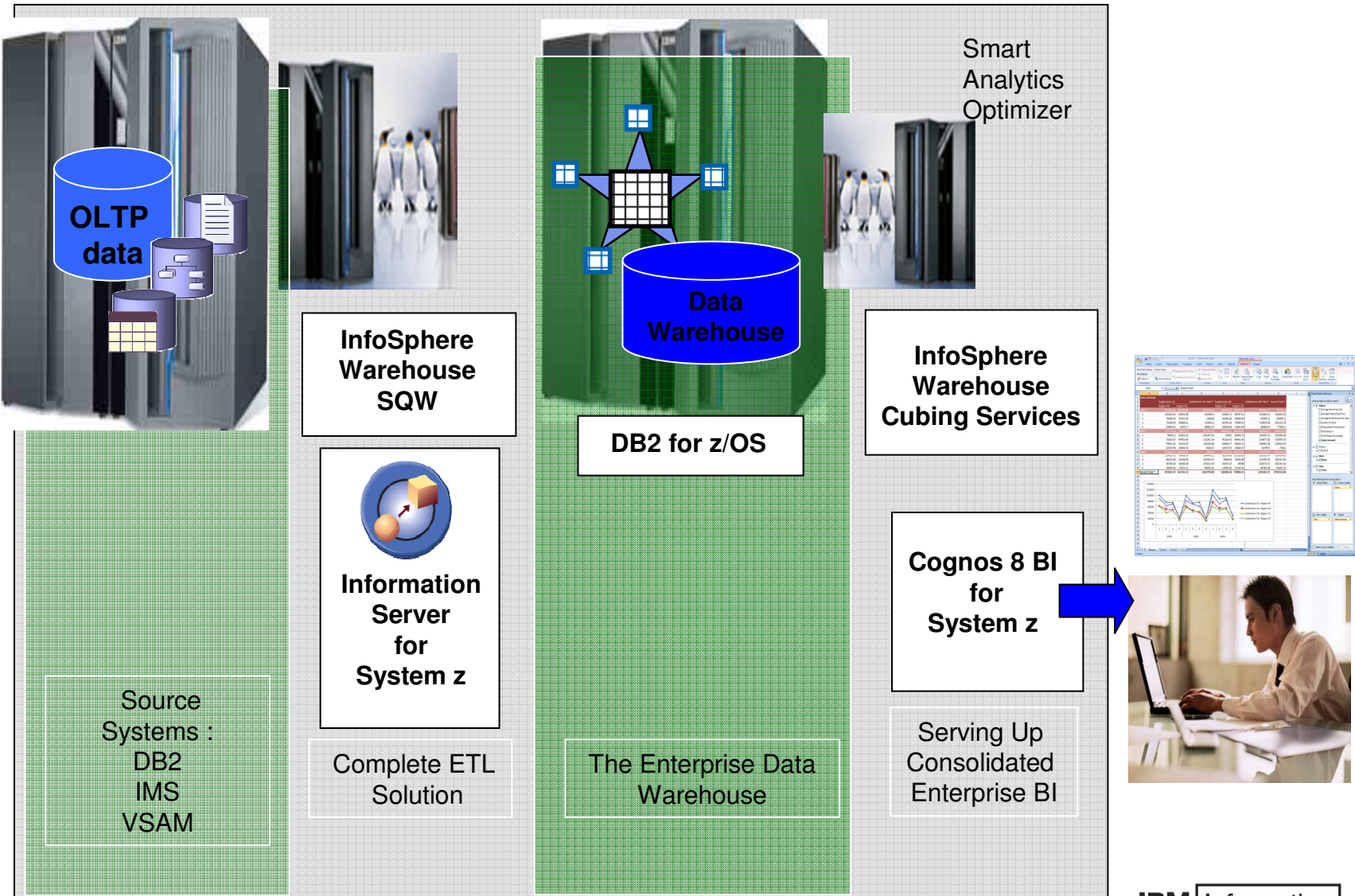
Summary

→ We have enhanced the System z portfolio significantly to accommodate requests for data warehousing

→ IBM's 2 major offerings on System z are

– Information Server

– InfoSphere Warehouse

→ Data Warehousing is typically driven by Business Intelligence requirements

→ DW is not a prerequisite for BI

→ Without trusted information there will be a justified opening for continued, random, end user initiatives "outside the box"

Thank You!