UNIVERSITÉ DU
MAINFRAME

WebSphere
Information Integration

Eric DERBANNE

4 mai 2006

ALLONZ-Y !

Agenda

- **Introduction to Enterprise Information Integration**
  - ▸ Big Picture
  - ▸ IBM Platform

- **Understand, Cleanse and Transform Enterprise Data**

- **Federate Enterprise Data**

- **Connect and Deliver Enterprise Data**
  - ▸ SQL and Q Replication
  - ▸ Event Publishing

- **Combining strategies**

- **Models of Data Integration**
  - ▸ Real time - pull and push
  - ▸ Near real time - replication
  - ▸ Scheduled - ETL and consolidation

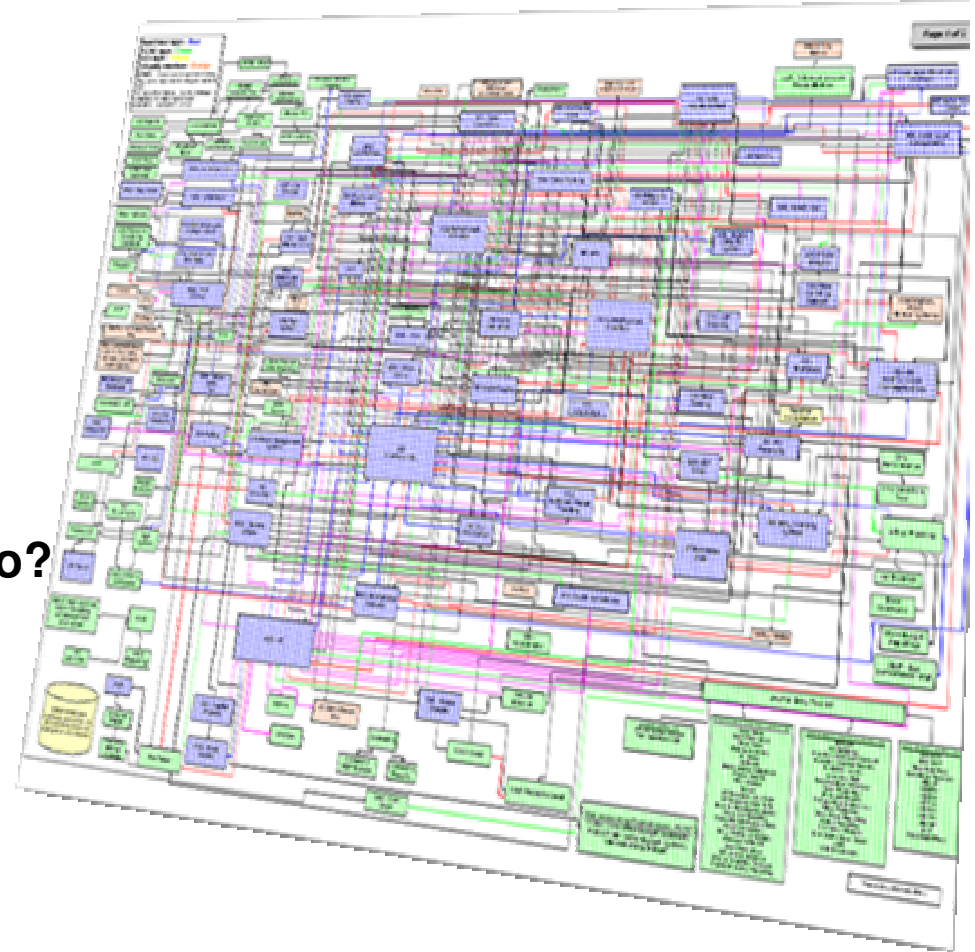# Businesses Expect Information Availability

The most anticipated impact of business investing in IT has become information availability

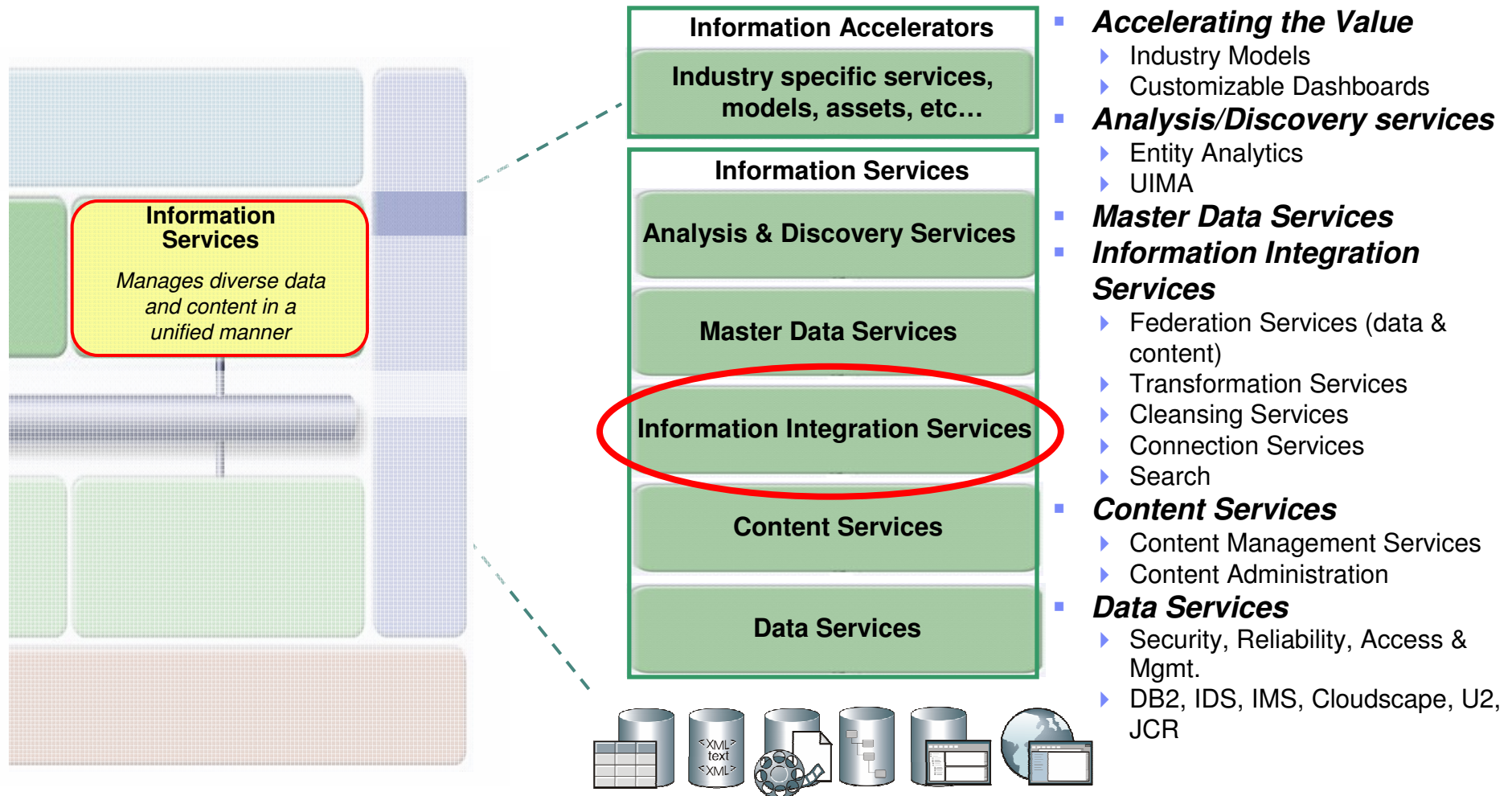| Top 3 Business Challenges	Top Capabilities (by spending)	Anticipated Business Impact
		Most Selected
• Streamline/improve efficiency of business processes	• Introduce new apps • Standardize, automate and integrate business processes • Align systems to business goals	• **Information Availability**
• Better understand and meet customer expectations	• Analyze and use information to make better business decisions	• **Information Availability**
• Increase employee productivity	• Enhance employees' skills • Empower employees with tools and info for decision-making	• **Information Availability** • Operational Stability • Productivity

Source: IBM Attributes & Capabilities Study, 2005

- **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?**

SOA Reference Architecture: Information Services
The Heart of Information On Demand

IBM

UNIVERSITÉ DU MAINFRAME

Information Services

Manages diverse data and content in a unified manner

Information Accelerators
Industry specific services, models, assets, etc…

Information Services
- Analysis & Discovery Services
- Master Data Services
- Information Integration Services
- Content Services
- Data Services

- **Accelerating the Value**
  - Industry Models
  - Customizable Dashboards
- **Analysis/Discovery services**
  - Entity Analytics
  - UIMA
- **Master Data Services**
- **Information Integration Services**
  - Federation Services (data & content)
  - Transformation Services
  - Cleansing Services
  - Connection Services
  - Search
- **Content Services**
  - Content Management Services
  - Content Administration
- **Data Services**
  - Security, Reliability, Access & Mgmt.
  - DB2, IDS, IMS, Cloudscape, U2, JCR

Service-Oriented Architecture

Understand

WebSphere. Information Analyzer

WebSphere. ProfileStage

Rational. Data Architect

Discover, define, model, and govern information quality and structure

Cleanse

WebSphere. QualityStage ™

Standardize, Merge, & Correct information

Transform

WebSphere. DataStage ®

WebSphere. DataStage ® TX

Transform & Enrich information

Federate

WebSphere. Information Integration Standard Edition

WebSphere. Information Integration Classic Federation

WebSphere. Information Integration Content Edition

Virtualize Access to disparate information

Integrated Metadata Management

Parallel Processing

Data ⟵———————— **Connect & Deliver** ————————⟶ Content

WebSphere. Information Integration Replication for z/OS

**WebSphere.** Information Integration Replication Edition

WebSphere. Information Integration Classic Event Publisher for z/OS

WebSphere. Information Integration Event Publisher Edition

Access, Publish and Replicate information

Agenda

- **Introduction to Enterprise Information Integration**
  - ▸ Big Picture
  - ▸ IBM Platform

- **Understand, Cleanse and Transform Enterprise Data**

- **Federate Enterprise Data**

- **Connect and Deliver Enterprise Data**
  - ▸ SQL and Q Replication
  - ▸ Event Publishing

- **Combining strategies**

- **Models of Data Integration**
  - ▸ Real time - pull and push
  - ▸ Near real time - replication
  - ▸ Scheduled - ETL and consolidation

UNIVERSITÉ DU MAINFRAME
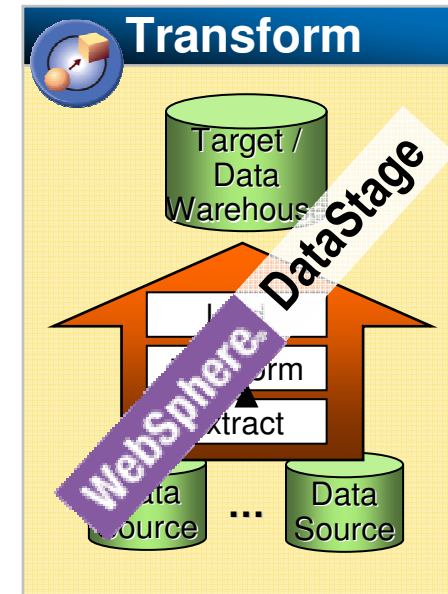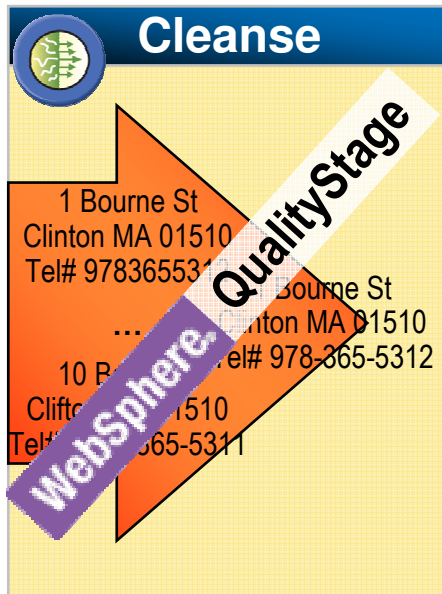

Model

Analyze
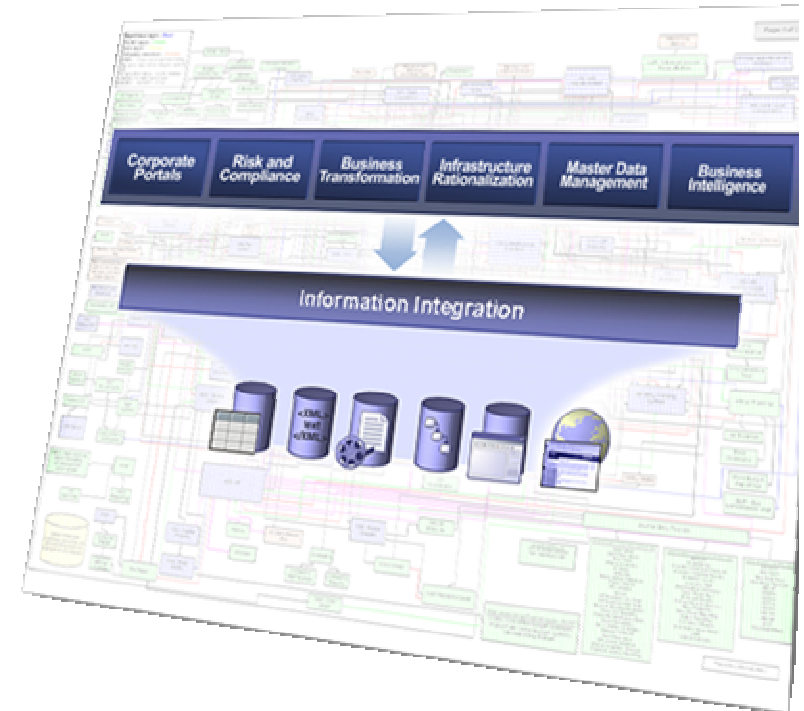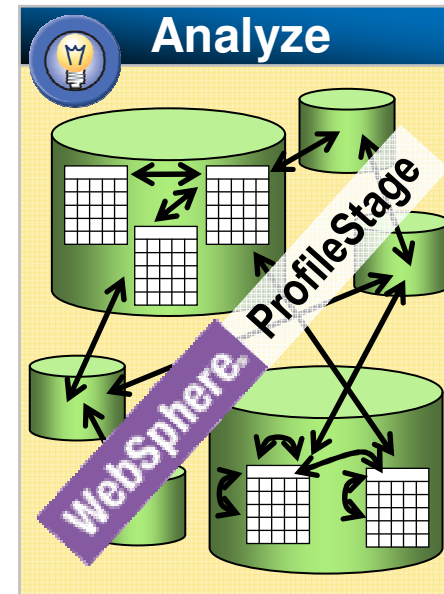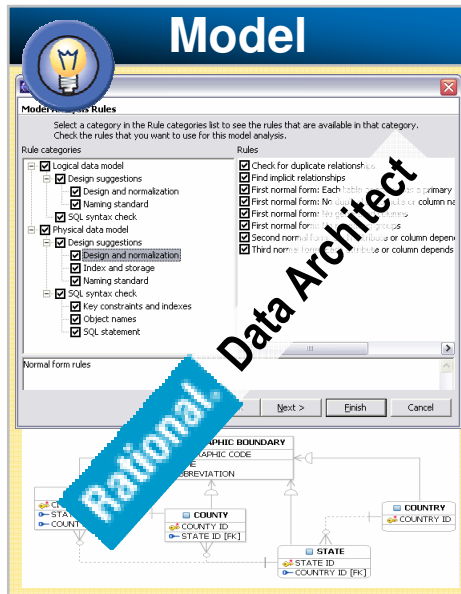
Legacy databases often not well understood…
➢ decades of application growth and adaptation to business needs results in limited knowledge

Understanding data & relationships is first step in leveraging legacy data assets
➢ How can you migrate to DB2 without understanding your VSAM file or IDMS database?
➢ How can you reuse your IMS databases without knowing what's in them?

UNIVERSITÉ DU MAINFRAME

Cleanse

QualityStage

WebSphere.

1 Bourne St
Clinton MA 01510
Tel# 97836553...

...Bourne St
...nton MA 01510
...Tel# 978-365-5312

10 B...
Clifto... 1510
Tel... 365-5311

Transform

DataStage

WebSphere.

Target / Data Warehous...

...orm
...xtract

...ata ...ource ... Data Source

Legacy data quality is often suspect…

➢ Decades of application growth, acquisitions, varying application quality, changing business rules, personnel changes impact quality

Legacy operational data is frequently mission-critical…

➢ Consistent business entity content is essential to effective, efficient and competitive business operations

Empower "BI on z"…

➢ Source mainframe data (Classic Fed) or pull from distributed data sources (WebSphere II) for DB2 on z data warehouses, data marts
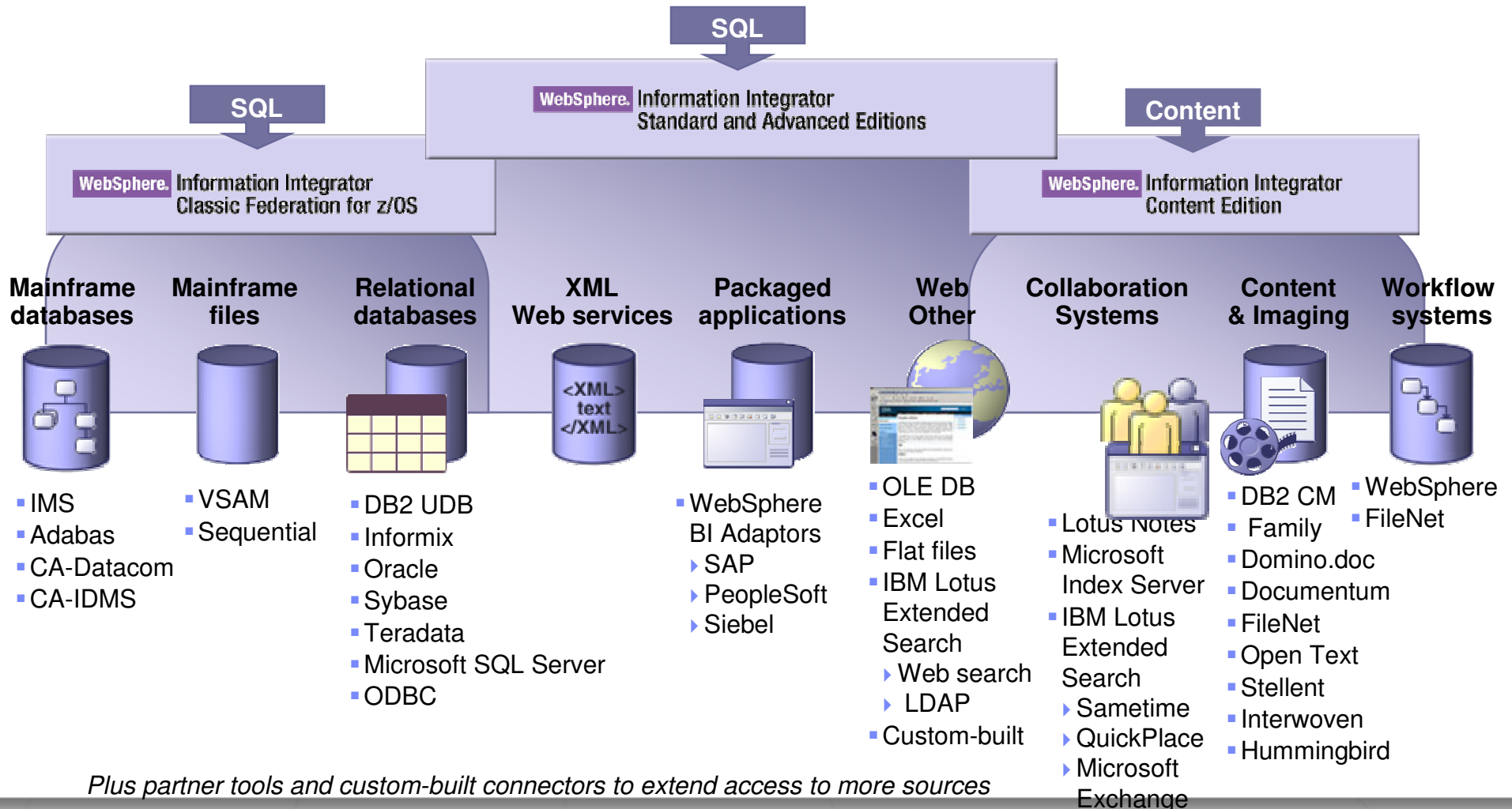
Enable master data management (MDM) …

➢ Critical information resides on "z"

➢ Rapid delivery of many-to-many transformations and mappings such as those associated with EDI, HIPAA and SWIFT implementations.

Agenda

- **Introduction to Enterprise Information Integration**
  - ▸ Big Picture
  - ▸ IBM Platform

- **Understand, Cleanse and Transform Enterprise Data**

- **Federate Enterprise Data**

- **Connect and Deliver Enterprise Data**
  - ▸ SQL and Q Replication
  - ▸ Event Publishing

- **Combining strategies**

- **Models of Data Integration**
  - ▸ Real time - pull and push
  - ▸ Near real time - replication
  - ▸ Scheduled - ETL and consolidation
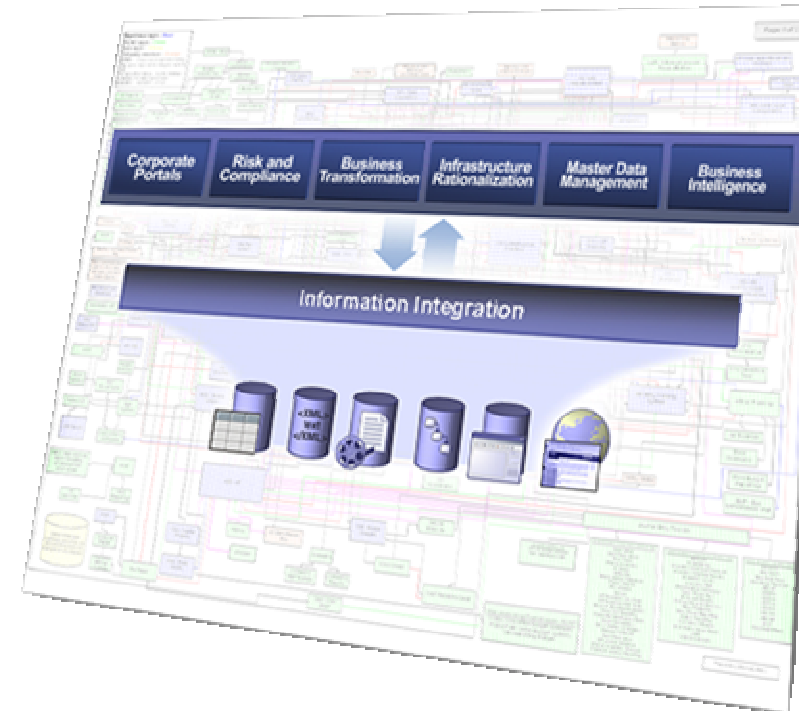
Access diverse and distributed information as if it were in one system
Single sign on, Unified views, Common language, Web services or Java API Query and update, Optimized access
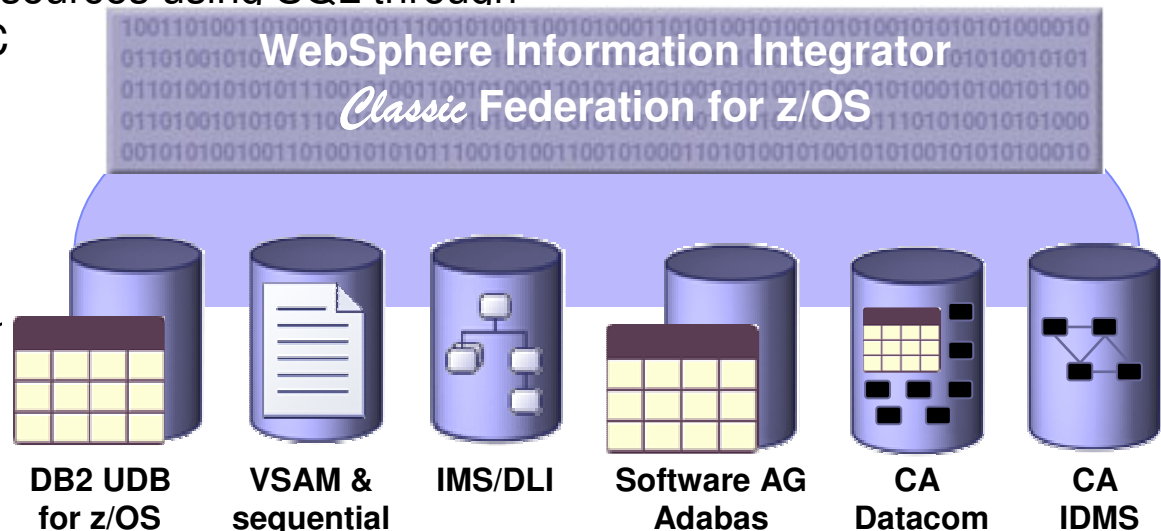
SQL

WebSphere. Information Integrator Standard and Advanced Editions

SQL

Content

WebSphere. Information Integrator Classic Federation for z/OS

WebSphere. Information Integrator Content Edition

Mainframe databases	Mainframe files	Relational databases	XML Web services	Packaged applications	Web Other	Collaboration Systems	Content & Imaging	Workflow systems

Mainframe databases
- IMS
- Adabas
- CA-Datacom
- CA-IDMS

Mainframe files
- VSAM
- Sequential

Relational databases
- DB2 UDB
- Informix
- Oracle
- Sybase
- Teradata
- Microsoft SQL Server
- ODBC

XML Web services
`<XML> text </XML>`

Packaged applications
- WebSphere BI Adaptors
  ‣ SAP
  ‣ PeopleSoft
  ‣ Siebel

Web Other
- OLE DB
- Excel
- Flat files
- IBM Lotus Extended Search
  ‣ Web search
  ‣ LDAP
- Custom-built

Collaboration Systems
- Lotus Notes
- Microsoft Index Server
- IBM Lotus Extended Search
  ‣ Sametime
  ‣ QuickPlace
  ‣ Microsoft Exchange

Content & Imaging
- DB2 CM
  Family
- Domino.doc
- Documentum
- FileNet
- Open Text
- Stellent
- Interwoven
- Hummingbird

Workflow systems
- WebSphere
- FileNet

Plus partner tools and custom-built connectors to extend access to more sources

WebSphere Information Integrator Classic Federation for z/OS
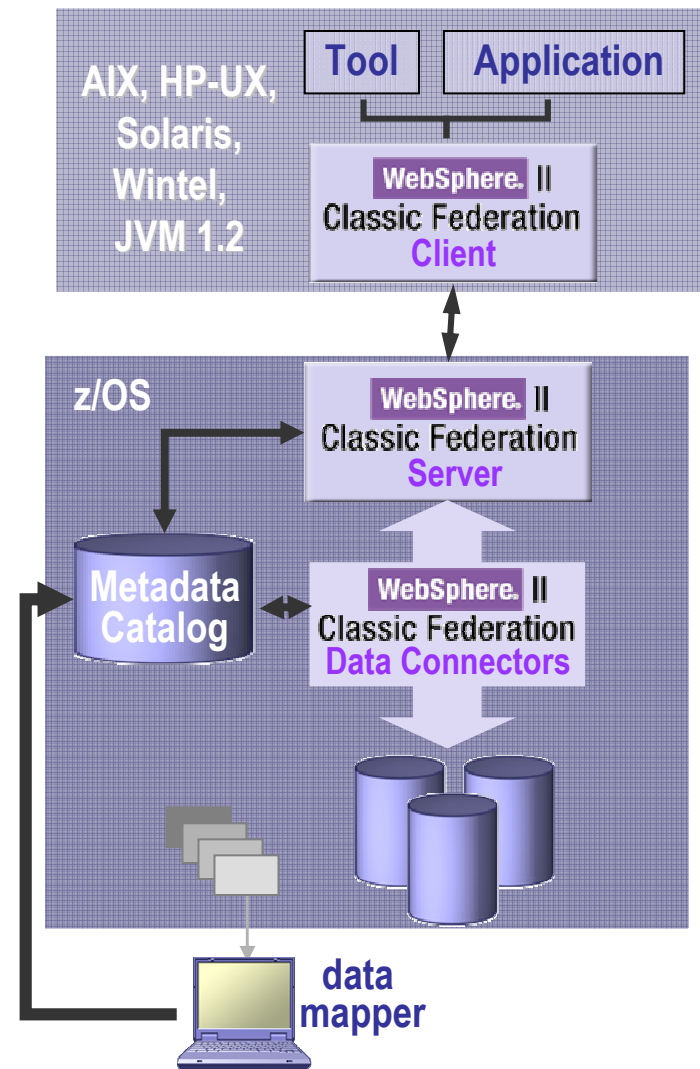
- **Typical Large IT Enterprise Environment**
  - ▶ Decades of heterogeneous technology investment :
    60% of data resides on mainframe and is growing 20% per year
  - ▶ Real-time access to mainframe–based data to remain competitive
  - ▶ High performance and scalability are mandatory

- **DB2 Information Integrator Classic Federation for z/OS**
  - ▶ Read/Write mainframe data sources using SQL through standard ODBC/CLI & JDBC
  - ▶ Native database connectors leverage power of each database/file accessed
  - ▶ Metadata-driven means:
    - • No mainframe programming required
    - • Fast installation, configuration & ease of maintenance

WebSphere Information Integrator
Classic **Federation for z/OS**

DB2 UDB for z/OS	VSAM & sequential	IMS/DLI	Software AG Adabas	CA Datacom	CA IDMS

WebSphere II Classic Federation Implementation

IBM

1) Create relational description of mainframe data sources by mapping the physical data definitions to logical tables and views

2) Mainframe Server and components act as a relational database engine

3) JDBC and/or ODBC drivers provide standardized interface for tools and applications

AIX, HP-UX, Solaris, Wintel, JVM 1.2

Tool   **Application**

WebSphere. II
Classic Federation
Client

z/OS

WebSphere. II
Classic Federation
Server

Metadata Catalog

WebSphere. II
Classic Federation
Data Connectors

data mapper

Self-service Application for Insurance Carrier

Pain associated with " *traditional* " implementations

▶ Option A : copy data to non-mainframe environments
  - Estimated cost $2M
  - Data refreshed every 30 hours or so

▶ Option B : integrate the IMS transactions
  - Estimated cost 10,000 man-hours per application

* Interactive Voice Response

Self-service Application for Insurance Carrier
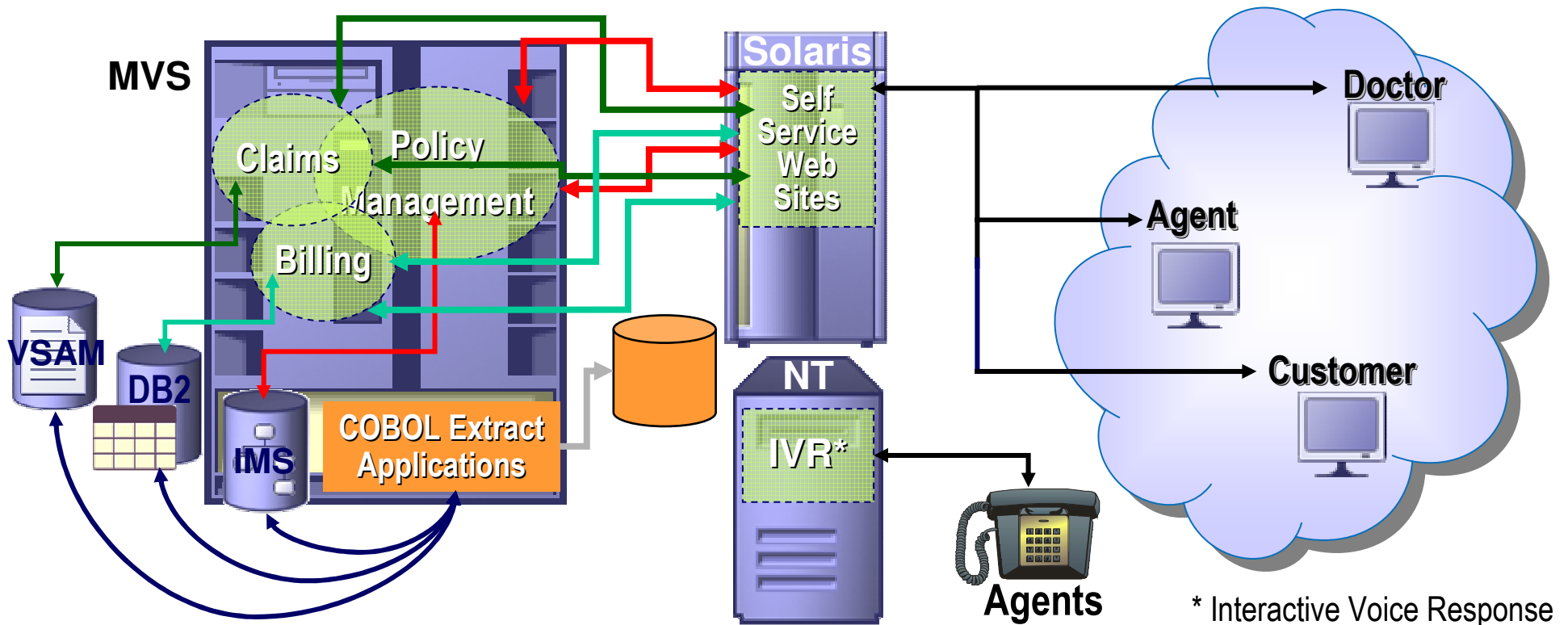
The IBM solution : **empower self-service environments**

▸ Provide up-to-the-minute policy, claims and accounting information

▸ Connect interactive voice response (IVR) system to IMS, VSAM & DB2
  - $250K versus $2M

▸ Connect operational data with self-service Web sites
  - 200 man-hours versus 10,000

* Interactive Voice Response

IBM

UNIVERSITÉ DU MAINFRAME

SQL

federated views (optional)

WS Information Integrator

DB2 Catalog

Stores information about
- Wrappers, servers, nicknames
- Server attributes
- Nickname attributes
- Remote functions

Nickname Nickname Nickname

Server   Server

Wrapper

Data at DB2 II instance:
- Global catalog
- User data
- MQTs
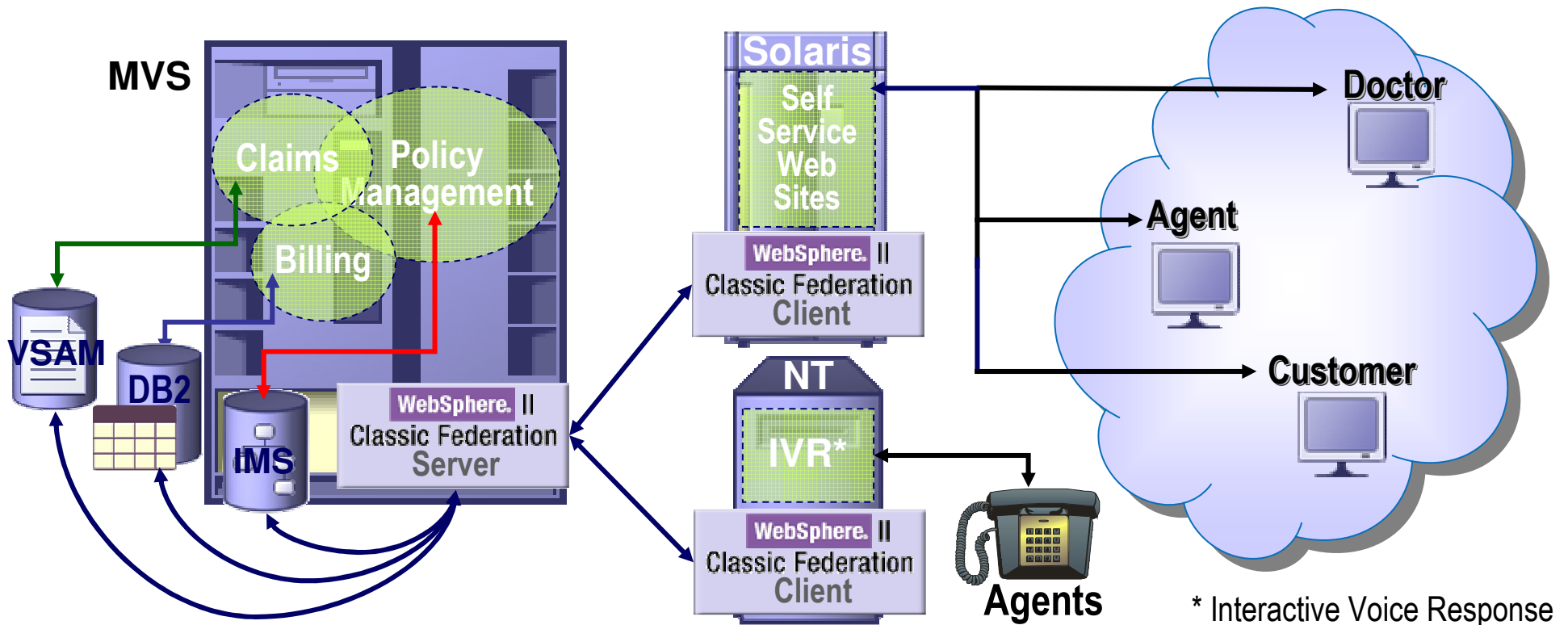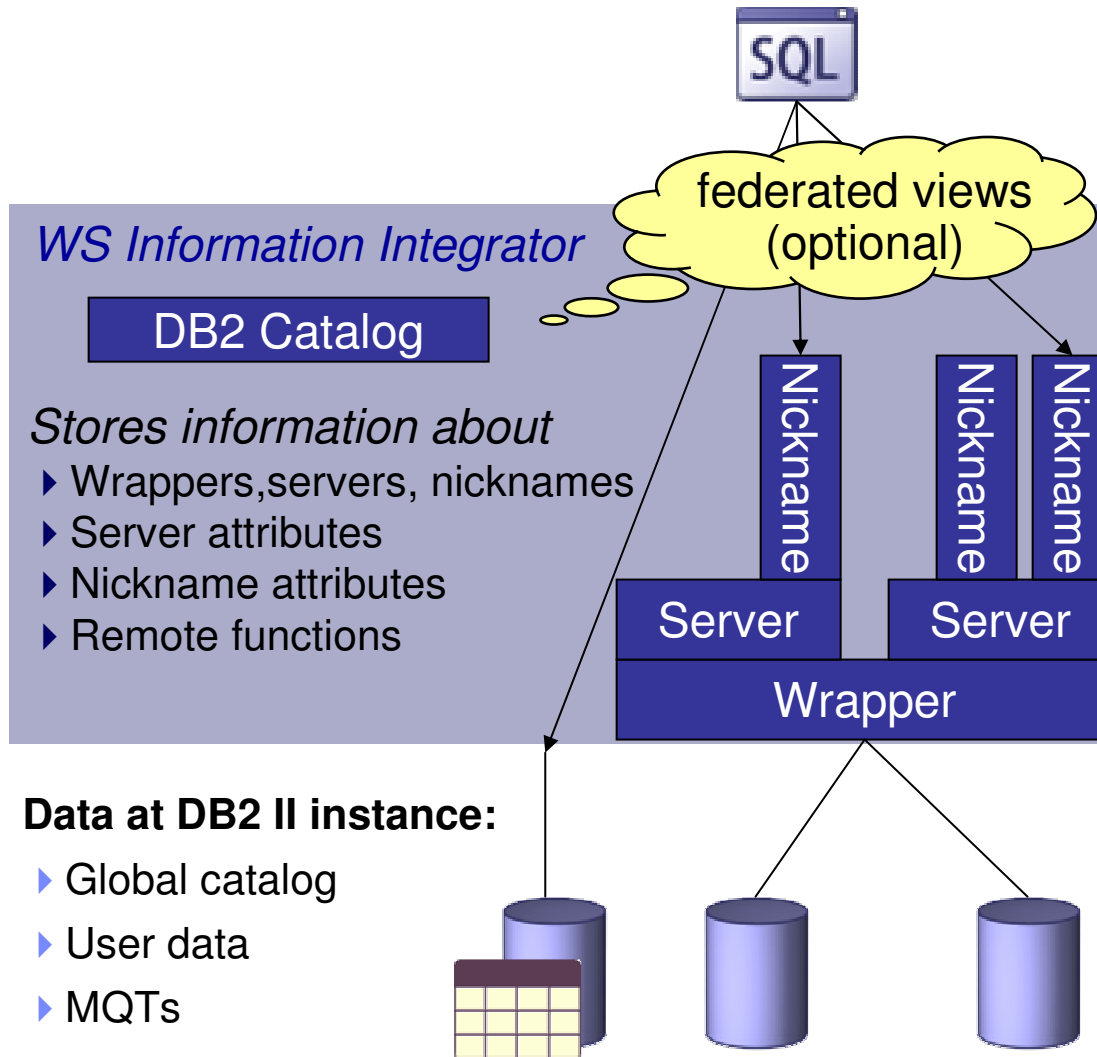
- **Wrapper:**
  a library allowing access to a particular class of data sources or protocols (Net8, DRDA, CTLIB...). Contains information about data source characteristics

- **Server:**
  represents a specific data source

- **Nickname:**
  a local alias to data on a remote server (mapped to rows and columns)

Business Scenario

➢ Access all kinds of DB2 II Data Sources from legacy Applications

➢ Single Site Update, Multi Site Read

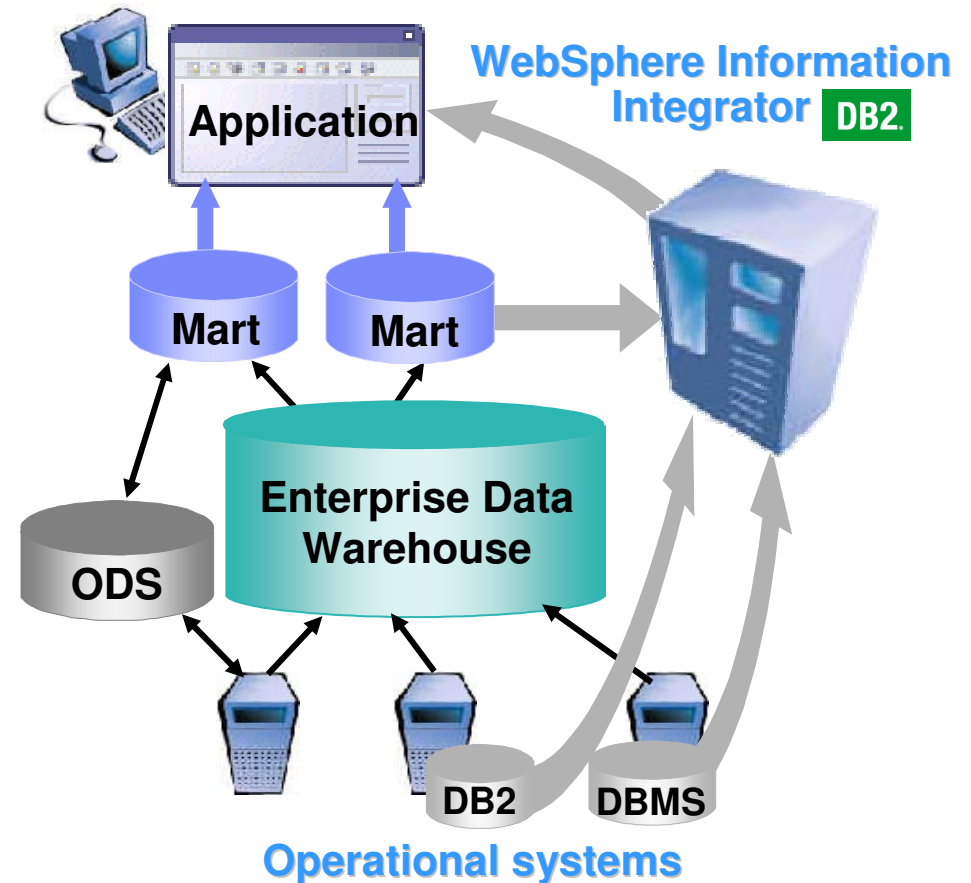➢ Distributed UOW with 2PC support in plan for following version of WebSphere II

CICS/IMS Applications on z/OS

Websphere Information Integrator Database configured as DB2 z/OS CDB Location

DB2

DRDA
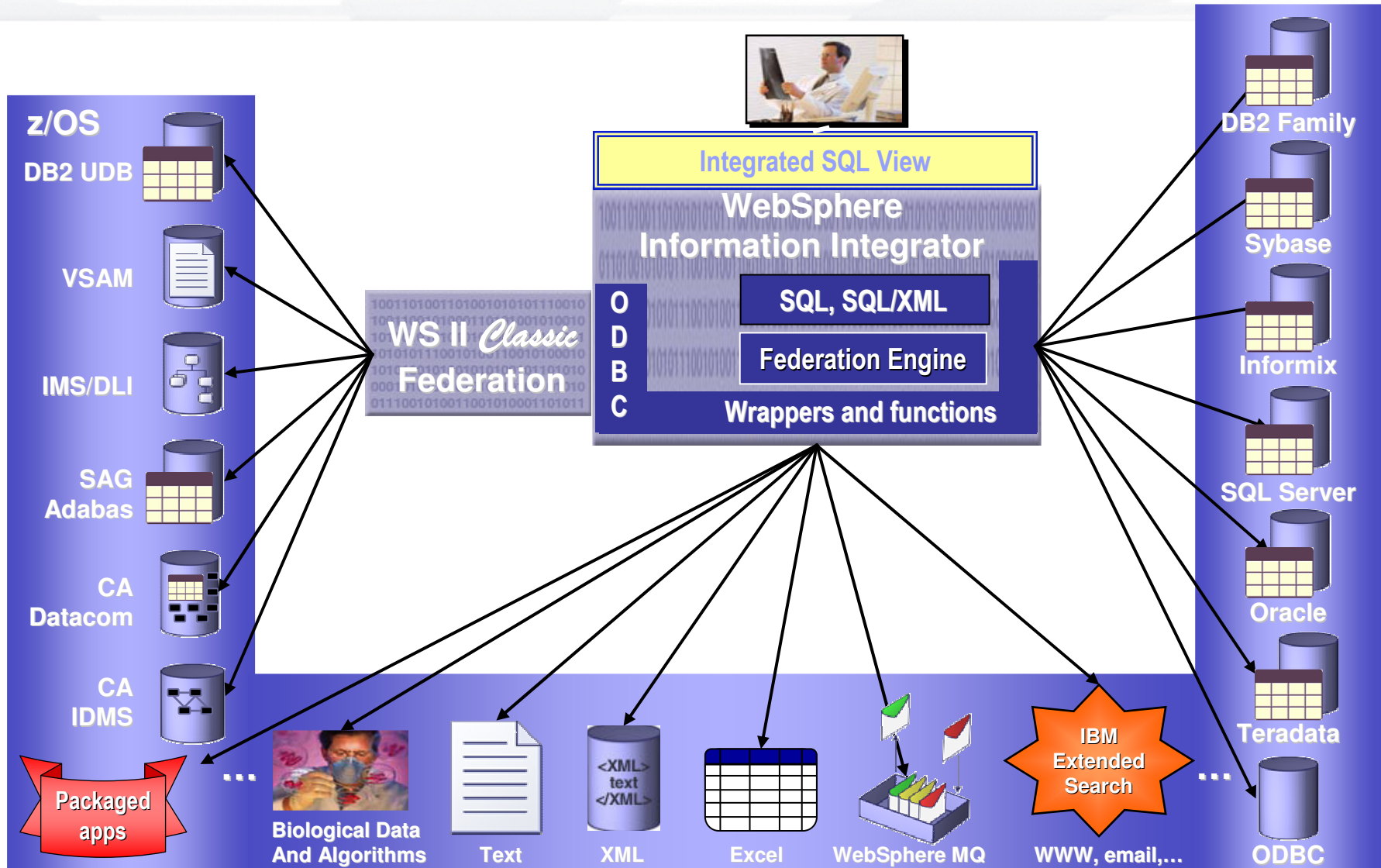
Distributed Hetero-geneous Data Sources are accessed (read/write) through DB2 Information Integrator Nicknames

DB2

WebSphere Information Integrator

DB2

ODBC

Microsoft

net8

ORACLE

Speeding Portal Development
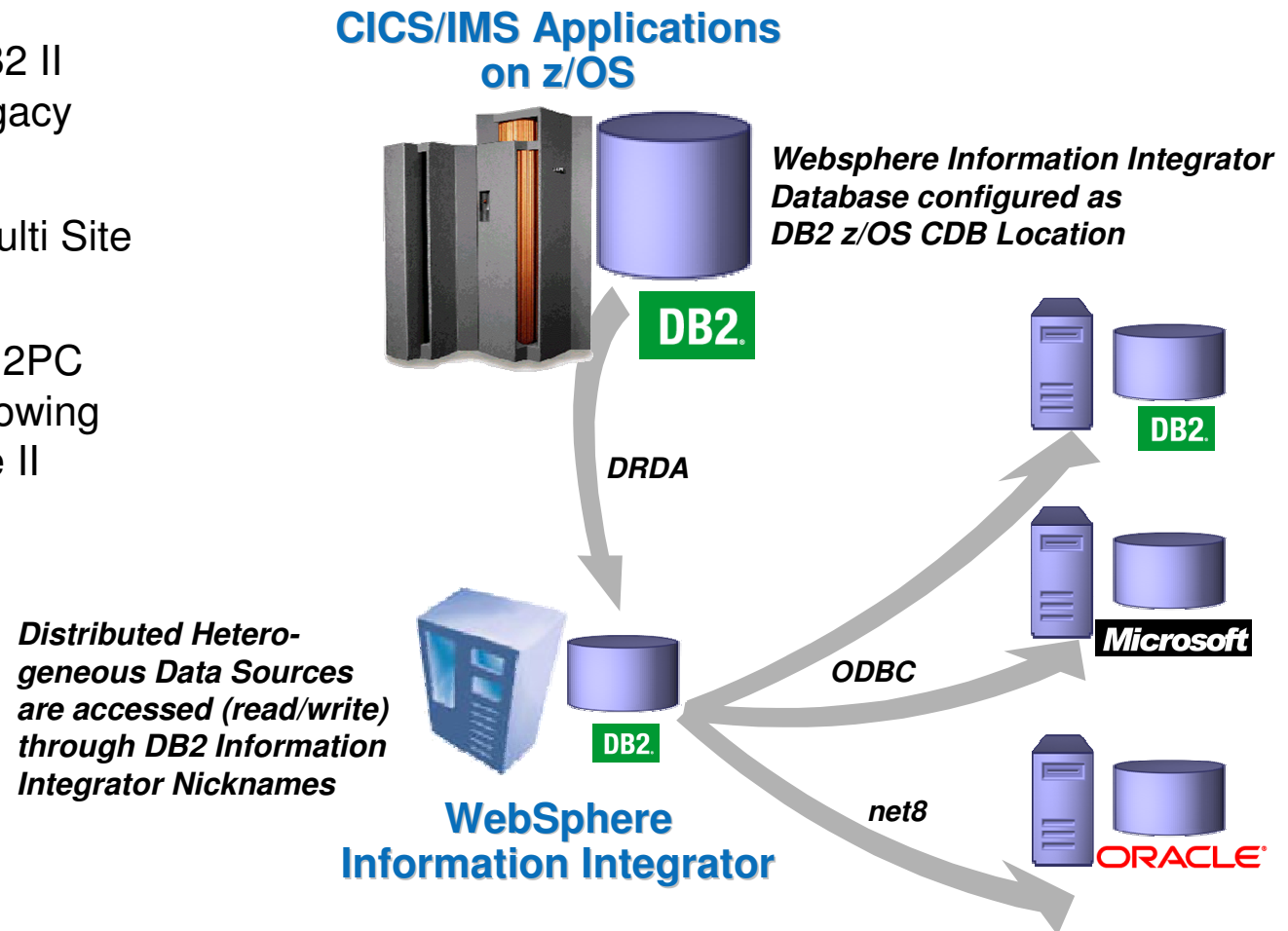
Business Scenario

➢ Providing Portal Developers with a single Interface for structured and unstructured Information

SQL
SQL/XML
XML-UDF's

Portal Application

Type 4 JDBC Driver
J2EE 1.4 / JDBC 3.0
DB2 .NET Data Provider
.NET UDF / SP

DB2.

Account Status

Docs, EMails

XML Repository
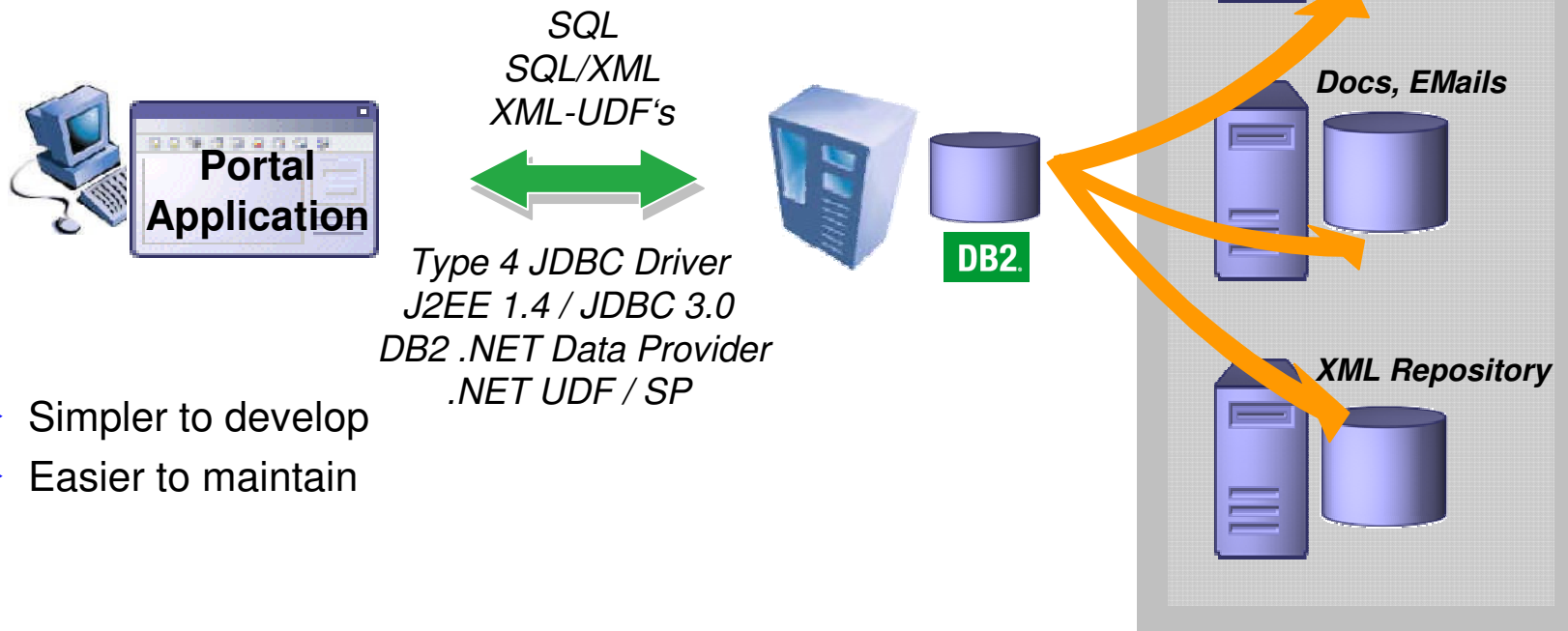
➢ Simpler to develop
➢ Easier to maintain

Extending Data Warehouses with Real-Time Data

Business Scenario

- ➤ Access current customer records from a call centre

- ➤ Access current stock levels from a supply chain data mart

- ➤ Basel II compliance

- ➤ Business activity monitoring – linking events to trends

Application

WebSphere Information Integrator DB2.

Mart **Mart**

ODS **Enterprise Data Warehouse**

DB2 **DBMS**

Operational systems

IBM

- Real-Time Access

- Global Optimization

- SQL Pushdown

- **<u>SQL Transparency</u>**
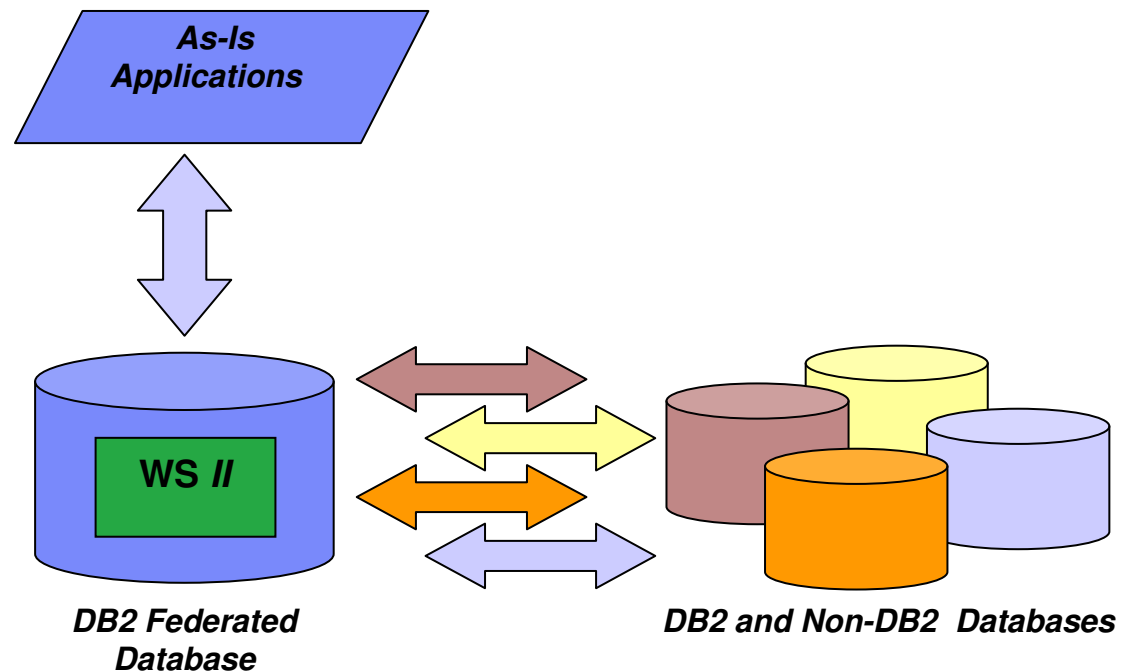
As-Is
Applications

WS *II*

DB2 Federated
Database

DB2 and Non-DB2 Databases

Agenda

- **Introduction to Enterprise Information Integration**
  - ▸ Big Picture
  - ▸ IBM Platform

- **Understand, Cleanse and Transform Enterprise Data**

- **Federate Enterprise Data**

- **Connect and Deliver Enterprise Data**
  - ▸ SQL and Q Replication
  - ▸ Event Publishing

- **Combining strategies**

- **Models of Data Integration**
  - ▸ Real time - pull and push
  - ▸ Near real time - replication
  - ▸ Scheduled - ETL and consolidation
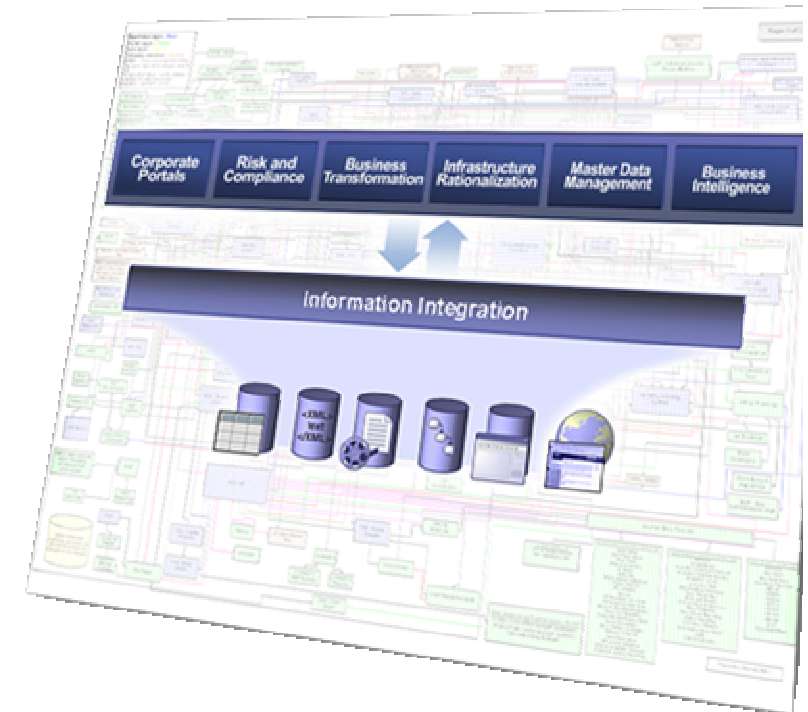
Why Replicate?

Distribution / Consolidation

- Move data between central to branches, branches to central, or both
- Federate or Replicate ?
  - where does the application need the data ?
  - does the data need to be real time ?
  - what is the change volume ?

Warehouse & Business Intelligence

- Move data to new platform/database, transform data
- ETL or Replicate ?
  - latency needs
  - change volume versus total volume
  - complexity of transformation and/or cleansing

Mobile Workforce

- Occasionally connected distribution/consolidation

Availability

- Scheduled outage, failover, disaster recovery
  - can use Hardware and/or Software
  - replication offers lower expense, faster restart, multi-purpose
  - Hardware offers simplicity of setup
- Move query/reporting work to separate system
  - other methods such as flash copy also possible
- Peer to peer - split workload
  - only possible through replication
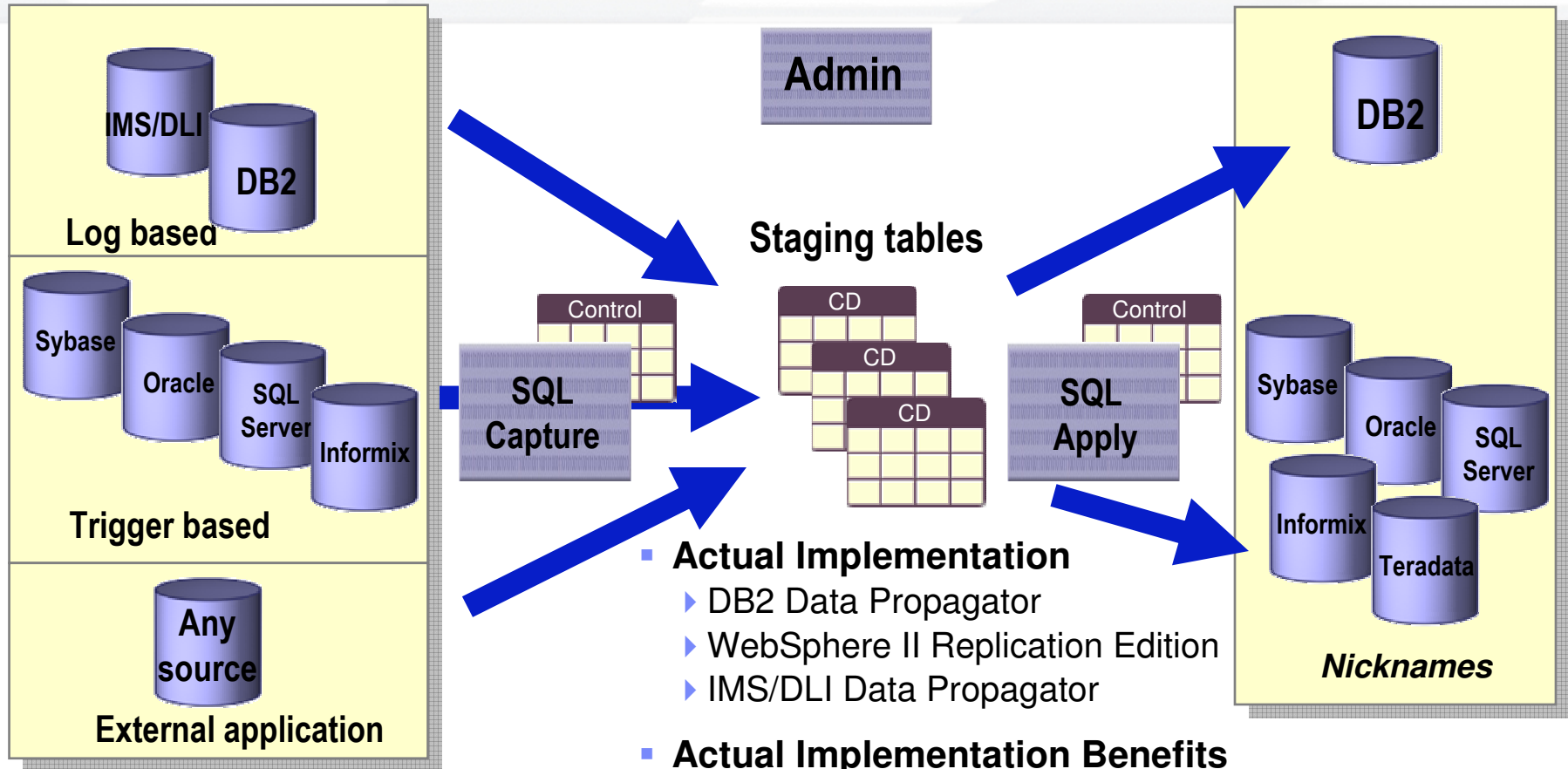 - requires serious planning and consideration

Replicate

Database

Database — Capture Apply

Database
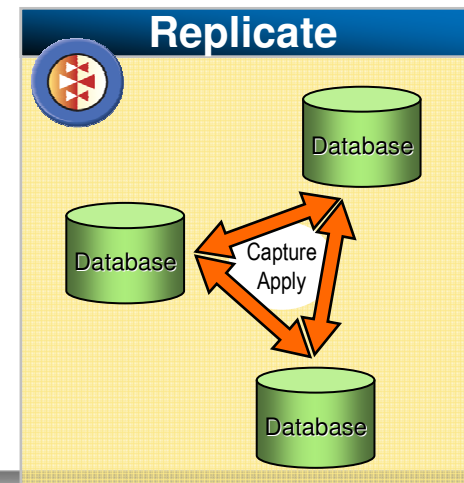
# IBM's Information Replication Architectures

- **SQL Replication (a.k.a. IBM DataPropagator)**
  - ▸ Log or trigger-based asynchronous changed-data capture
  - ▸ Apply to DB2 staging tables
  - ▸ Additional transformations and fan-out from the staging tables
  - ▸ Versatile replication architecture for both homogeneous (DB2 Family) and heterogeneous replication

- **Q-based Replication**
  - ▸ Log-based asynchronous changed-data capture
  - ▸ Data distribution via WebSphere MQ message queues
  - ▸ Highly parallel apply mechanism at target platform
  - ▸ High-volume, low-latency architecture
  - ▸ Uni-directional, bi-directional and peer-to-peer
 - ▸ Proven successes with about a year of general availability

SQL Replication Architecture

IMS/DLI

DB2

Log based

Sybase

Oracle

SQL Server

Informix

Trigger based

Any source

External application

Admin

Staging tables

Control

SQL Capture

CD

CD

CD

Control

SQL Apply

DB2

Sybase

Oracle

SQL Server

Informix

Teradata

Nicknames

- **Actual Implementation**
  - ▸ DB2 Data Propagator
  - ▸ WebSphere II Replication Edition
  - ▸ IMS/DLI Data Propagator

- **Actual Implementation Benefits**
  - ▸ Extremely flexible and resilient
  - ▸ Very easy to set up transformations
  - ▸ Scales well to reach multiple targets
  - ▸ Homogeneous & Heterogeneous Sources

UNIVERSITÉ DU MAINFRAME

Websphere Information Integrator introduces new replication architecture for delivering extremely low latency replication
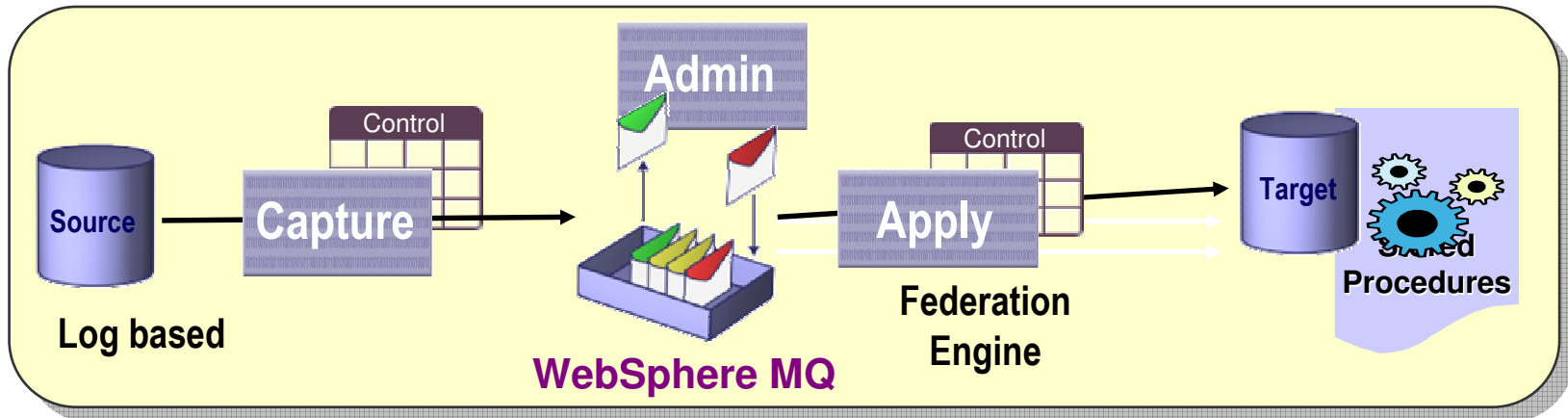
- Each message represents a transaction

- Highly parallel apply process
  - Non dependent transactions re-parallelized at the target

- Differentiated conflict detection and resolution

- Integrated infrastructure for replication and publishing

- DB2 to DB2 today

- Data Integrity
  - Persistent messaging with WebsphereMQ
  - Detects missing messages

- Data transformation
  - Triggers on the target table
  - Stored Procedures called by Apply at the row level
  - Publish Event to user application

- DB2 to DB2 today
  - Staged availability of heterogeneous support
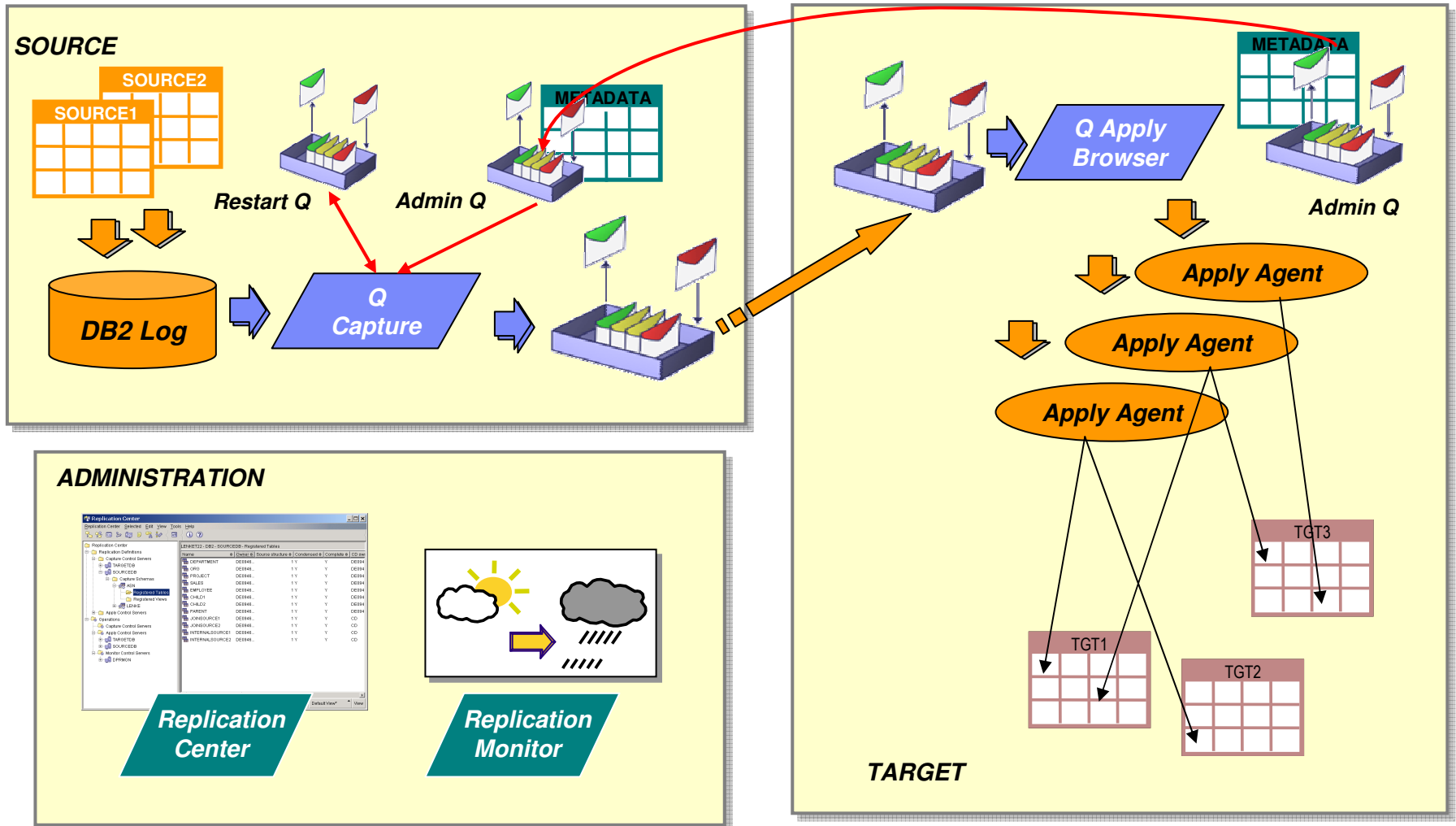
Q Replication – Defining Subsets or Filters

- **Subset data**
  - Subset of rows through Q Capture predicate on subscription/publication
  - Subset of columns through subscription/publication definition
  - Signal (IGNORETRANS) defined to allow user selected transactions to be ignored
  - Subscription/publication send options
    - *Change Only*: Publish only columns that have changed vs all columns in the row
    - *All Changed Rows*: Publish a row if any column changes (subscribed or not)
    - *Suppress Delete*: Do not publish row deletes
    - *Before/After values*: Publish before values as well as after values

- **Predicate examples**
  - Based on values in the row data itself
    ```
    WHERE :LOCATION ='EAST' AND :SALES  > 100000
    ```
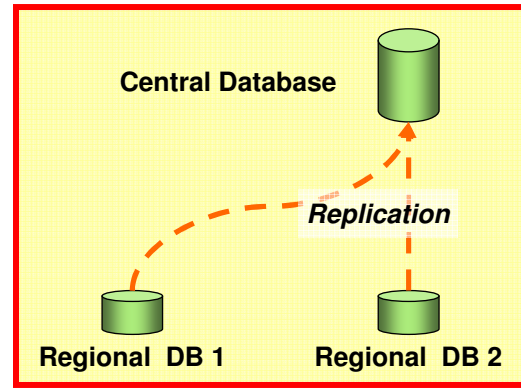
  - Based on values in data in other tables
    ```
    WHERE :LOCATION ='EAST' AND :SALES > (SELECT SUM(expense)
    FROM STORES WHERE stores.deptno = :DEPTNO)
    ```

Many Models of Q-Replication

High Availability
(bi-directional)

Production   Live Backup

Replication →

Reporting Tools

Rollup
(many to 1)

Central Database

Replication

Regional DB 1   Regional DB 2

Peer To Peer
(multi-directional)

Production 1   Production 2

Replication →

Balanced Workload

Distribution
(1 to many)

DB2

Replication

Oracle   MS SQL   Informix   Sybase
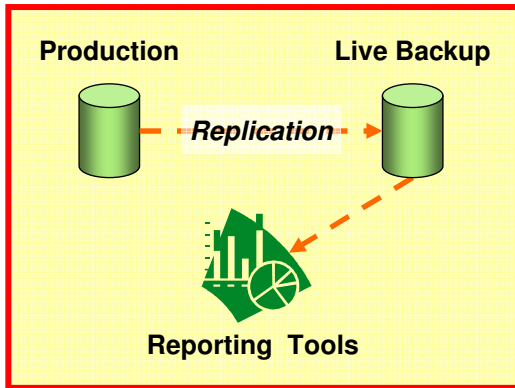
- Provide continuous availability for critical applications – zero downtime for planned and unplanned outages, automatic conflict detection and resolution

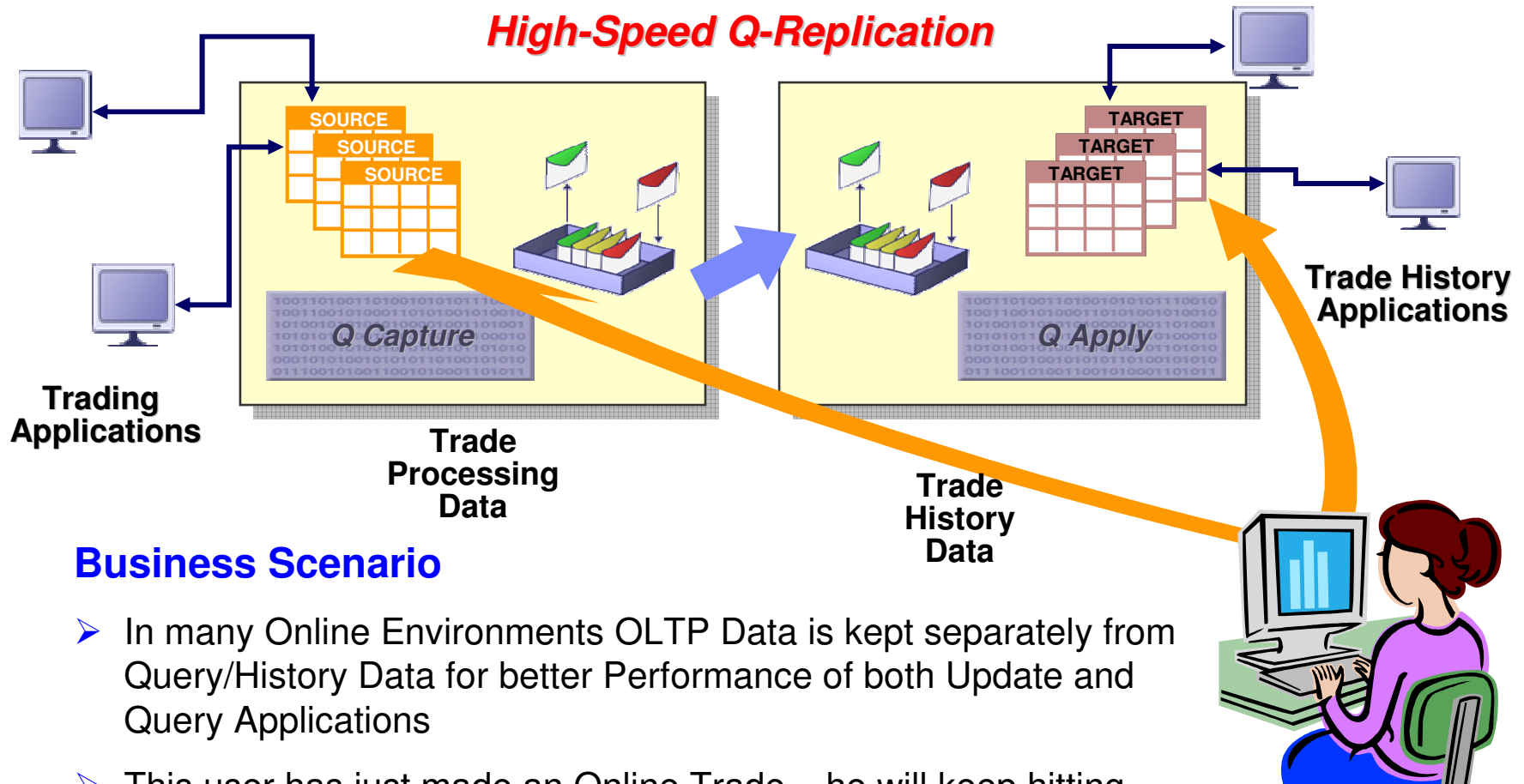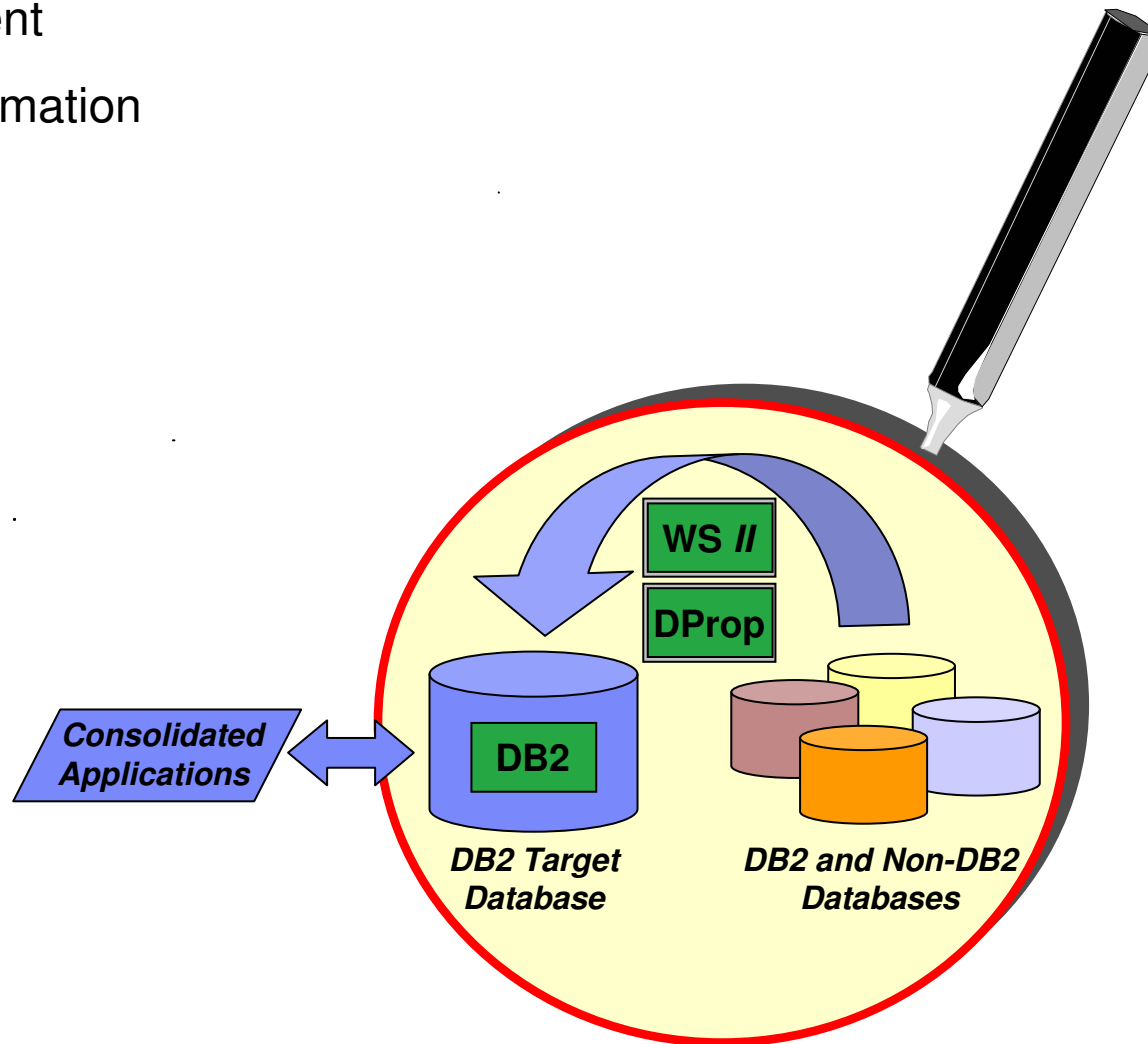- Spread application load across multiple servers or data centers with low-latency data synchronization
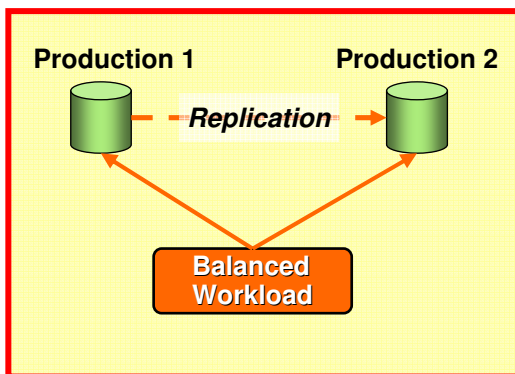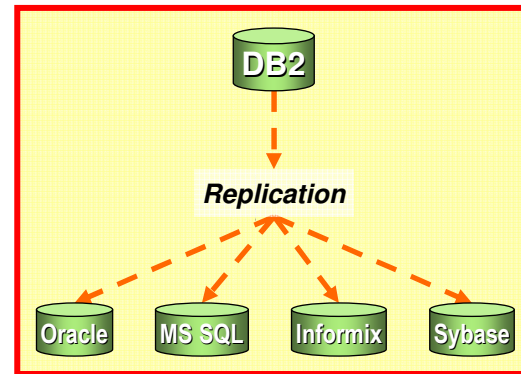
- Provide consistent, timely information to users and applications for better decision-making while offloading critical application servers

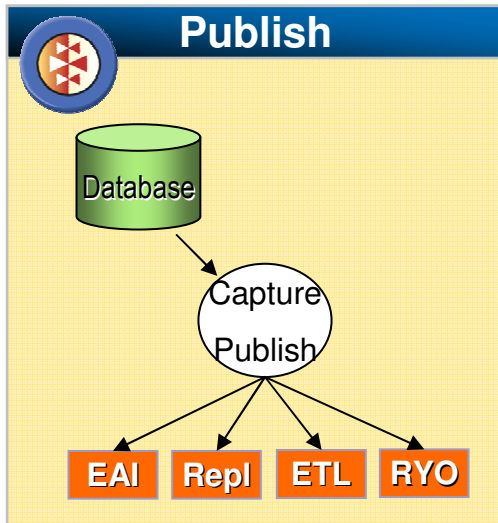- Automatically distribute data to many locations, reducing development costs and effort

High-Speed Q-Replication

Trading Applications

SOURCE
SOURCE
SOURCE

Q Capture

TARGET
TARGET
TARGET

Q Apply

Trade History Applications

Trade Processing Data

Trade History Data

Business Scenario

➢ In many Online Environments OLTP Data is kept separately from Query/History Data for better Performance of both Update and Query Applications

➢ This user has just made an Online Trade – he will keep hitting Enter until he sees that the Trade is complete, in this Case meaning it has been replicated to the Trade History Database

- Data Movement

- Data Transformation

- **<u>Autonomy</u>**

WS *II*

DProp

DB2

Consolidated Applications

DB2 Target Database

DB2 and Non-DB2 Databases

Publish

Database

Capture

Publish

| EAI | Repl | ETL | RYO |

- Database to Application Messaging
  - ▸ Drive downstream applications, based on transactional data events
  - ▸ Reduce application development and maintenance
  - ▸ Reduce performance impact to source applications
  - ▸ Reduce availability impact to source applications

- Meet Auditing Requirements
  - ▸ Capture and store information regarding what changes were made to critical business data and by whom

- Event Notification
  - ▸ Stream changed data information to Web interfaces
  - ▸ Stream only particular events of interest (filter data)

- Warehouse / Business Intelligence
  - ▸ Integrate captured changed data with an ETL tool
  - ▸ Perform very complex transformations, use a specific transaction format to update target

- Integration is independent of the source applications
  - ▸ Relatively straight forward to find data items rather than every business rule
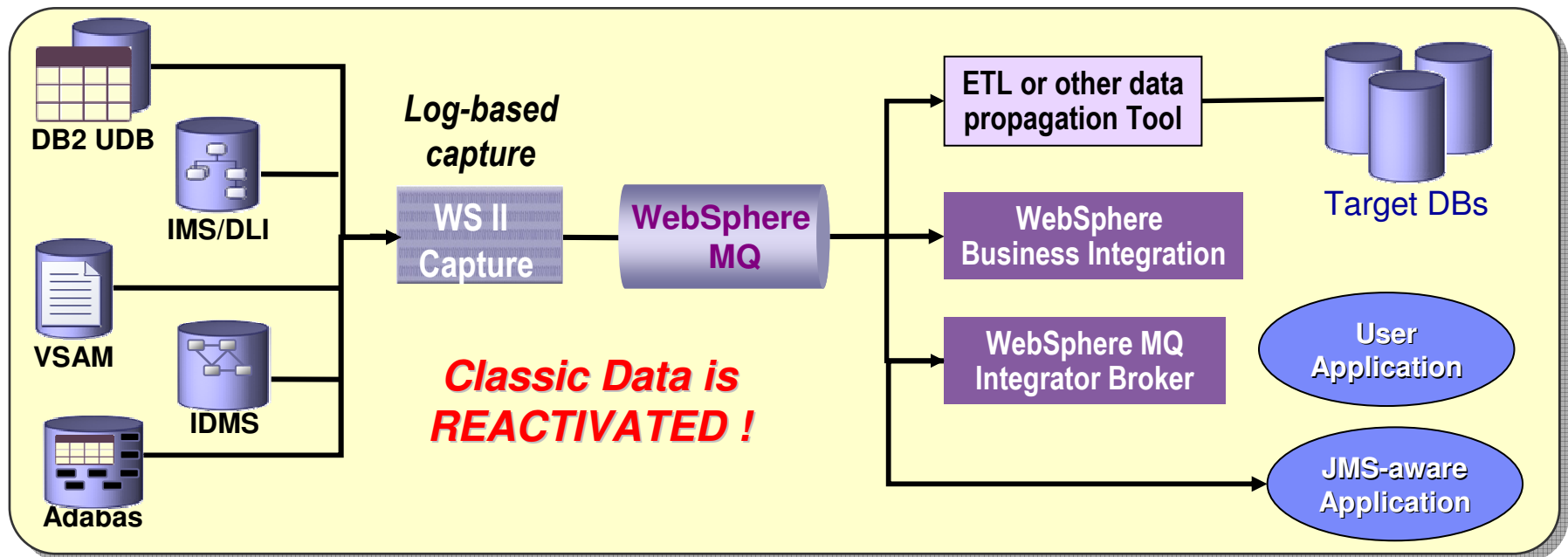 - ▸ Applications grow and evolve with minimal impact on the integration

Publishing data events to facilitate business integration

IBM   **UNIVERSITÉ DU MAINFRAME**

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

Function

➢ Publish events to a message queue

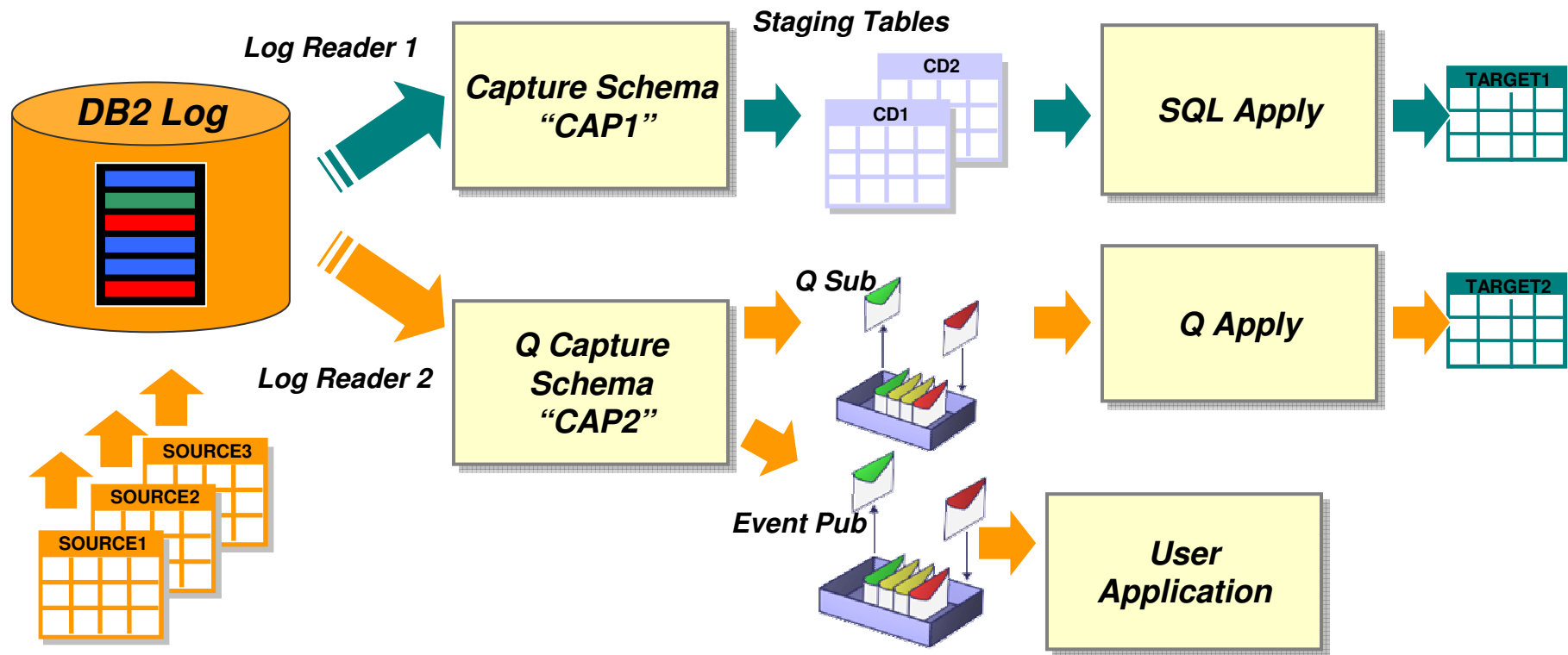➢ XML self-describing format

➢ Wizard-driven configuration

Usage

➢ Application to application messaging

➢ Event streaming

➢ Source for ETL tool

DB2 UDB

IMS/DLI

VSAM

IDMS

Adabas

Log-based capture

WS II Capture

WebSphere MQ

Classic Data is REACTIVATED !

ETL or other data propagation Tool

WebSphere Business Integration

WebSphere MQ Integrator Broker

Target DBs

User Application

JMS-aware Application

Two Event Publisher infrastructures

▸ DB2 UDB for z/OS based on WebSphere II Replication

▸ WebSphere II Classic Event Publisher (based on Classic Federation)

UNIVERSITÉ DU MAINFRAME

Staging Tables

Log Reader 1

DB2 Log

Capture Schema "CAP1"

CD2

CD1

SQL Apply

TARGET1

Q Sub

Q Apply

TARGET2

Log Reader 2

Q Capture Schema "CAP2"

SOURCE3

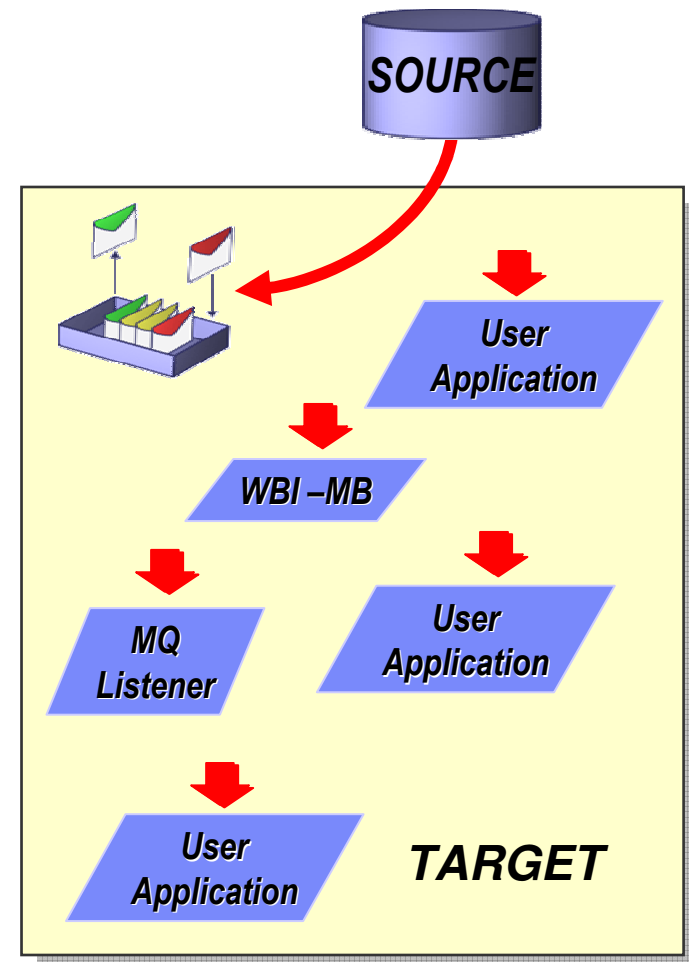SOURCE2

SOURCE1

Event Pub

User Application

SQL Replication and Q Replication can co-exist

➢ Managed at source by using multiple capture schemas

➢ One Q Capture can handle both Publications and Subscriptions

- **Integration is independent of the source applications**

  ▸ Applications grow and evolve with
  " minimal impact on the integration "

  ▸ Relatively straight forward to find data items
  " rather than every business rule "

  ▸ e.g.many applications change inventory
  but the data values trigger re-stocking

  ▸ e.g. many transactions impact Claim status
  changes to status value drives workflow

  ▸ e.g. order data is needed by a CRM
  CRM has no dependence on ordering process

SOURCE

User Application

WBI –MB

MQ Listener

User Application

User Application

TARGET

Event Publishing for Business Intelligence

Feeding Changed Data to :

➤ Data Warehouse

➤ Datamart

➤ Operational Data Store (ODS)

**Optimize Resource Utilization**

➤ Minimize Bandwidth Requirements

➤ Maximize Data Currency

**Complements with
WebSphere II Federation**

➤ Data feed using Event *Publishers*

➤ Real-time extensions using *Federation*

Q-enabled ETL-Tool

WebSphere Information Integrator Event Publishers for z/OS

DB2 UDB for z/OS   **VSAM**   **IMS/DLI**   **CA-IDMS**

# Event Publisher simplifies cross-application synchronization
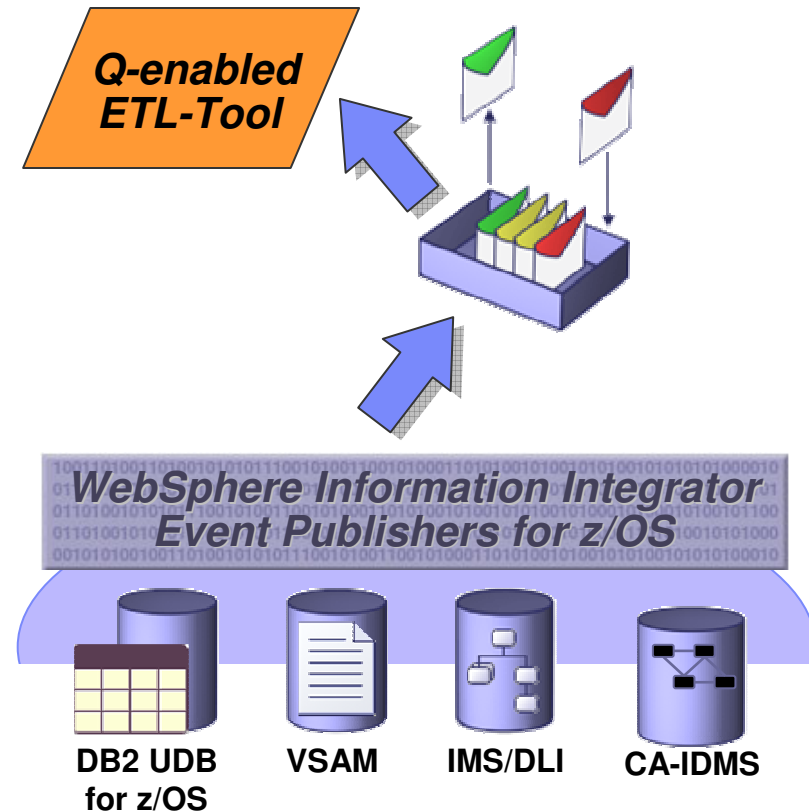
- Near real-time cross-silo data synchronization
  - e.g. New order data is automatically pushed to a CRM application
  - e.g. VSAM employee data updates are pushed to SAP payroll
  - ▶ Loosely coupled integration
 - ▶ Minimizes development effort
  - ▶ Simplifies maintenance
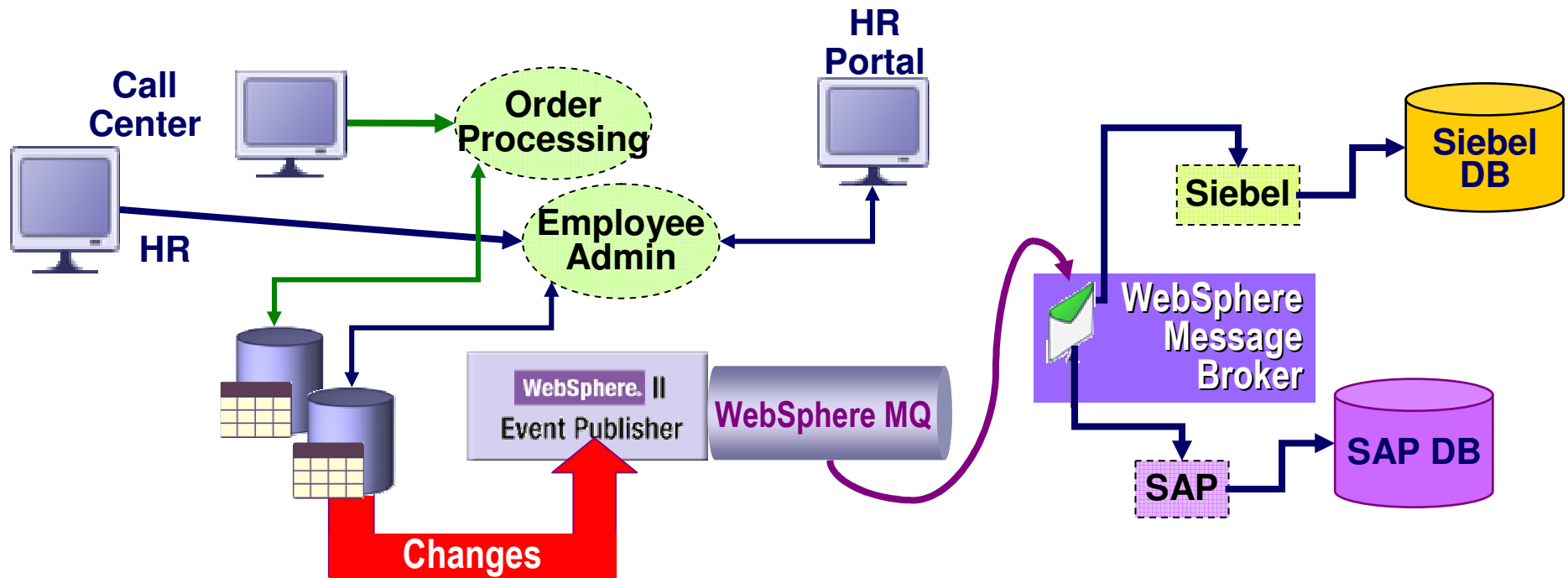
Some Rules – When to choose What

- Low-Latency, High-Volume Replication
- Source and Target Tables of similar Structure
- Bi-directional Replication (e.g. for Hot-Standby Purposes)
- Peer-2-Peer – Splitted Workload
- Huge Number of Tables (e.g. Siebel)
- ✓ Replication across DB2-Family
- ✓ Replication from DB2 to federated Targets
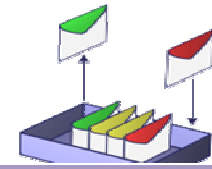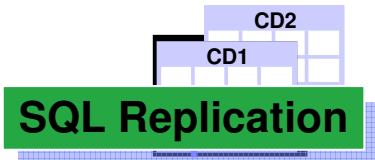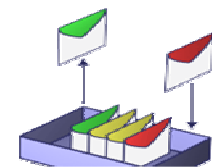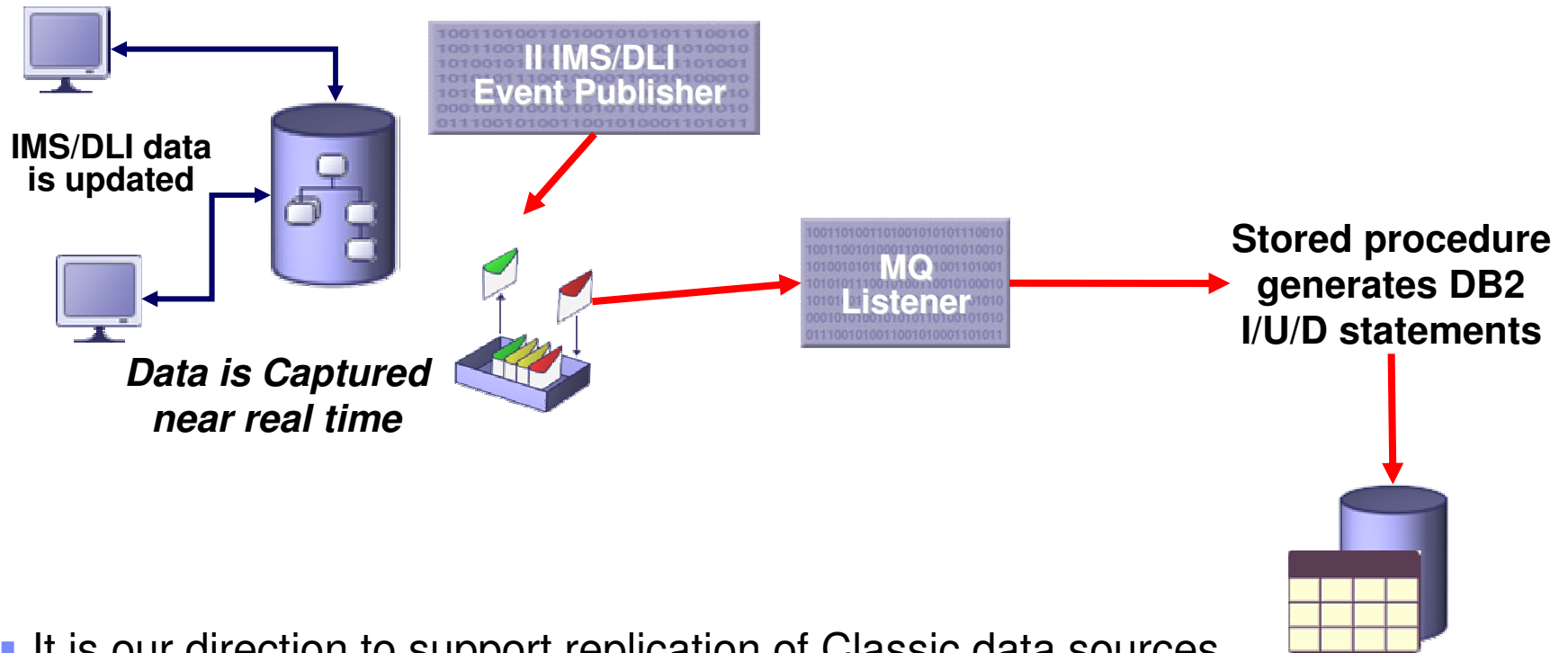
Q Replication

- Fan-Out to huge Number of Targets
- Multi-Tier Staging via CCD
- Source and Target Tables not of the same Structure (which requires Source Views, Joins or SQL Expressions)
- Replication from federated Sources

CD2
CD1
SQL Replication

- Processes triggered by Data Events
- Kick-Off of Workflows due to changed Data
- Seamless Integration of DB2 and Classic Data (IMS, VSAM, IDMS)
- Delivery of Data Events from various Origins to a single Application or Message Broker
- Real-Time ETL

Event Publishing

II IMS/DLI Event Publisher

IMS/DLI data is updated

Data is Captured near real time

MQ Listener

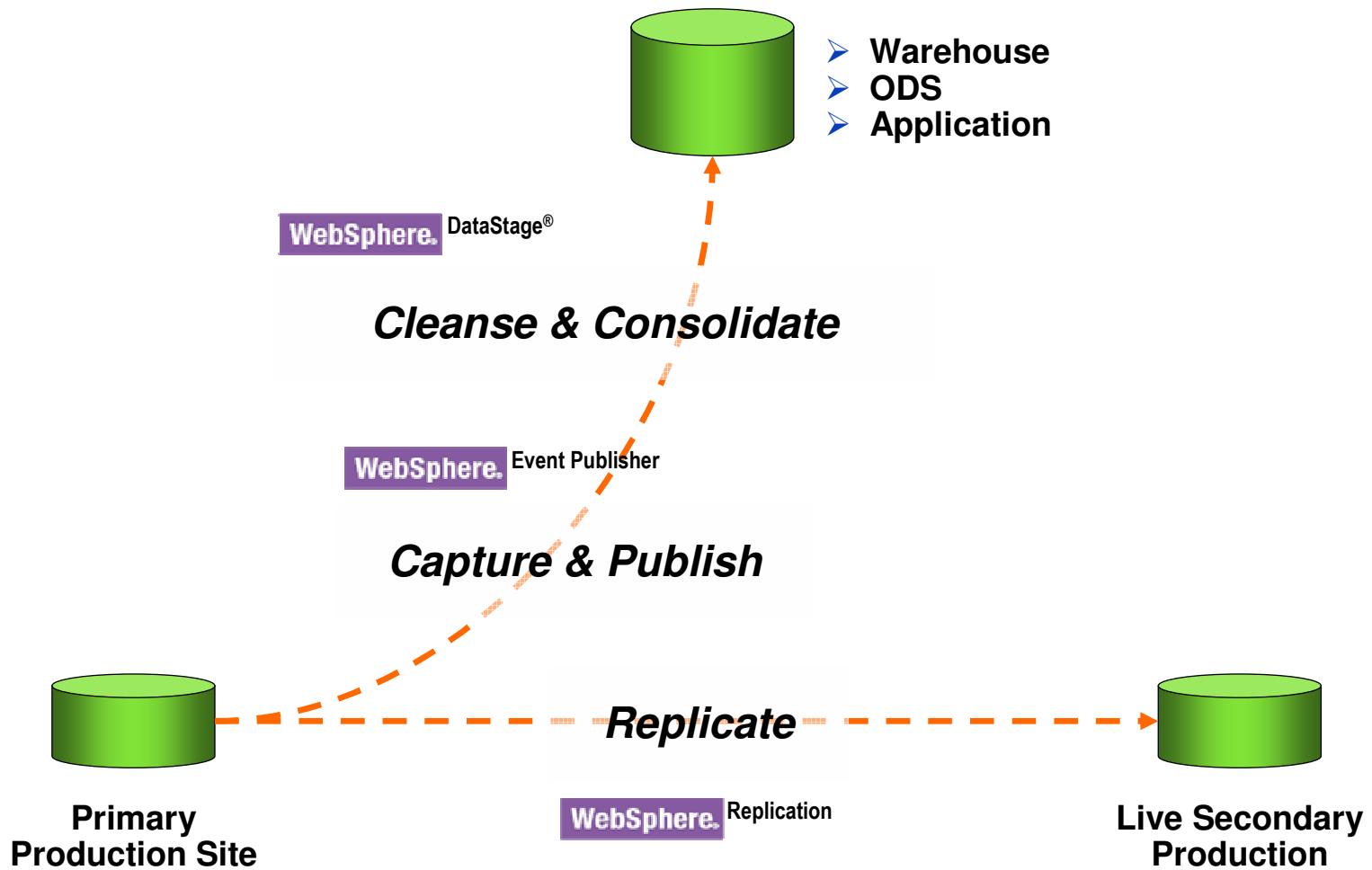**Stored procedure generates DB2 I/U/D statements**

- It is our direction to support replication of Classic data sources using the highly parallel Q Apply of DB2 II Replication.

- As an interim suggestion, SQL stored procedures can be used to apply the data captured through the II Classic Event Publisher

Agenda

- **Introduction to Enterprise Information Integration**
  - ▸ Big Picture
  - ▸ IBM Platform

- **Understand, Cleanse and Transform Enterprise Data**

- **Federate Enterprise Data**

- **Connect and Deliver Enterprise Data**
  - ▸ SQL and Q Replication
  - ▸ Event Publishing

- **Combining strategies**

- **Models of Data Integration**
  - ▸ Real time - pull and push
  - ▸ Near real time - replication
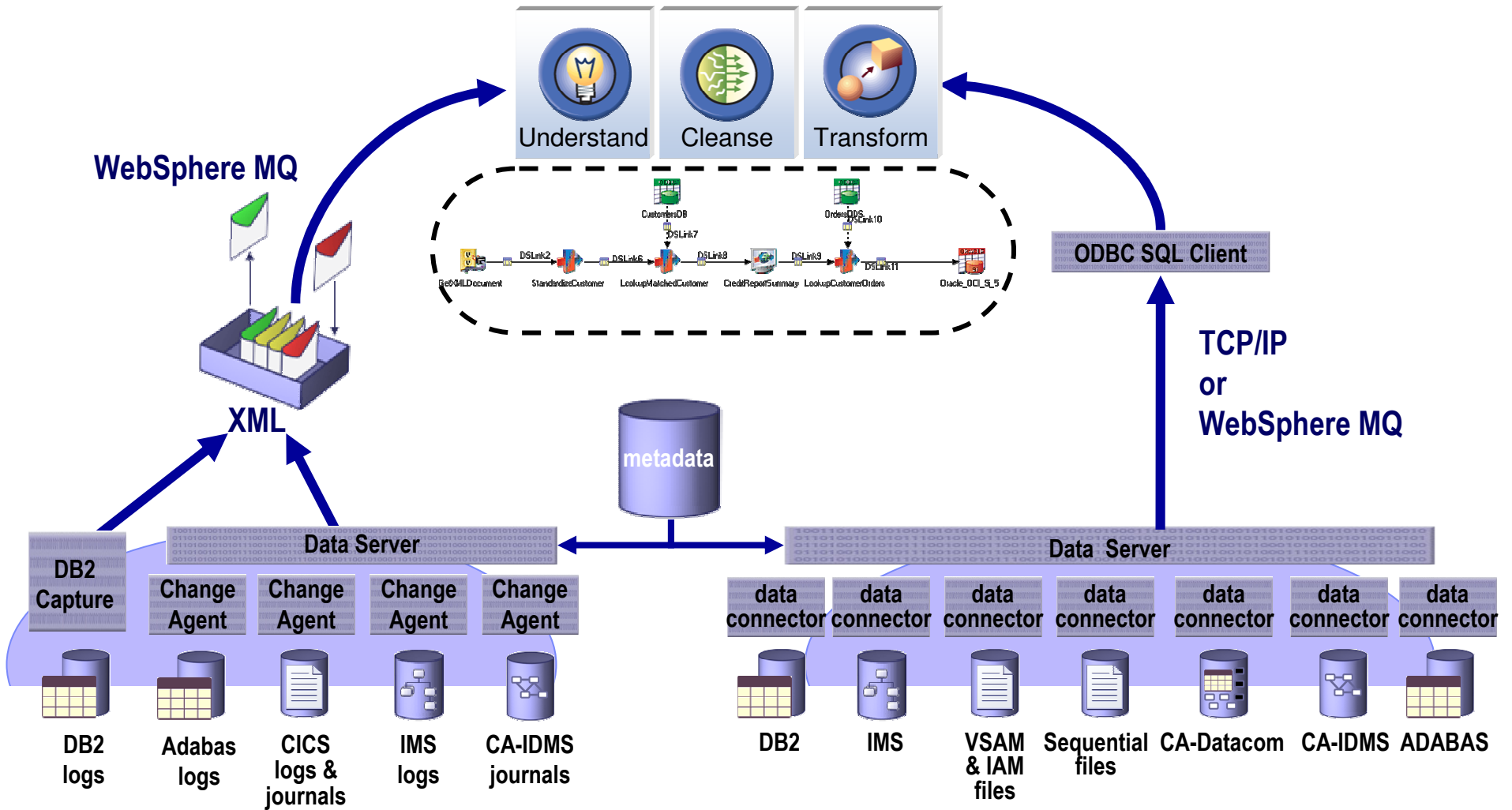 - ▸ Scheduled - ETL and consolidation
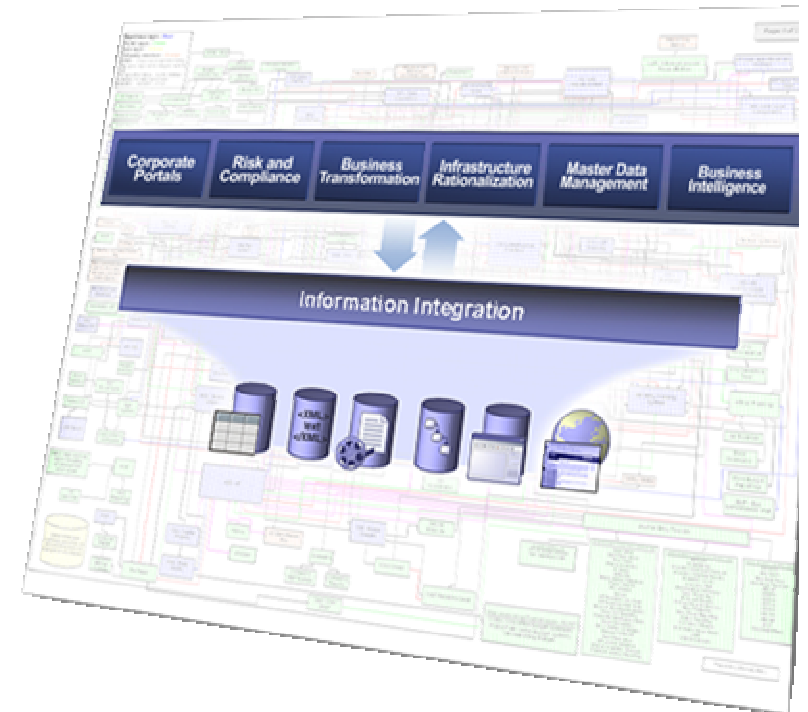
Combining Models - *use II Infrastructure to publish to the warehouse while maintaining a live secondary backup production system*

UNIVERSITÉ DU
MAINFRAME

➢ **Warehouse**
➢ **ODS**
➢ **Application**

WebSphere. DataStage®

Cleanse & Consolidate

WebSphere. Event Publisher

Capture & Publish

Replicate

**Primary
Production Site**

WebSphere. Replication

**Live Secondary
Production**

zSeries Data behind:
ProfileStage, QualityStage & DataStage

| Understand | Cleanse | Transform |

WebSphere MQ

CustomersDB
DSLink7
Orders.DDS
DSLink10

GetXMLDocument — DSLink2 — StandardizeCustomer — DSLink6 — LookupMatchedCustomer — DSLink8 — CreditReportSummary — DSLink9 — LookupCustomerOrders — DSLink11 — Oracle_OCI_Sj_5

XML

ODBC SQL Client

TCP/IP or WebSphere MQ

metadata

Data Server

Data Server

| DB2 Capture | Change Agent | Change Agent | Change Agent | Change Agent |

| data connector | data connector | data connector | data connector | data connector | data connector | data connector |

| DB2 logs | Adabas logs | CICS logs & journals | IMS logs | CA-IDMS journals |

| DB2 | IMS | VSAM & IAM files | Sequential files | CA-Datacom | CA-IDMS | ADABAS |

Agenda

- Introduction to Enterprise Information Integration
  - ▸ Big Picture
  - ▸ IBM Platform

- Understand, Cleanse and Transform Enterprise Data

- Federate Enterprise Data

- Connect and Deliver Enterprise Data
  - ▸ SQL and Q Replication
  - ▸ Event Publishing

- Combining strategies

- Models of Data Integration
  - ▸ Real time - pull and push
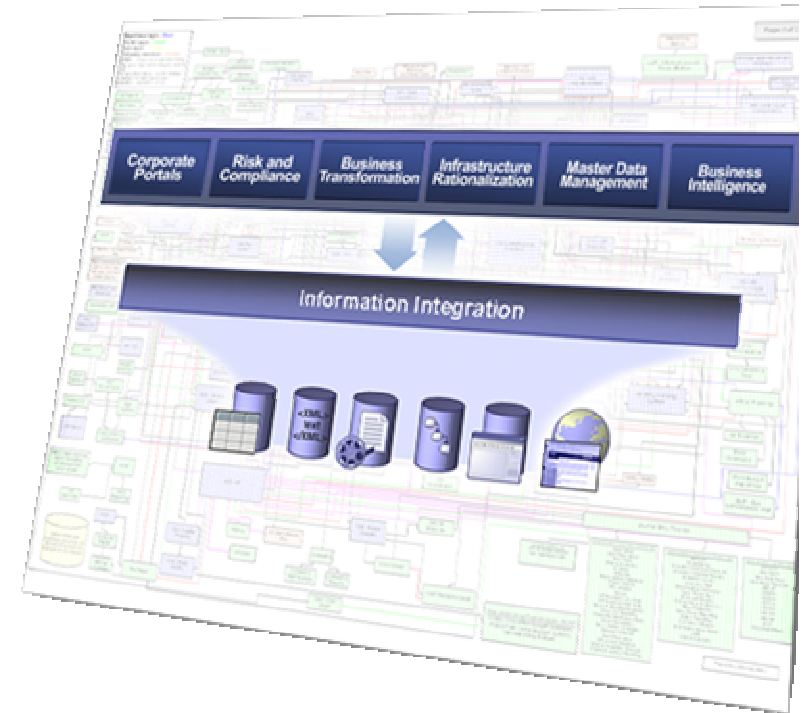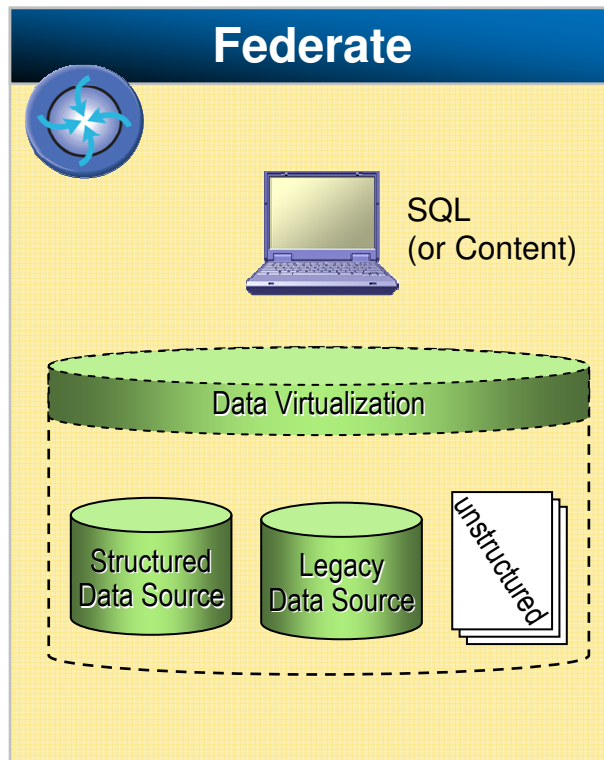 - ▸ Near real time - replication
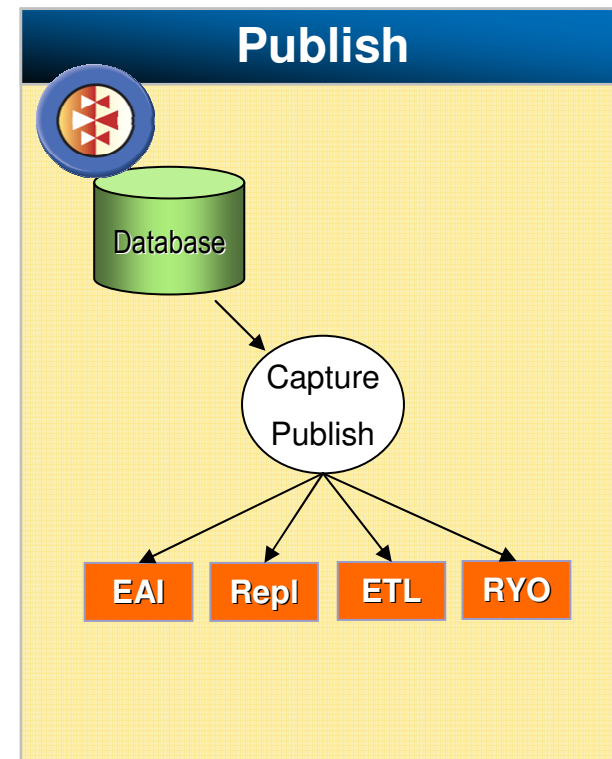 - ▸ Scheduled - ETL and consolidation

Data delivery models behind Information Integration

Real-time Pull

Federate

SQL (or Content)

Data Virtualization

Structured Data Source

Legacy Data Source

unstructured

Real-time information access

➤ Federate data from multiple sources
➤ Dynamic drill down
➤ Structured, semi-structured & unstructured

Real-time Push

Publish

Database

Capture
Publish

| EAI | Repl | ETL | RYO |

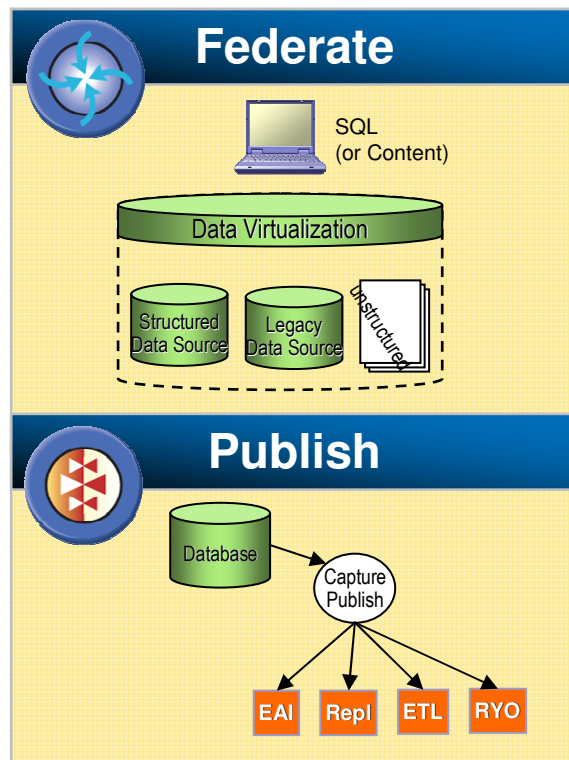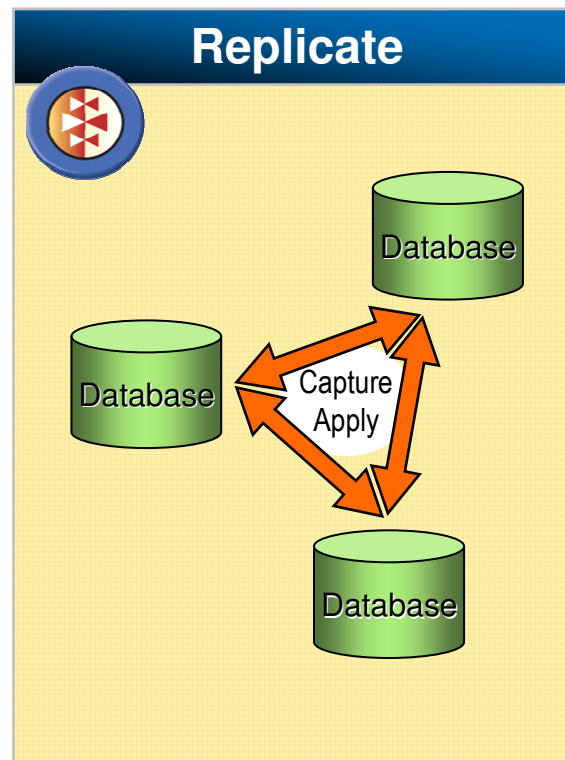**Message-based publishing**

➤ Data event (i.e. change) capture
➤ Push database changes/events via messaging
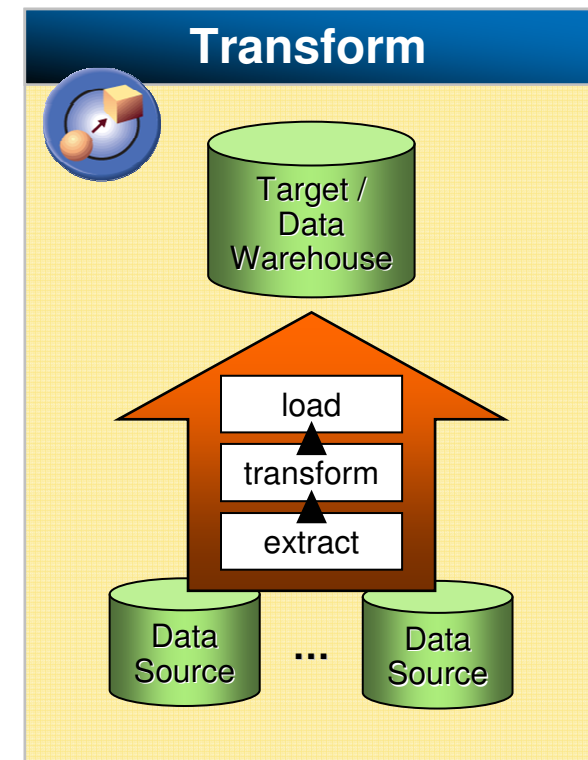➤ Leverage XML for standardized output

Data delivery models behind Information Integration

Real-time Pull/Push

Federate

SQL (or Content)

Data Virtualization

Structured Data Source

Legacy Data Source

Unstructured

Publish

Database → Capture Publish

EAI | Repl | ETL | RYO
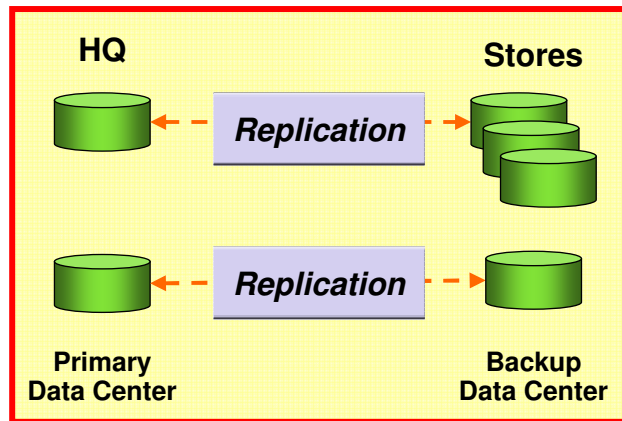
Near Real-time
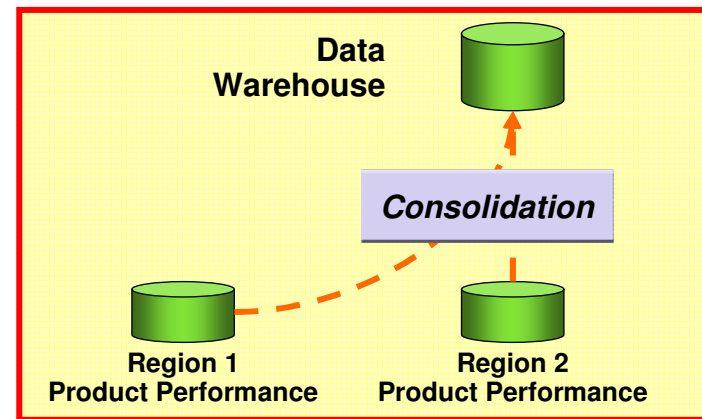
Replicate

Database

Database ↔ Capture Apply ↔ Database

➢ Distribution, consolidation or synchronization between "similar" databases
➢ Transactional integrity
➢ Multidirectional

Scheduled

Transform

Target / Data Warehouse

load

transform

extract

Data Source ... Data Source

➢ Bulk data integration
➢ Set-based & hierarchical transforms
➢ High scale, batch-oriented delivery
➢ Unidirectional

IBM

UNIVERSITÉ DU MAINFRAME

Replication

HQ — Stores

Replication

Replication

Primary Data Center — Backup Data Center

ETL Consolidation

Data Warehouse

Consolidation

Region 1 Product Performance — Region 2 Product Performance

Characteristics:

➤ Low latency

➤ Small amount of transformation (or none)

➤ One or many targets

➤ Supports multidirectional data movement
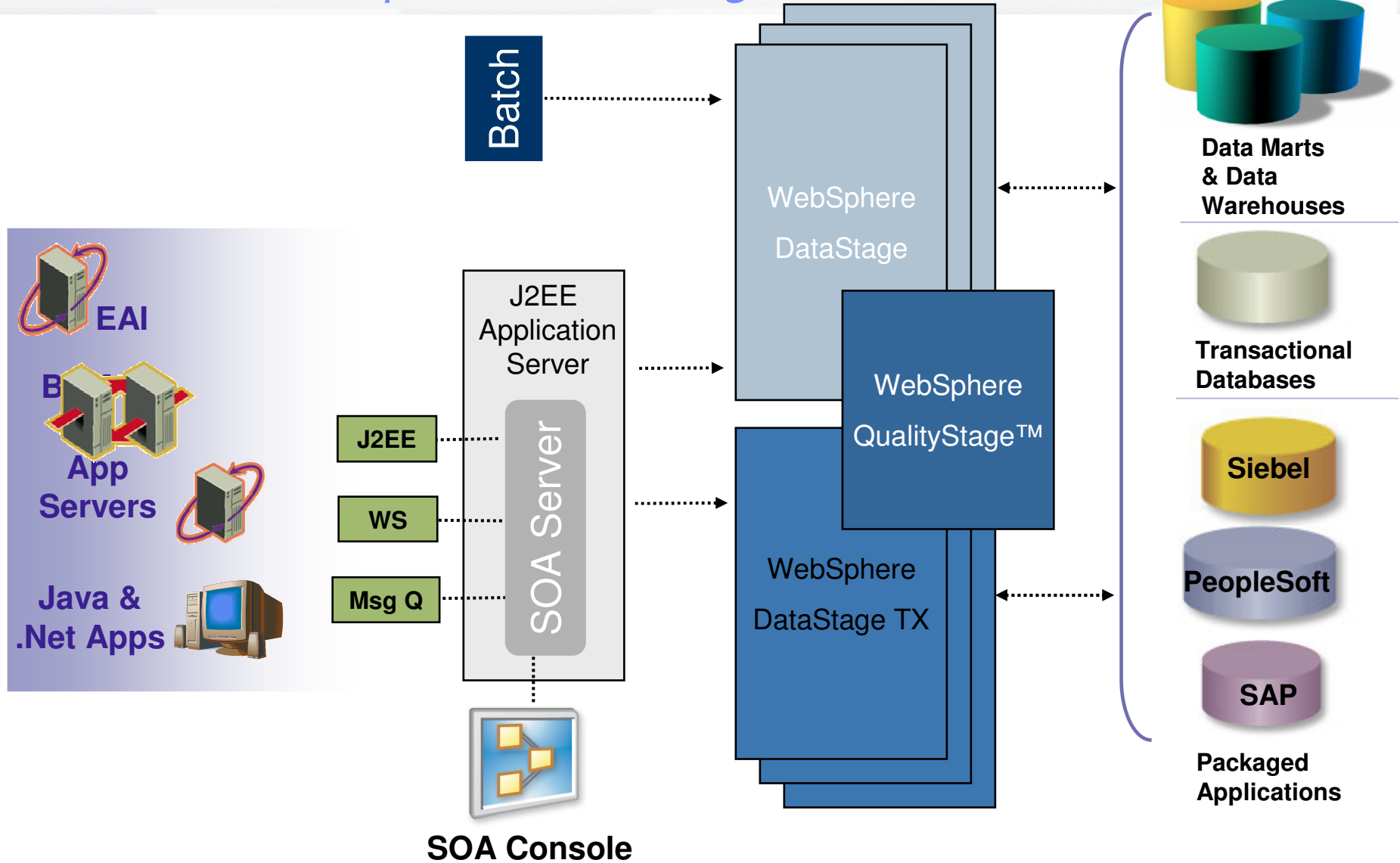 • Provides Conflict Detection and Resolution

Characteristics:

➤ Typically batch type processing

➤ Supports all transformations
  • Versioning
  • Aggregations
  • Cleansing

➤ Typically supporting small number of target databases

➤ Unidirectional data movement

Batch

WebSphere
DataStage

J2EE
Application
Server

SOA Server

J2EE

WS

Msg Q

WebSphere
QualityStage™

WebSphere
DataStage TX

EAI

App
Servers

Java &
.Net Apps

SOA Console

Data Marts & Data Warehouses

Transactional Databases

Siebel

PeopleSoft

SAP

Packaged Applications

Other services   Other services

ESB

WebSphere.
Process Server

Web Services   **Java**   **Message Queues**   **SQL**   **SCA/SDO**   **BPEL**

IBM Information Server

Data   **Content**   **Warehouse**   **Transform**   **Cleanse**   **Move**   **Master**   **Analyze** . . . .

Summary
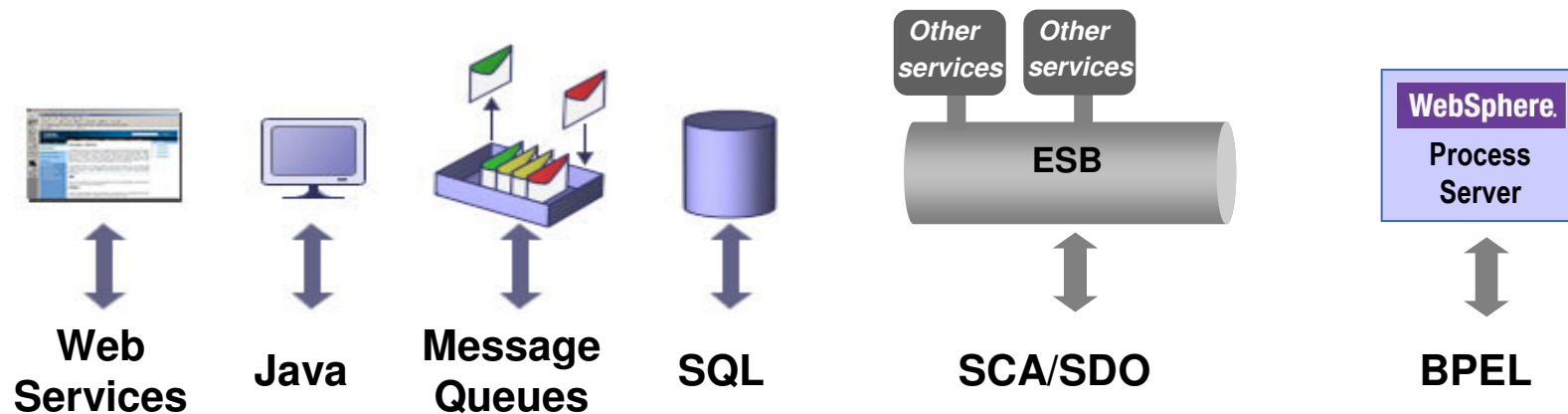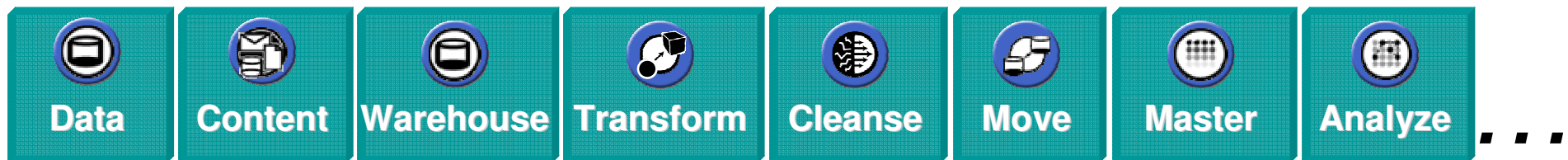
- Information integration is a foundation for companies to build an On Demand Operating Environment enabling them to align their IT infrastructure to business priorities

- WebSphere Information Integration Solutions provide access to diverse, distributed, and real-time data as if it were a single source, no matter where it resides.

- WebSphere Information Integration Solutions will help businesses
  - Optimize IT investments given more choice in data access
  - Improve productivity and application efficiency
  - Enable greater return on existing assets

- Rely on IBM's proven technology and support for open standards

Online Resources

- WebSphere Information Integration customer web site on WWW:
  - ▶ http://www.ibm.com/software/data/integration/

- Demos
  - ▶ http://db2premier.dfw.ibm.com/wps/portal/

- White Papers and Brochures
  - ▶ http://www-306.ibm.com/software/data/integration/library.html

IBM

UNIVERSITÉ DU
MAINFRAME

Questions - Réponses

MERCI