IBM Software Group

IBM WebSphere Information Integration

" IBM Information Integration offering on z/OS"

WebSphere. Information Integration Software

ON DEMAND BUSINESS

Eric Derbanne

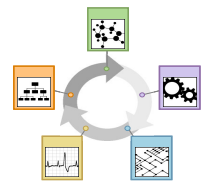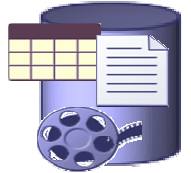**IBM France Software Group**

eric.derbanne@fr.ibm.com

© 2005 IBM Corporation

Defining Business Integration

Integration

**The efficient and flexible combination of resources
to optimize operations across and beyond the enterprise**

People

➢ Portal – personalized information
➢ Collaboration technology
➢ Adaptable workplace
➢ Consistent rule-based experience across devices

Processes

➢ Workflow management
➢ B2B connectivity
➢ Messaging infrastructure (EAI)
➢ e-Business transactions
➢ Service-oriented architecture

Information

➢ Federation & data placement (replication, ETL, caching)
➢ XML (store, query, webservices)
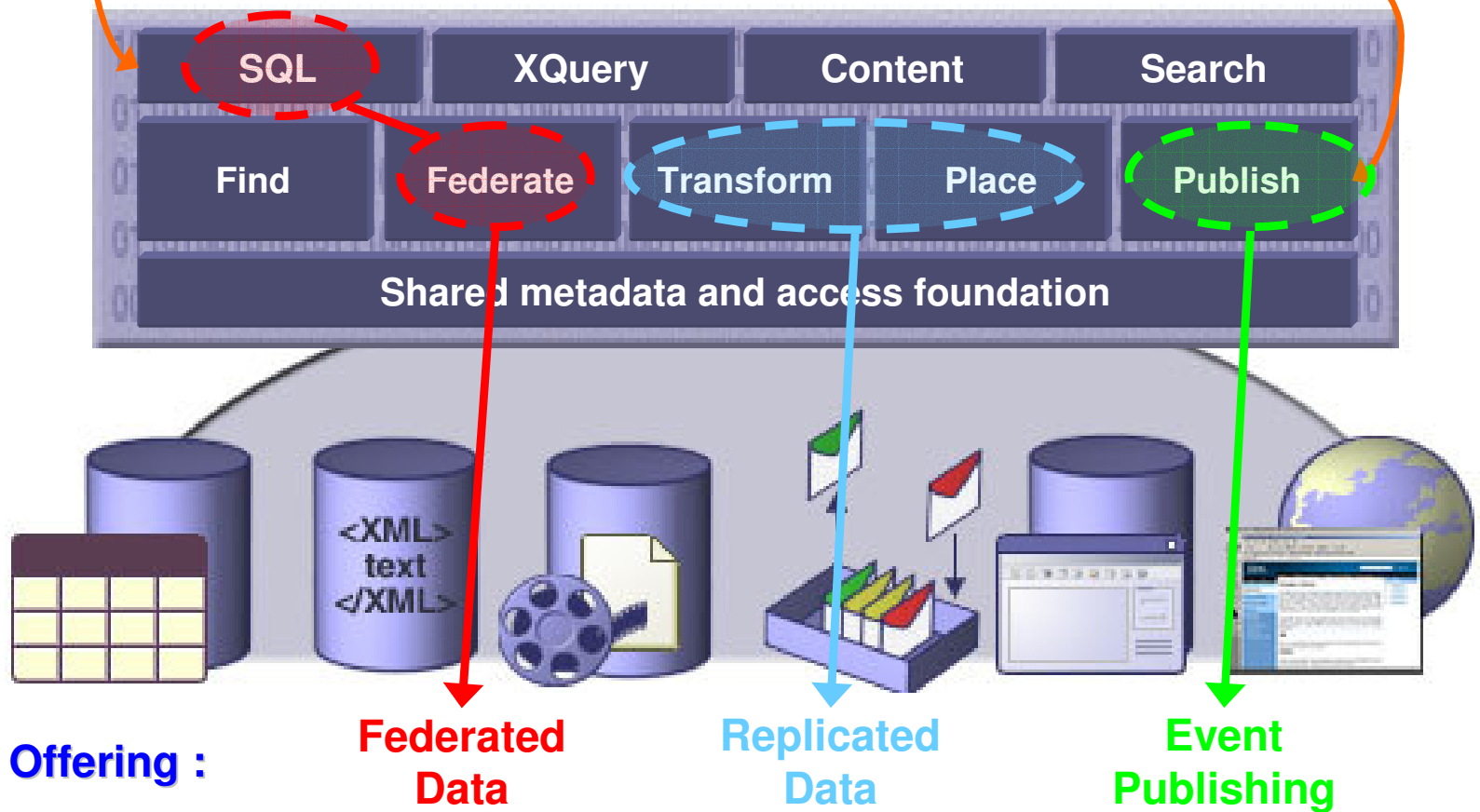➢ Meta-data management
➢ Text Search and analytics

40% of people's time is spent searching for relevant information	*40% of IT budgets may be spent on integration*	*30-50% of design time is copy management*	*85% of information is unstructured*

☹ *for each $1 spent for a packaged application, customers spend $5 to $9 on the labor for integration !*
(IBM Customer Surveys, 2001, 2002)

Can I create business value from my existing IT systems? ➡ *People, Process,* **Information**

Can users react in real-time to the most recent information? ➡ *People,* **Information**

Are business operations fully integrated for optimal efficiency? ➡ *Process,* **Information**

IBM Information Integration Vision

Any Data

➢ *Multiple access paradigms*

➢ *Multiple integration disciplines*

SQL	XQuery	Content	Search

Find	Federate	Transform	Place	Publish

Shared metadata and access foundation

zSeries II Offering :

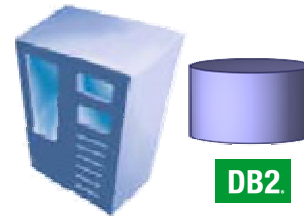**Federated Data**

Replicated Data

Event Publishing

<XML> text </XML>

Using Websphere Information Integrator as GRID Enabler

- Leverage IT Infrastructure

- Accessing Information regardless of Database Management System and Structure

Federation instead of Centralization

WebSphere Information Integrator

Information Virtualization

File System / Storage Virtualization
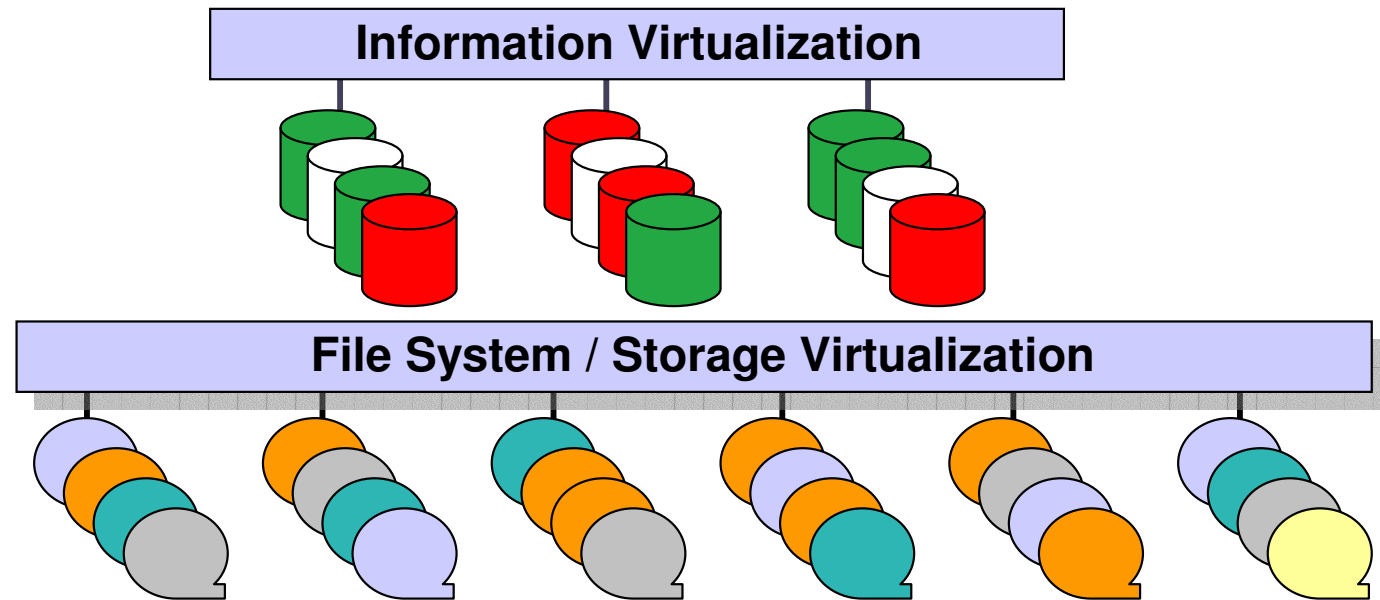
Federated Data Server

➤ *WebSphere Information Integrator Classic Federation – z/OS Platforms*

➤ *WebSphere Information Integrator - LUW Platforms*

WebSphere Information Integrator Classic Federation for z/OS

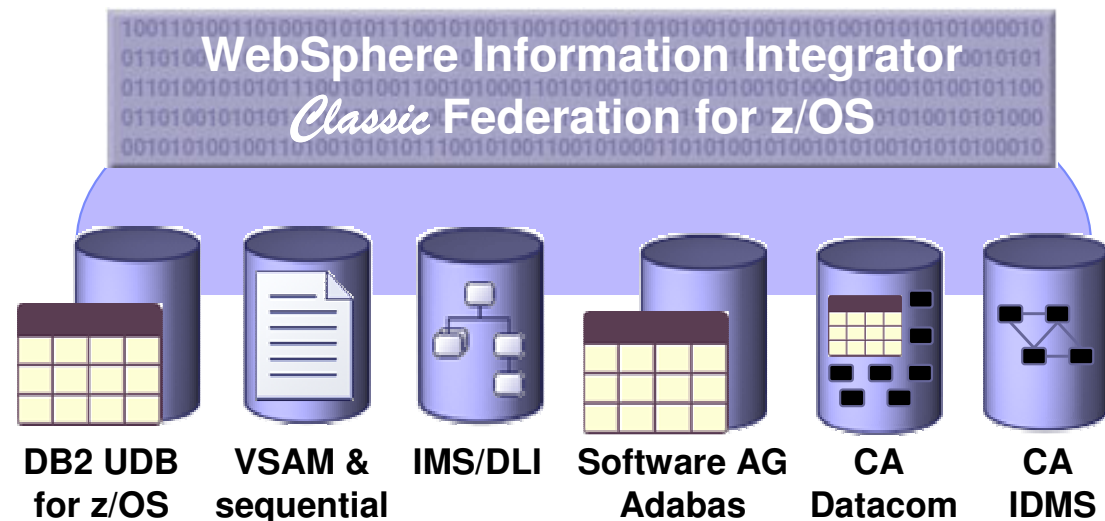- **Typical Large IT Enterprise Environment**

  ▶ 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, JDBC, CLI

  ▶ 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 |

WdebSphere Information Integratror *Classic* Federation

AIX, HP-UX, Solaris, JVM 1.2, Widows NT, 2000, XP, z/OS, OS/390

| Portal | BI Tool | Servlet | Servlet | Client class |

ODBC Client

EJB **JDBC Client**

z/OS OS/390

Metadata Catalog

USE Grammar
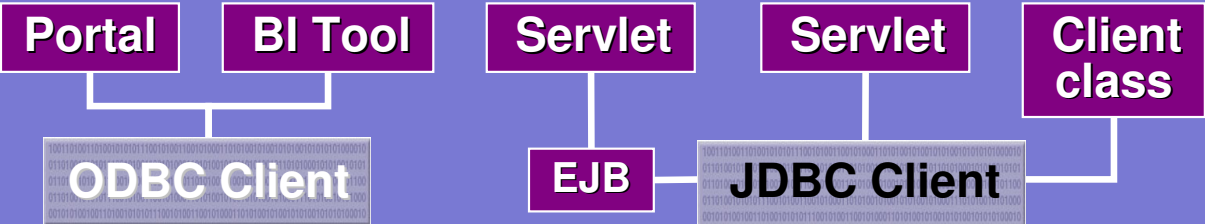
WebSphere Information Integration *Classic* Federation Integration Server

Copybooks, DBDs, …

| Data Connector | Data Connector | Data Connector | Data Connector | Data Connector | Data Connector |

| DB2 UDB z/OS | VSAM | IMS/DLI | Adabas | CA-Datacom | CA-IDMS |

DataMapper
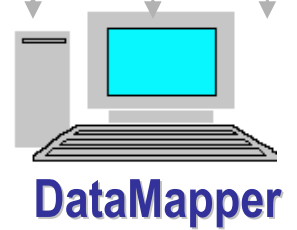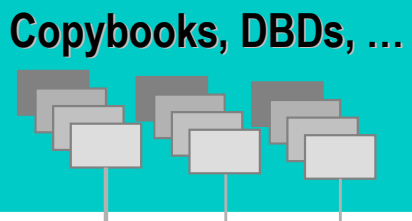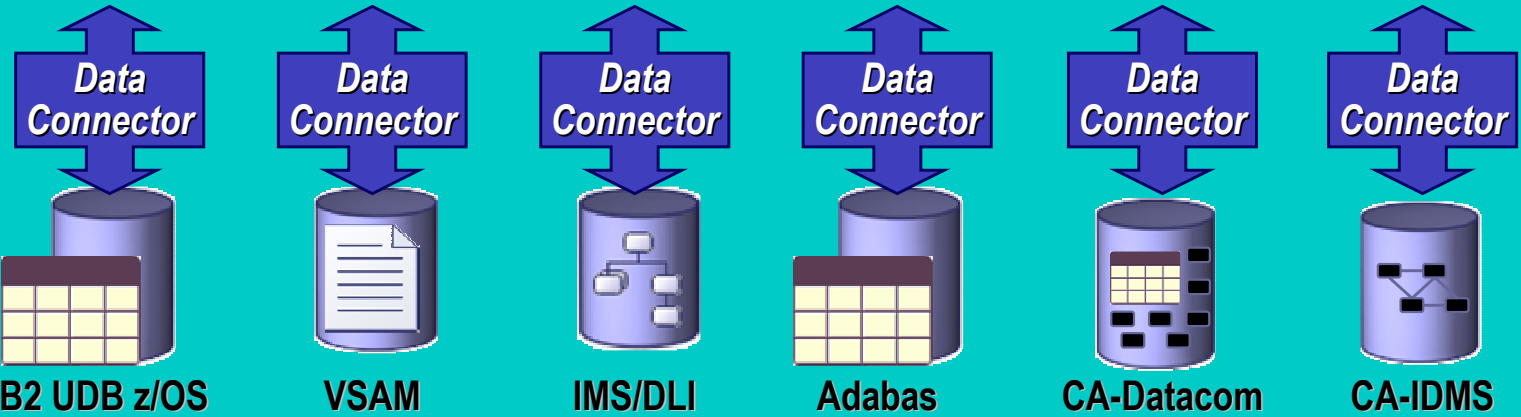
WebSphere and II Classic Federation

WebSphere Application Server

WebSphere Portal Server

| JDBC Portlet | SQL Portlet | Custom Portlet |

Classic **Federation JDBC Client**

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

DB2 UDB for z/OS | VSAM & sequential | IMS/DLI | Software AG Adabas | CA Datacom | CA IDMS
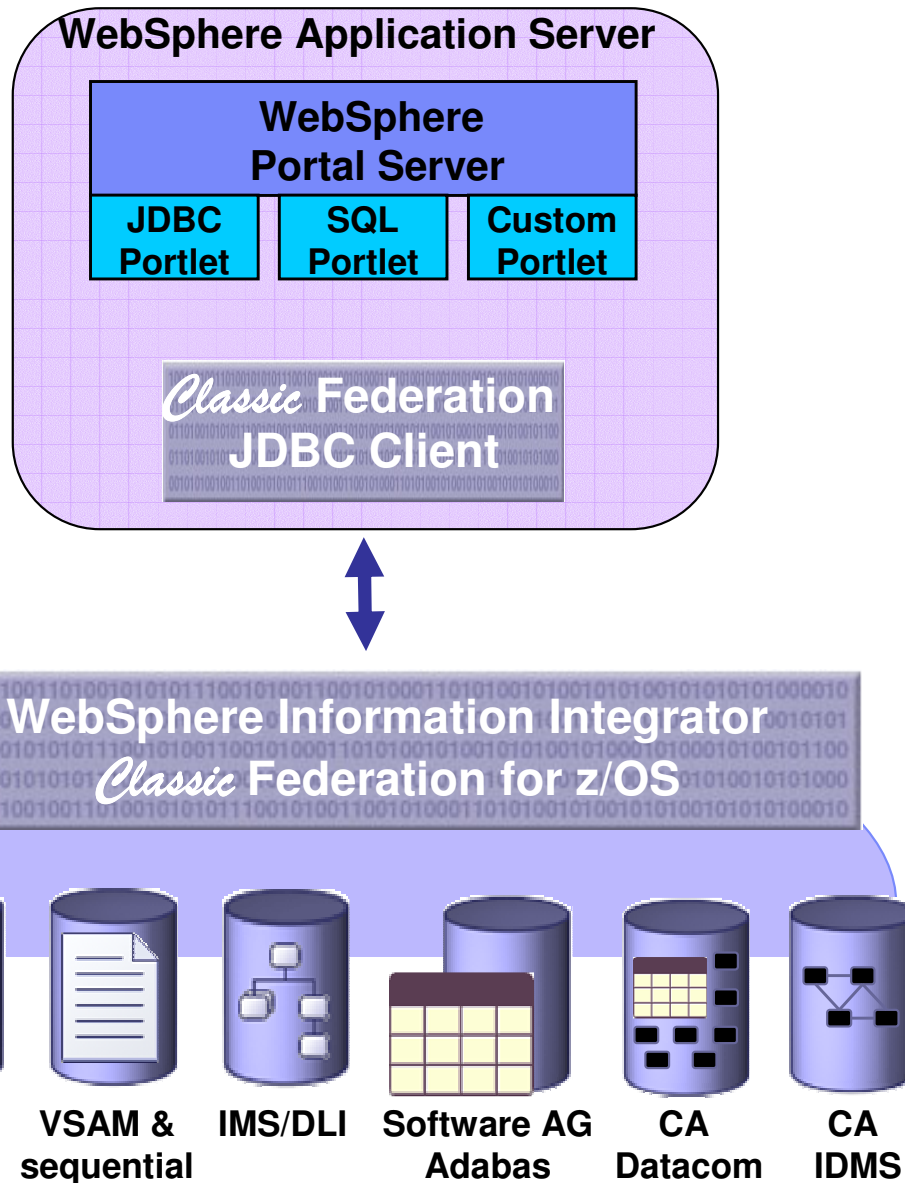
- Integrate mainframe data directly with Web applications, Portals, EAI platforms

- Supports JDBC 2.1 Core APIs Plus
  - Scrollable/Updatable ResultSets
  - Commit – Rollback – AutoCommit - XA
  - Stored Procedure to mainframe programs
  - Parameter Markers
  - Metadata commands for tables, columns, keys, procedures, …
  - Connection Pooling Support via Relational Resource Adapter (RRA)
  - SQLBatch Operations
  - Statement commands e.g. re-execution of prepared statements

Integration in Action – European Catalog House

Seamlessly share order processing data and algorithms between:

- ➢ Legacy call-center systems
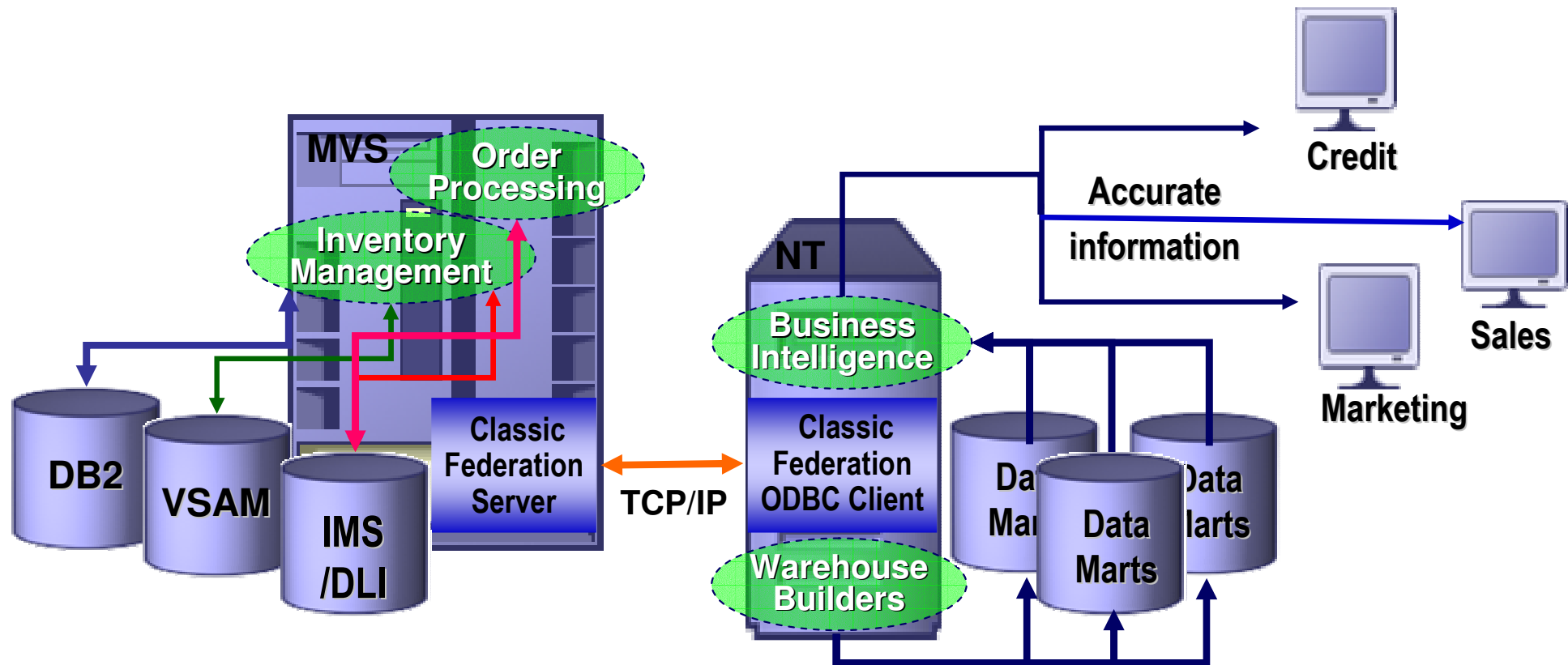- ➢ New e-commerce applications
- ➢ No mainframe skills required for e-commerce site development
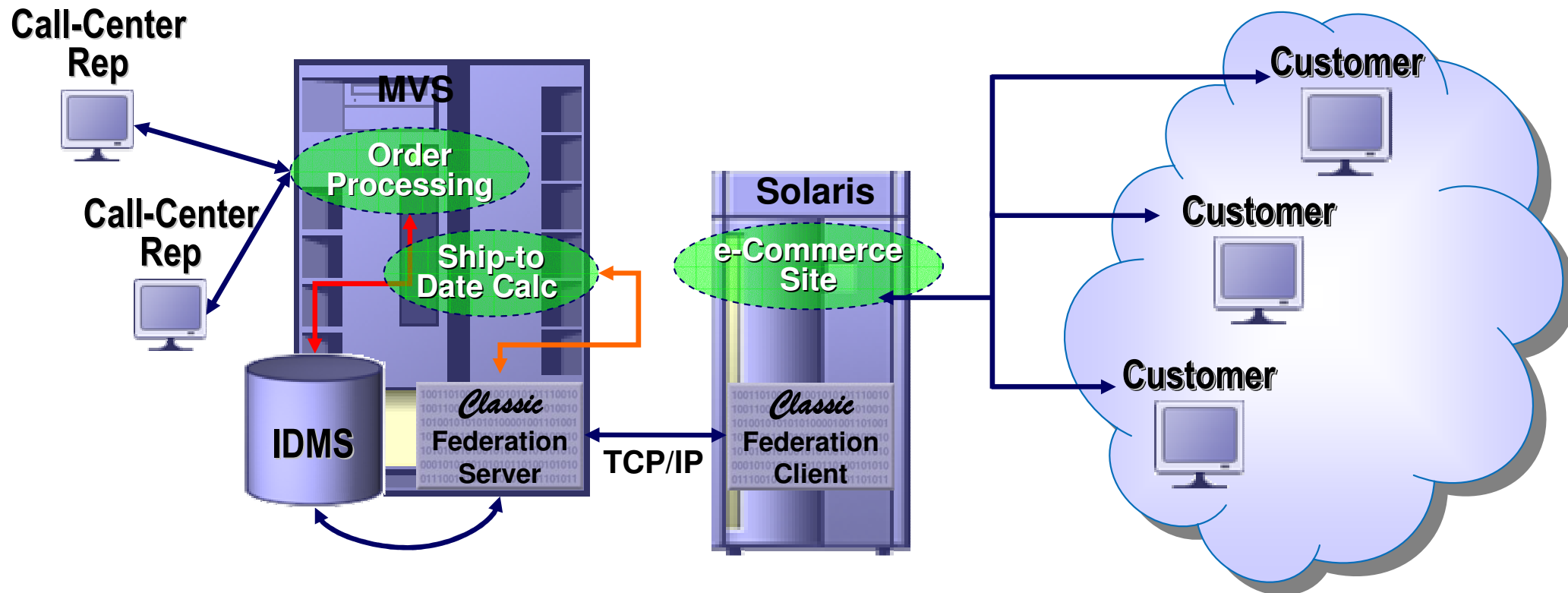
Integration in Action – Recreation Vehicle Manufacturer

Provide accurate inventory and sales data to data warehouse

➢ Dynamically connect data warehouse tool with mainframe data

➢ Cut development time in half

➢ Accelerate product delivery with warehouse "pull" of new shipment data

Integrating Enterprise Data

z/OS
- DB2 UDB
- VSAM
- IMS/DLI
- SAG Adabas
- CA Datacom
- CA IDMS

Packaged apps

DB2 II *Classic* Federation

Integrated SQL View

DB2 Information Integrator

O
D
B
C

SQL, SQL/XML

Federation Engine

Wrappers and functions

New

Biological Data And Algorithms

Text

<XML> text </XML>

XML

Excel

WebSphere MQ

IBM Extended Search

WWW, email,…

DB2 Family
Sybase
Informix
SQL Server
Oracle
Teradata
ODBC
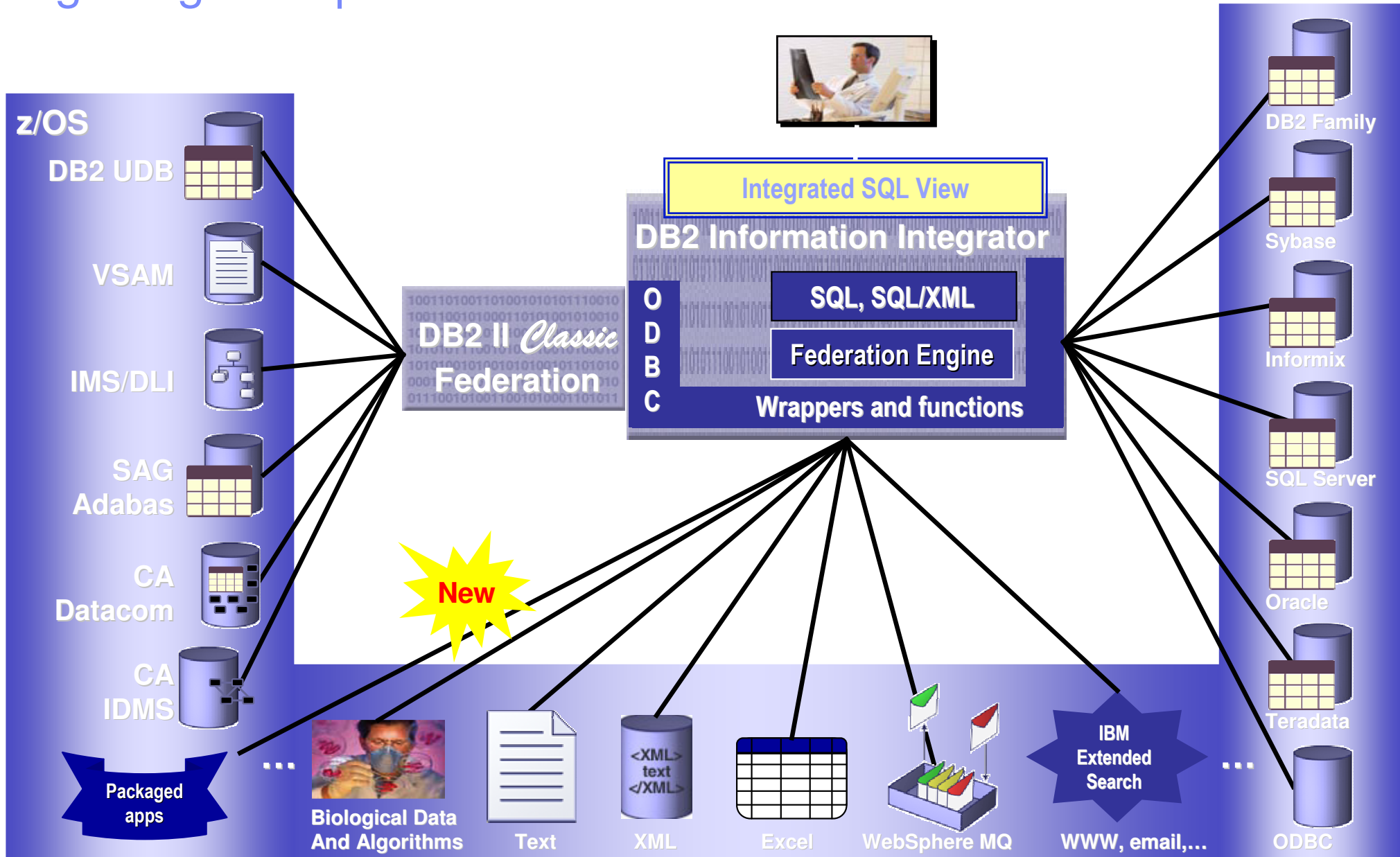
Access Distributed Heterogeneous Data from CICS Applications
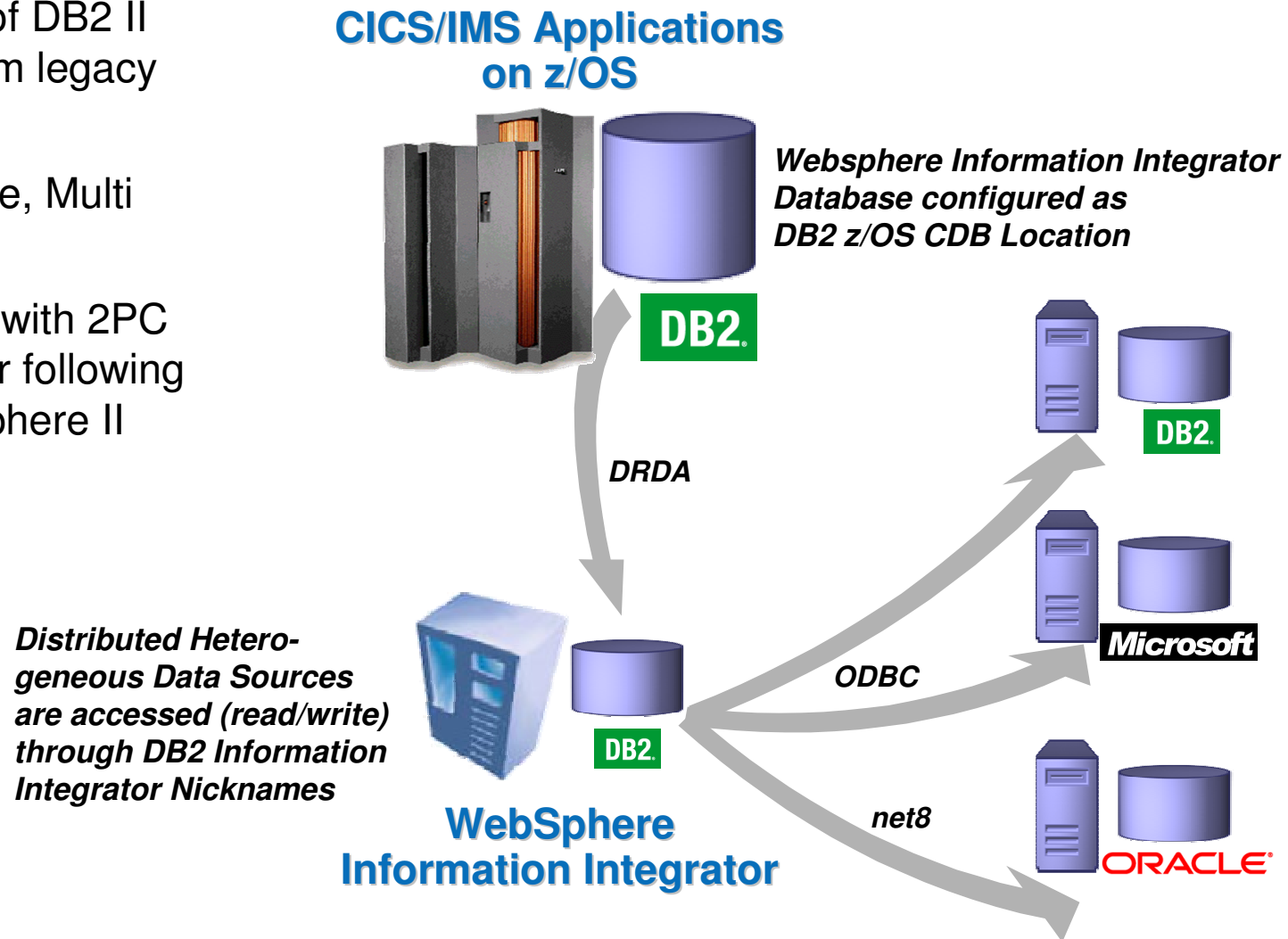
Business Scenario

➤ 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.

Microsoft

ODBC

ORACLE

net8

Information and Process Integration with WebShpere

Business Scenario

➢ Enterprise Applications provide APIs for Business Object/Component Retrieval

➢ Enterprise Business Components can be mapped into relational Format using WebSphere components :
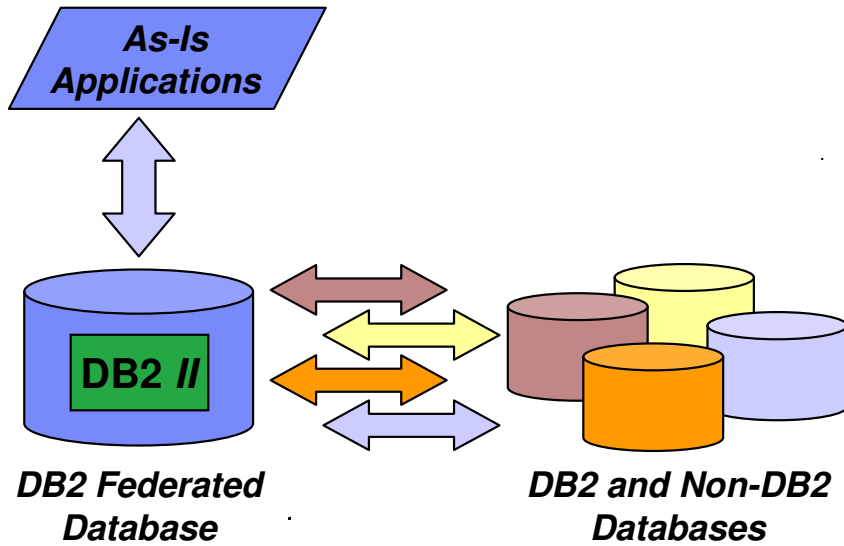 - ✓ Information Integration
  - ✓ Business Adapters

➢ Business Objects can be joined with other relational / non-relational Information

SQL, SQL/XML

Model
Manage   Transform
Interact
Integrate
Accelerate

WebSphere Information Integrator

WebSphere Business Integrator

Wrappers for Integrated Business Objects

Application Access Services

SAP   SIEBEL   PEOPLE Soft

Enterprise Applications
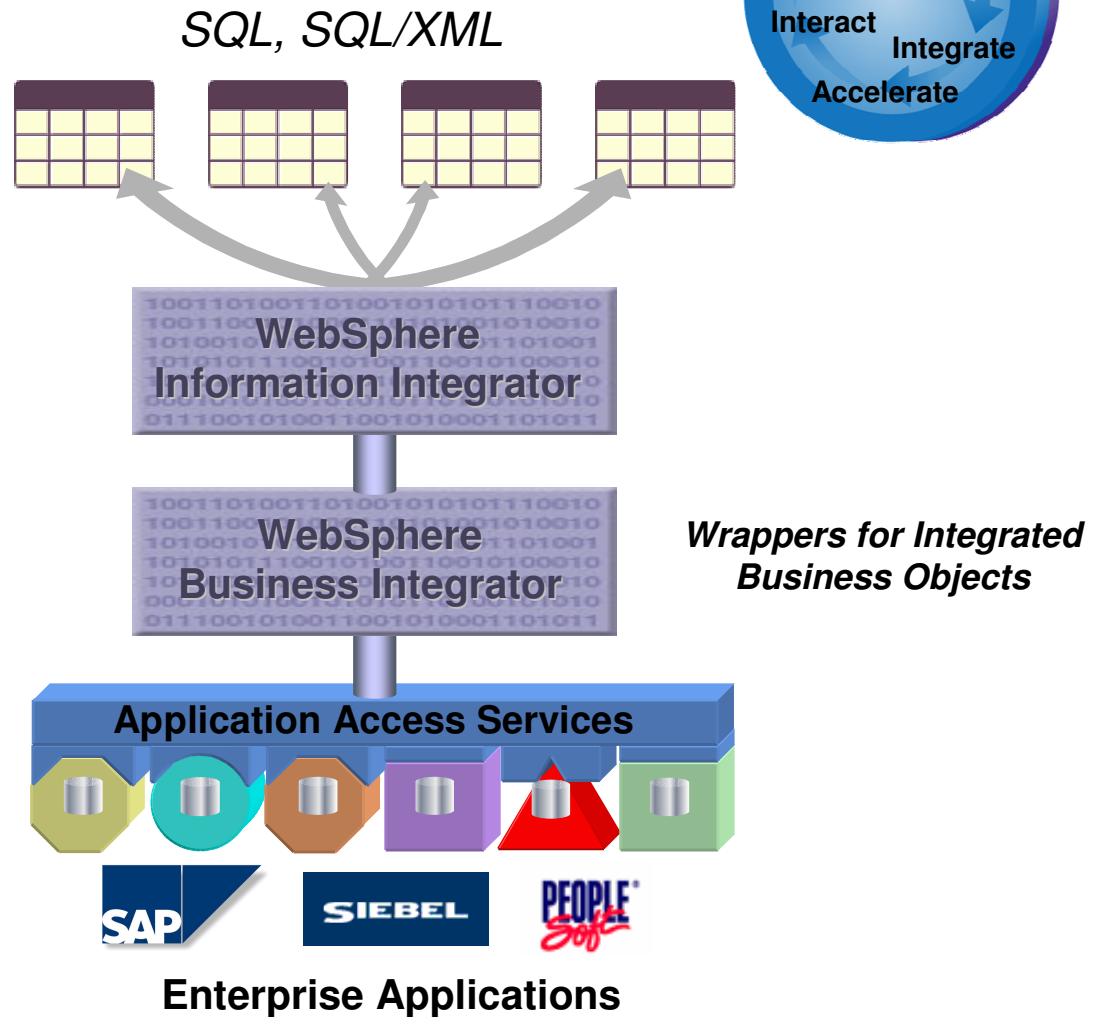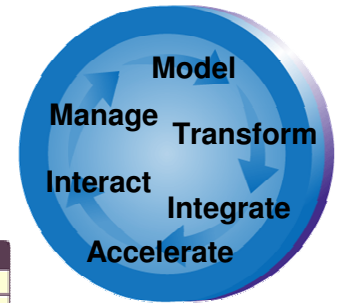
Federate or Replicate – That is the Question !!

As-Is Applications

DB2 *II*

DB2 Federated Database

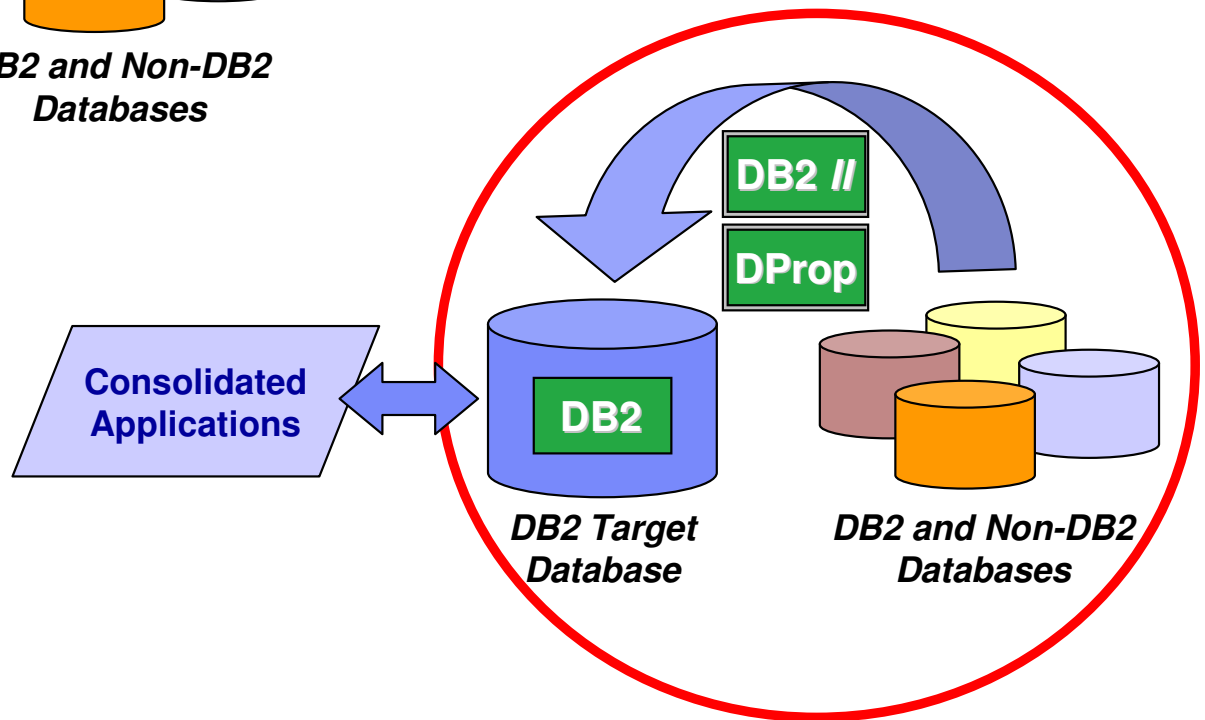*DB2 and Non-DB2 Databases*

Data Federation
➢ SQL Transparency
➢ Real-Time Access
➢ Global Optimization
➢ SQL Pushdown

Data Replication
➢ Data Movement
➢ Data Transformation
➢ Autonomy

Consolidated Applications

DB2 *II*

DProp

DB2

DB2 Target Database

DB2 and Non-DB2 Databases

Replicated Data

➢ *SQL Replication*

➢ *Q Replication*

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 ?

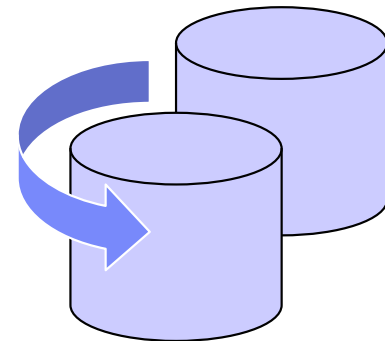- **Warehouse & Business Intelligence**
  - ▸ Move data to new platform/database, transform data
  - ▸ ETL or Replicate?
    - • latency needs
    - • change volume versus total volume
    - • complexity of transformation and/or cleansing

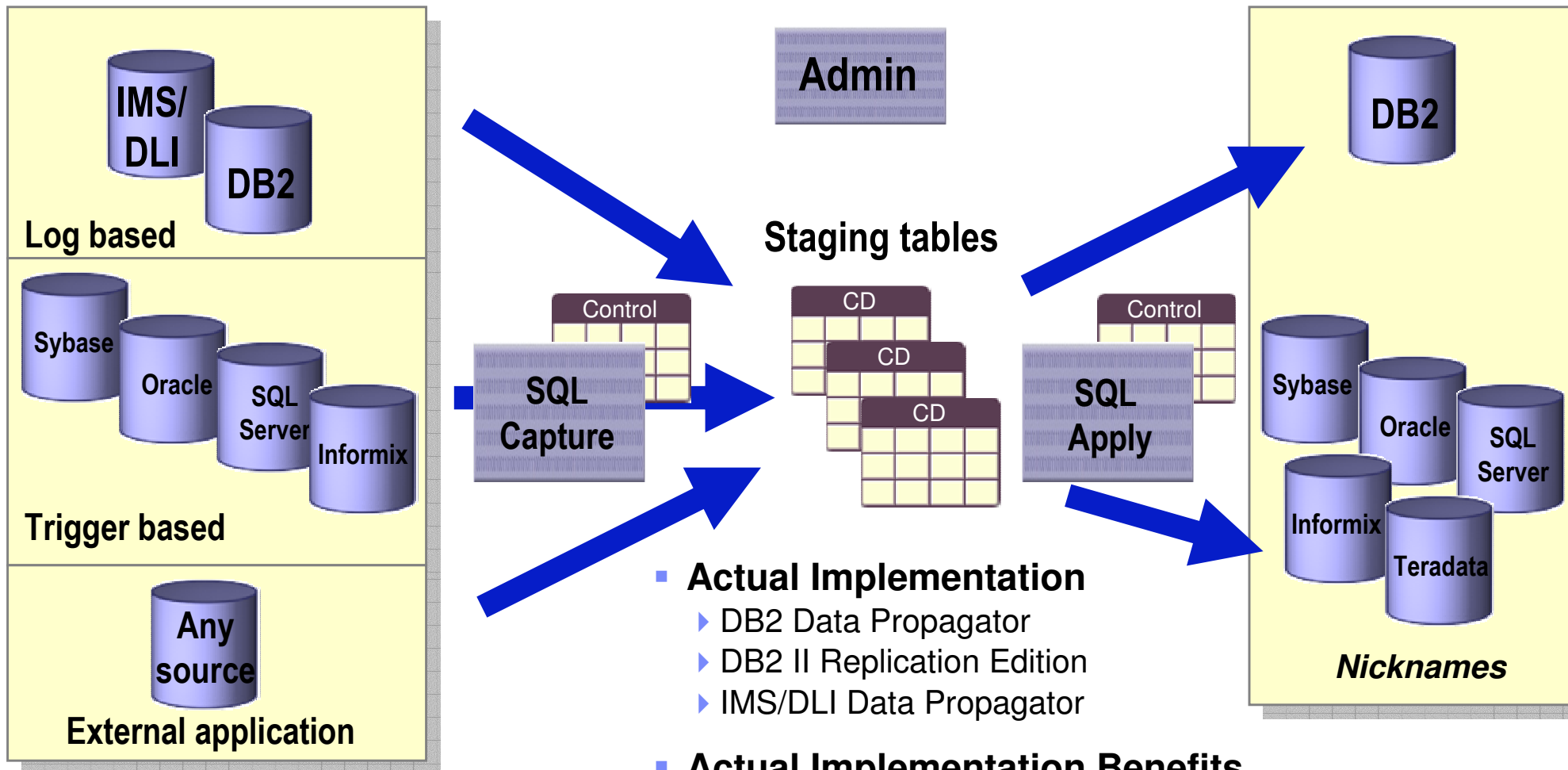- **Mobile Workforce**
  - ▸ Occasionally connected distribution/consolidation

- **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 or reporting work to a separate system
    - • other methods such as flash copy also possible
  - ▸ Peer to peer - split workload
    - • only possible through replication
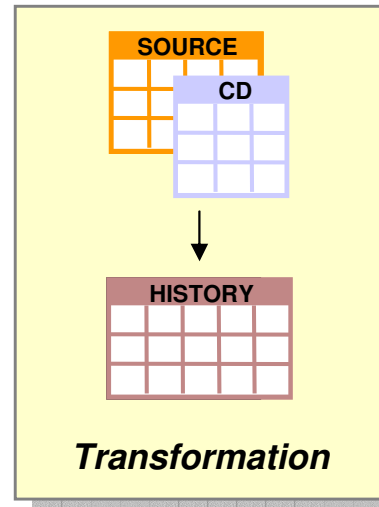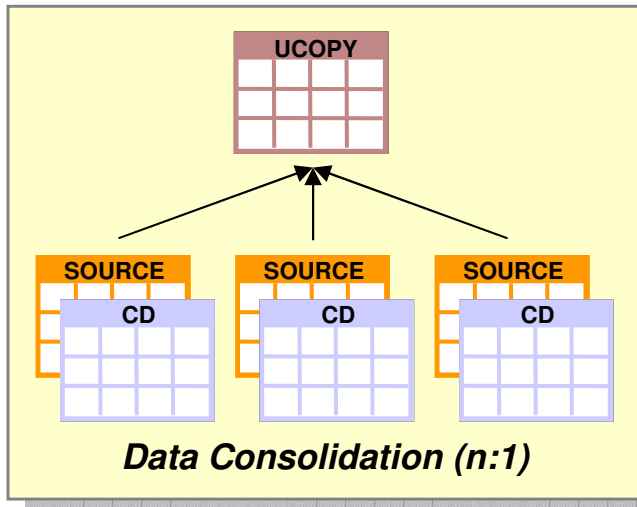   - • requires serious planning and consideration

SQL Replication Architecture

Log based

IMS/DLI

DB2

Trigger based

Sybase

Oracle

SQL Server

Informix

Any source

External application

Admin

SQL Capture

Control

Staging tables

CD
CD
CD

SQL Apply

Control

DB2

Sybase

Oracle

SQL Server

Informix

Teradata

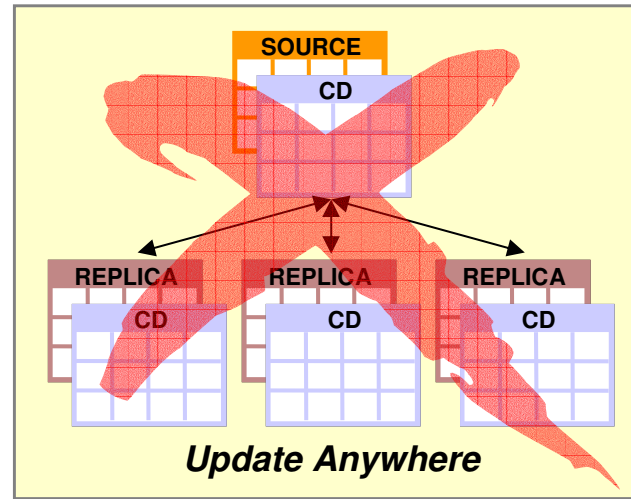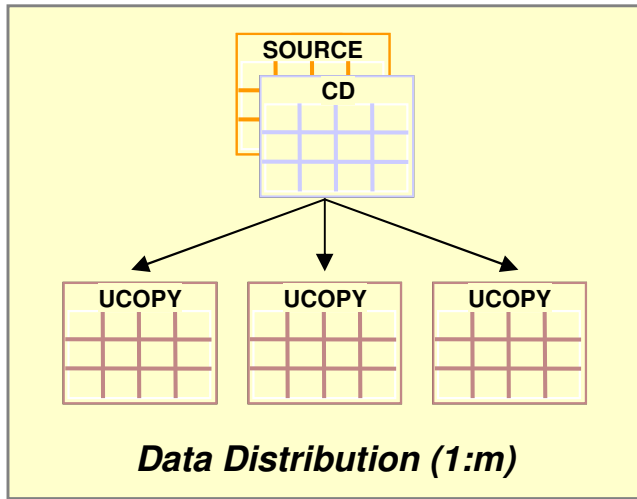*Nicknames*

- **Actual Implementation**
  - ▸ DB2 Data Propagator
  - ▸ DB2 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

Sample SQL Replication Scenarios

SOURCE
CD
UCOPY **UCOPY** **UCOPY**

Data Distribution (1:m)

SOURCE
CD
REPLICA **REPLICA** **REPLICA**
CD **CD** **CD**

Update Anywhere

SOURCE
CD
STAGING **STAGING**
UCOPY **UCOPY**
UCOPY **UCOPY**

Multi-Tier Staging

UCOPY
SOURCE **SOURCE** **SOURCE**
CD **CD** **CD**

Data Consolidation (n:1)

SOURCE
CD
HISTORY
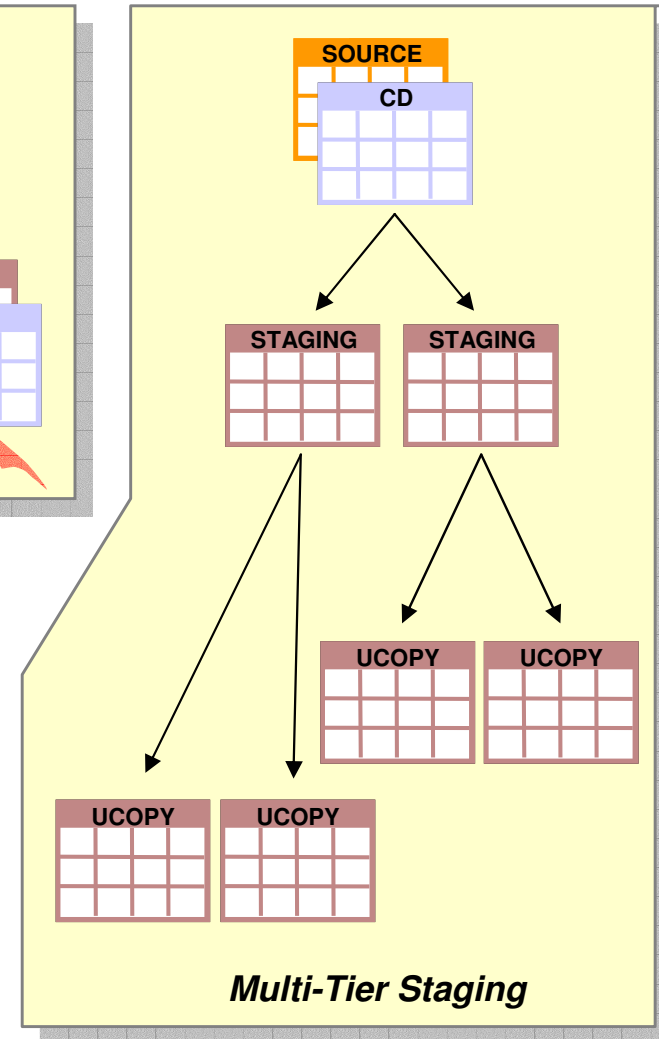
Transformation

- Subsets
- SQL Transformations
- Updateable Predicates
- Updateable Primary Keys

DB2 Data Replication TO / FROM Federated SOURCES

SOURCE

SOURCE1 SOURCE2

CAPTURE CONTROL

CONTROL UOW

Triggers

DProp Capture

CD1 CD2

NICKNAME

SOURCE-NICK

CNTL-NICK

CD-NICK

APPLY CONTROL

SPILL1 SPILL2

CONTROL

DProp Apply

ADMINISTRATION

Replication Center

Replication Monitor

(currently heterogeneous replication with SQL Replication only)

NICKNAME

COPY-NICK

TARGET COPY-NICK
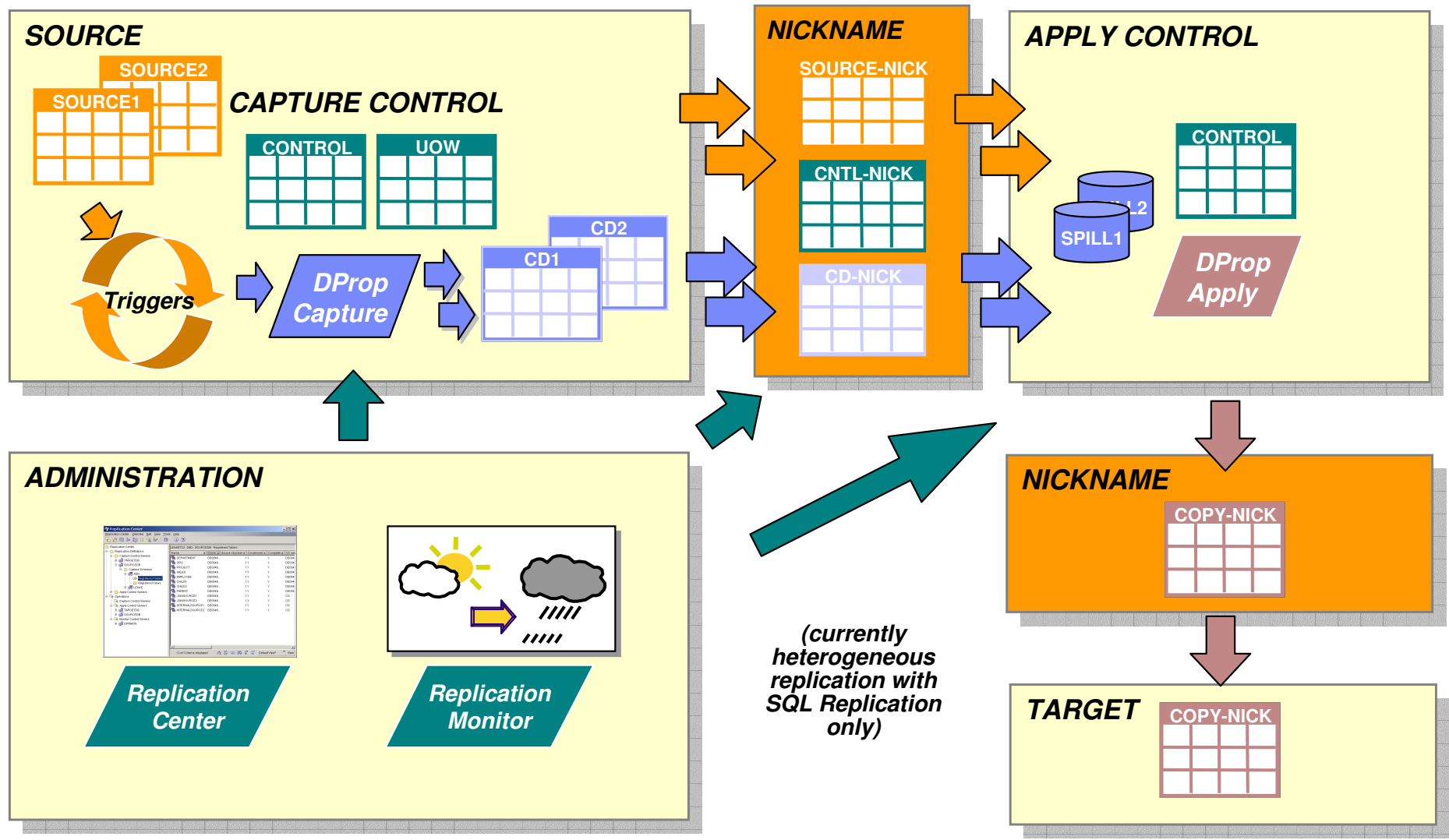
Federated Sources & Targets

➢**DB2** ➢**Informix**
➢**Oracle** ➢**Sybase**
➢**SQL/Server** ➢**Teradata (Target only)**

Why Create Another Replication Architecture?

- **Performance**
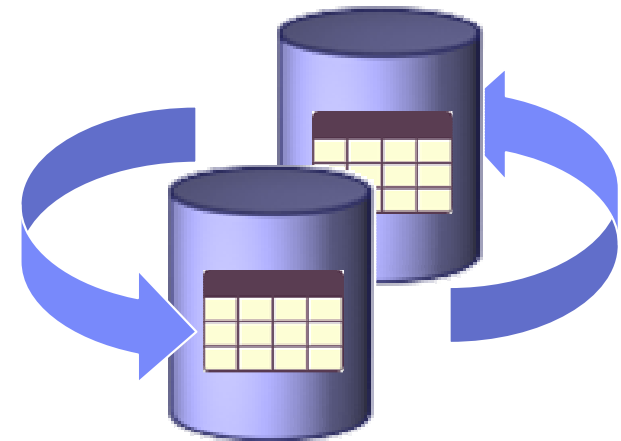 - ▶ Combine high Throughput with low Latency

- **New Function**
  - ▶ Event Publishing from DB2 and Classic Sources

- **Capability**
  - ▶ Significantly improve multi-directional Replication Support
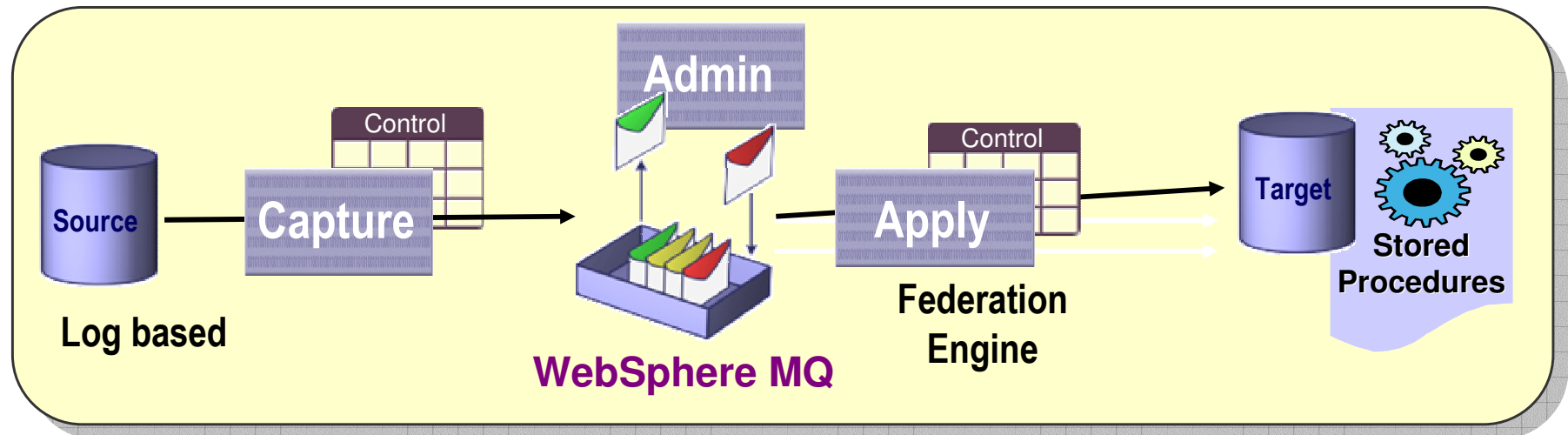
- **Manageability**
  - ▶ Reduce the number of Replication Objects to be defined and managed

  - ▶ Ease the Definition Process with new Replication Center Wizards

Q Replication Architecture

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

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
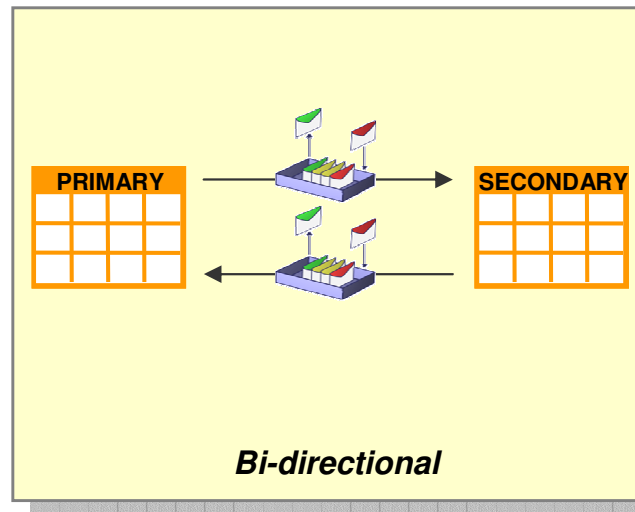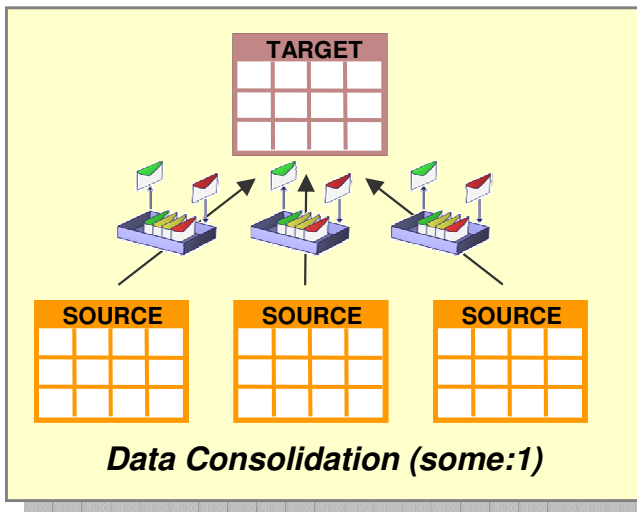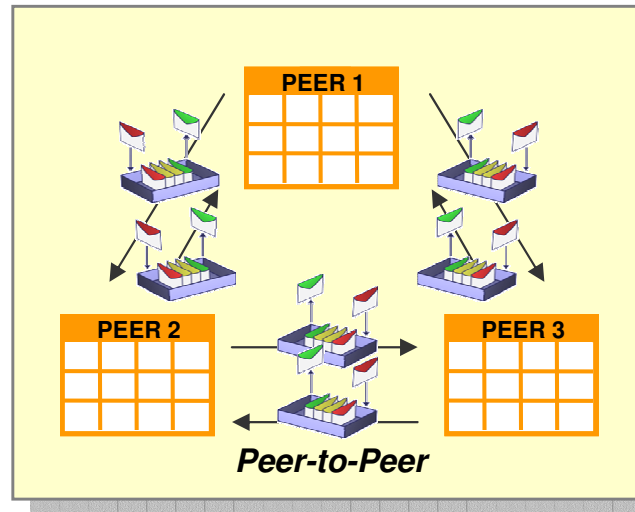
- **Predicate examples**
  - ▸ Based on values in the row data itself
    ```
    WHERE :LOCATION ='EAST' AND :SALES  > 100000
    ```

  - ▸ Based on values in data in other tables
    ```
    WHERE :LOCATION ='EAST' AND :SALES > (SELECT SUM(expense)
    FROM STORES WHERE stores.deptno = :DEPTNO)
    ```

Sample Q-Replication Scenarios

Data Distribution (1:some)

Peer-to-Peer

Data Consolidation (some:1)
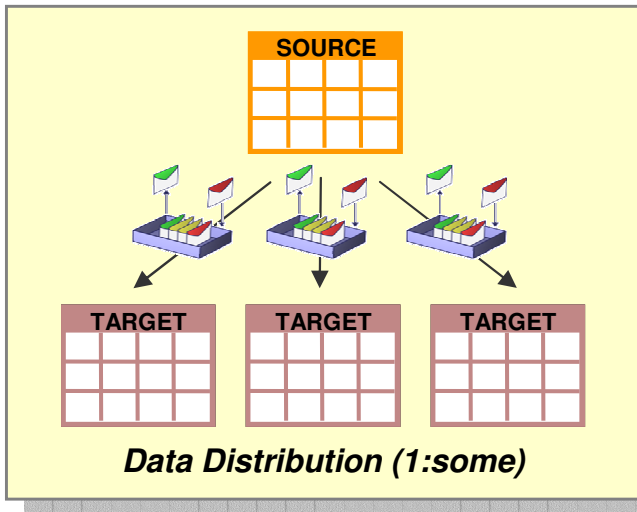
Bi-directional

Key Scenarios:

➢ Low-Latency Replication

➢ Geographically dispersed Applications with distributed Databases

➢ Bi-directional Replication with Conflict Checking, Handling, and Notification

➢ Software-based Hot-Standby

➢ Cross DB2-Family

Feeding Trade-History Database with Q-Replication
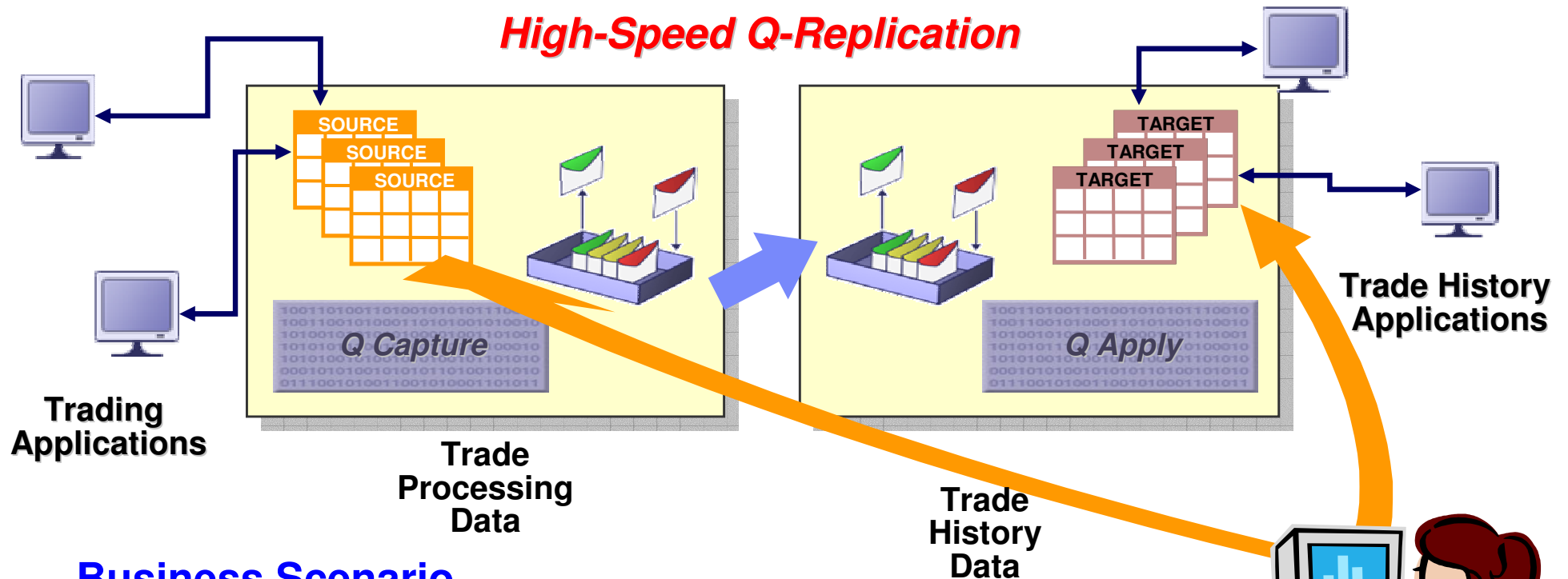
High-Speed Q-Replication

SOURCE
SOURCE
SOURCE

Q Capture

TARGET
TARGET
TARGET

Q Apply

Trade History Applications

Trading Applications

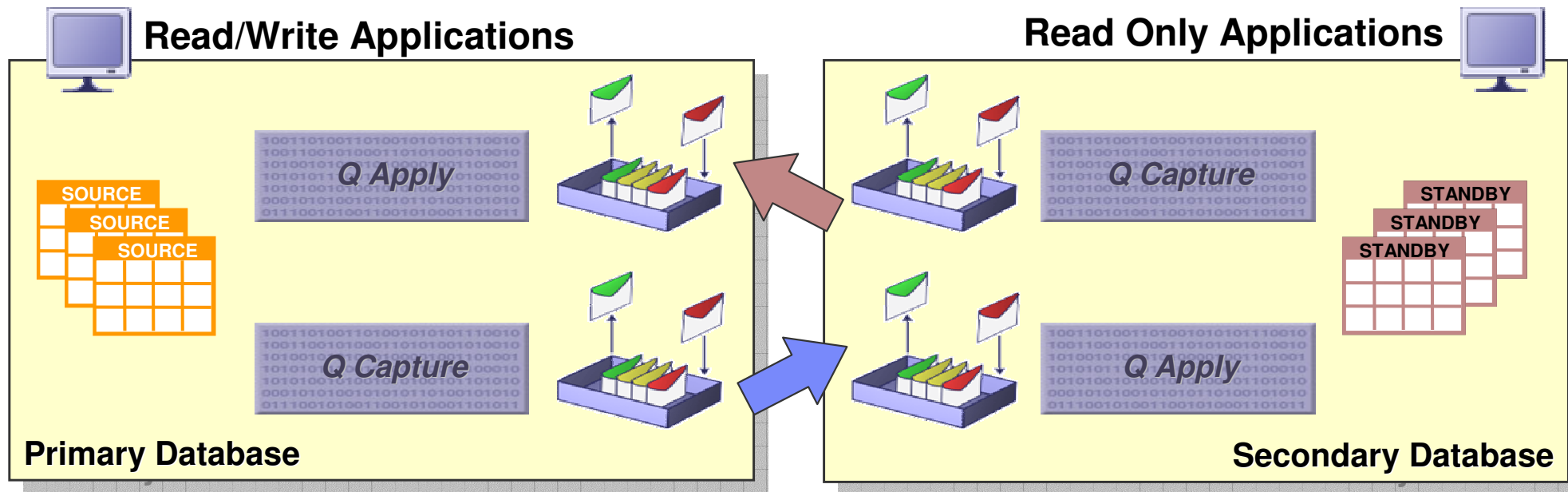**Trade Processing Data**

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

High-Availability Solution built upon Q-Replication

Read/Write Applications

Q Apply

SOURCE
SOURCE
SOURCE

Q Capture

Primary Database

Read Only Applications

Q Capture

Q Apply

STANDBY
STANDBY
STANDBY

Secondary Database

Business Scenario

➢ Replication Processes and Subscriptions are defined in both Directions, but Data mainly flows in one Direction at a Time

➢ Recursion is stopped by Capture, which reads special logged Events created by Apply

➢ Data at the Secondary System is transactionally consistent and is available for "read only" Applications permanently

➢ Procedures for Failover and Switchback will depend on which Options have been selected for Conflict Detection
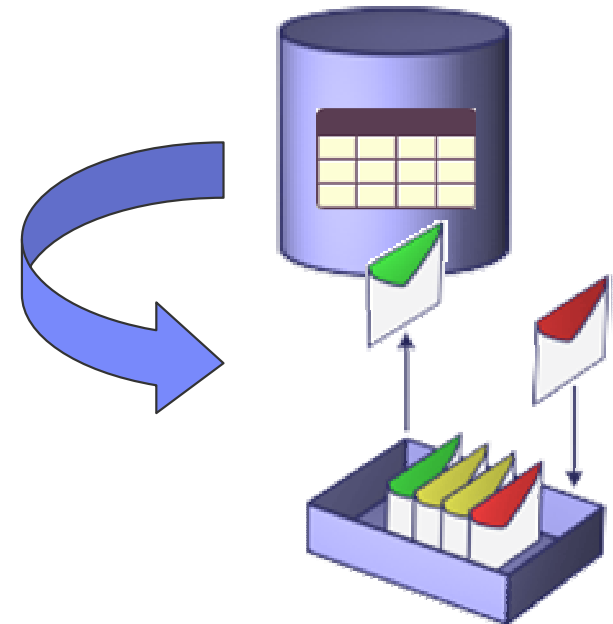
Event Publishing

Why Publish Data?

- **Database to Application Messaging**
  - ▸ Drive downstream Applications or APIs based on the Transactional Data of the changed Database Events

- **Event Notification**
  - ▸ Stream changed Data Information to Web Interfaces
  - ▸ Stream only particular Events of Interest (filter Data)

- **Data Warehouse / Business Intelligence**
  - ▸ Integrate captured Changed Data with an ETL Tool
  - ▸ Perform complex Transformations with custom Logic
  - ▸ Use a specific Transaction Format to update Target

- **MQ provides guaranteed delivery**
  - ▸ Avoids the need for 2-Phase-Commit (2PC)
  - ▸ Works even when the Target is not available

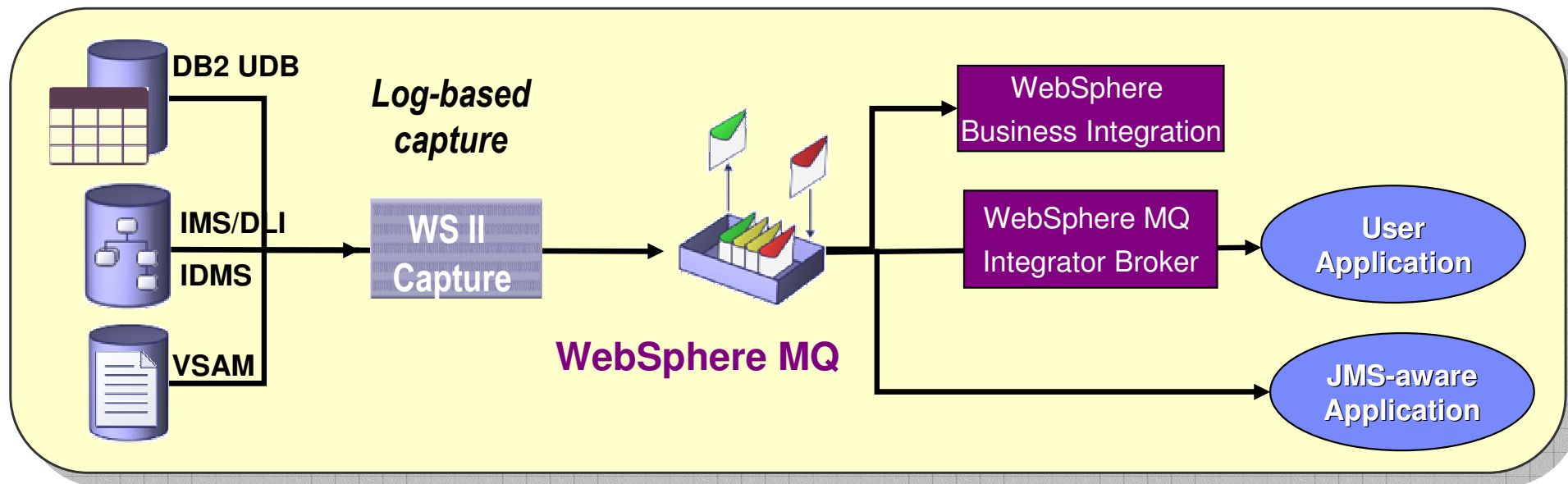# Publishing data events to facilitate business integration

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

Classic Data is REACTIVATED !

Event Publishing - Publication Options

- **Format**
  - ▶ Only data from committed transactions is published
  - ▶ Data is self describing with XML tags
  - ▶ Row based = one row per message
  - ▶ Transaction based = one transaction per message

- **Row Content**
  - ▶ Subset by column
  - ▶ Subset by predicate
  - ▶ Changed column values only or all column values
  - ▶ New data values only or include old values

Combining SQL and Q Replication with Event Publishing
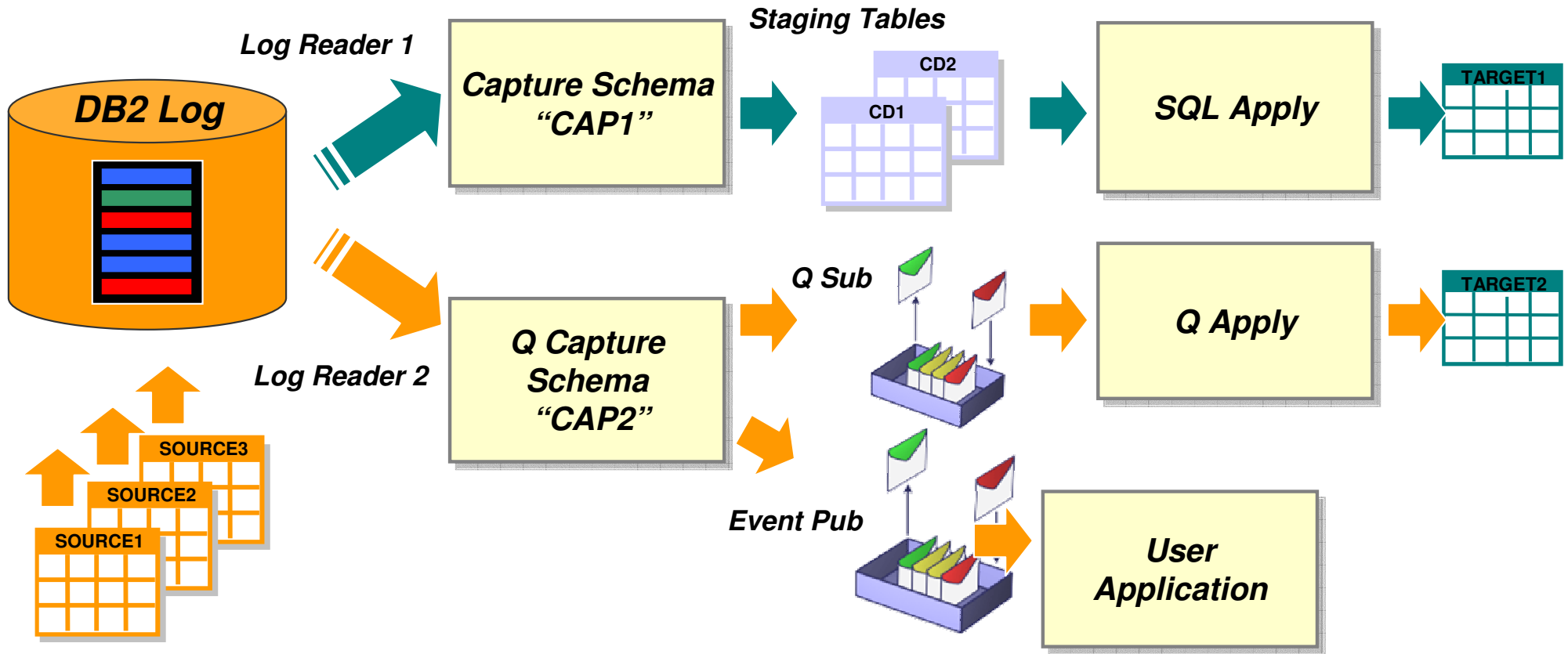
SQL Replication and Q Replication can co-exist

➢ Managed at source by using multiple capture schemas

➢ One Q Capture can handle both Publications and Subscriptions

DB2 II Event Publisher & Business Intelligence

Feeding Changed Data to :

➢ Data Warehouse

➢ Datamart

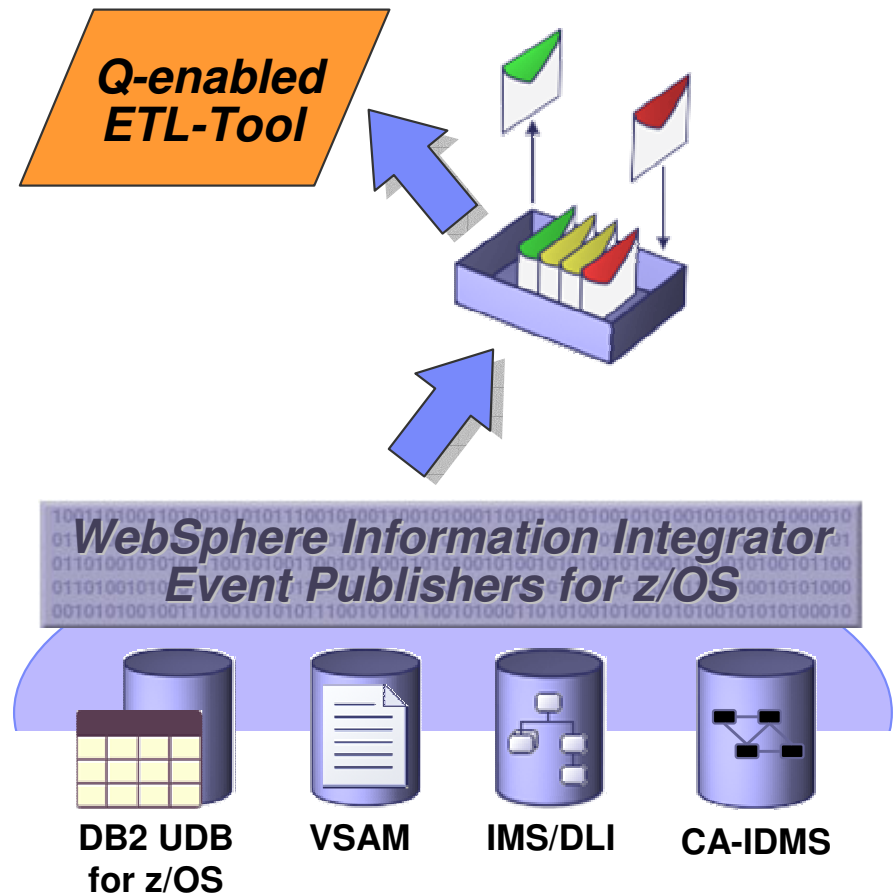➢ Operational Data Store (ODS)

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**

DB2 II Event Publisher & Business Integration

Data "Events" drive business integration

➤ Seamless integration with EAI via WebSphere MQ

Web Service

SeeBeyond

WebSphere

SIEBEL

Data used to drive EAI workflow

➤ Inventory update hits threshold…
triggering restocking process

➤ Addition of new customer:
  - ✓ Initiates welcome email
  - ✓ Credit verification
  - ✓ Accounting updates
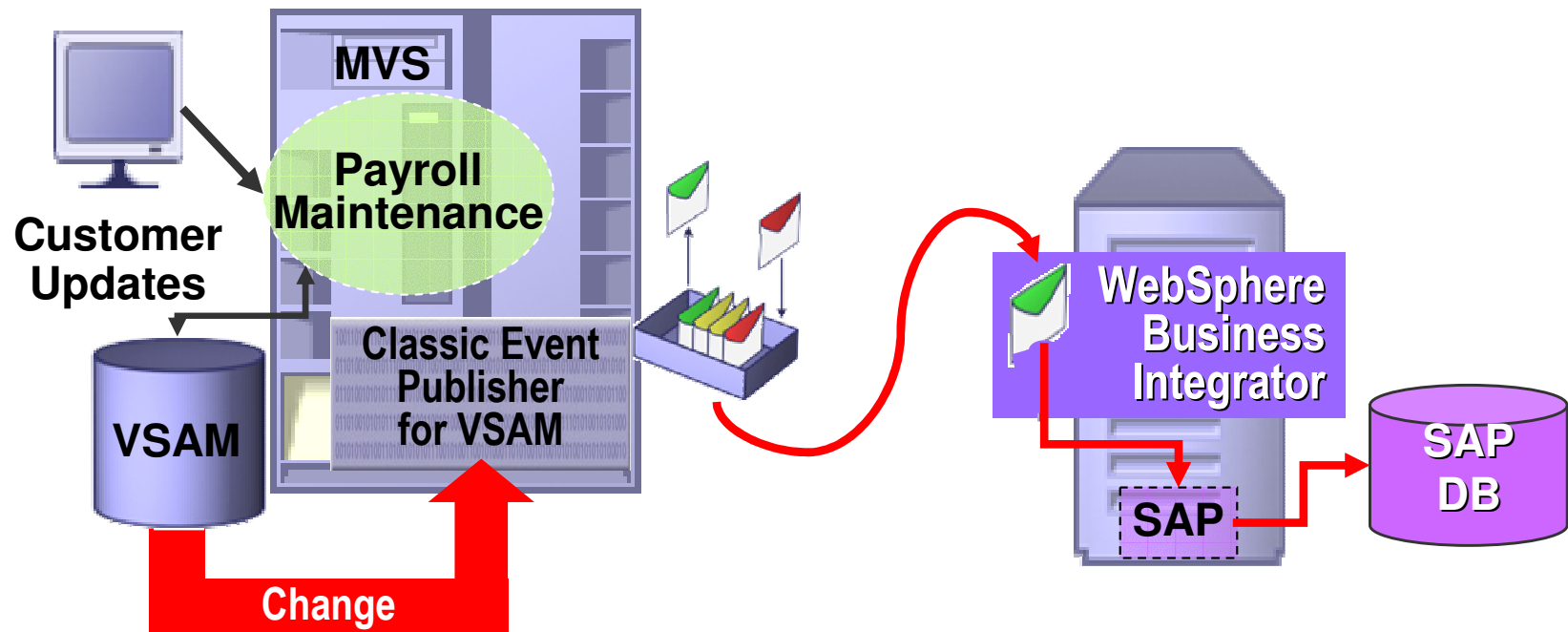 - ✓ …

Cross-silo data synchronization

➤ Synchronize mainframe updates with:
  - ✓ CRM
  - ✓ ERP
  - ✓ HR, etc.

WebSphere Information Integrator Event Publishers for z/OS

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

Sample Application

- Near real-time cross-silo data synchronization
  - ▶ Loosely coupled integration
  - ▶ Minimizes development effort
  - ▶ Simplifies maintenance

- *e.g. New order data is automatically pushed to a CRM application*
  e.g. VSAM employee data updates are pushed to SAP payroll

Customer Updates

MVS

Payroll Maintenance

VSAM

Classic Event Publisher for VSAM

Change

WebSphere Business Integrator

SAP

SAP DB

Order Processing – Exploiting II Event Publishing

Order Entry Applications

Order Processing Data

SOURCE
SOURCE
SOURCE

Q Capture

WBI Event Broker

Billing Request Creation

Shipping Request Creation
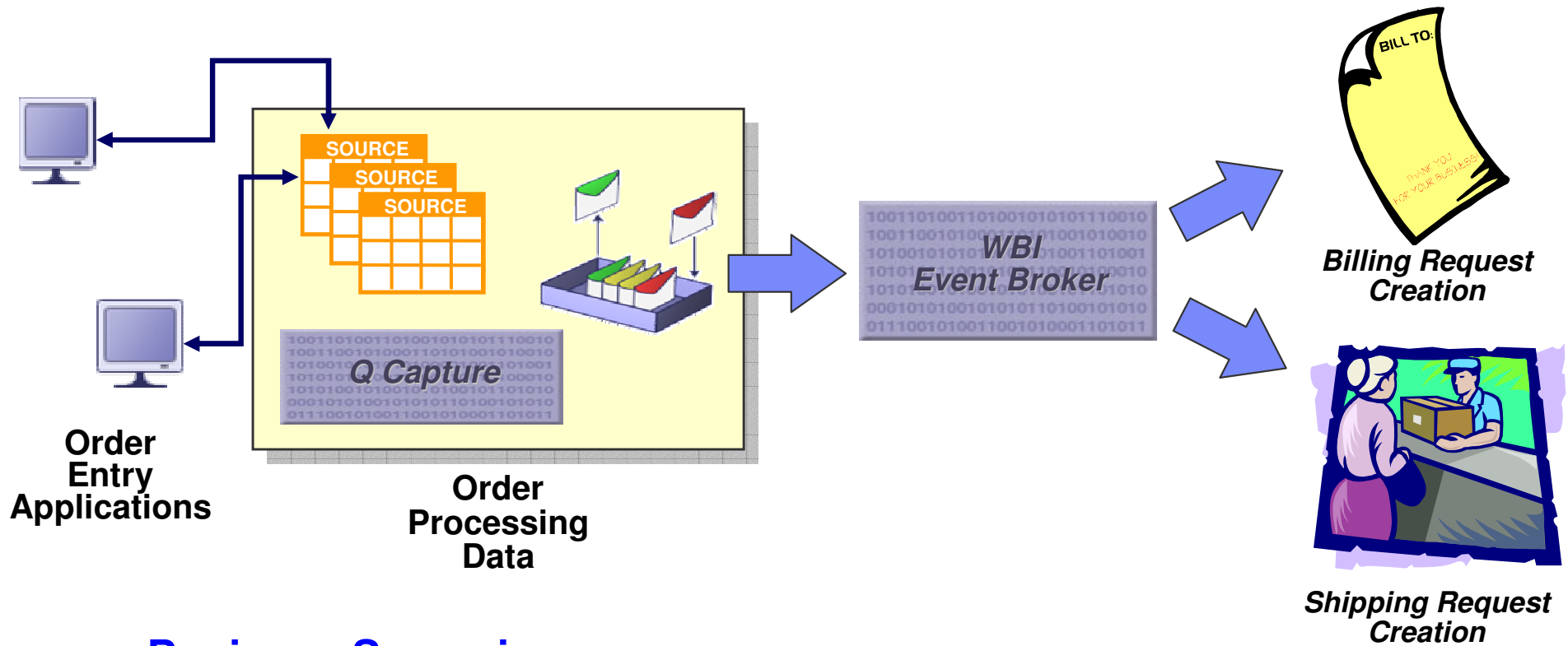
Business Scenario

➢ As new Orders are entered into the Order Entry System, the pertinent Data is captured and published into a Queue

➢ The Websphere MQ Integrator Broker processes the queued Data

➢ A billing Transaction is created and queued in one System and a Shipping Transaction is created and queued in another System
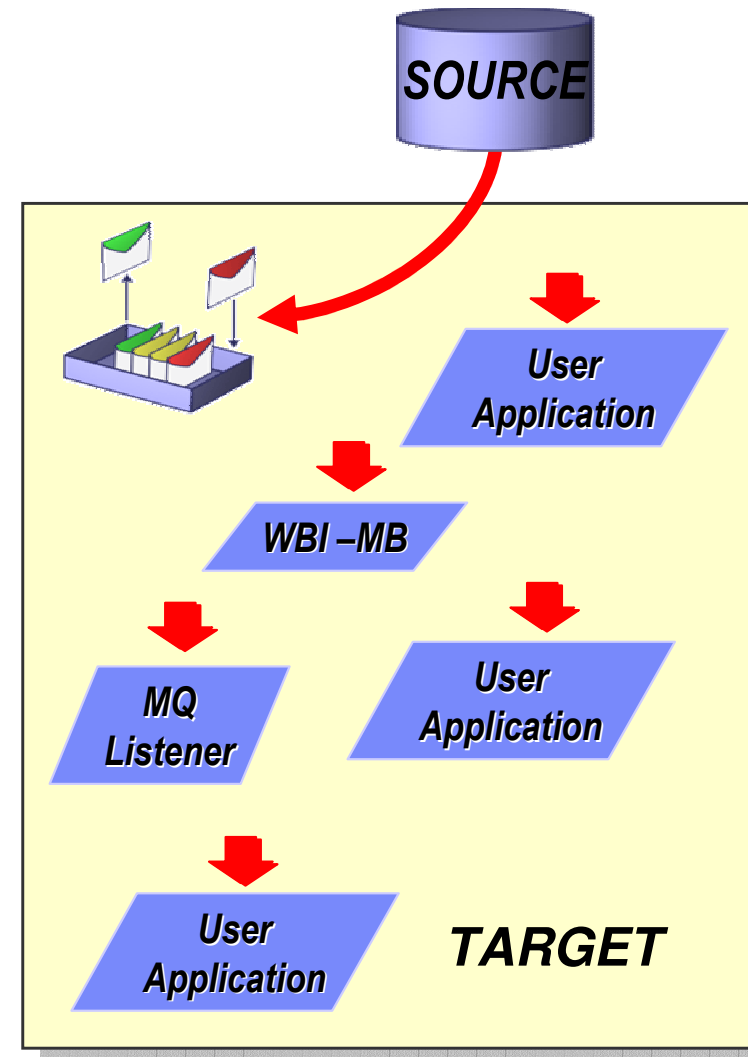
Sample Application

- **Event Notification**
  - ▶ Receivable balances approaching credit limit pushed to a credit analyst
    - Threshold is independent of order processing and accounting applications
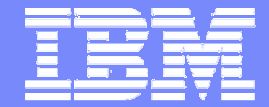   - No "hard-hooks" in OLTP applications necessary
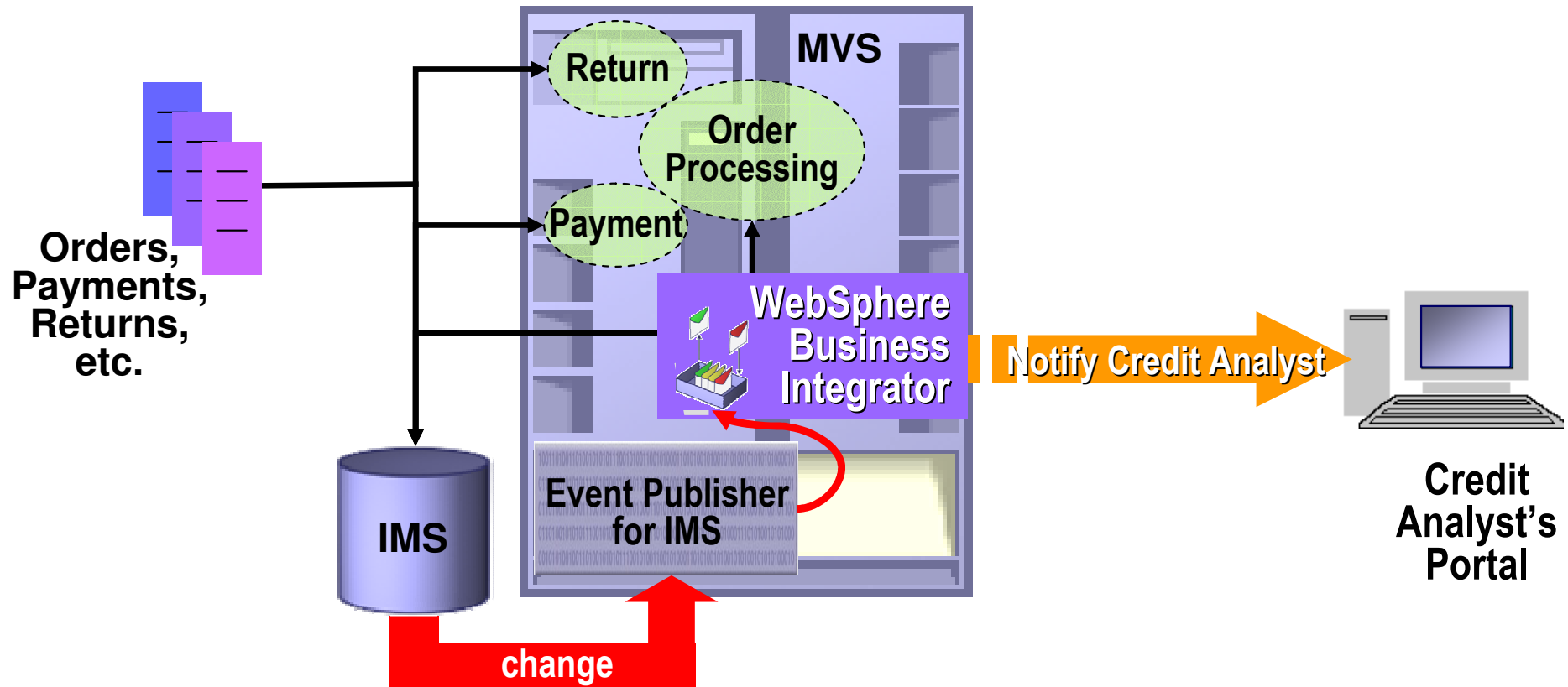
Why data events versus application events?

- **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

Summary

Using Classic II Event Publishing for Replication

IMS/DLI data is updated

Data is Captured near real time

II IMS/DLI Event Publisher

MQ Listener

Stored procedure generates DB2 I/U/D statements
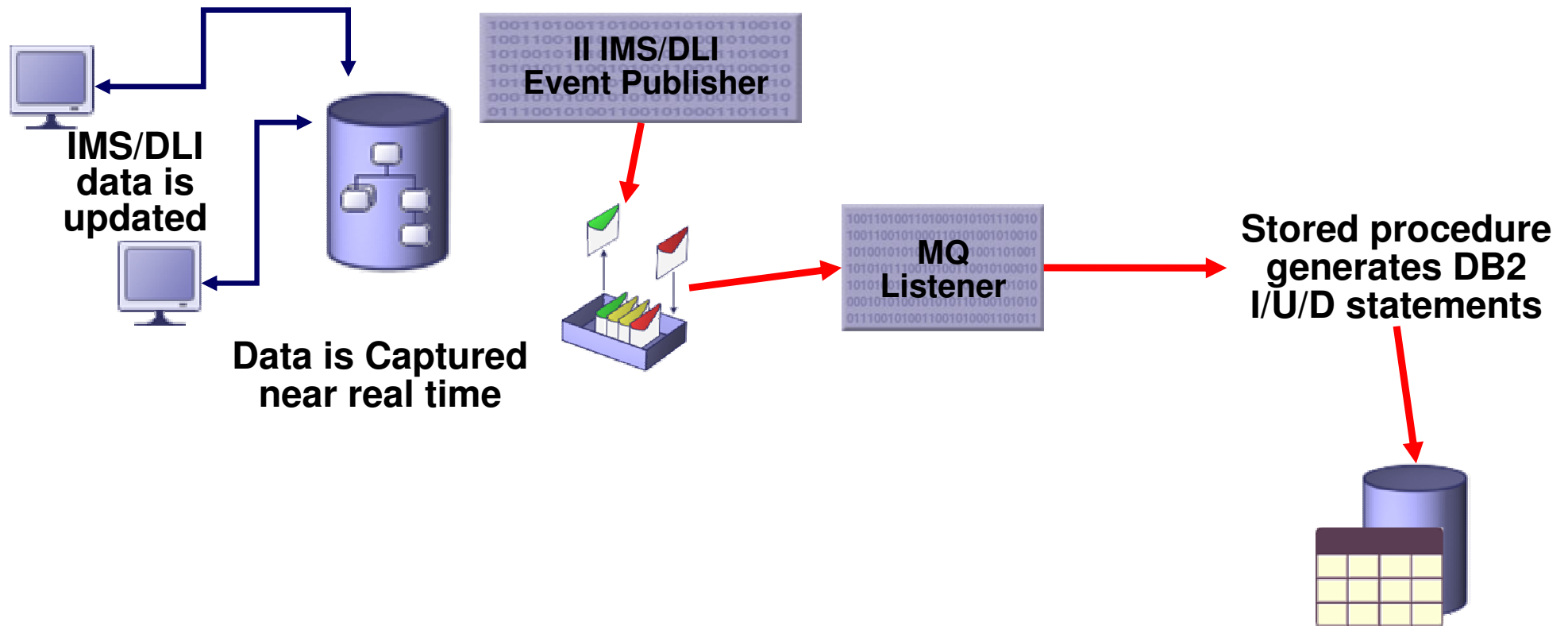
- It is our direction to support replication of Classic data sources using the highly parallel Q Apply of DB2 II Replication.

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

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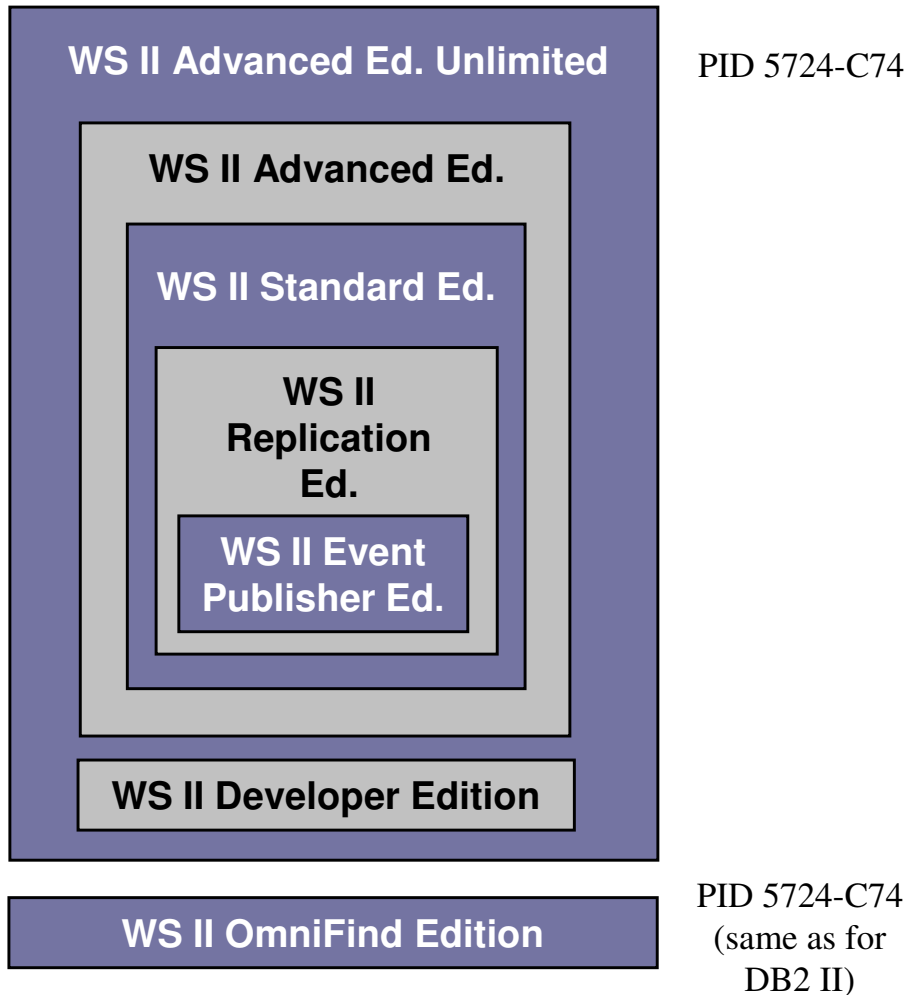
- WebSphere Information Integrator provides access to diverse, distributed, and real-time data as if it were a single source, no matter where it resides.

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

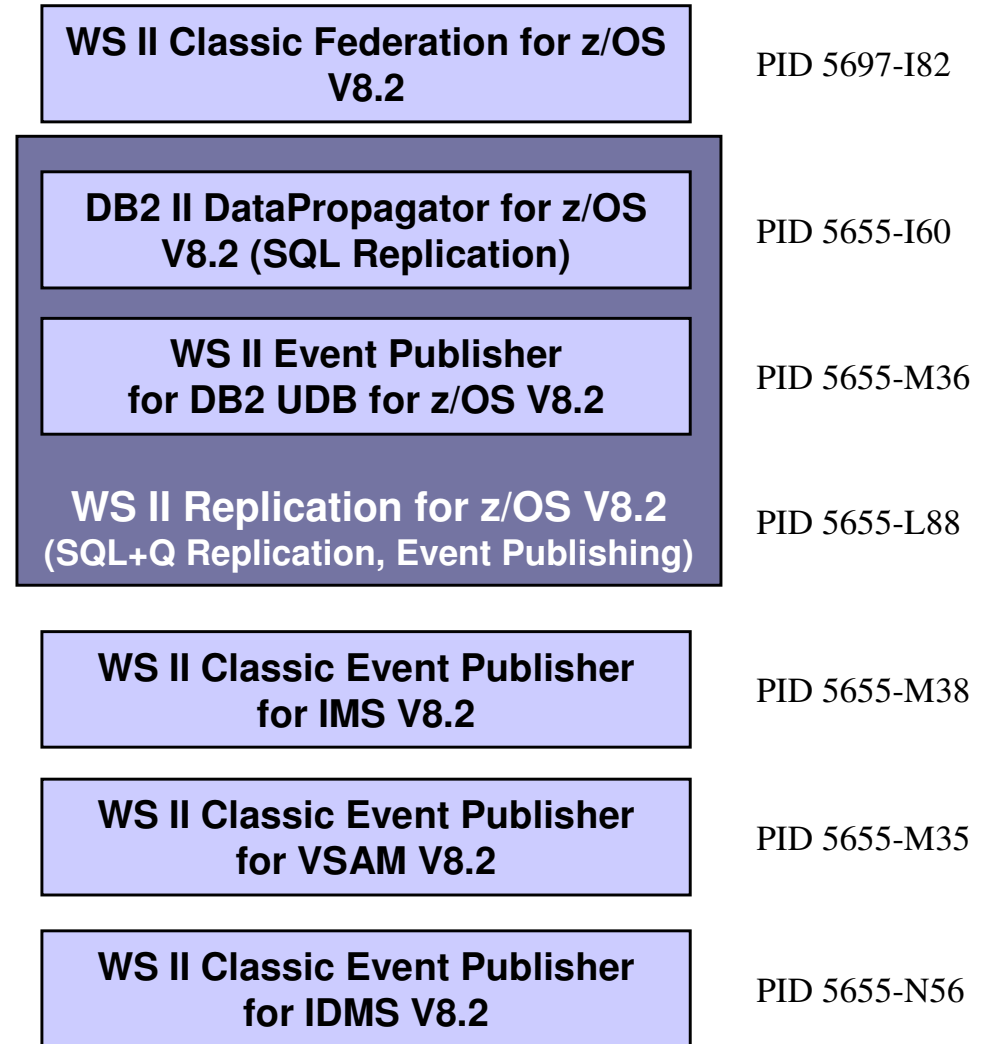- Rely on IBM's proven technology and support for open standards

IBM

WebSphere Information Integrator Packaging and PIDs

Distributed (Linux, UNIX, Windows)

WS II Advanced Ed. Unlimited   PID 5724-C74

WS II Advanced Ed.

WS II Standard Ed.

WS II Replication Ed.

WS II Event Publisher Ed.

WS II Developer Edition

WS II OmniFind Edition   PID 5724-C74 (same as for DB2 II)

- Processor-based pricing except for Developer Edition which is priced by user
- Priced Connectors to access non-IBM sources

Mainframe (IBM eServer zSeries)

WS II Classic Federation for z/OS V8.2   PID 5697-I82

DB2 II DataPropagator for z/OS V8.2 (SQL Replication)   PID 5655-I60

WS II Event Publisher for DB2 UDB for z/OS V8.2   PID 5655-M36

WS II Replication for z/OS V8.2 (SQL+Q Replication, Event Publishing)   PID 5655-L88

WS II Classic Event Publisher for IMS V8.2   PID 5655-M38

WS II Classic Event Publisher for VSAM V8.2   PID 5655-M35

WS II Classic Event Publisher for IDMS V8.2   PID 5655-N56

- Value Unit pricing Model