

***Il Mondo dei Partner***  
***INNOVARE E CRESCERE. INSIEME*** **2006**  
*Milano 19 - 20 Ottobre* ***Technical World***

**Soluzioni IBM per la gestione integrata delle informazioni**

***Paolo Crivelli***

[paolo.crivelli@it.ibm.com](mailto:paolo.crivelli@it.ibm.com)

**IBM**

# Corporate View of Information Architecture Is Changing

- **Information is the key to Business Innovation**
  - Organizations highly **effective** at driving information integration are five times more likely to drive value creation
  - **Information architecture can't exist in a vacuum – it needs to be tied to enterprise architecture**



*87% of CEOs believe fundamental **change** is required in next two years to drive innovation*

*Over 60% of CEOs believe their organizations need to do a better job leveraging information*

*Source: 2006 IBM Global CEO Survey*

# What is Driving the Change? – Gartner Perspective

## Efficiency

### Process Simplification

- Eliminate redundancy
- Drive to standardization
- Promote reuse and data quality

### Compliance

- Reduce risks with conflicting sources
- Make information transparent

### "Infoglut"

- Manage expanding volume and velocity
- Control unstructured content

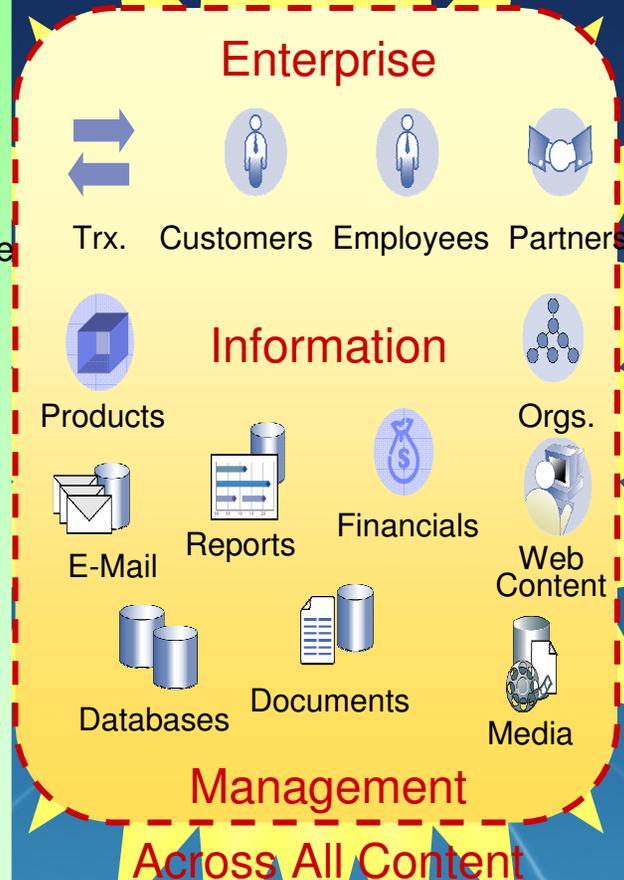
### Vendor Consolidation

- Spend less on same functionality/technology

### M&A

- Reduce integration burdens

Across the Enterprise



## Differentiation

### Enterprise Agility

- Sense and respond
- Provide consistency, accuracy
- Support continuous information flows
- Rapid orchestrate processes

### Real Time

- Enable closed-loop analytics
- Immediately integrate with partners, suppliers

### Single View

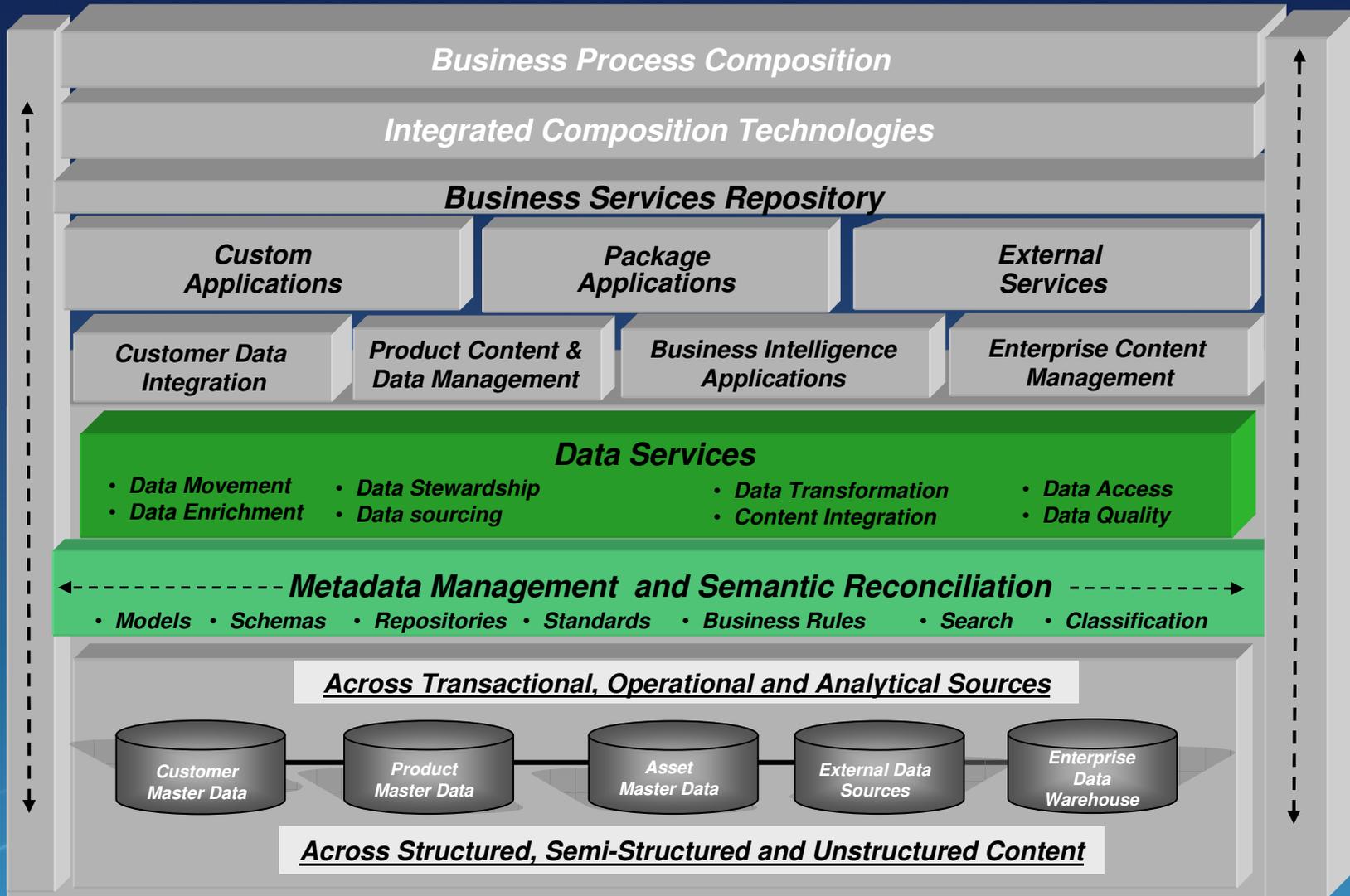
- Create consistent and holistic view across all channels
- Manage relationships

### Revenue Optimization

- Support top-line growth on cross-sell/upsell
- Leverage global purchasing power

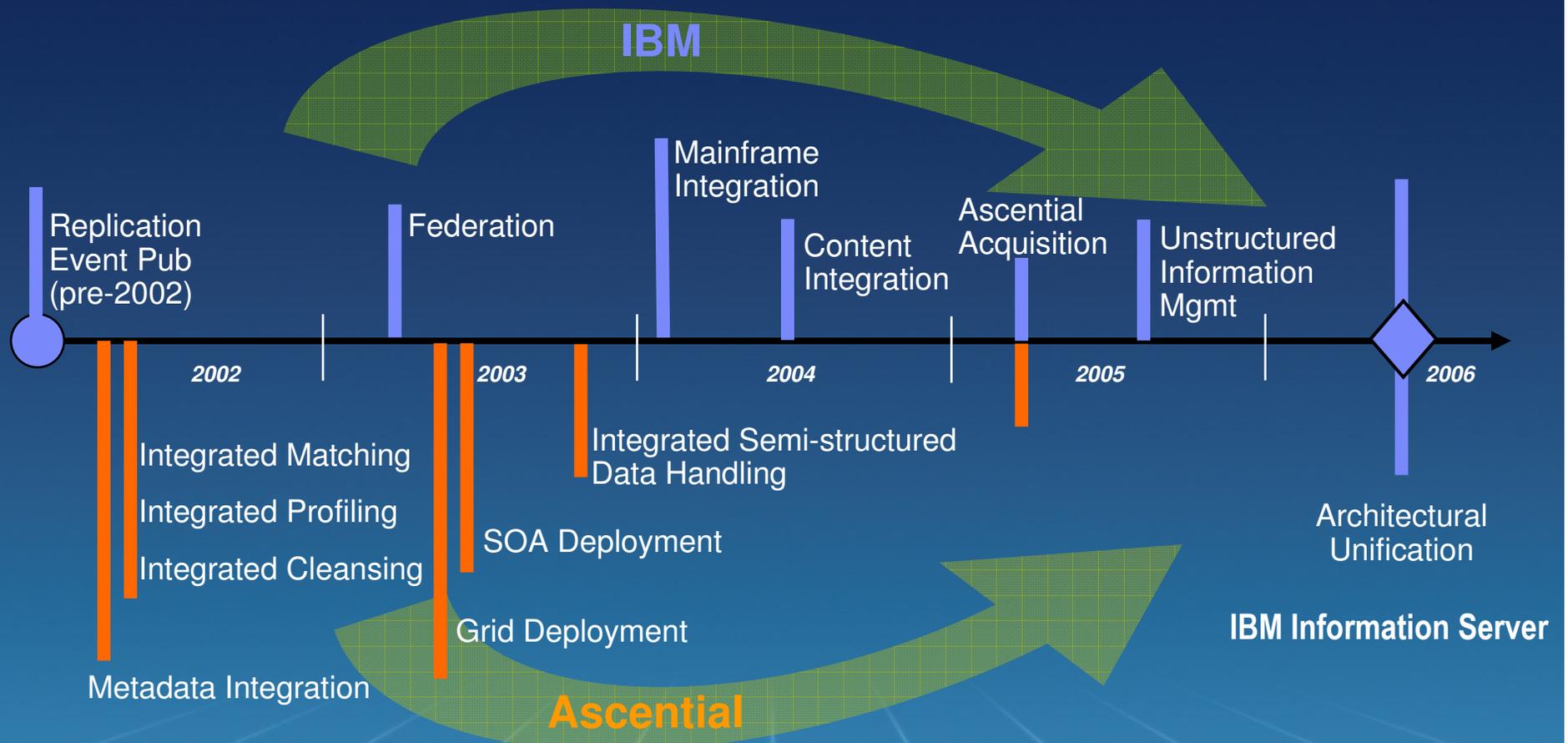
Enterprise Information Management: Getting Value From Information Assets Gartner Business Intelligence Summit 2006 David Newman 6-8 March 2006

# How Gartner Defines the Requirement



Enterprise Information Management: Getting Value From Information Assets Gartner Business Intelligence Summit 2006 David Newman 6-8 March 2006

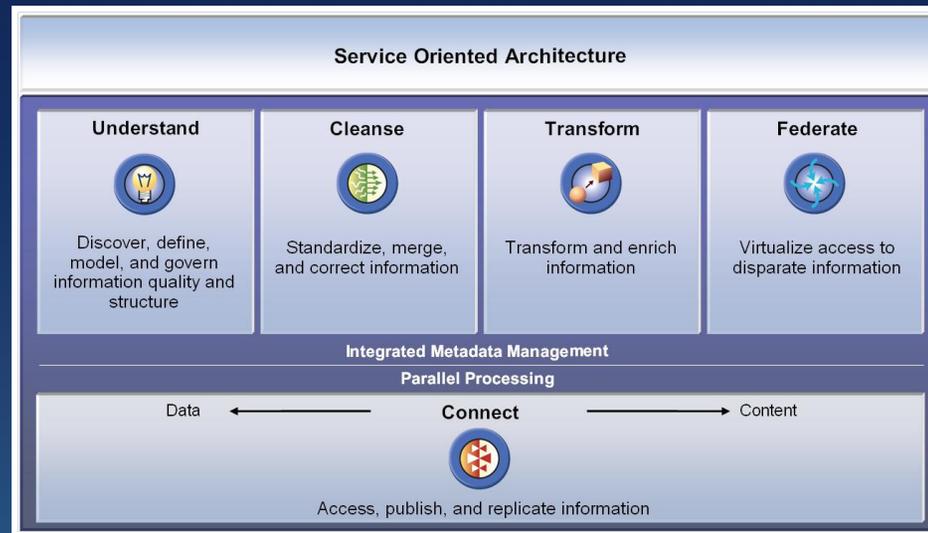
# The Construction of Our Platform



# The IBM Solution, STEP ONE: WebSphere Information Integration Platform

*Real-time, access-in-place technology and event-driven data movement for any data and content*

*High-speed, high-volume, parallel data movement, transformation, and data quality.*



- *Enterprise search*
- *Text analysis*
- *Data and content federation*
- *Replication and event publishing*



WebSphere Information Integrator



WebSphere Data Integration Suite (Ascential)

- *Data profiling*
- *Data quality*
- *Data transformation*
- *Data extraction and loading*

# The IBM Solution, STEP TWO: IBM Information Server

*Delivering information you can trust*

## IBM Information Server

### Unified Deployment

#### Understand



Discover, model, and govern information structure and content

#### Cleanse



Standardize, merge, and correct information

#### Transform



Combine and restructure information for new uses

#### Deliver



Synchronize, virtualize and move information for in-line delivery

### Unified Metadata Management



Parallel Processing



Rich Connectivity to Applications, Data, and Content

# The IBM Solution: IBM Information Server

*Delivering information you can trust*

## IBM 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



Synchronize, virtualize and move information for in-line delivery

## Platform Services

### Parallel Processing Services



### Connectivity Services



### Metadata Services



### Administration Services



### Deployment Services



# The IBM Solution: IBM Information Server

*Delivering information you can trust*

## IBM 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



Synchronize, virtualize and move information for in-line delivery

## Platform Services

### Parallel Processing Services



### Connectivity Services



### Metadata Services



### Administration Services

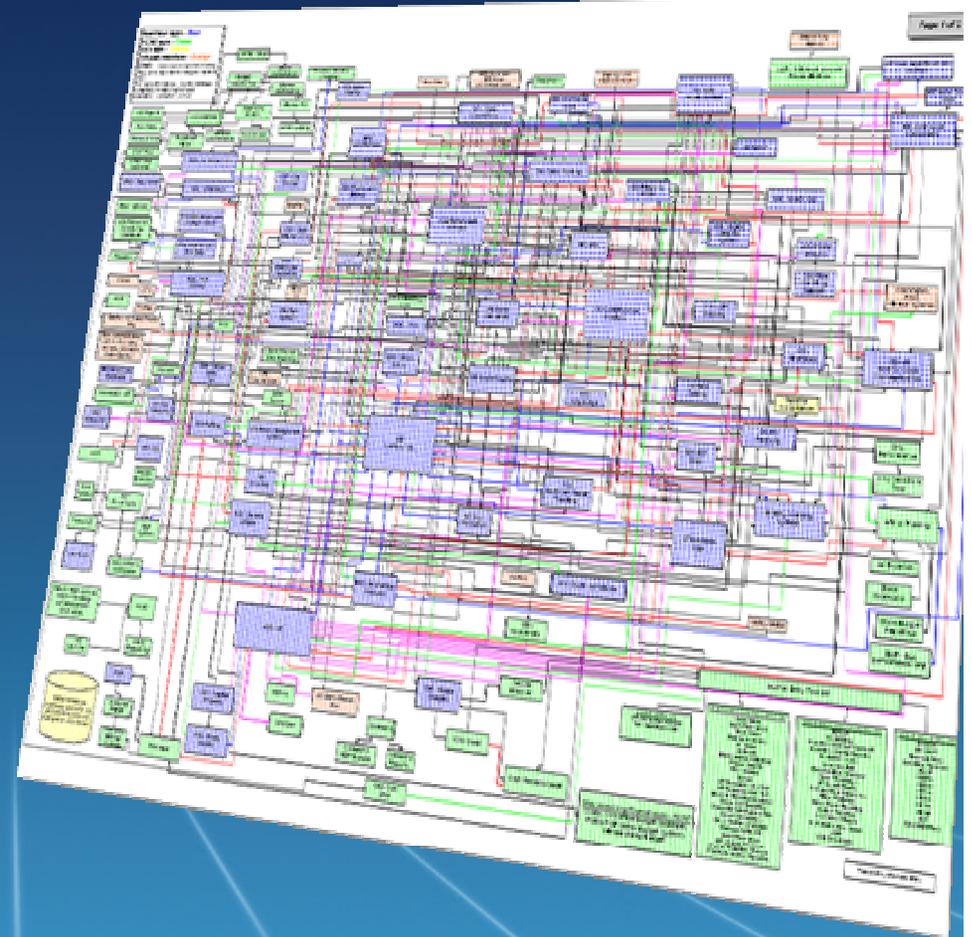


### Deployment Services



# Why Is it Important to Start with Understanding?

- Where is my information?
- How do I get it when I need it?
- What does it mean?
- Can I trust it?
- How do I get it in the form I need?
- How do I get it where it needs to go?
- How do I control it?



# What is WebSphere Information Analyzer?

- **Provides in-depth analysis of existing systems**
  - Data-centric analysis of application, database, and file-based sources for content, quality, and structure
  - Secure, detailed profiling of fields, and relationship analysis across fields and across sources
- **Enables ongoing measurement and baseline reporting of information quality**
- **Creates metadata that describes where information is managed across systems**
  - Provides an understanding of the fitness of specific sources and highlights data that may need downstream attention

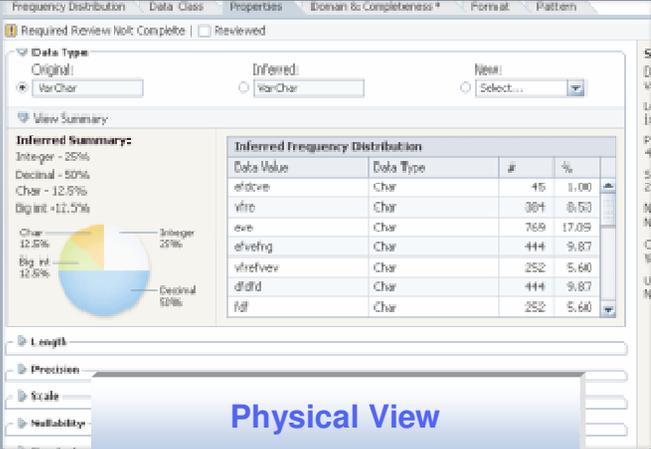
**Subject Matter Experts**      **Data Analysts**

**Understand**



**IBM Information Analyzer**

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



**Physical View**

Data Value	Data Type	#	%
efdcve	Char	415	1.00
vire	Char	384	0.50
eve	Char	769	17.09
efvefng	Char	444	9.87
virefvev	Char	252	5.60
dldld	Char	444	9.87
fdl	Char	252	5.60

# What Makes WebSphere Information Analyzer Different?

Comprehensive, Collaborative Analysis	<u>Benefits</u>
Complete analysis, including extremely detailed field-level analysis, cross-field analysis, and cross-source analysis	<b>Ensures a complete understanding of the structure, content, and quality of sources</b>
Intuitive, methodology-driven design environment, with collaboration capabilities and over 30 built-in reports	<b>Speeds analysis time and allows subject matter experts to be involved, to further improve results</b>
Intuitive, graphical analysis and reporting, with drill through capabilities	<b>Makes it easier to identify issues and to involve business users in analysis</b>
Analysis and results flow downstream to development tasks	<b>Accelerates downstream development</b>
Handles unicode data	<b>Supports global data</b>
Ongoing Quality Measurement	
Baseline measurement and comparison capabilities	<b>Allows data quality to be measured and managed over time</b>
Analysis can be scheduled, or triggered from job sequences	<b>Allows analysis to be automated</b>
Secure Data Analysis	
Individual fields can be secured within the analysis process	<b>Ensures data privacy</b>

# What is WebSphere Business Glossary?

- Enables Web-based authoring, sharing & managing 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

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.



**Subject Matter Experts**



**Business Users**



**Understand**

**WebSphere Business Glossary**

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



# What Makes WebSphere Business Glossary Different?

Provides Tools for the Business	<u>Benefits</u>
Business users can create and collaborate on terms, definitions, synonyms, usage examples, and taxonomies that reflect the business viewpoint.	<b>Enables business users to actively participate in project specification, Solutions align better to the business</b>
Provides rich tools for working with business metadata, including extensive reporting, search, search by synonym, impact analysis, annotations, and more.	<b>Approachable &amp; easy to use for the business, Provides a living record of business rules that encourages reuse</b>
Facilitates Business & IT Collaboration	
Terms can be categorized and taxonomies managed across different business units, teams, product lines, or divisions appropriate to the business needs.	<b>Provides a mediation tool for different business perspectives that records all viewpoints &amp; decisions</b>
Business metadata is actively managed in the shared repository.	<b>Allows business terms to become part of overall metadata analysis</b>
Annotations can be created on terms and categories.	<b>Facilitates consensus building</b>
Allows business terms to be linked to columns or tables.	<b>Shortens analysis time, Creates better results</b>
Manages Data Stewardship	
Allows data stewards to be assigned that own terms.	<b>Supports governance &amp; accountability</b>

# The IBM Solution: IBM Information Server

*Delivering information you can trust*

## IBM 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



Synchronize, virtualize and move information for in-line delivery

## Platform Services

### Parallel Processing Services



### Connectivity Services



### Metadata Services



### Administration Services



### Deployment Services



# 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

# What is WebSphere QualityStage?

- Provides specialized data quality processing**
  - Ensures clean, standardized, **de-duplicated** 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**



**Subject Matter Experts**



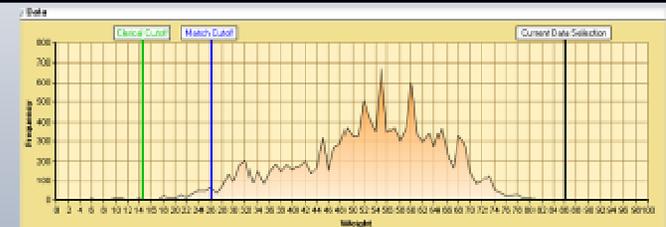
**Data Analysts**

**Cleanse**



**WebSphere QualityStage™**

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



id	PersonType	PersonNumber	Age	Gender	First Name	Last Name
4111	DA	1	05 57	IM	OTIS	GARLAND
4111	DA	1	05 57	IM	OTIS	GARLAND
5011	DA	1	06 08	IM	NICHOLAS	T
5011	DA	1	06 08	IM	NICHOLAS	T
3420	DA	1	01 68	IM	CLAUD	LADALE
3420	DA	1	01 68	IM	CLAUD	LADALE
3420	DA	1	01 68	IM	CLAUD	LADALE
4826	DA	1	01 68	IM	CLIFF	CLIFF
4826	DA	1	01 68	IM	CLIFF	CLIFF
1328	DA	1	01 68	IM	CHIN	CHIN
1328	DA	1	01 68	IM	CHIN	CHIN
2382	DA	1	01 68	IM	GILLIS	GILLIS
2382	DA	1	01 68	IM	GILLIS	GILLIS
6979	DA	1	01 68	IM	SCHULLE	SCHULLE
6979	DA	1	01 68	IM	SCHULLE	SCHULLE

**Visual Match Rule Design**

# What Makes WebSphere QualityStage Different?

Industry-Leading Parsing & Matching	<u>Benefits</u>
Probabilistic matching engine	<b>Produces the highest match rates and the lowest number of false positives</b>
Intuitive match designer, with data sampling, visual fine tuning, baseline comparison, match score comparison, and extensive reporting	<b>Speeds design of data quality logic, produces better results, and allows rules to be more easily tuned over time</b>
Parsing handles any number of fields & free-form fields	<b>Supports more uses &amp; better results</b>
Unified parallel processing framework	<b>Massive scalability, Lower support costs</b>
Seamless Integration of Data Quality	
Single design paradigm across data quality and ETL	<b>Developer productivity, lower costs</b>
Granular design integration of data quality logic with data processing functions, connectors, and transformations.	<b>Provides more flexibility &amp; control for quality logic, Reduces maintenance cost</b>
Unified metamodel with ETL and profiling, active metadata sharing, and comprehensive metadata analysis	<b>Speeds project delivery, reduces project risk, and improves collaboration</b>
Easy SOA deployment of data quality logic	<b>Improves the reach of data quality</b>
Global Data Support	
Global support & ability to verify 200+ country addresses	<b>Better utility, broader applicability</b>

# The IBM Solution: IBM Information Server

*Delivering information you can trust*

## IBM 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



Synchronize, virtualize and move information for in-line delivery

## Platform Services

### Parallel Processing Services



### Connectivity Services



### Metadata Services



### Administration Services

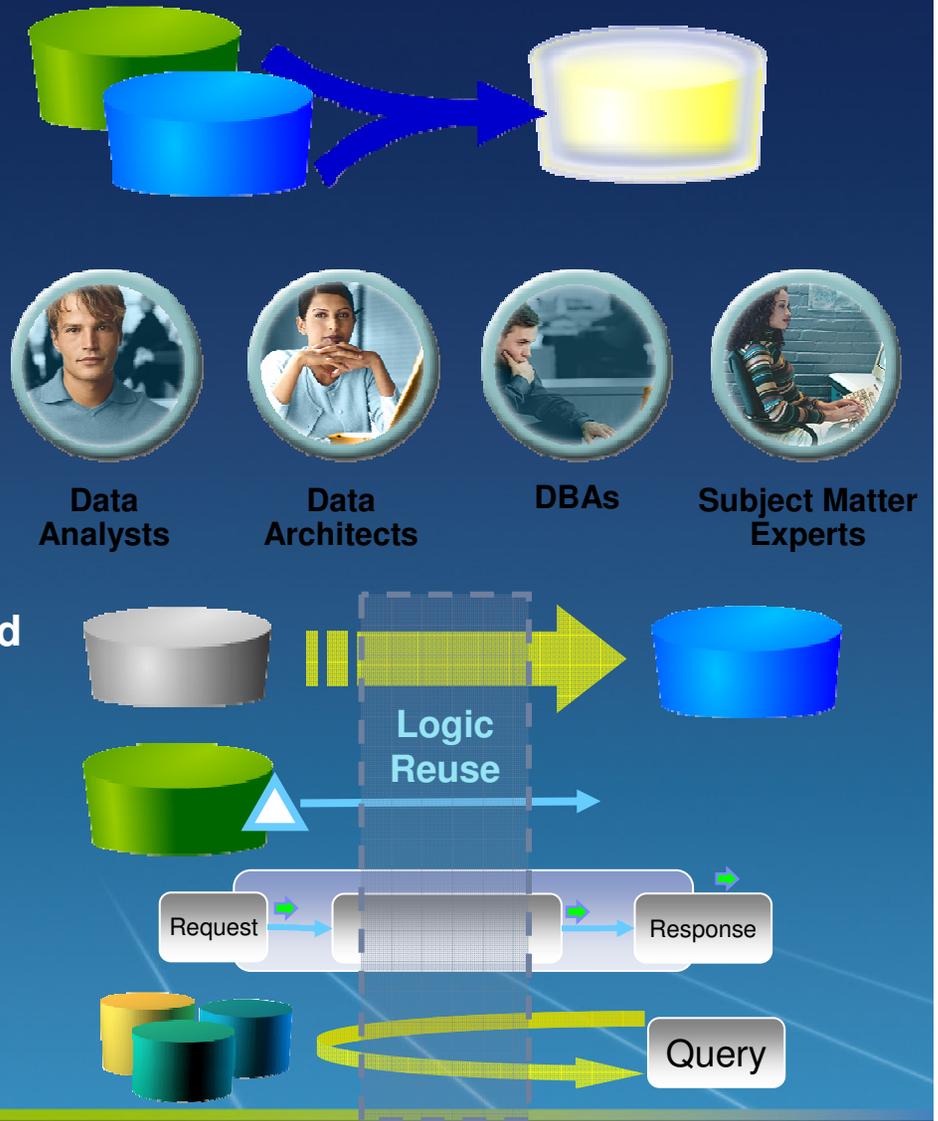


### Deployment Services



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



# What is WebSphere 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



Developers



Architects

Transform

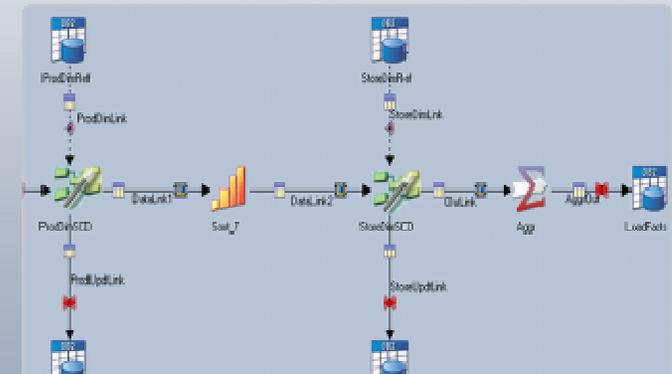


Deliver



WebSphere DataStage®

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



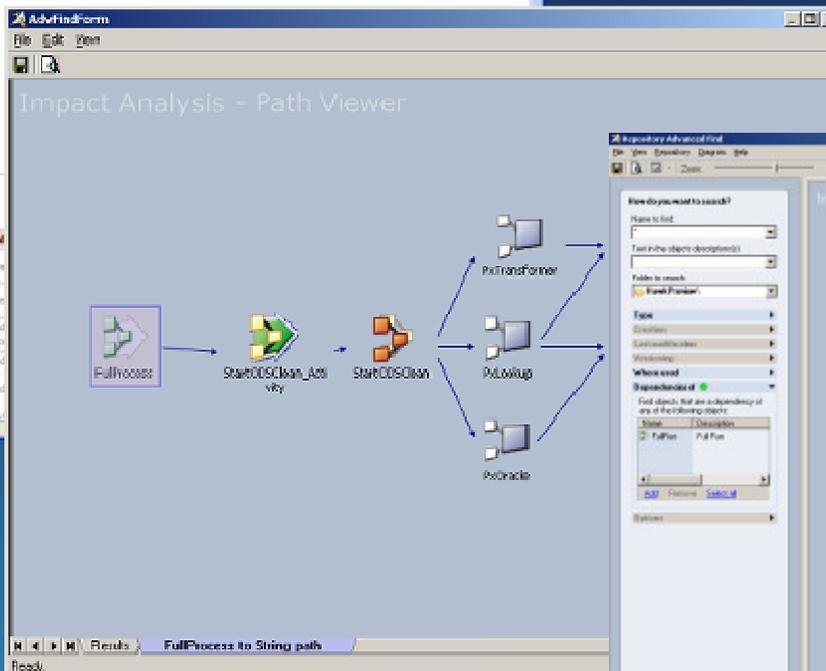
Hundreds of Built-in Transformation Functions

# Graphical Impact Analysis and Lineage Provide Trust

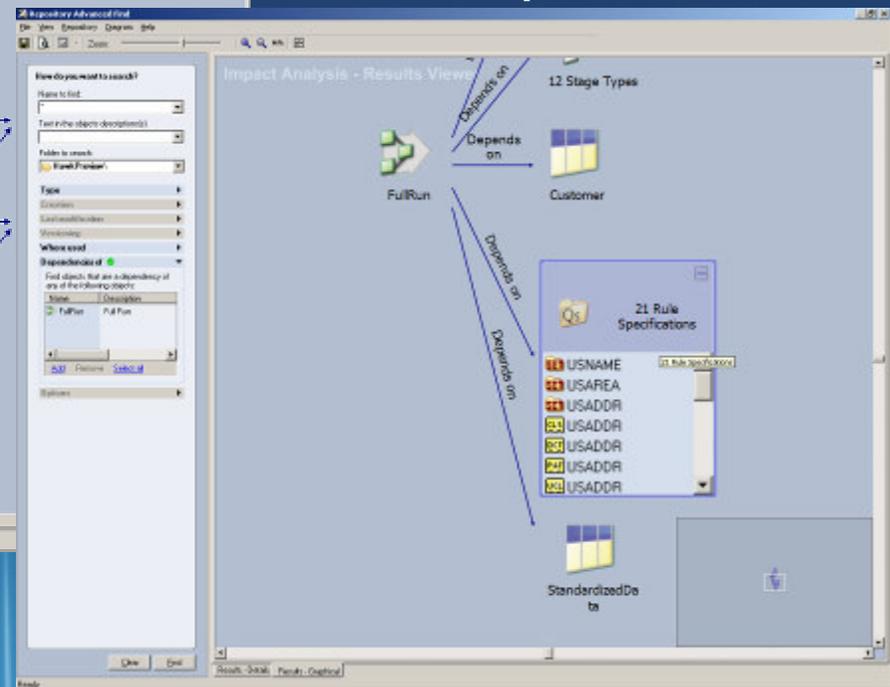
## HTML View

Name	Depend
Transformer	process_orders -> Job -> Co
ODBC	process_orders -> Job -> rdb
Sequential File	process_orders -> Job -> Out process_orders -> Job -> Wro datage0714 EXAMPL12 process_orders -> Job -> Out datage0714 EXAMPL12 process_orders -> Job -> Out

## Path View



## Graphical Tree View



# What Makes WebSphere DataStage Different?

Easy Design of Complex Data Processing	<u>Benefits</u>
Graphical, top-down design metaphor, with extensive library of pre-built functions & graphical sequencing	<b>Faster time to market, Low cost to develop skills, Lower maintenance costs</b>
Extensible, component-based architecture	<b>Lower risk, Better capitalizes on existing investments</b>
Strong reuse capabilities, including shared containers, routines, connection objects, and reusable services	<b>Better consistency, faster time to market, stronger project leverage</b>
Broad and deep connectivity, with bulk connectivity, changed data capture, and dynamic connectivity options	<b>Better utility, better project flexibility, faster time to market</b>
Rapid SOA deployment capability	<b>Better utility, broader applicability</b>
Massive Scalability	
Design serially, deploy in parallel	<b>Able to deal with any data volume without logic changes, Greater utility</b>
Metadata-driven Integration	
Unified metamodel across IBM Information Server	<b>Speeds project delivery, Improves collaboration, Produces better results</b>
Active metadata analysis, including diff, impact, and lineage	<b>Better productivity, reduced risk</b>

# The IBM Solution: IBM Information Server

*Delivering information you can trust*

## IBM 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



Synchronize, virtualize and move information for in-line delivery

## Platform Services

### Parallel Processing Services



### Connectivity Services



### Metadata Services



### Administration Services



### Deployment Services



# What is IBM WebSphere Federation Server?

- **Provided access to diverse & distributed information as if it were in one system**
  - Single SQL query access to diverse sources
  - Provides visual tools for defining federated queries
- **Includes industry-leading query optimization with single sign-on, unified views, and function compensation**
- **Supports transactional write capabilities across heterogeneous sources**
- **Enables bi-directional data access services to be published in a SOA**

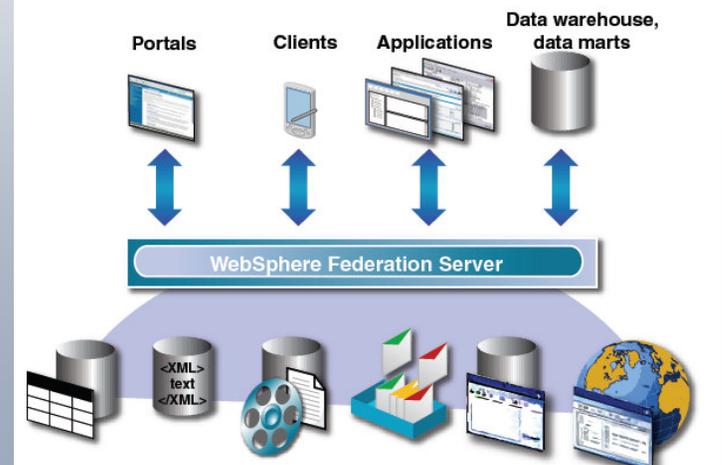
Deliver



## 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



# What Makes WebSphere Federation Server Different?

Industry-Leading Query Optimization	<u>Benefits</u>
Industry-leading Cost-based push-down optimization of queries	<b>Improves query performance, leverages latent processing power of DBMS's</b>
Data caching	<b>Improves query performance</b>
Function compensation ensures that even non-DBMS sources support standard SQL functions	<b>Improves query performance, Allows a more seamless SQL response</b>
Parallel processing capable	<b>Massive scalability</b>
Broadest Data Reach	
Support for many sources, including relational, mainframe, content, packaged applications, Web services, & XML	<b>Better utility, Seamless combination of all data across the enterprise</b>
Federated stored procedures	<b>Reuses existing logic, leverages skills</b>
Extensible connectivity via toolkit	<b>Enables it to work with any data source</b>
Read & Write Support	
Federated two-phase commit	<b>Better utility, ensures the integrity of information across heterogeneous sources</b>

# Federated Queries Make Integration as Easy as SQL

```
SELECT  parameters_return_billto_key as BILL_TO_KEY,  
        billto_company_name,  
        parameters_return_shipto_key as SHIP_TO_KEY,  
        CASES_SHIPPED,  
        GROSS_SALES,  
        URL  
FROM    GETKEYSSOAP_GETKEYSREALTIME_NN,  
  
        GLOBAL_SALES_TRAN_NN,  
  
        BILLTO_DIMENSION,  
  
        URL_INVOICES  
WHERE   getkeysrealtime_ship_to_number = '13546'  
        and getkeysrealtime_ship_to_number = URL_INVOICES.shipno  
        and ltrim(rtrim(translate(ship_to_number, ' ', x'0a')))  
            = getkeysrealtime_ship_to_number  
        and parameters_return_billto_key = billto_key  
        and ltrim(rtrim(translate(sales_order_number, ' ', x'0a')))  
            = URL_INVOICES.orderno;
```

## Single SQL Query Joins:

- ← Web Service
- ← XML Documents
- ← Data Warehouse
- ← Unstructured Data

# The IBM Solution: IBM Information Server

*Delivering information you can trust*

## IBM 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



Synchronize, virtualize and move information for in-line delivery

## Platform Services

### Parallel Processing Services



### Connectivity Services



### Metadata Services



### Administration Services



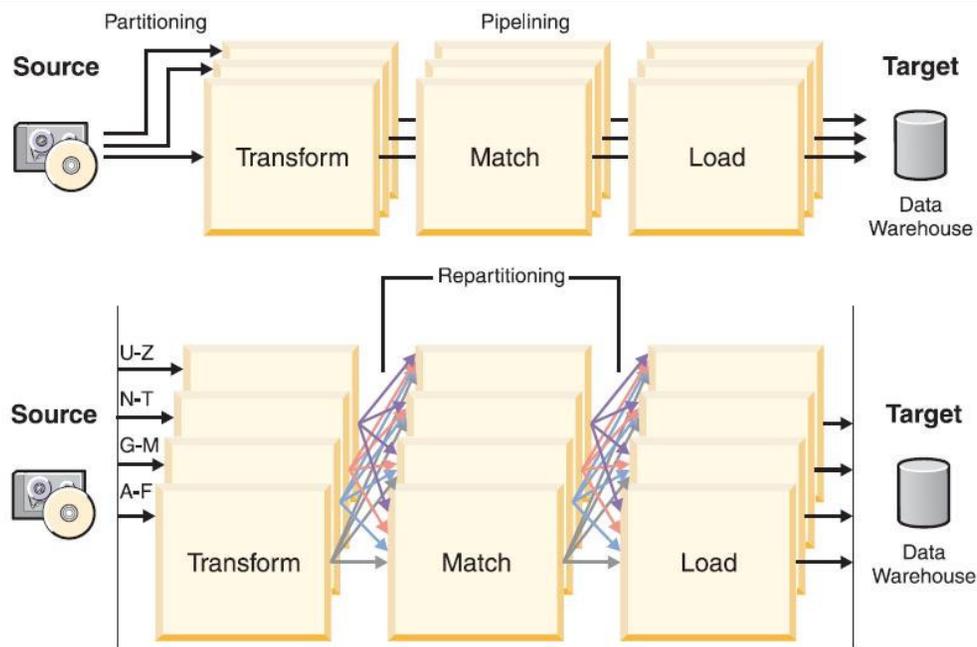
### Deployment Services



# What Makes Parallel Processing Services Different?

Industry-Leading Parallelism	<u>Benefits</u>
Pipelining support	<b>Ensures constant streaming of data</b>
Partitioning & repartitioning in memory	<b>Faster, Less resource utilization, Allows true leverage of bulk load utilities</b>
Design serially, deploy sequentially	<b>Allows scaling to be done easily, without impacting data flow design</b>
Extensibility of parallelism to third-party components	<b>Eliminates potential bottlenecks</b>
Grid support, leveraging grid schedulers to manage parallel configuration	<b>Provides for low-cost, massive scalability solutions</b>
Performance Control	
Performance Tuning	<b>Easy analysis and tuning of data flows for optimal performance</b>
Resource Estimation	<b>Makes it easy to estimate how new loads will perform and how to scale</b>
Common Parallel Framework	
Parallelism across IBM Information Server	<b>Ensures massive scalability across all operations</b>

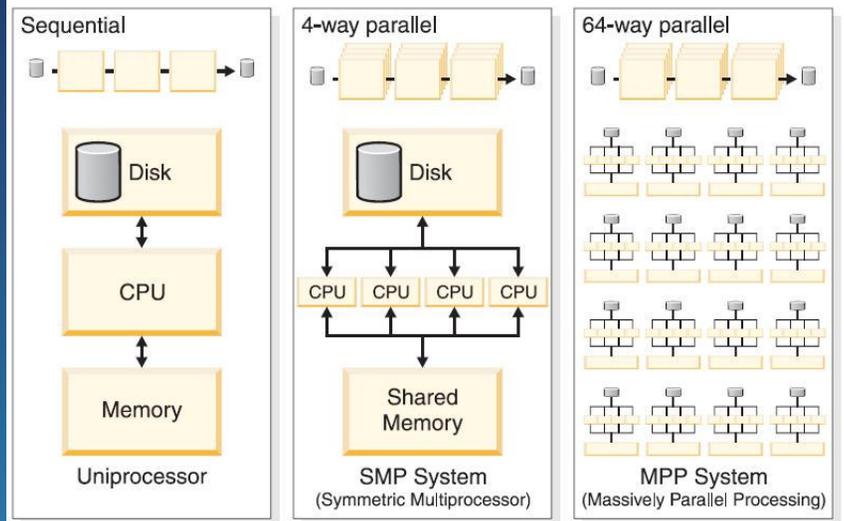
# Parallel Processing Services: Scalability, Performances, Portability



## Application Assembly: One dataflow graph



## Application Execution: Sequential or parallel



# The IBM Solution: IBM Information Server

*Delivering information you can trust*

## IBM 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



Synchronize, virtualize and move information for in-line delivery

## Platform Services

### Parallel Processing Services



### Connectivity Services



### Metadata Services



### Administration Services



### Deployment Services



# What Makes Connectivity Services Different?

Broadest Range of Connectors	<u>Benefits</u>
Bulk and direct connections to databases, files, and other data sources	<b>Enterprise coverage, Ability to handle large data volumes quickly</b>
Easy metadata-driven connection to enterprise applications	<b>Easy inclusion of enterprise data</b>
Changed data capture enables change awareness & streaming	<b>Allows response to data changes</b>
Mainframe data sources	<b>Ensures enterprise coverage</b>
Unstructured data sources through IICE	<b>Allows for new classes of solutions</b>
Advanced Connectivity	
New “rich” connectors	<b>Reduced development burden, faster performance, easier configuration</b>
Support for third-party code through Java, C or COBOL	<b>Extensibility, Reuse of existing assets</b>
Easier, consistent connectivity configuration interface	<b>Lowers development learning curve</b>
Shared Connectivity Across the Platform	
Connector reuse	<b>Speeds development</b>
Connectivity shared by DS & QS	<b>Lowers learning curve</b>

# Connectivity Services: Rich Connectivity

<b>General Access</b>	<b>Real Time</b>	<b>Enterprise Applications</b>	<b>Connect</b>	<b>MS Analysis</b>
Sequential File	WebSphere MQ	JD Edwards OneWorld (direct)	Alibase/SOL	Normal
Complex Flat File	SeeBeyond	Oracle Applications (Direct, Hierarchical)	Cache	Non StopSQL
File Set	Java Messaging Services (JMS)	PeopleSoft (Direct, Trees)	C-ISAM	Nucleus
Data Set	Java (Client & Transformer)	SAP BW (BAD, IDOC)	Datacom/DB	ODBC
Named Pipe	XML (Read / Write)	SAP R/3 (AEAP, EAPI, IDOC)	DB2 UDB	OLAP Services
FTP (standard, secured)	YSL-T / YSL-T Transformer	Siebel (EIM) Business	DB2/400	Oracle
Compressed / Encoded Data	Web Services (SOAP)		DBMS	Progress

**Rich Connectivity**

**Shared, easy to use connectivity infrastructure**

**Best-of-breed, metadata-driven connectivity to enterprise applications**

**High volume, parallel connectivity to databases and file systems**

**Event-driven, real-time, and batch connectivity**

**Frictionless Connectivity**

# The IBM Solution: IBM Information Server

*Delivering information you can trust*

## IBM 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



Synchronize, virtualize and move information for in-line delivery

## Platform Services

### Parallel Processing Services



### Connectivity Services



### Metadata Services



### Administration Services



### Deployment Services



# What Makes Metadata Services Different?

Active Metadata Across the Platform	<u>Benefits</u>
Unified Metamodel	<b>Facilitates leverage across product modules, Speeds development</b>
Metadata flow from product module to product module	<b>Increases collaboration, Creates better results, Speeds development</b>
Seamless in-tool metadata reporting and access	<b>Reduces Risk, Speeds Development</b>
Operation & Design metadata managed holistically	<b>Improves manageability, Reduces Cost</b>
Role-Based Collaboration	
Accommodates multiple role-bases perspectives	<b>Improves collaboration, Produces better results</b>
Allows for annotations on any object	<b>Reduces errors, Improves collaboration</b>
Enterprise-Class Metadata Management	
Over 20 metabrokers and bridges	<b>Allows for leverage of existing work</b>
Scalable metadata infrastructure	<b>Ensures the platform will meet enterprise requirements</b>

# The IBM Solution: IBM Information Server

*Delivering information you can trust*

## IBM 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



Synchronize, virtualize and move information for in-line delivery

## Platform Services

### Parallel Processing Services



### Connectivity Services



### Metadata Services



### Administration Services



### Deployment Services



# What Makes Administration Services Different?

Common Security Model	<u>Benefits</u>
LDAP and Active Directory integration	<b>Reduces risk, Leverages existing investments, Reduces administration burden</b>
Strong, granular security control	<b>Reduces risk, Improves control</b>
Unified Administration	
Unified logging	<b>Reduces debugging time, Reduces administration costs</b>
Unified user management	<b>Reduces risk, Reduces administration costs</b>
Powerful Reporting	
Common graphical reporting tool with options for on-screen, HTML, or PDF formats	<b>Provides stronger collaboration capabilities, Reduces training requirements</b>

# The IBM Solution: IBM Information Server

*Delivering information you can trust*

## IBM 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



Synchronize, virtualize and move information for in-line delivery

## Platform Services

### Parallel Processing Services



### Connectivity Services



### Metadata Services



### Administration Services



### Deployment Services



# What is WebSphere Information Services Director?

- Packages information integration logic as services that insulate developers from underlying sources
- Allows these services to be invoked as Enterprise Java Beans or Web services
- Provides load balancing & fault tolerance for requests across multiple Information Servers
- Provides foundation infrastructure for Information Services



Developers



Architects

## WebSphere Information Services Director

Flexibly deploy and manage reusable information services without hand coding

Overview Service 01

Bindings

Binding Settings

▼ EJB

Enable Binding

JNDI Name:

Package Name:

SOAP over HTTP

▼ SOAP over JMS

Enable Binding

Activation Spec: JNDI Name:  SOAP Style:  Priority:

Description:  Connect:

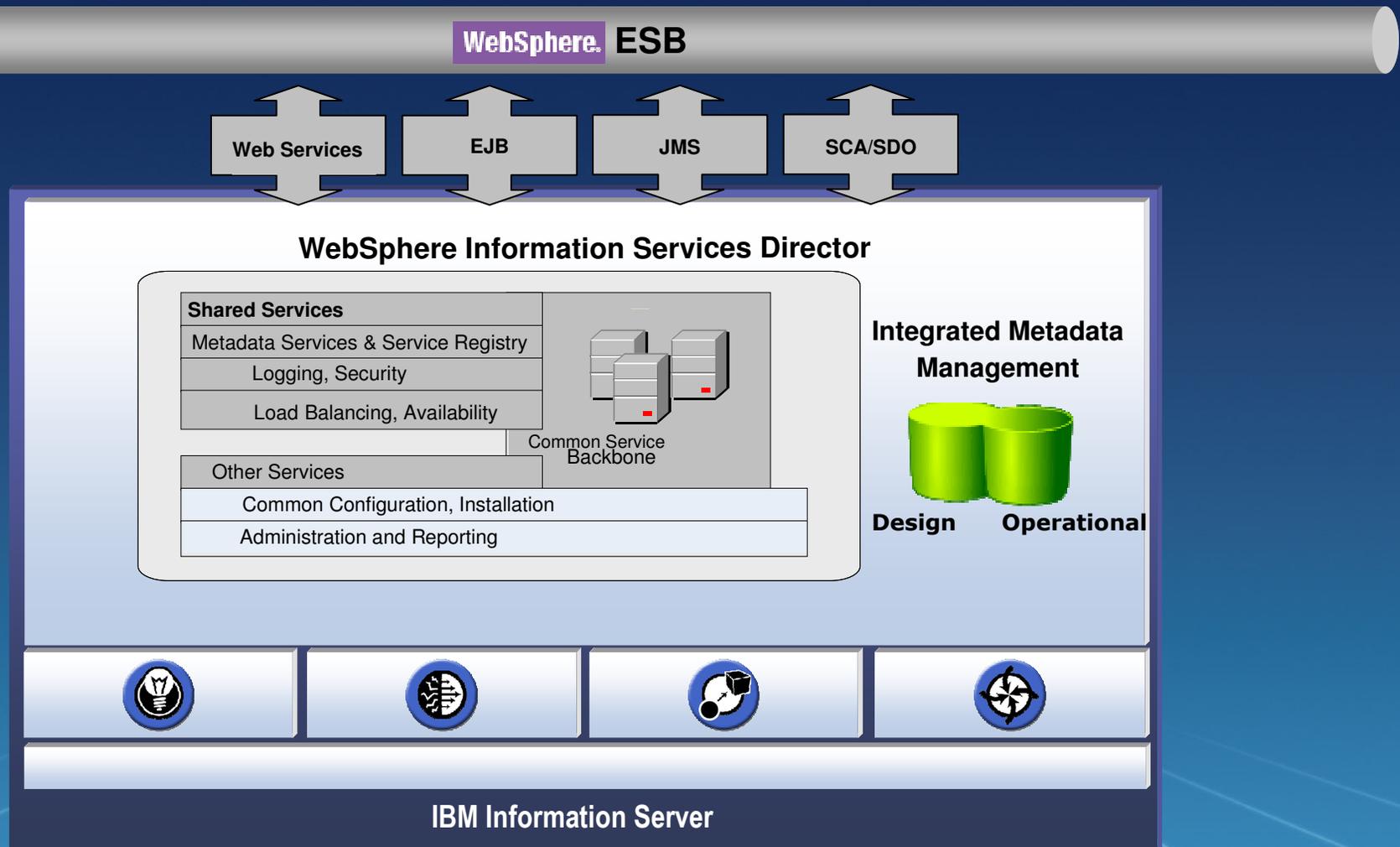
Destination:  Message:

**Rapid SOA Deployment**

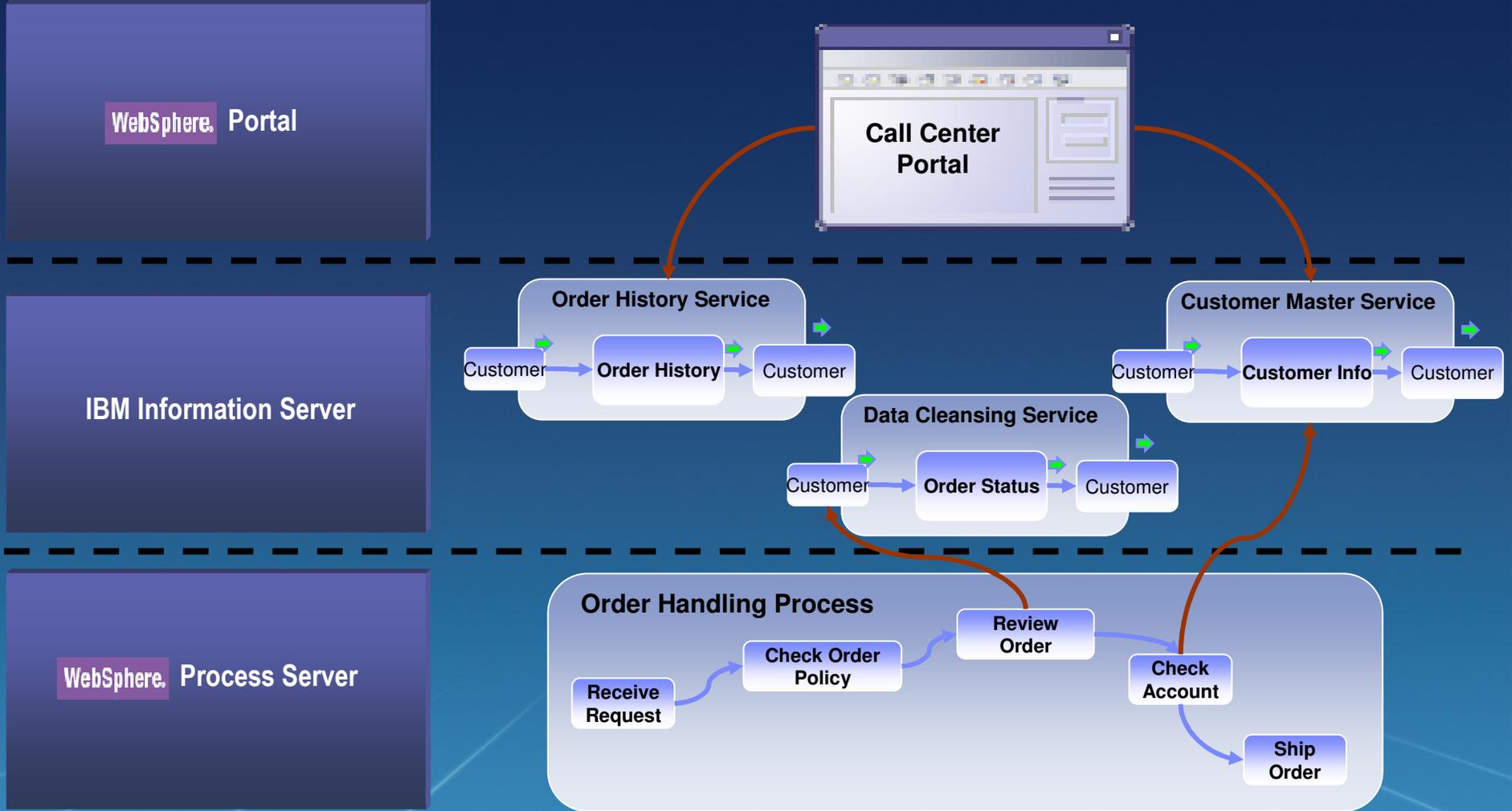
# What Makes WISD Different?

Unified Information Service Deployment	<u>Benefits</u>
Common mechanism for publishing and managing services across DataStage, QualityStage, and Federation	<b>Higher utility, Lower learning curve, Lower administration and debugging costs</b>
Supports direct publishing of services for DB2 databases	<b>Higher utility</b>
Supports Classic Federation & IICE through Federation Server	<b>Broader reach – services for any kind of data</b>
Service Metadata	
Provides active service metadata management across design and operational metadata	<b>Better control, Reduced administration costs</b>
Reports show impact analysis and lineage all the way out to services	<b>Better control, Reduced project risk, Higher service reuse potential</b>
Easy, Flexible Service Deployment	
Easy, quick service deployment	<b>Faster development, Lower learning curve, Leverages existing skills</b>
Supports Web services and EJB bindings with a single service definition, single point of maintenance	<b>Higher reuse potential, Higher utility</b>
Flexibility in defining service interface, including support for arrays and field defaults and overrides	<b>Lower development costs, Higher reuse potential, Better developer adoption</b>

# Common Programming Model



# Actionable Information Services

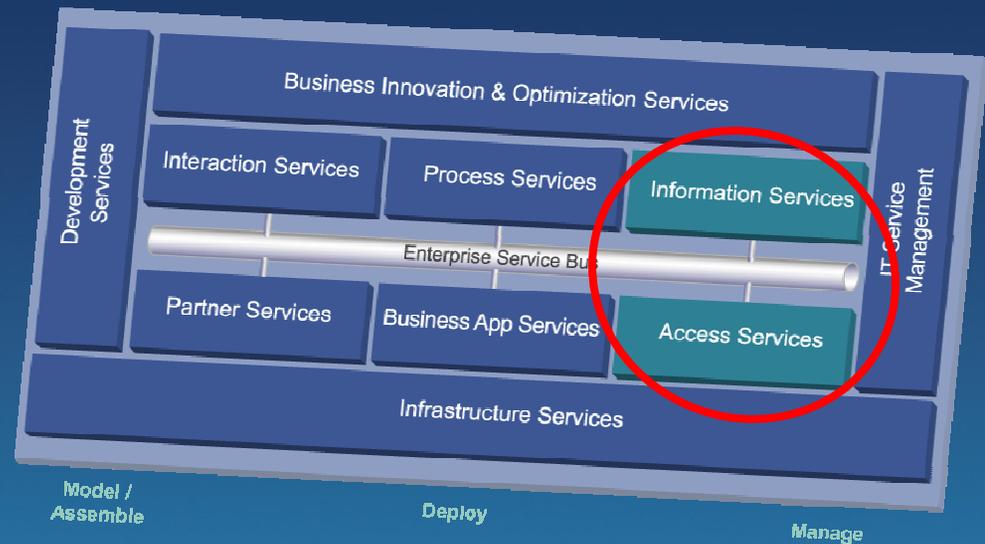
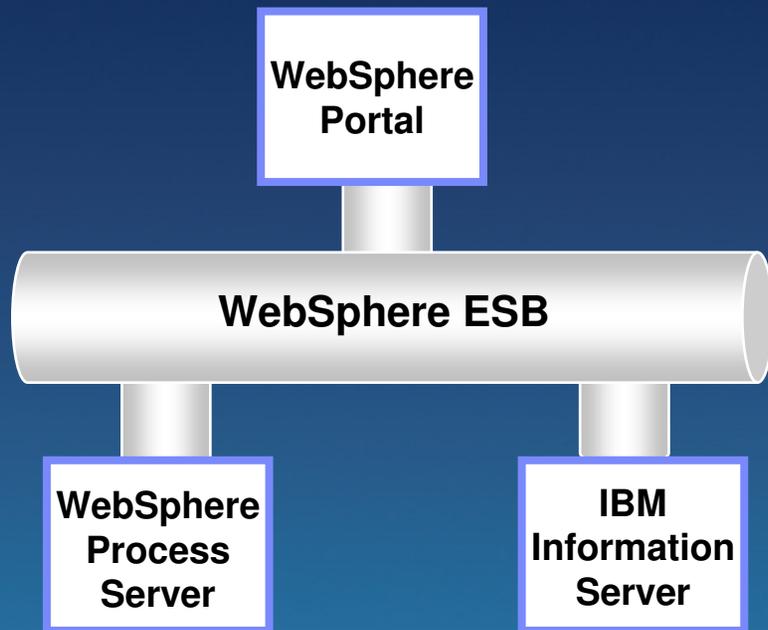


# Service Oriented Architecture

*Information as a Service is Key*

*Getting the right data quickly and consistently for all applications continues to be a key challenge for many enterprises.*

**Forrester, January 2006**

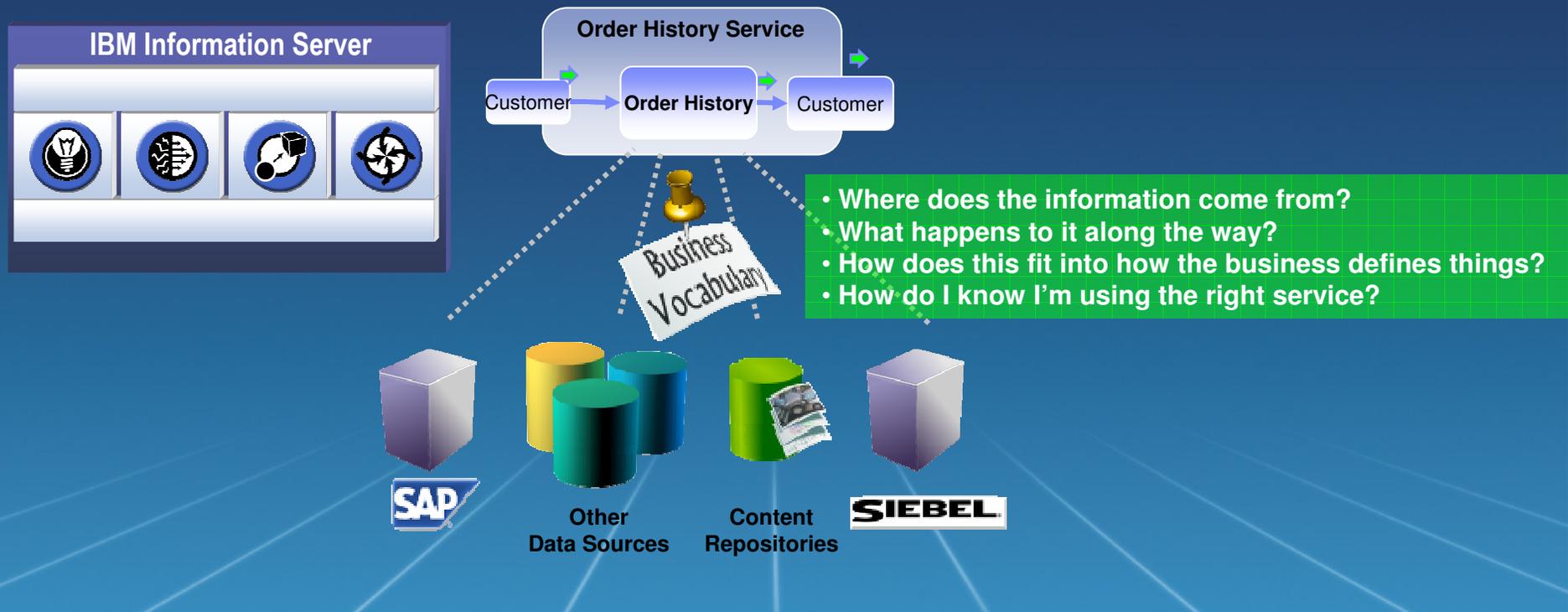


*You will waste your investment in SOA unless you have enterprise information that SOA can exploit.*

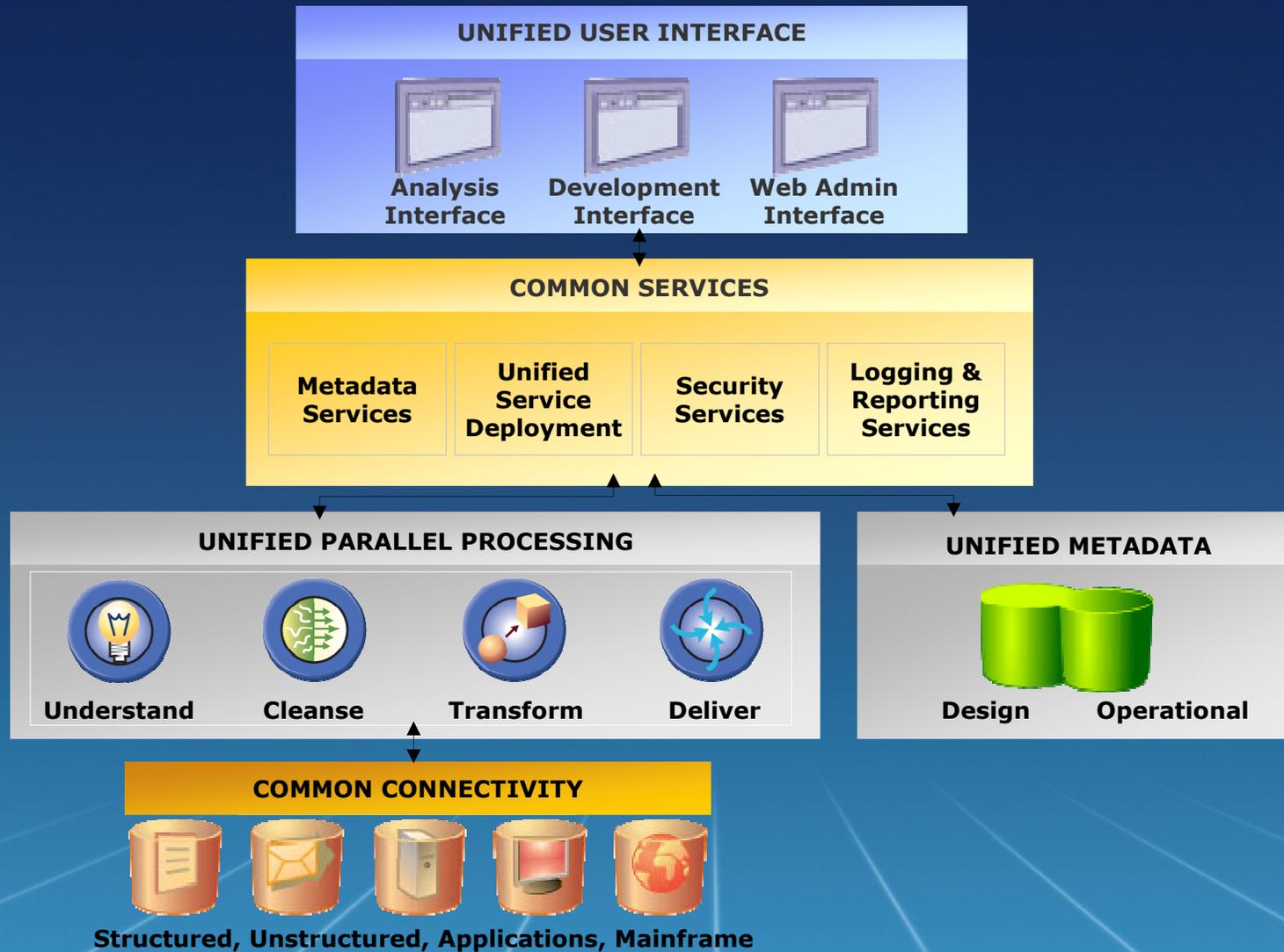
**Gartner, March 2005**

# Actionable Information Services

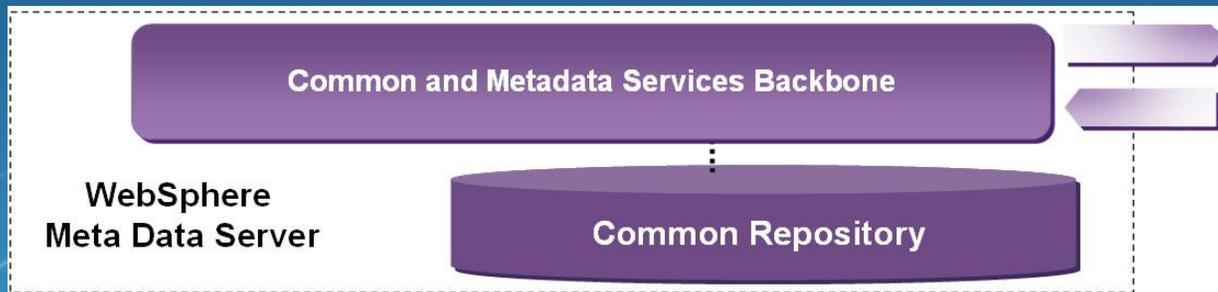
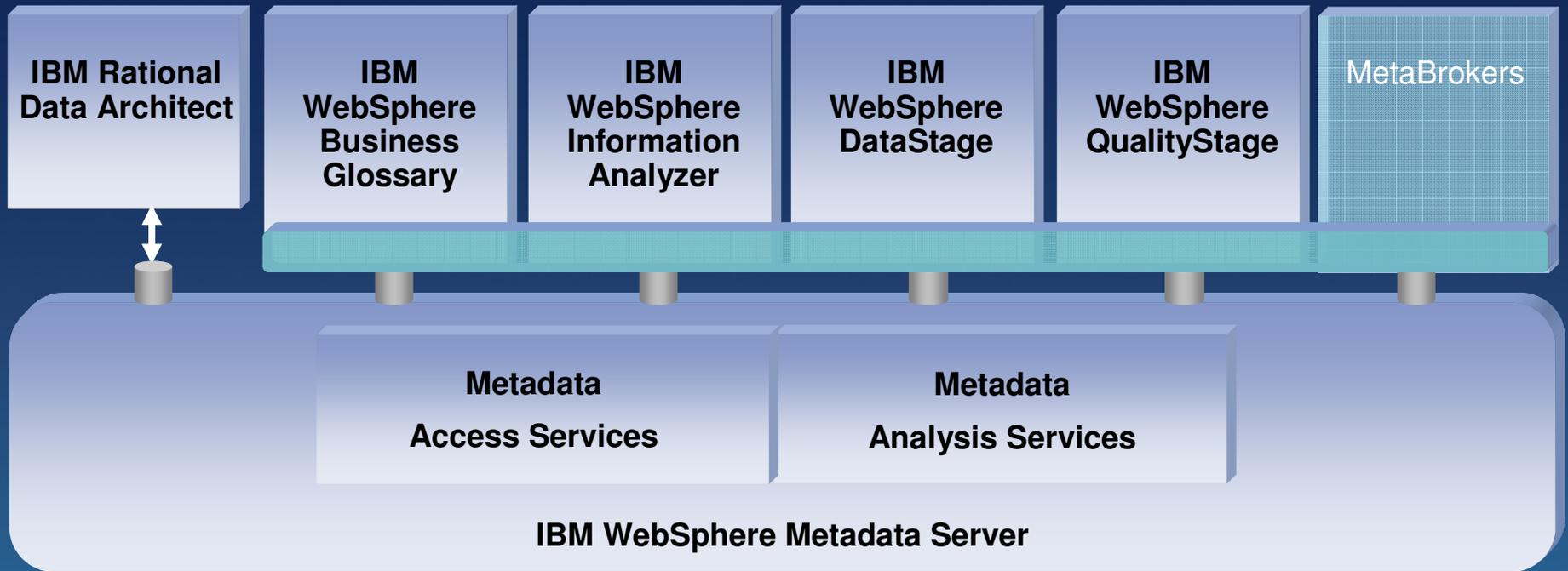
Information Services provide a basis for trust in information – providing visibility into lineage, relationships to other systems, and business definition



# IBM Information Server Architecture



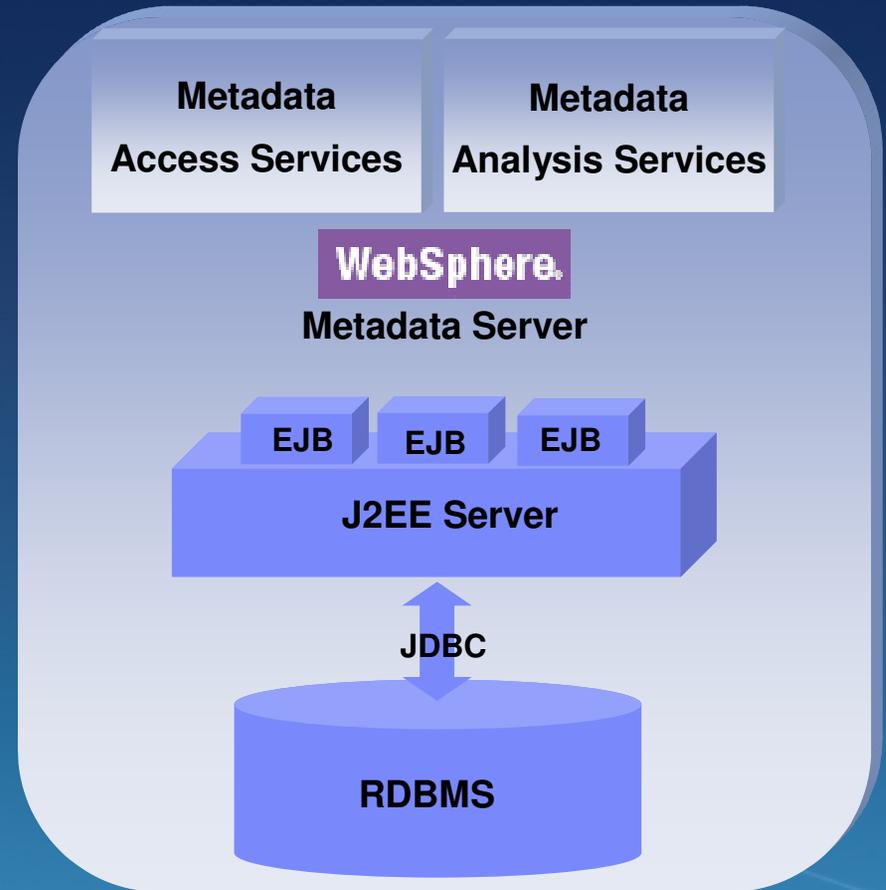
# IBM WebSphere Metadata Server – at the Core of IBM Information Server



# Shared Metadata Architectures

## Shared Metadata Infrastructure

- Common metadata services
  - Find & visualize information
  - Automate connection to information
  - Cross-tool lineage and impact analysis
  - Understanding of information relationships
- Extensible EMF-based model
- Common semantics and metadata understanding



# The IBM Solution: IBM Information Server

*Delivering information you can trust*

## IBM 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



Synchronize, virtualize and move information for in-line delivery

## Platform Services

### Parallel Processing Services



### Connectivity Services



### Metadata Services



### Administration Services



### Deployment Services



# 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

Thank  
you

- **Data Replication**
- **Data Event Publishing**

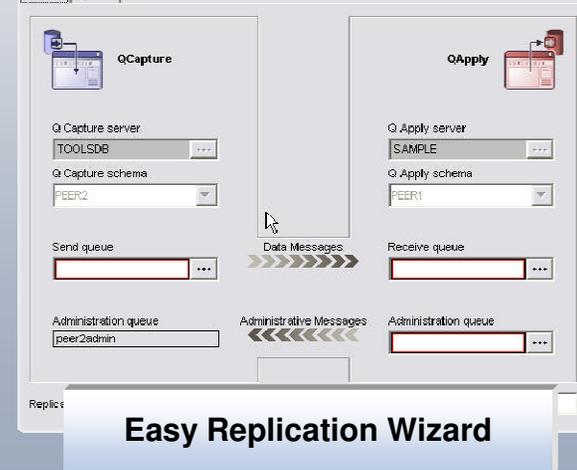
# Data Synchronization: IBM WebSphere Replication Server

- **Distribution, consolidation or synchronization of information in different databases**
- **Multi-directional delivery**
- **Ease-of-use features:**
  - Integrated monitoring and statistics
  - Changed data histories
  - Configuration options:
    - Wizard-driven GUI
    - Command-line processor
    - Script-driven processor

Deliver



**IBM WebSphere Replication Server**  
Replicate and synchronize between databases in high-speed based on data events for high availability & disaster recovery, data synchronization, and data distribution

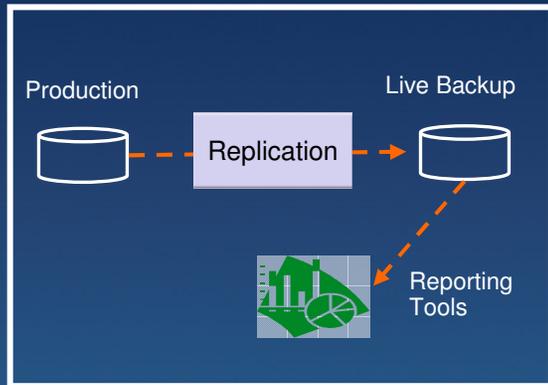


# Data Synchronization: IBM WebSphere Replication Server

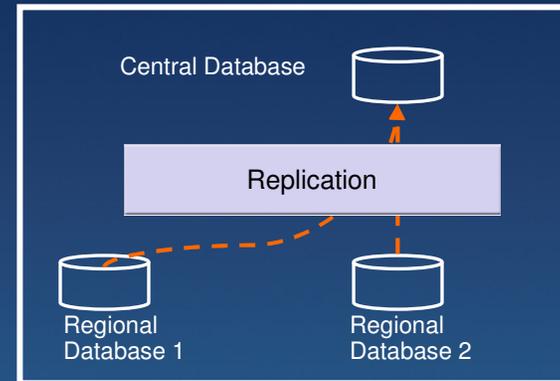
- **Replication can automatically keep multiple data locations consistent, and each target can be different to match the users needs. This includes different latency or differing timeliness of the data.**
  - replication can be by time interval, event driven or continuous
  - different enhancements (derivations, summarization, transformations)
  - different formats to each target
  
- **Data Replication**
  - High availability of production applications
  - Distribution of data to other locations
  - Consolidation of data from other locations (Data Warehouse and ODS applications);
    - Data Replication as part of the ETL process
  - Bidirectional exchange of data with other locations
  - Some variation or combination of the above

# Many Usage Scenarios For Replication

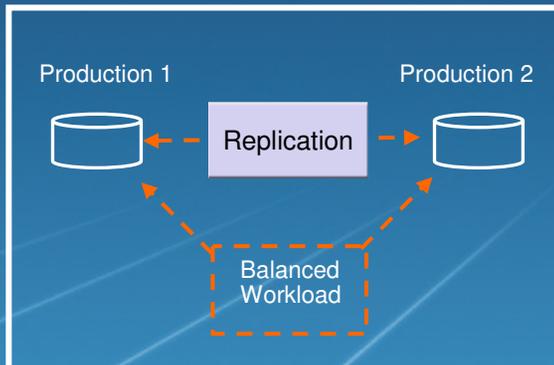
## High Availability



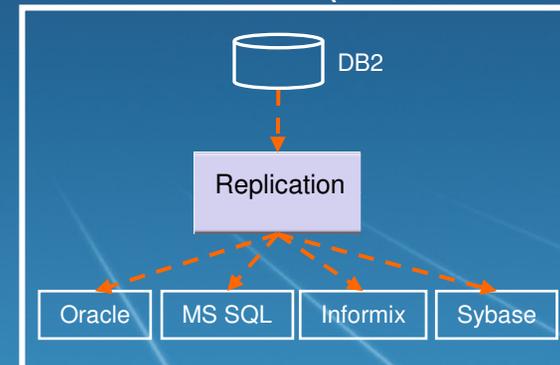
## Rollup (many to 1)



## Peer To Peer



## Distribution (1 to many)



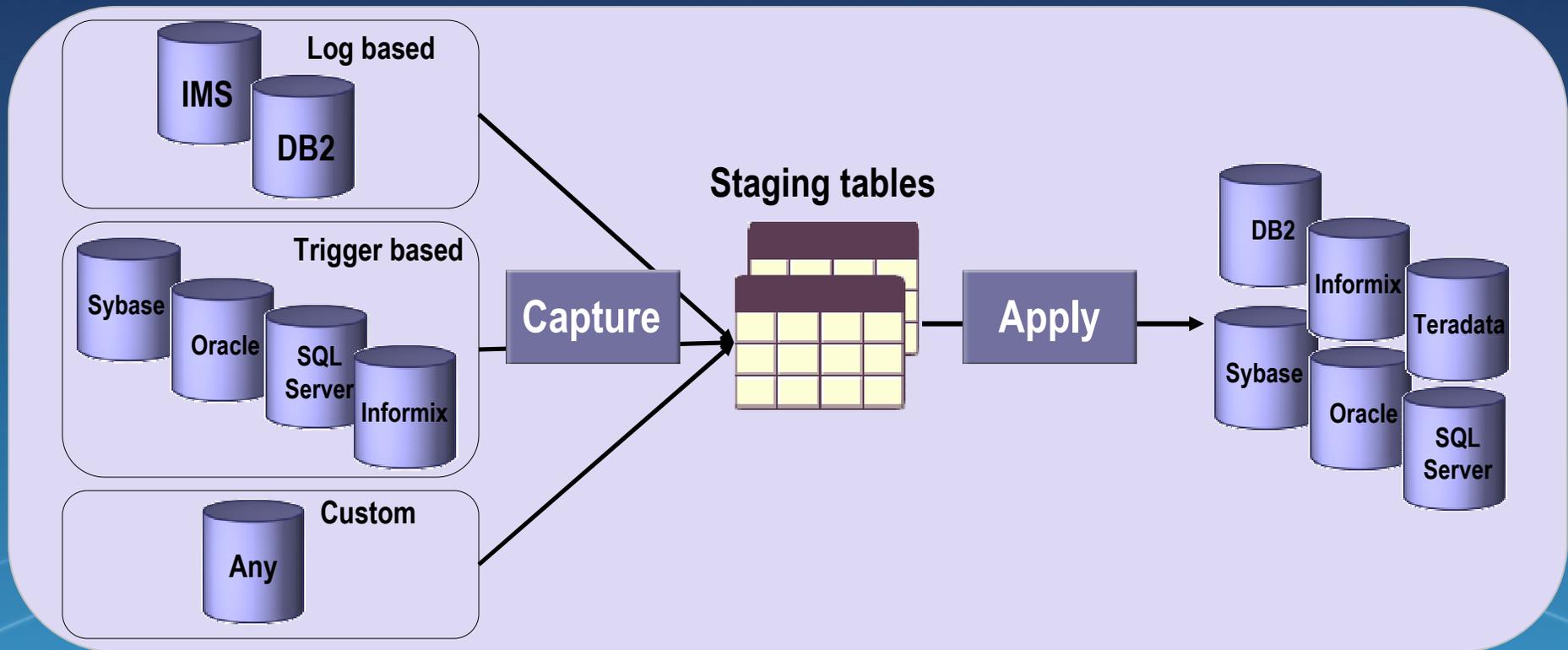
# WebSphere Replication Server: SQL Replication

## Function

- Filter and transform, Apply by table or by transaction
- Choose latency by schedule, interval, event, or continuous
- Replicate point-to-point, for distribution, or for consolidation
- Maintain snapshots, simple copies, histories, or aggregates

## Usage

- Business intelligence
- Distribution and consolidation
- Application integration



# WebSphere Replication Server: Q-Replication

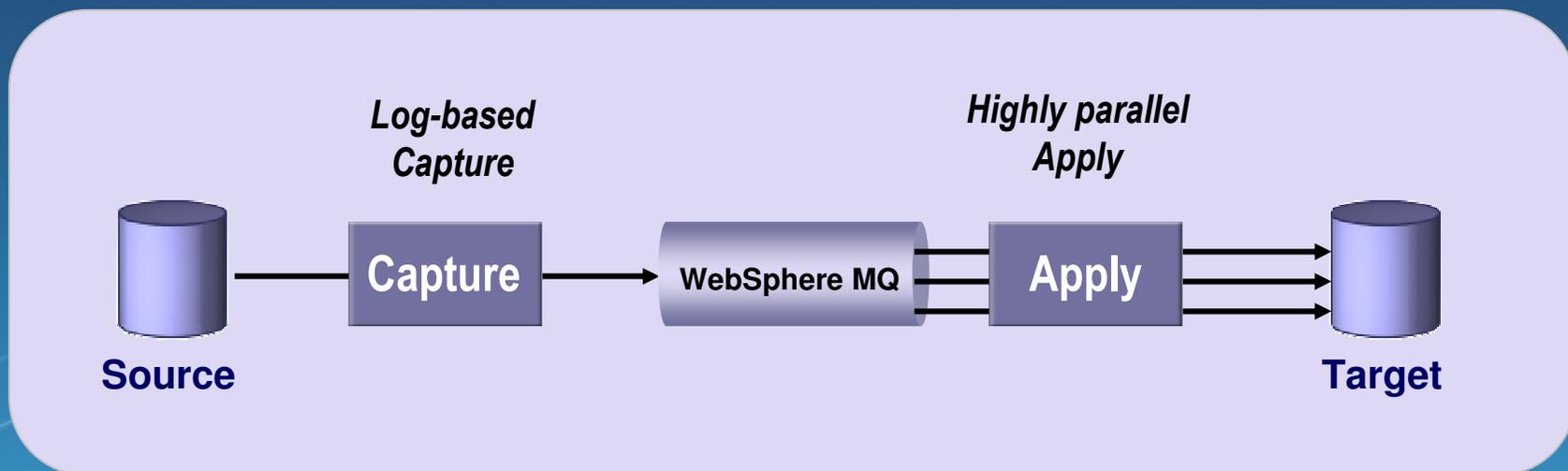
*New replication architecture for delivering extremely low latency replication for peer-to-peer environments*

## Function

- Replicate rows or transactions
- Filter and transform data
- Detect and resolve conflict
- Configure and monitor environment

## Usage

- High availability
- Workload distribution
- Application integration



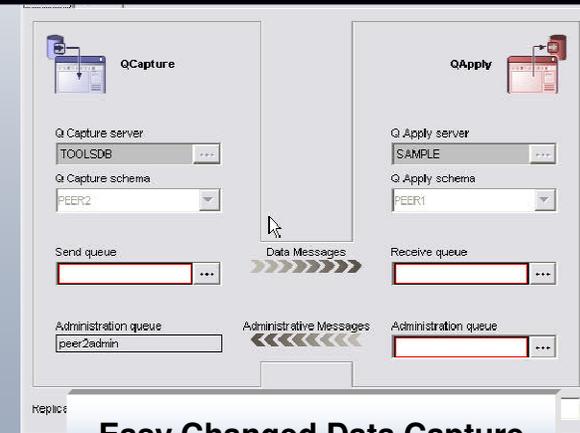
# Changed-data Publishing: IBM WebSphere Data Event Publishing

- Integration using data events rather than application development
- Mainframe data access without changes to applications
- Features for flexibility and efficiency:
  - Low-latency or scheduled data capture
  - Publication in a consistent, relational structure
  - Open XML delivery format
  - Recoverability
  - Assured delivery

Deliver



**IBM WebSphere Data Event Publisher**  
Detect and respond to data changes in source systems, and publish changes to subscribed systems, or feed changed data into other modules for event-based processing



# WebSphere Data Event Publishing

## ■ What is *Event Publishing* ?

- Capture changed-data from DB2, IMS and CICS/VSAM
- Correlate by transactions within single database
- Extract to consistent and documented XML format
- “Publish” to WebSphere MQ queue
- Received & Processed by any MQ “listener”

## ■ Why *Publish* data?

- Application to Application Messaging:
  - Stream changed data information to Web interface
  - Stream only particular events of interest (filter data)
- Event Notification
  - Stream changed data information to Web interface
  - Stream only particular events of interest (filter data)
- Warehouse / Business Intelligence
  - Integrate captured changed data with an ETL tool
  - Perform very complex transformations
- MQ provides guaranteed delivery
  - Works even when the target is not available



# WebSphere Data Event Publishing

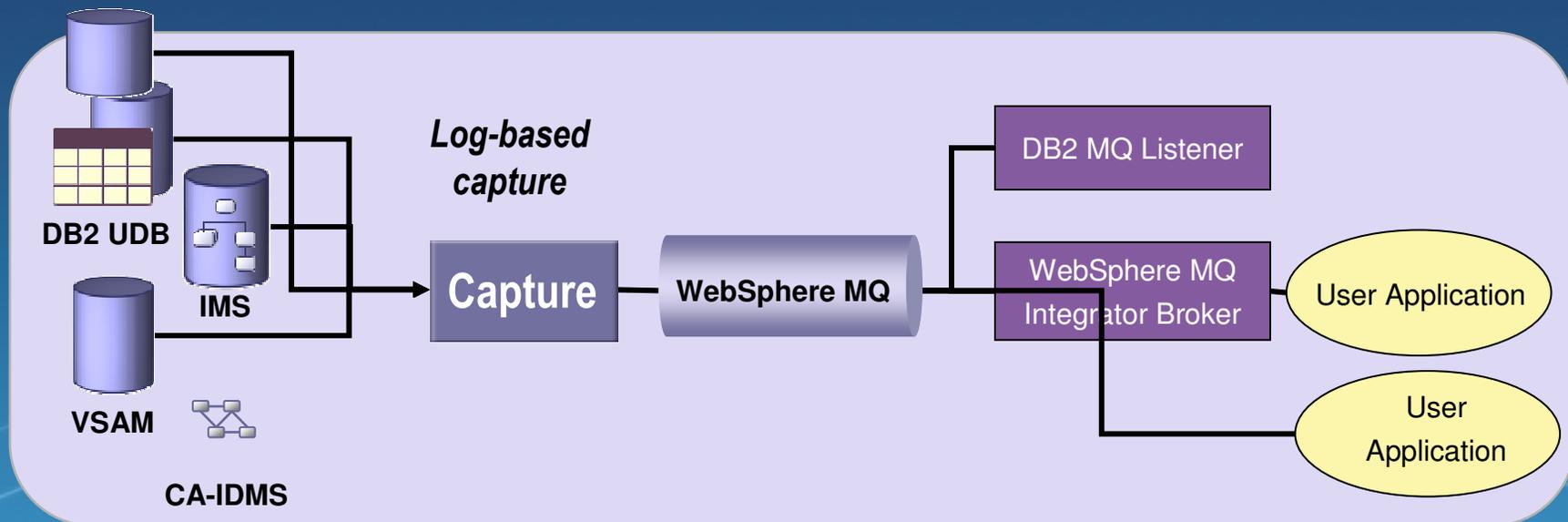
*Capture database changes and publish them as XML messages to WebSphere MQ*

## Function

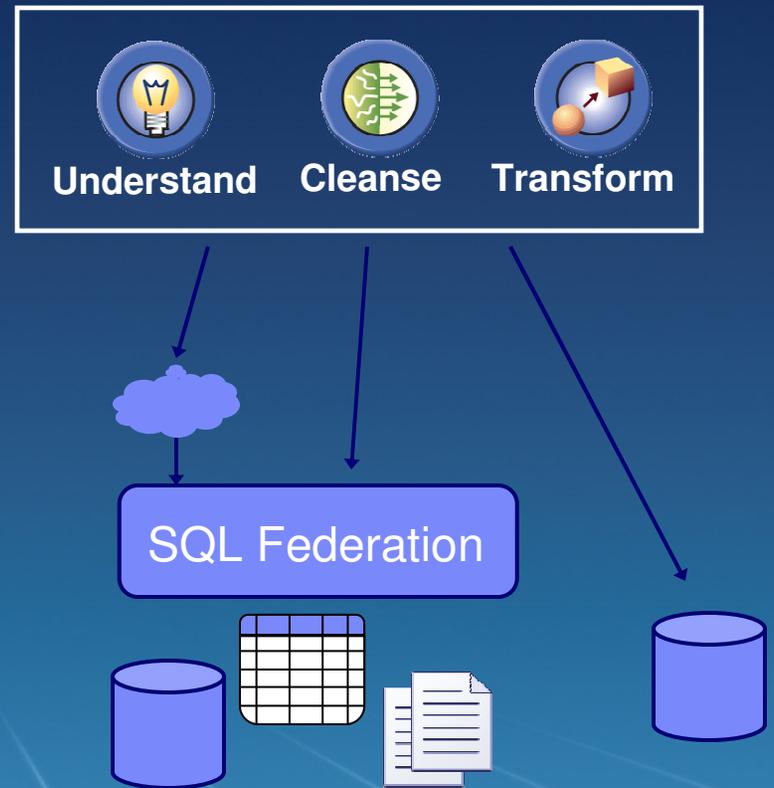
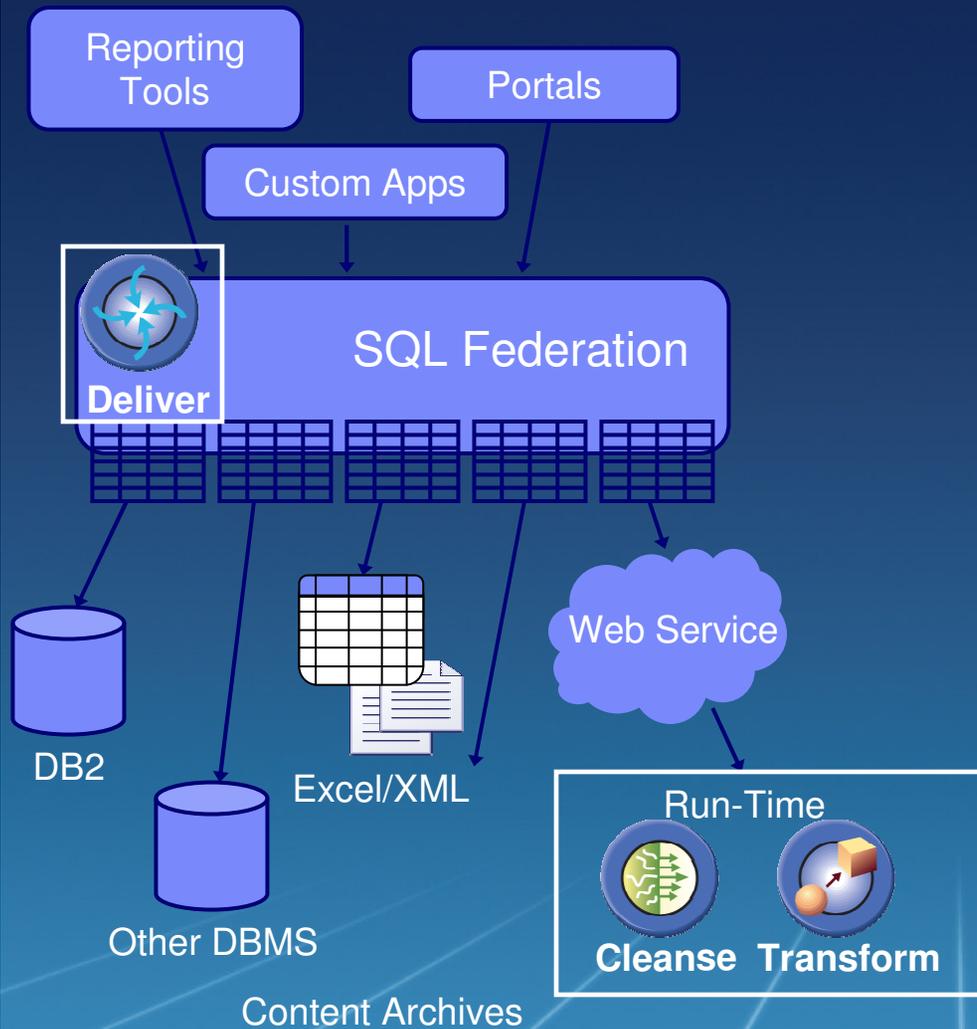
- Publish events to a message queue
- XML self-describing format
- Wizard-driven configuration

## Usage

- Application to application messaging
- Initiate business processes
- Source for ETL tool



# Federation, Analysis, Cleansing and Transformation: can work together !!!



# Combine Event-Driven Processing & Transformation

*Reduce latency for tactical decision making*

