



IBM Software Group

WebSphere Information Integration

“ IBM Information Integration offering on z/OS ”

Eric Derbanne – eric.derbanne@fr.ibm.com

WebSphere Information Integration Software

ON DEMAND BUSINESS

© 2005 IBM Corporation

Defining Business Integration



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

<p>People</p> <ul style="list-style-type: none"> ➤ Portal – personalized information ➤ Collaboration technology ➤ Adaptable workplace ➤ Consistent rule-based experience across devices 	<p>+</p> <p>Processes</p> <ul style="list-style-type: none"> ➤ Workflow management ➤ B2B connectivity ➤ Messaging infrastructure (EAI) ➤ e-Business transactions ➤ Service-oriented architecture 	<p>+</p> <p>Information</p> <ul style="list-style-type: none"> ➤ Federation & data placement (replication, ETL, caching) ➤ XML (store, query, webservice) ➤ Meta-data management ➤ Text Search and analytics
--	--	---

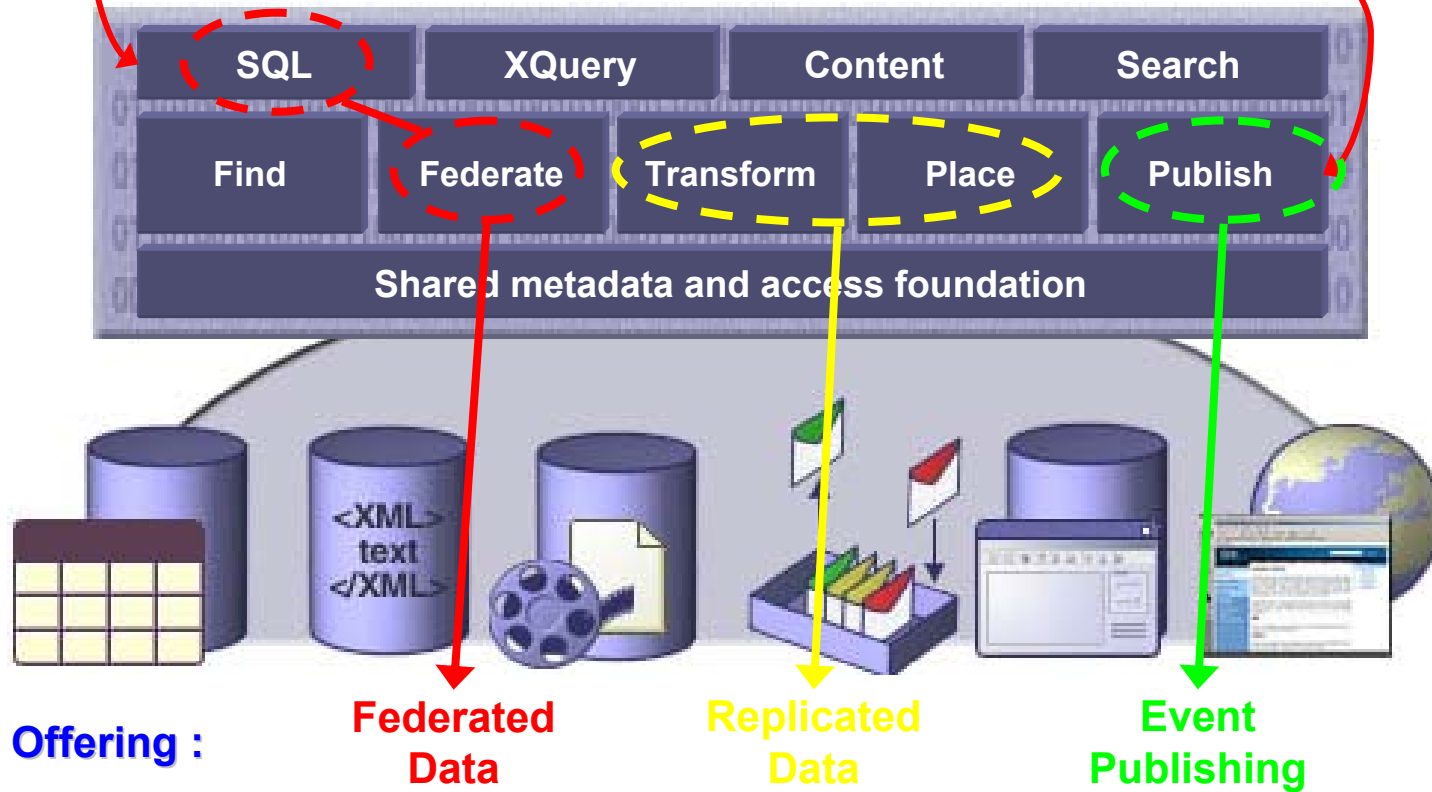
<i>40% of people's time is spent searching for relevant information</i>	<i>40% of IT budgets may be spent on integration</i>	<i>30-50% of design time is copy management</i>	<i>85% of information is unstructured</i>
<p>☹ <i>for each \$1 spent for a packaged application, customers spend \$5 to \$9 on the labor for integration !</i> <i>(IBM Customer Surveys, 2001, 2002)</i></p>			

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**



zSeries II Offering :

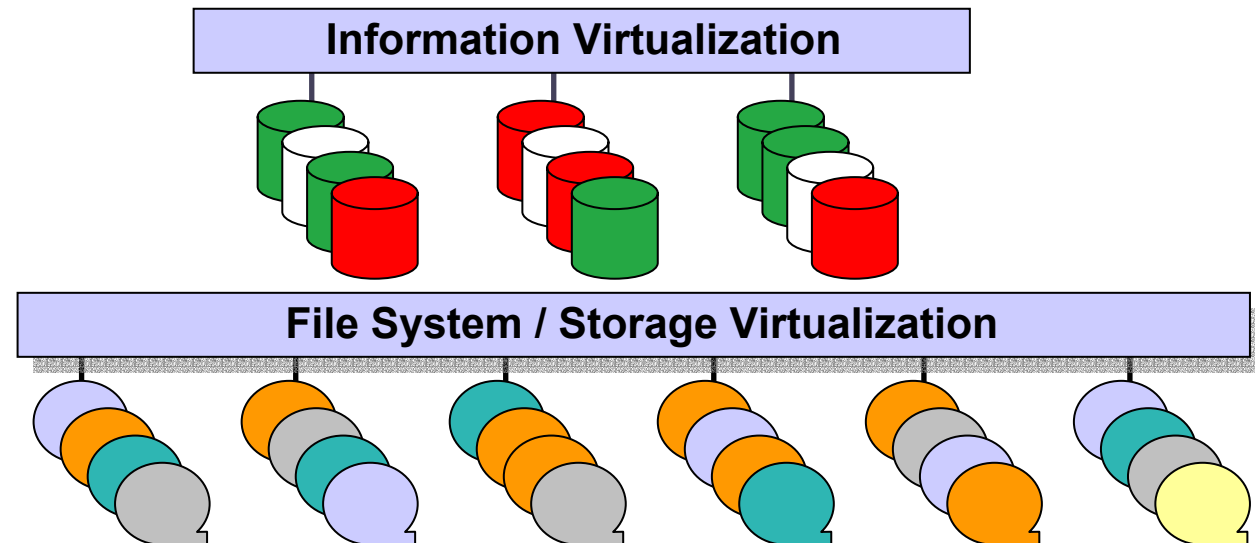
Using Websphere Information Integrator as GRID Enabler

- Leverage IT Infrastructure
- Accessing Information regardless of Database Management System and Structure



***Federation
instead of
Centralization***

WebSphere Information Integrator





Federated Data Server

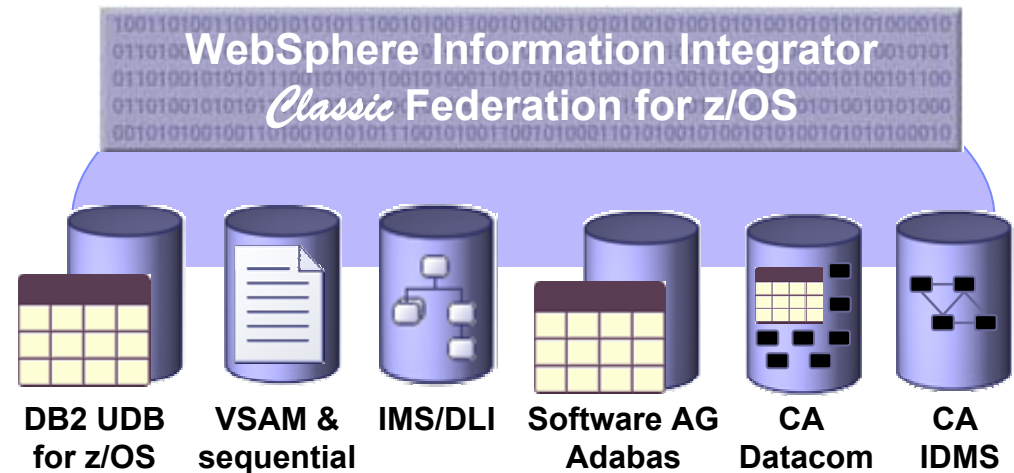
- *WebSphere Information Integrator Classic Federation – z/OS Platforms*
- *WebSphere Information Integrator - LUW Platforms*



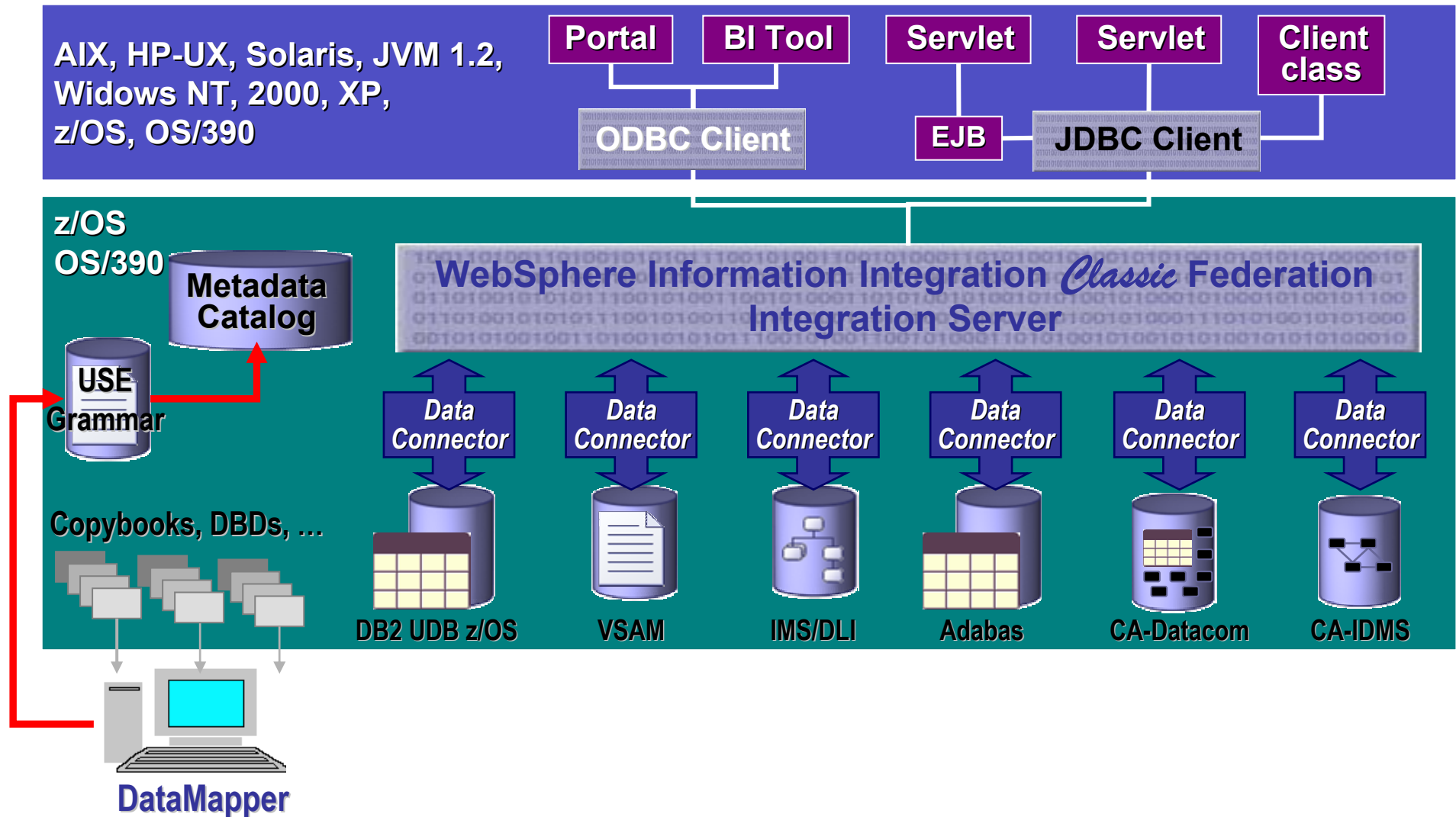
WebSphere Information Integrator Classic Federation for z/OS

- 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

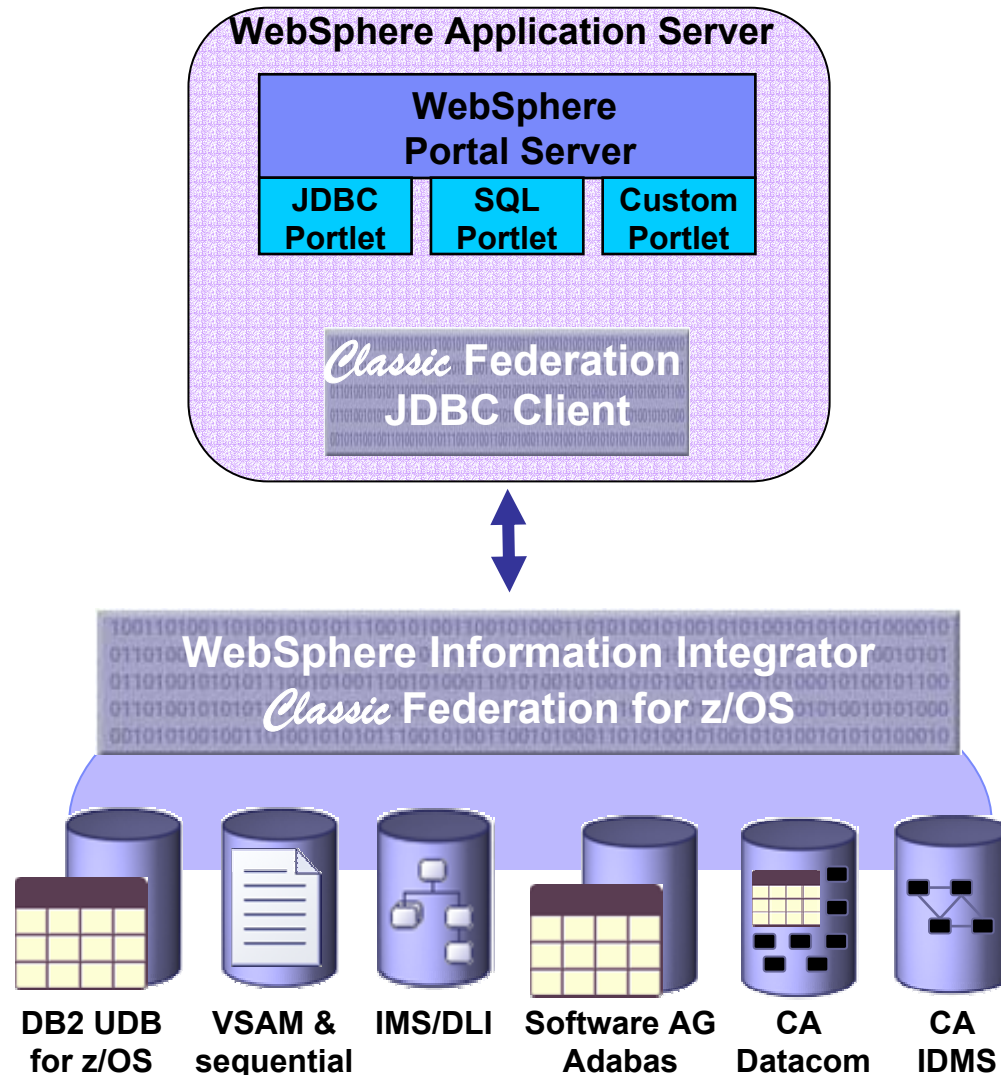
- DB2 Information Integrator Classic Federation for z/OS
 - ▶ Read/Write mainframe data sources using SQL through **standard ODBC, JDBC, CLI**
 - ▶ Native database connectors leverage power of each database/file accessed
 - ▶ Metadata-driven means:
 - No mainframe programming required
 - Fast installation,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



WebSphere Information Integrator *Classic* Federation



WebSphere and II Classic Federation

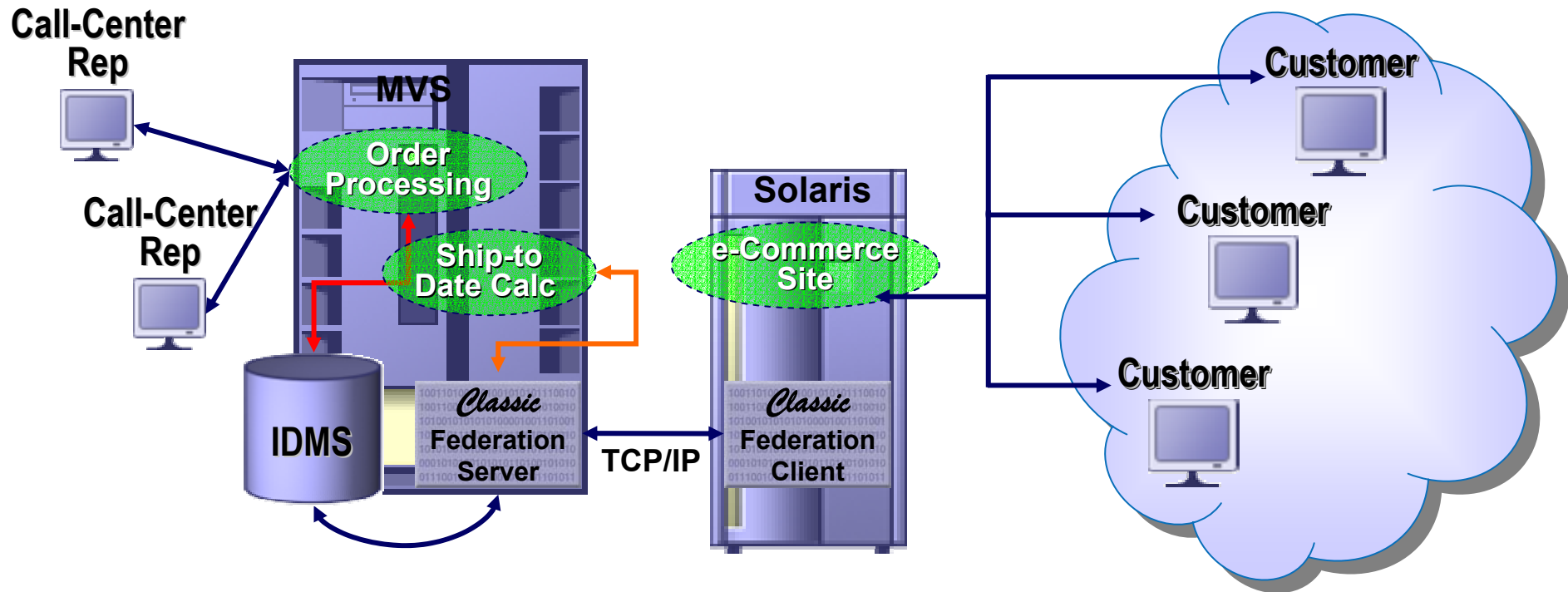


- Integrate mainframe data directly with Web applications, Portals, EAI platforms
- Supports JDBC 2.1 Core APIs Plus
 - ▶ Scrollable/Updatable ResultSets
 - ▶ Commit – Rollback – AutoCommit - XA
 - ▶ Stored Procedure to mainframe programs
 - ▶ Parameter Markers
 - ▶ Metadata commands for tables, columns, 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:

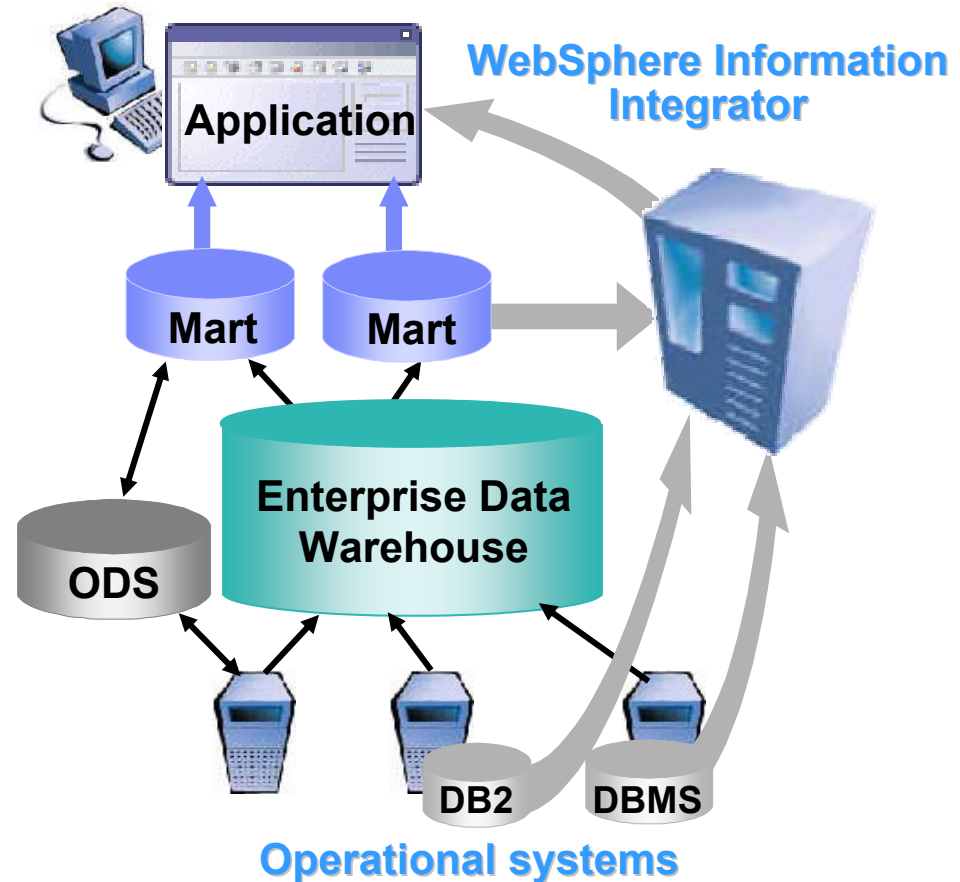
- Legacy call-center systems
- New e-commerce applications
- No mainframe skills required for e-commerce site development



Extending Data Warehouses with Real-Time Data

Business Scenario

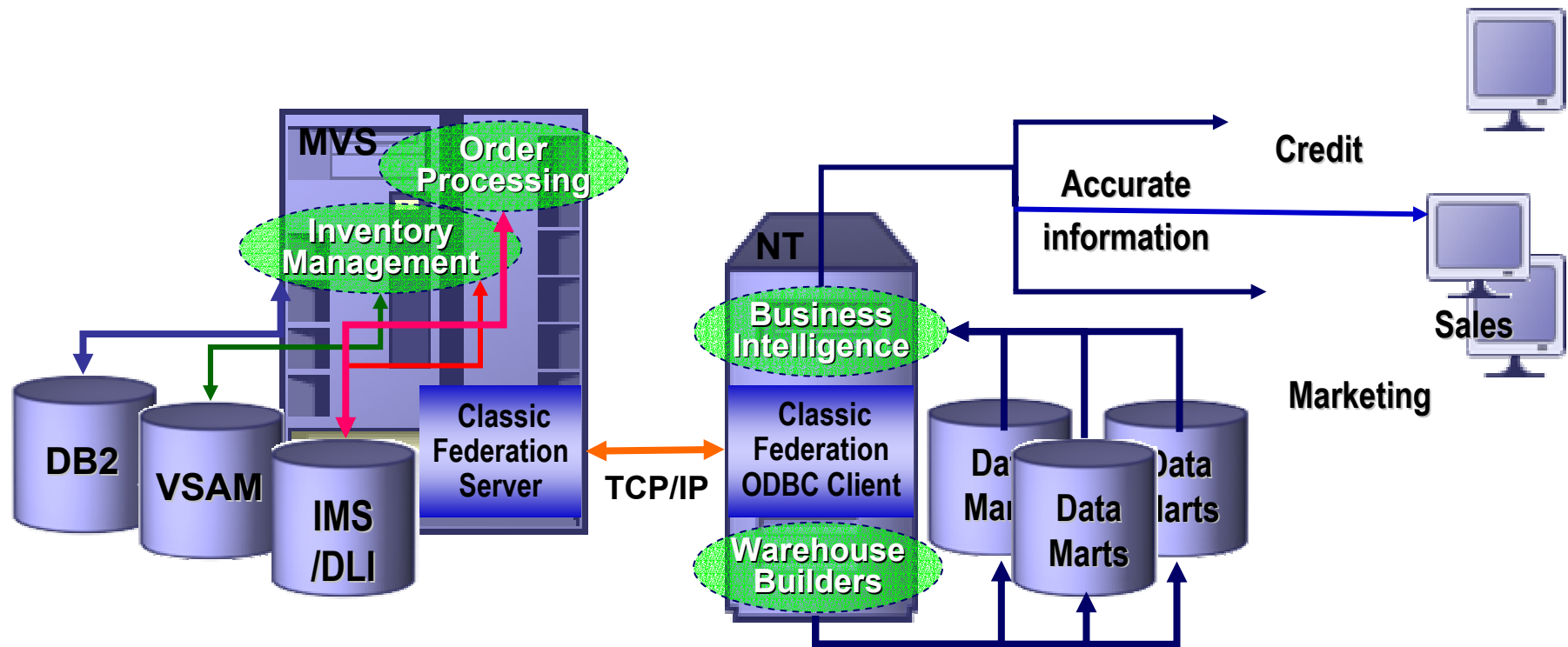
- Access current customer records from a call centre
- Access current stock levels from a supply chain data mart
- Basel II compliance
- Business activity monitoring – linking events to trends



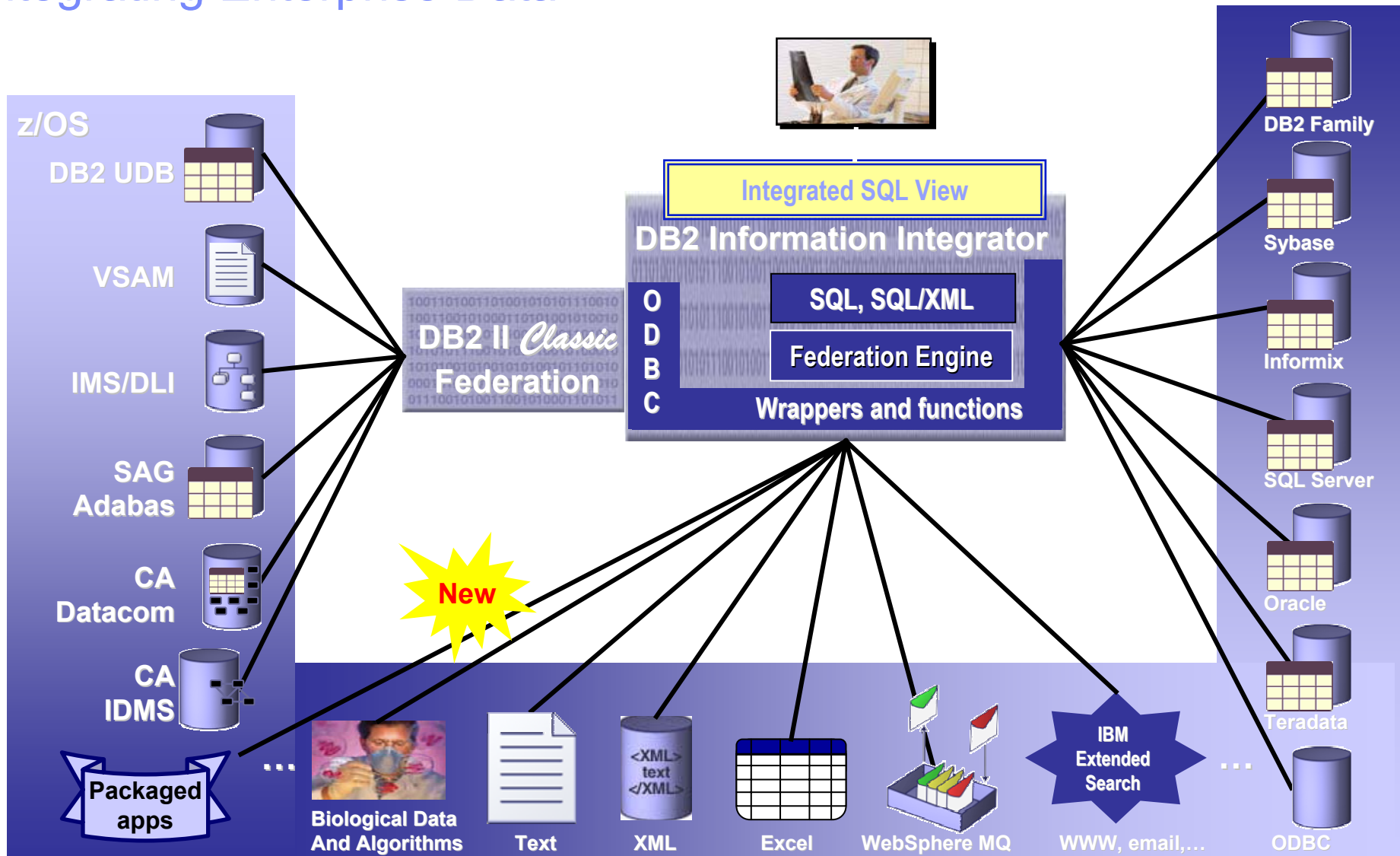
Integration in Action – Recreation Vehicle Manufacturer

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



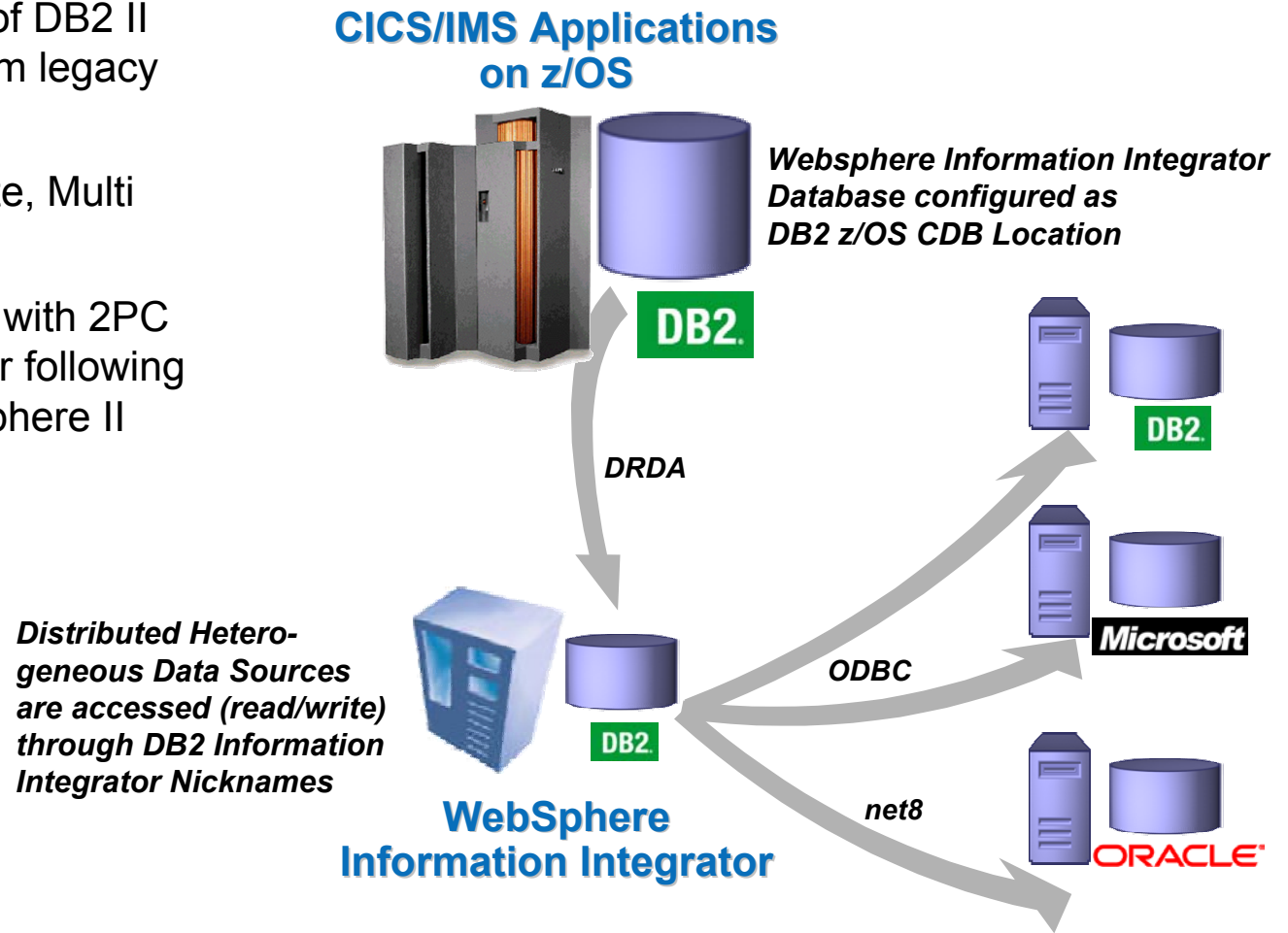
Integrating Enterprise Data



Access Distributed Heterogeneous Data from CICS Applications

Business Scenario

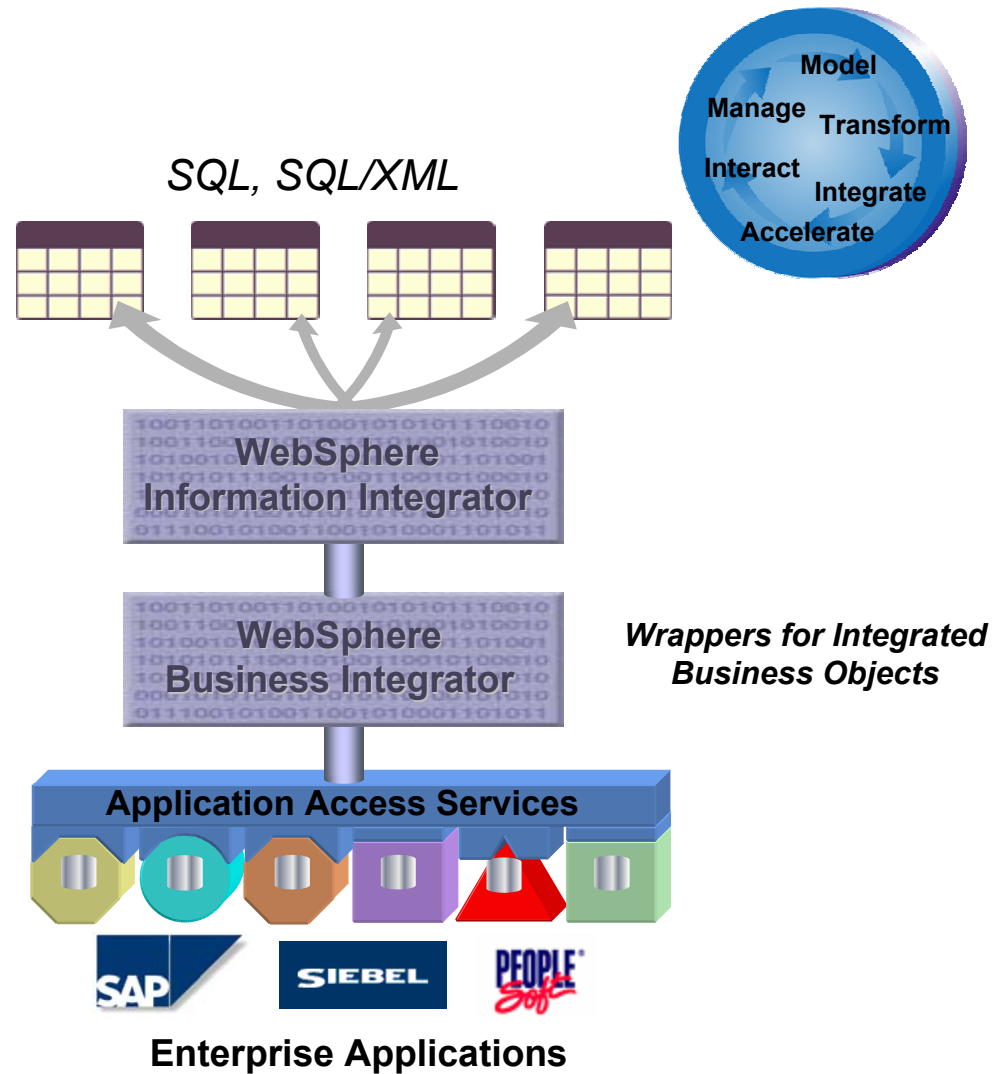
- Access all kinds of DB2 II Data Sources from legacy Applications
- Single Site Update, Multi Site Read
- Distributed UOW with 2PC support in plan for following version of WebSphere II



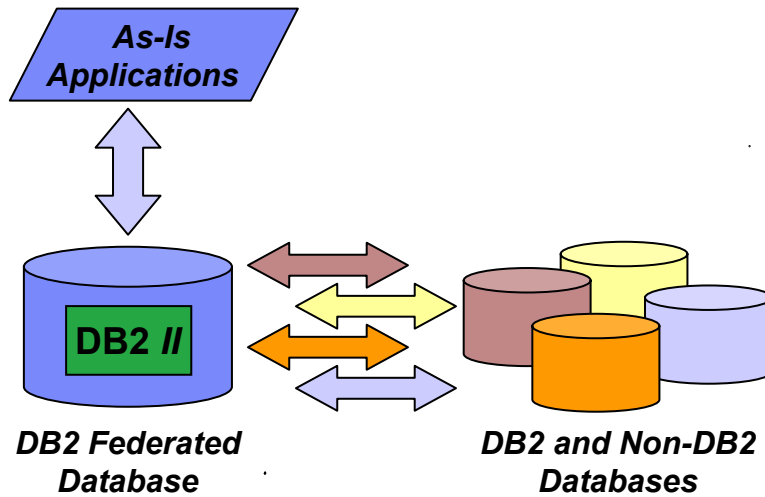
Information and Process Integration with WebSphere

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



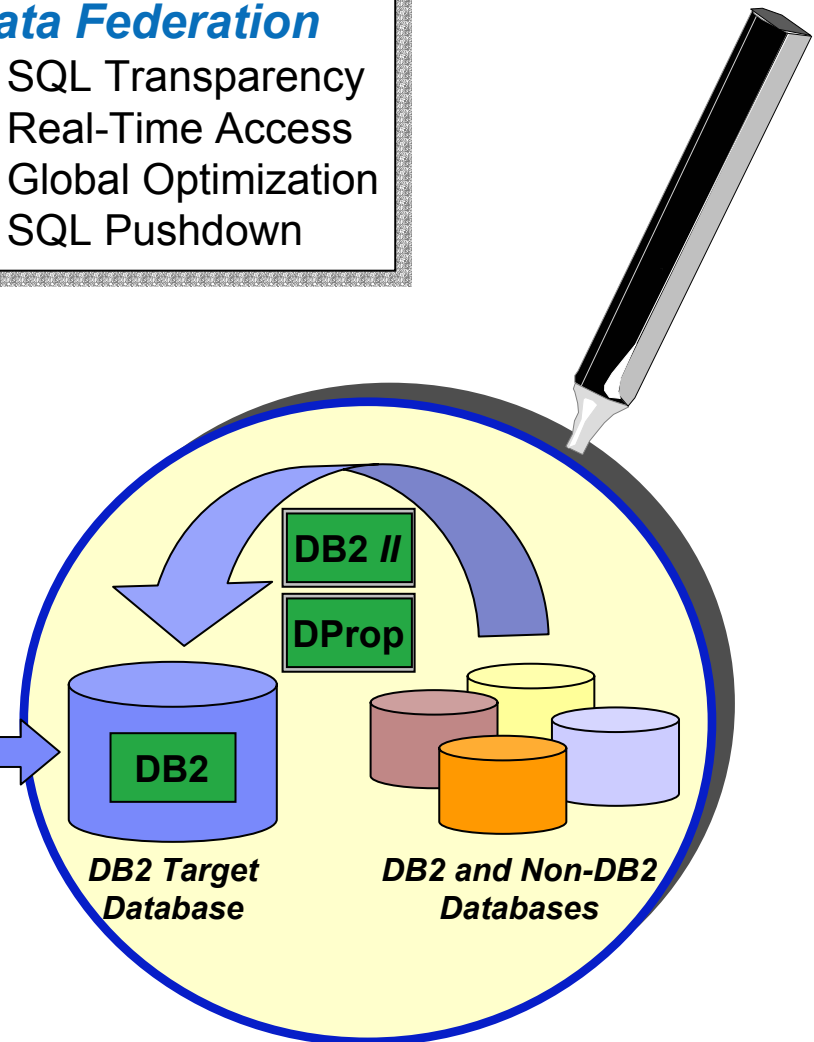
Federate or Replicate – That is the Question !!



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

- Data Replication**
- Data Movement
 - Data Transformation
 - Autonomy

Consolidated Applications





IBM Software Group | WebSphere Information Integration Software

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

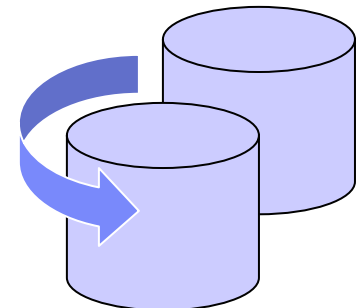
- ▶ Move data to new platform/database, transform data
- ▶ ETL or Replicate?
 - latency needs
 - change volume versus total volume
 - complexity of transformation and/or cleansing

▪ Mobile Workforce

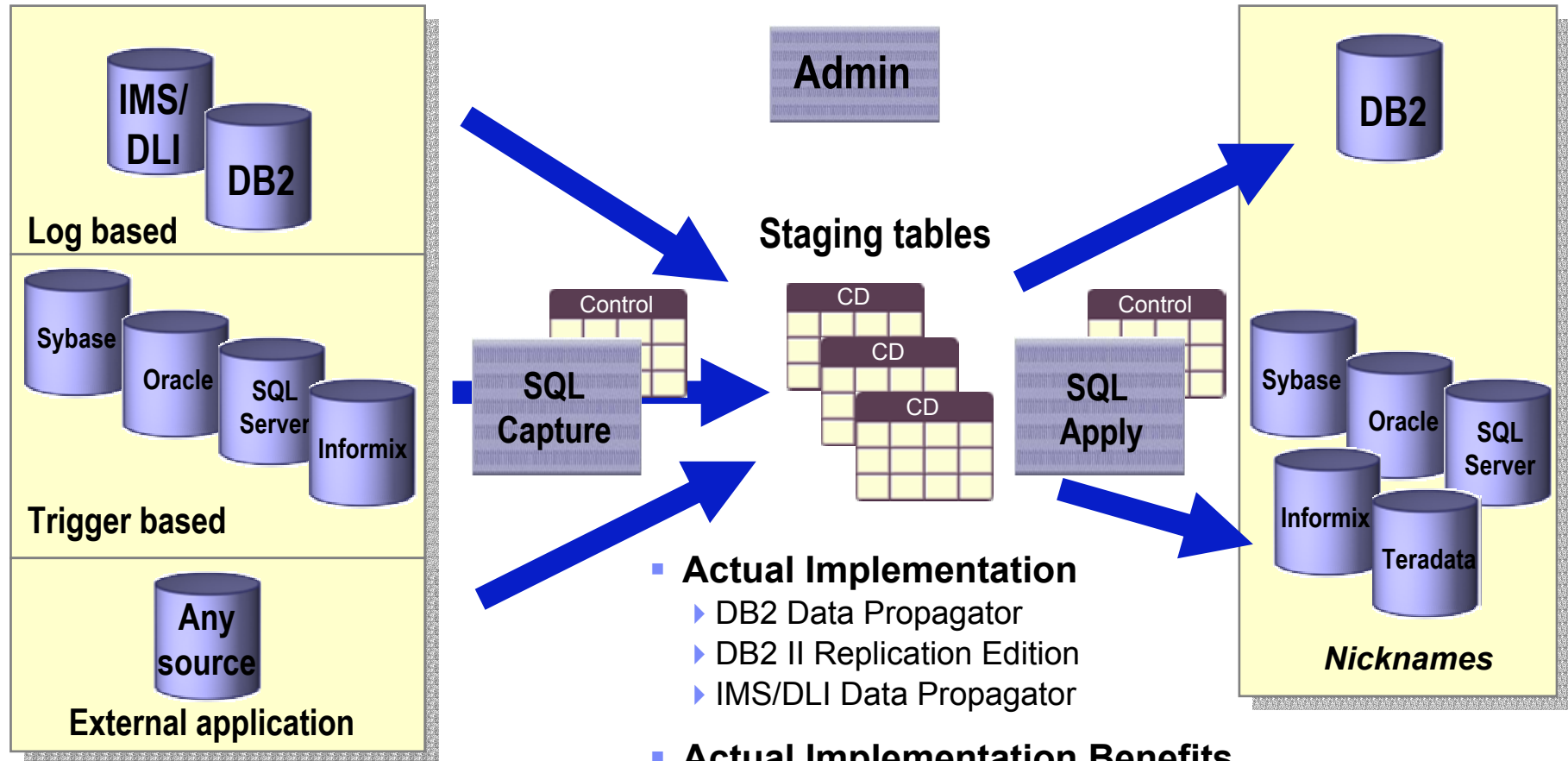
- ▶ Occasionally connected distribution/consolidation

▪ 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

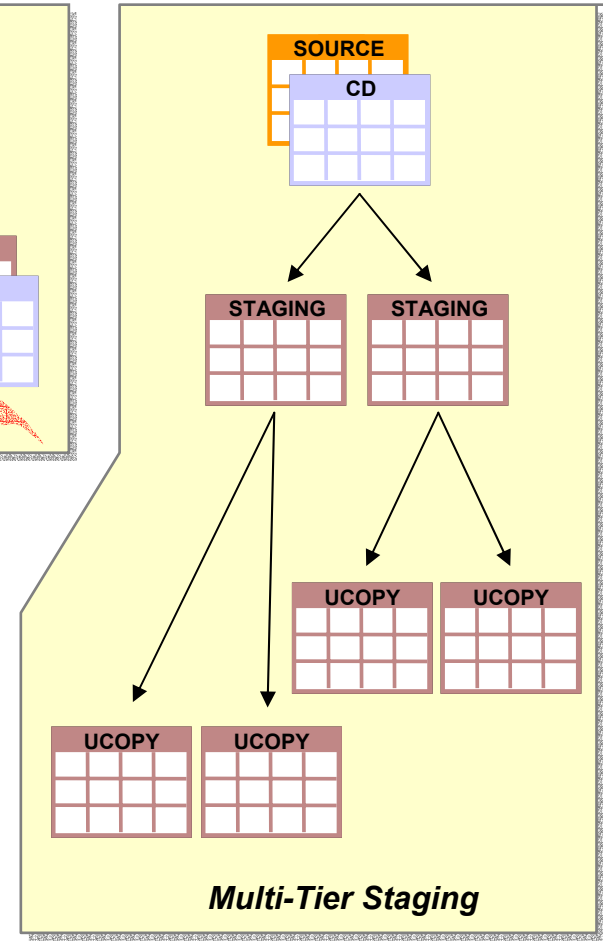
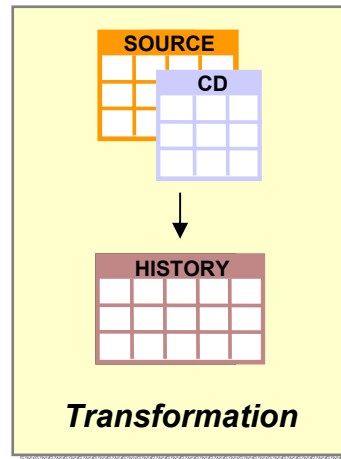
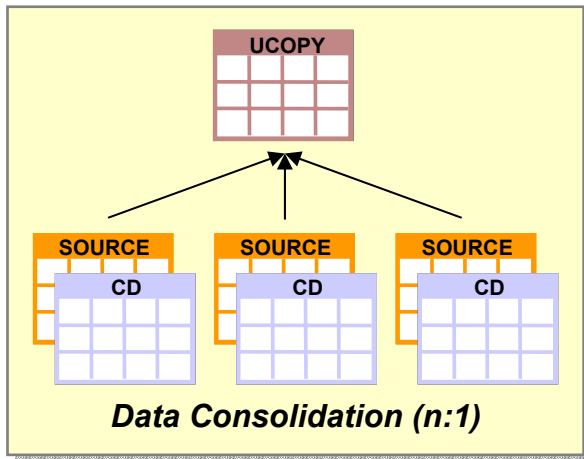
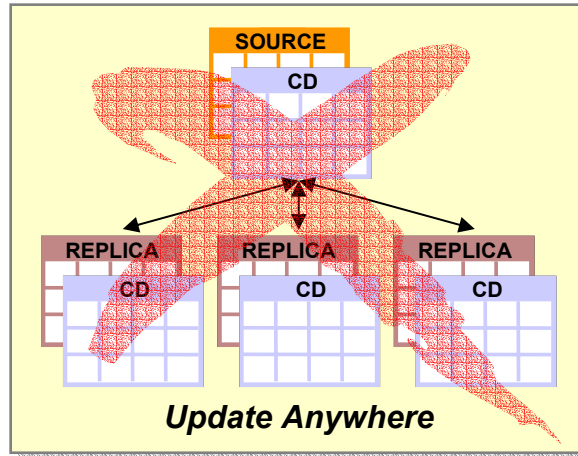
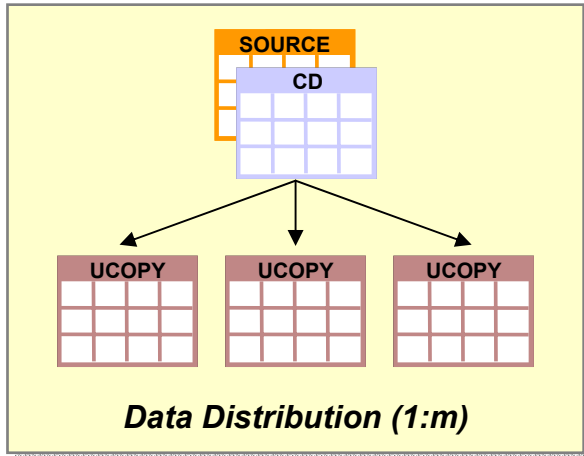


SQL Replication Architecture



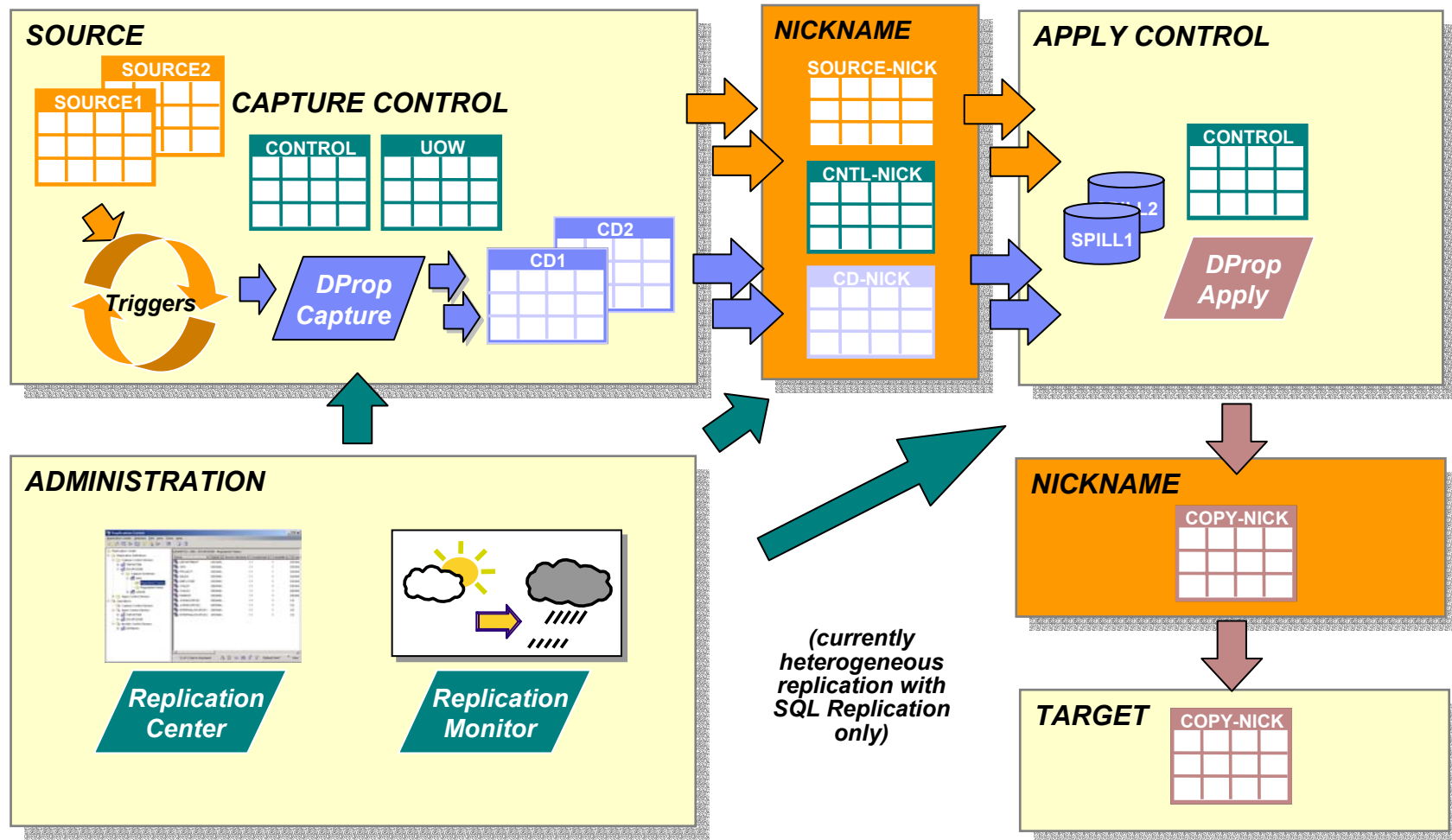
- **Actual Implementation**
 - DB2 Data Propagator
 - DB2 II Replication Edition
 - IMS/DLI Data Propagator
- **Actual Implementation Benefits**
 - Extremely flexible and resilient
 - Very easy to set up transformations
 - Scales well to reach multiple targets
 - Homogeneous & Heterogeneous Sources

Sample SQL Replication Scenarios



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

DB2 Data Replication from Federated SOURCES



(currently heterogeneous replication with SQL Replication only)

Federated Sources & Targets

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

Why Create Another Replication Architecture?

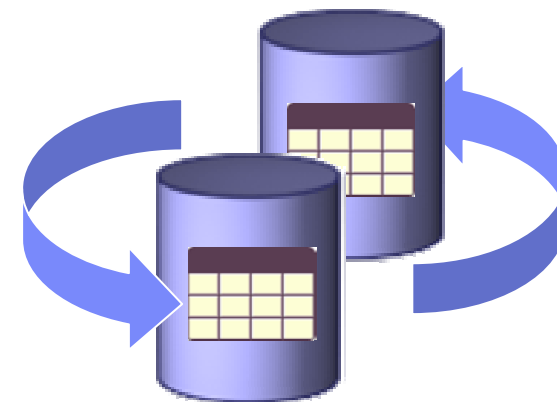
- **Performance**
 - ▶ Combine high Throughput with low Latency

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

- **Capability**
 - ▶ Significantly improve multi-directional Replication Support

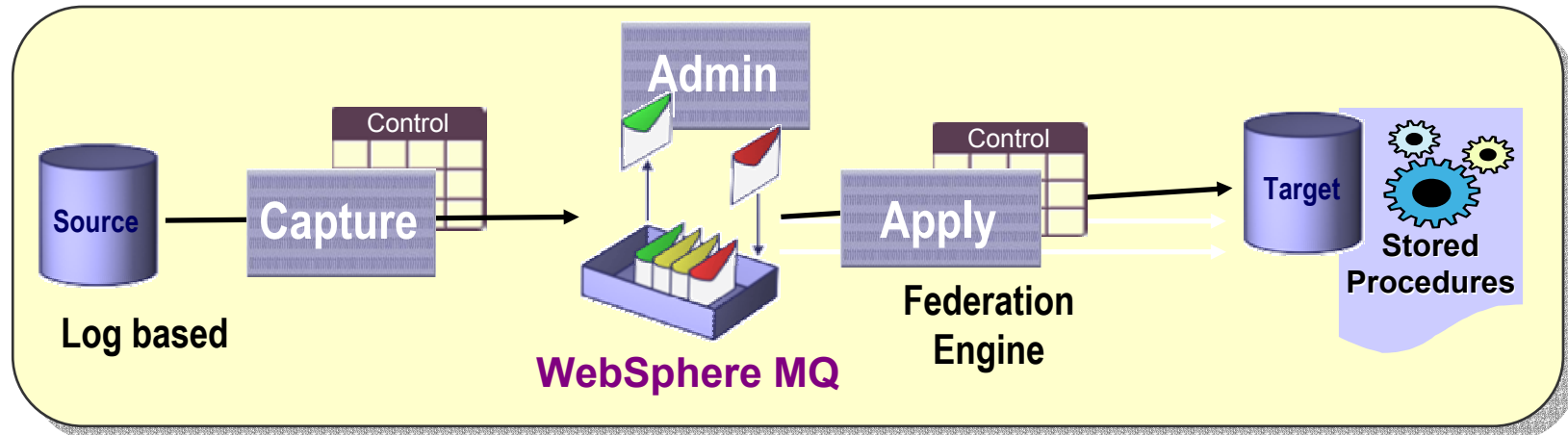
- **Manageability**
 - ▶ Reduce the number of Replication Objects to be defined and managed

 - ▶ 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

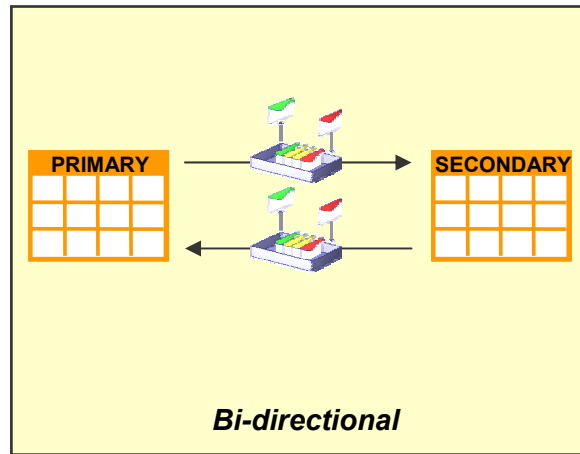
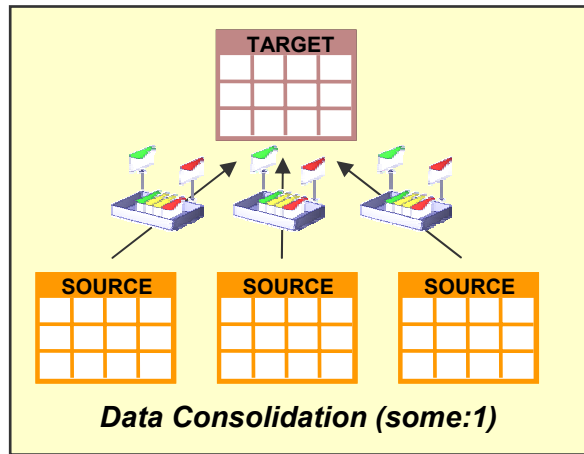
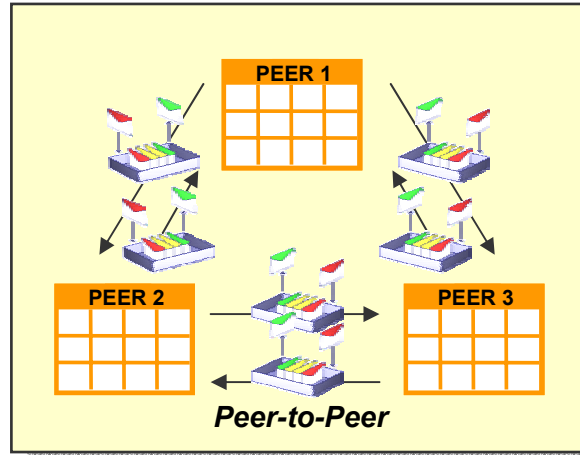
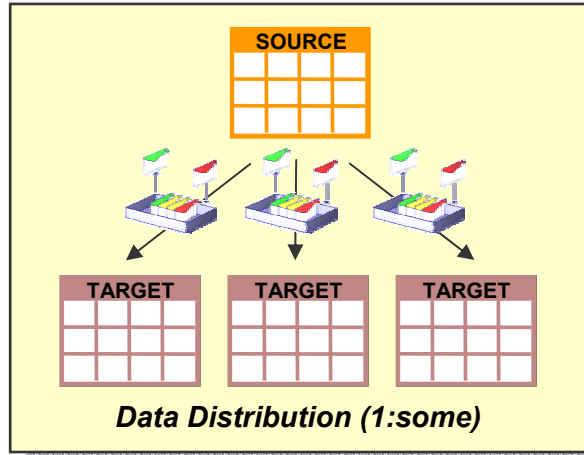
- Predicate examples
 - ▶ Based on values in the row data itself

```
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)
```

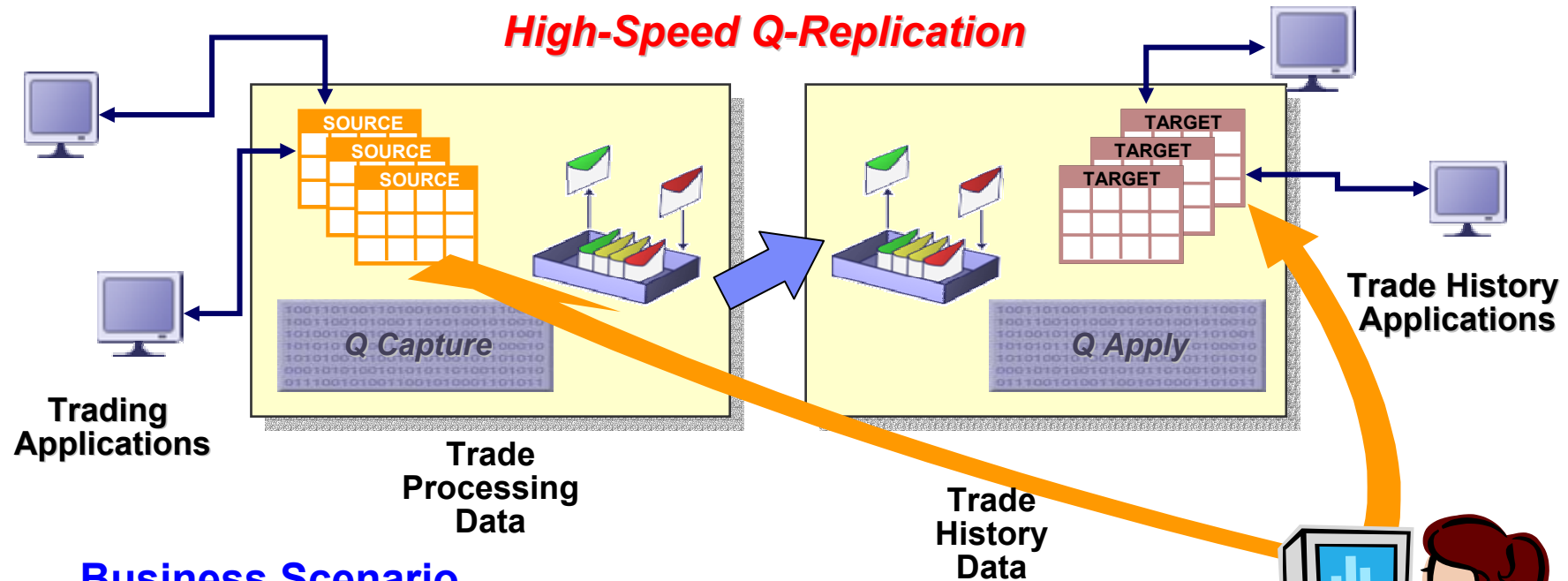
Sample Q-Replication Scenarios



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

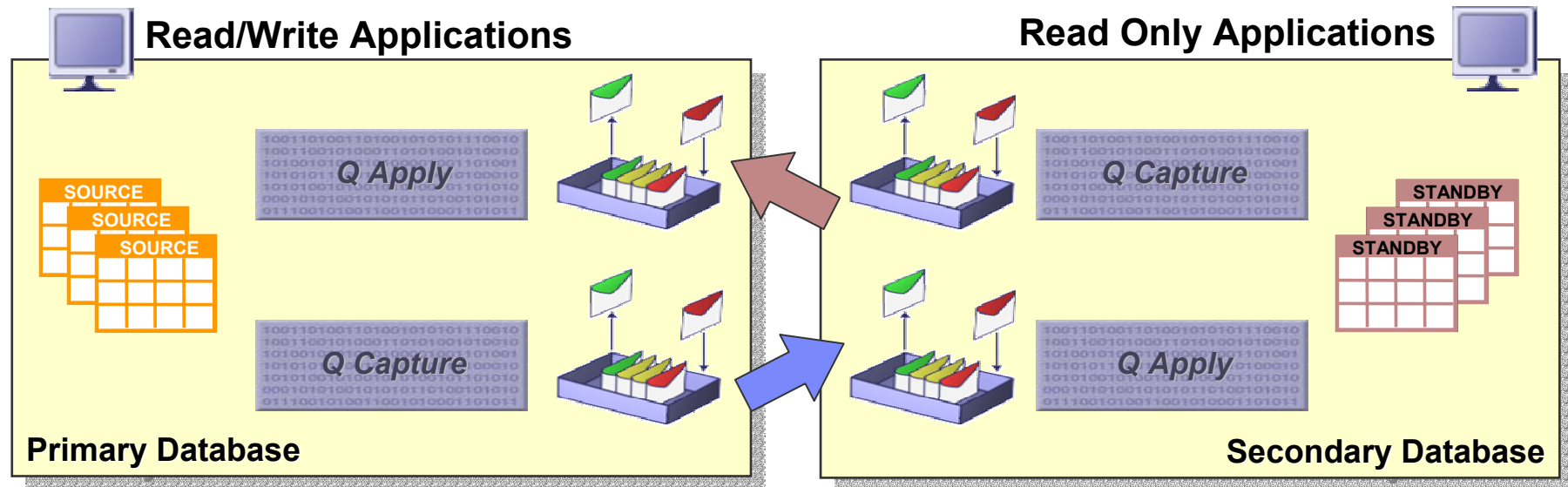


Business Scenario

- 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



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



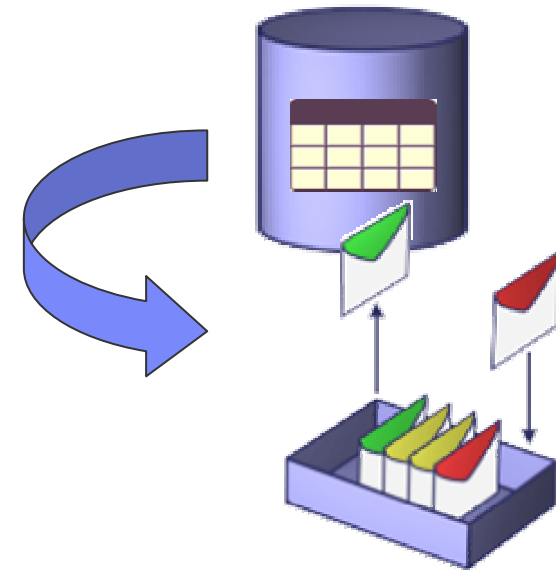
IBM Software Group | WebSphere Information Integration Software

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

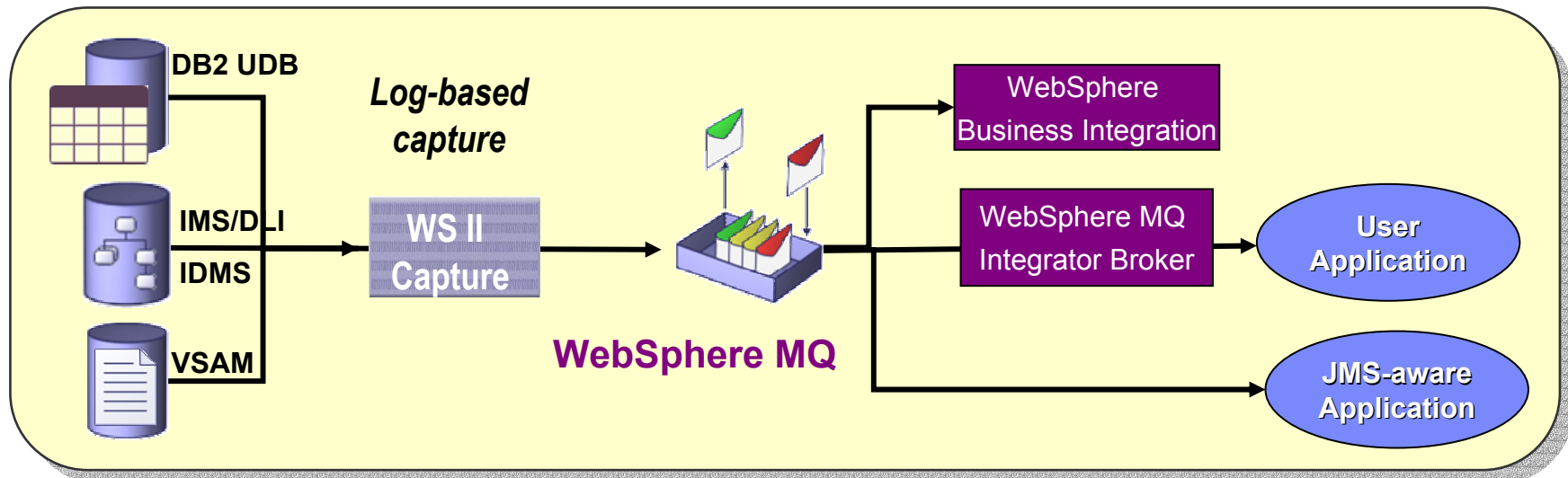
Capture database changes as XML messages and publish them to WebSphere MQ

▪ 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

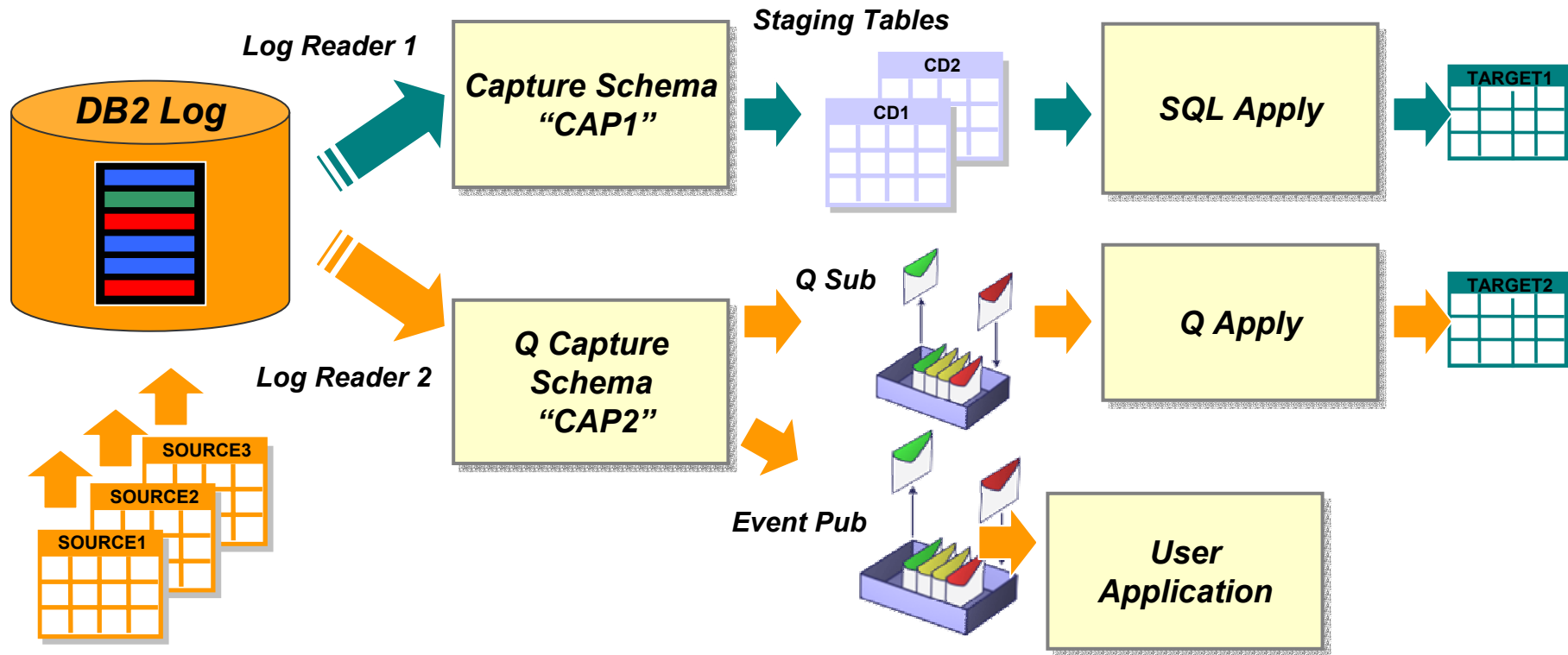
- **Format**

- ▶ 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 :

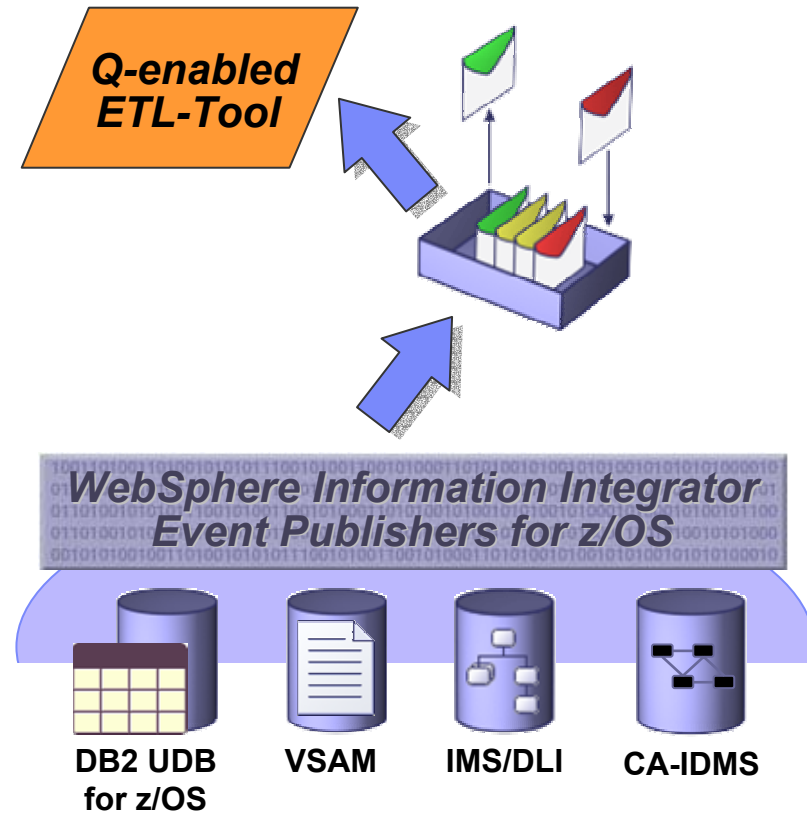
- Data Warehouse
- Datamart
- Operational Data Store (ODS)

Optimize Resource Utilization

- Minimize Bandwidth Requirements
- Maximize Data Currency

Complements with WebSphere II Federation

- Data feed using Event *Publishers*
- Real-time extensions using *Federation*



DB2 II Event Publisher & Business Integration

Data “Events” drive business integration

- Seamless integration with EAI via WebSphere MQ

Data used to drive EAI workflow

- Inventory update hits threshold... triggering restocking process
- Addition of new customer:
 - ✓ Initiates welcome email
 - ✓ Credit verification
 - ✓ Accounting updates
 - ✓ ...

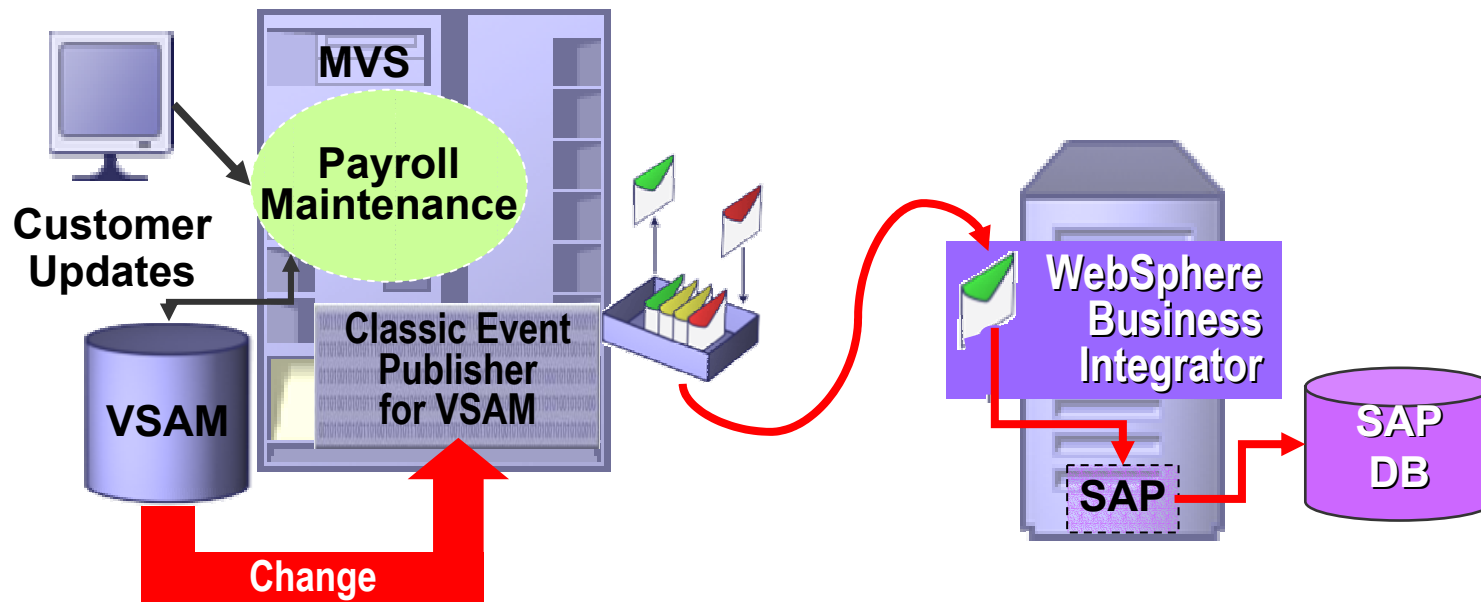
Cross-silo data synchronization

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

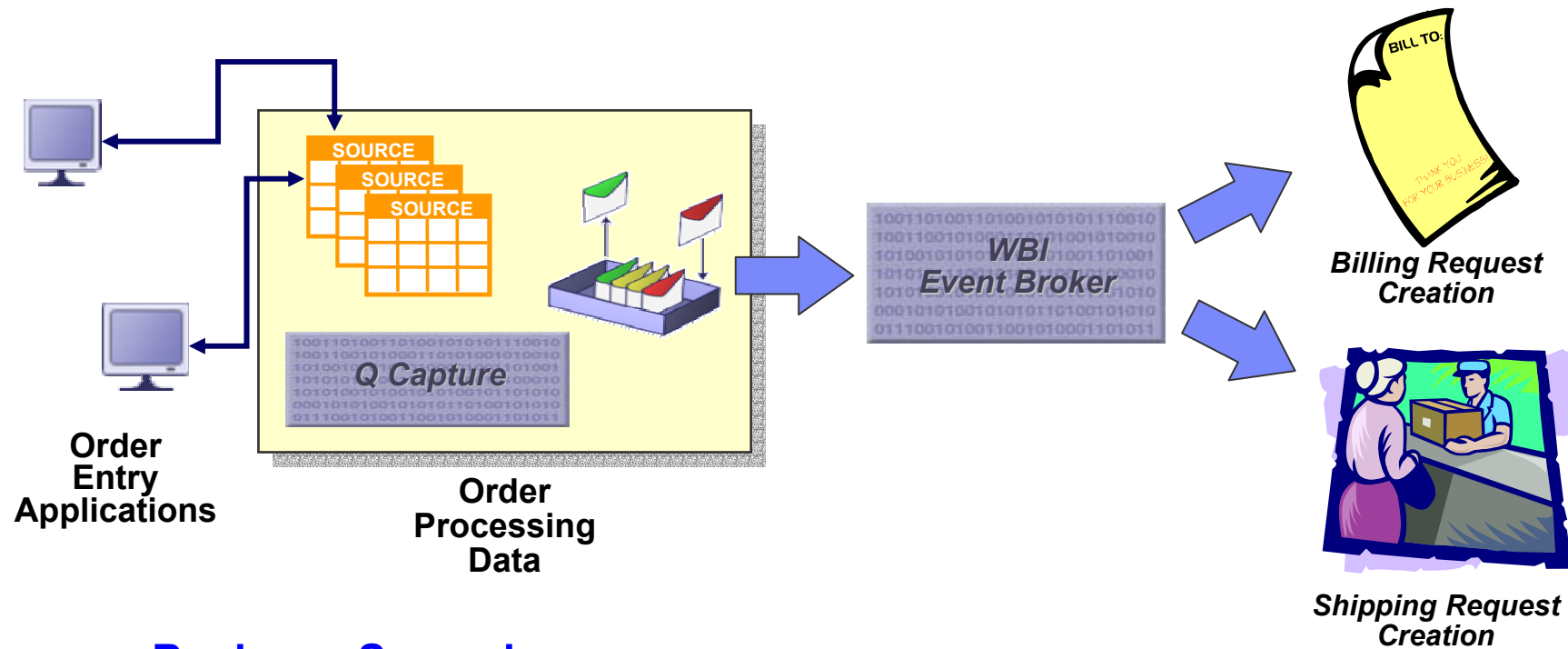


Sample Application

- 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



Order Processing – Exploiting II Event Publishing

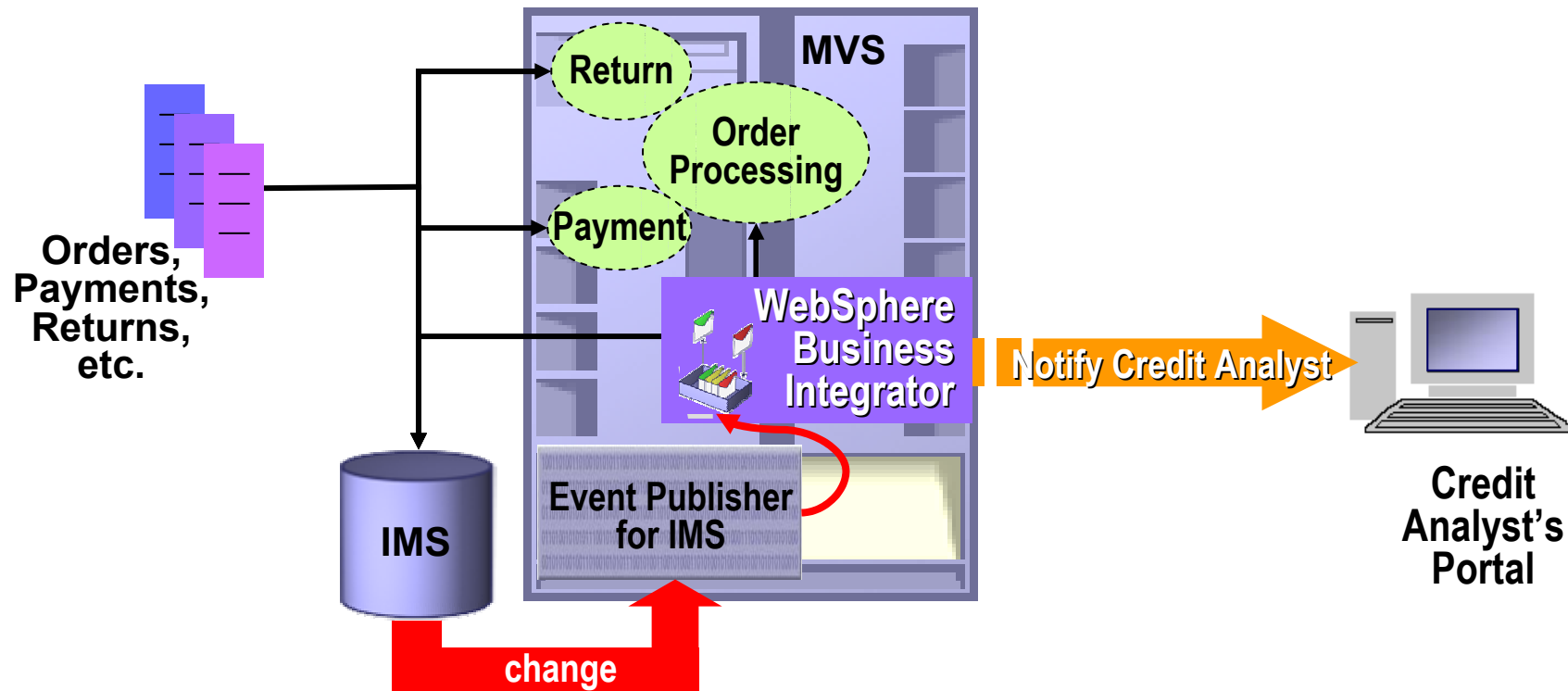


Business Scenario

- 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

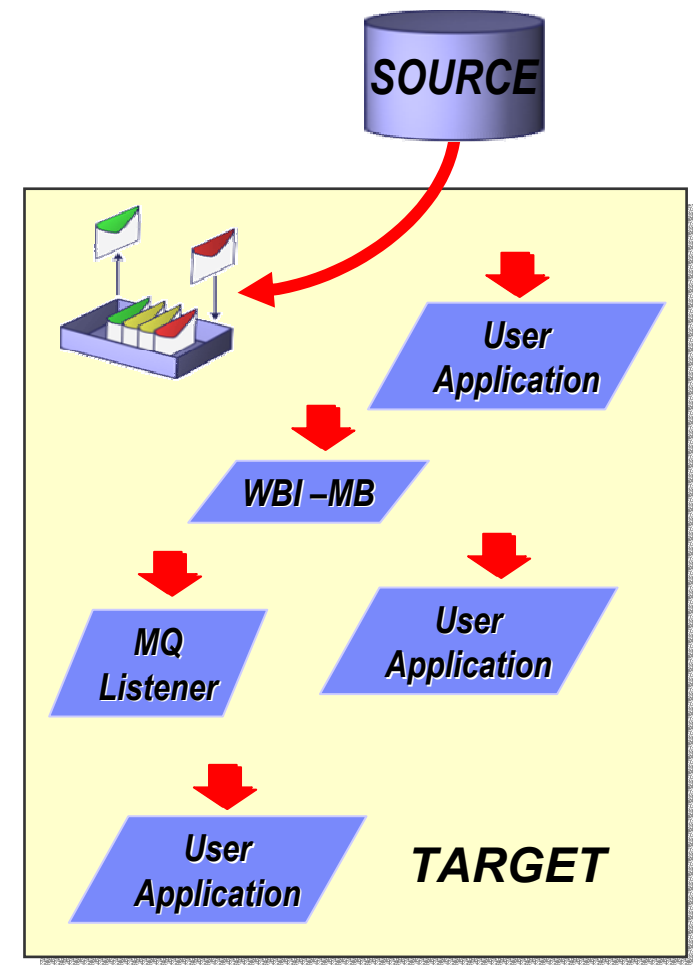
- Event Notification
 - ▶ Receivable balances approaching credit limit pushed to a credit analyst
 - Threshold is independent of order processing and accounting applications
 - No “hard-hooks” in OLTP applications necessary



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***



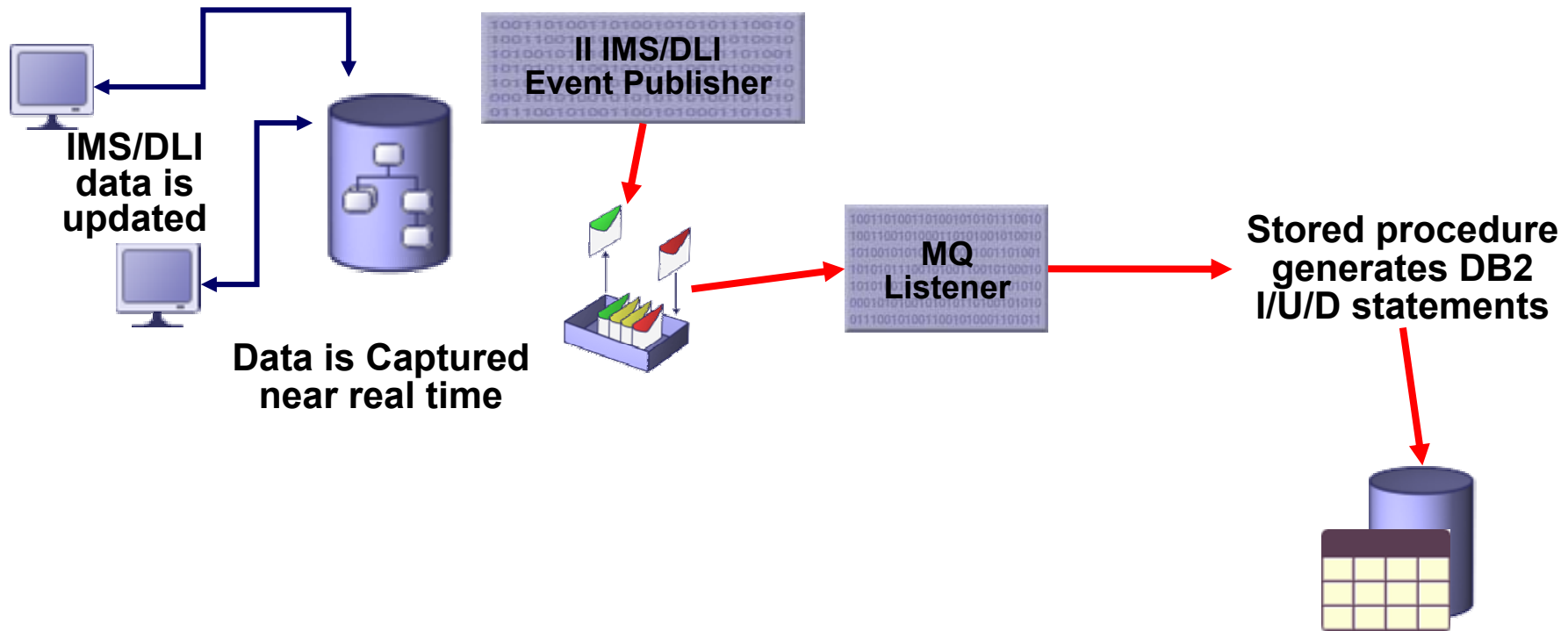


IBM Software Group | WebSphere Information Integration Software

Summary



Using Classic II Event Publishing for Replication



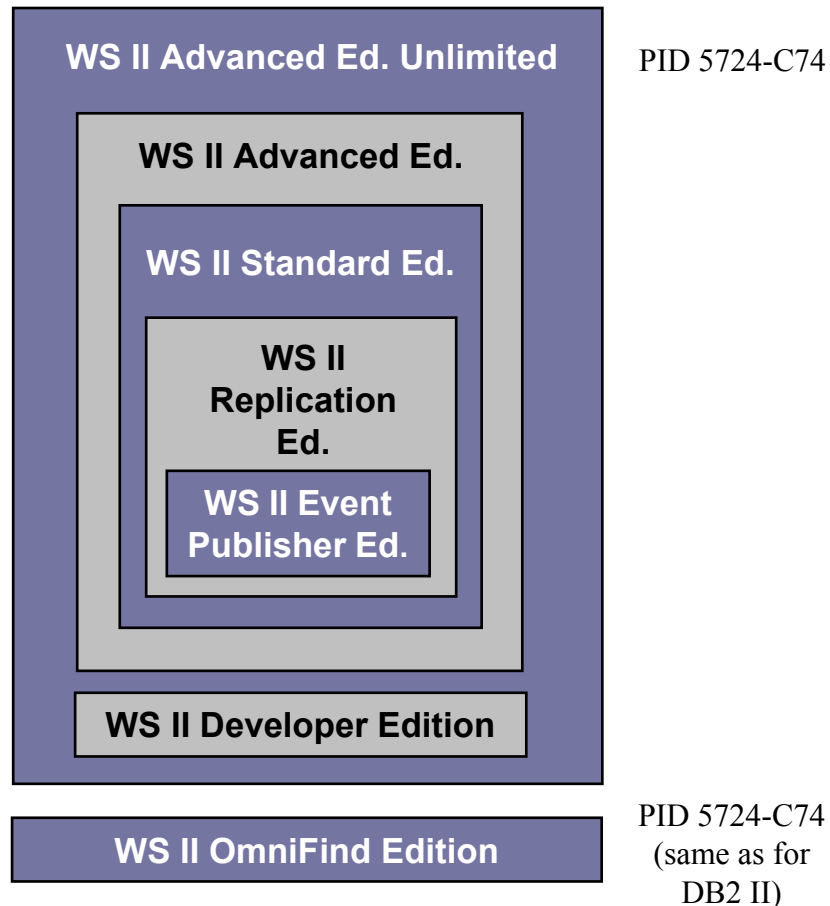
- It is our direction to support replication of Classic data sources using the highly parallel Q Apply of DB2 II Replication.
- As an interim suggestion, SQL stored procedures can be used to apply the data captured through the II Classic Event Publisher

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

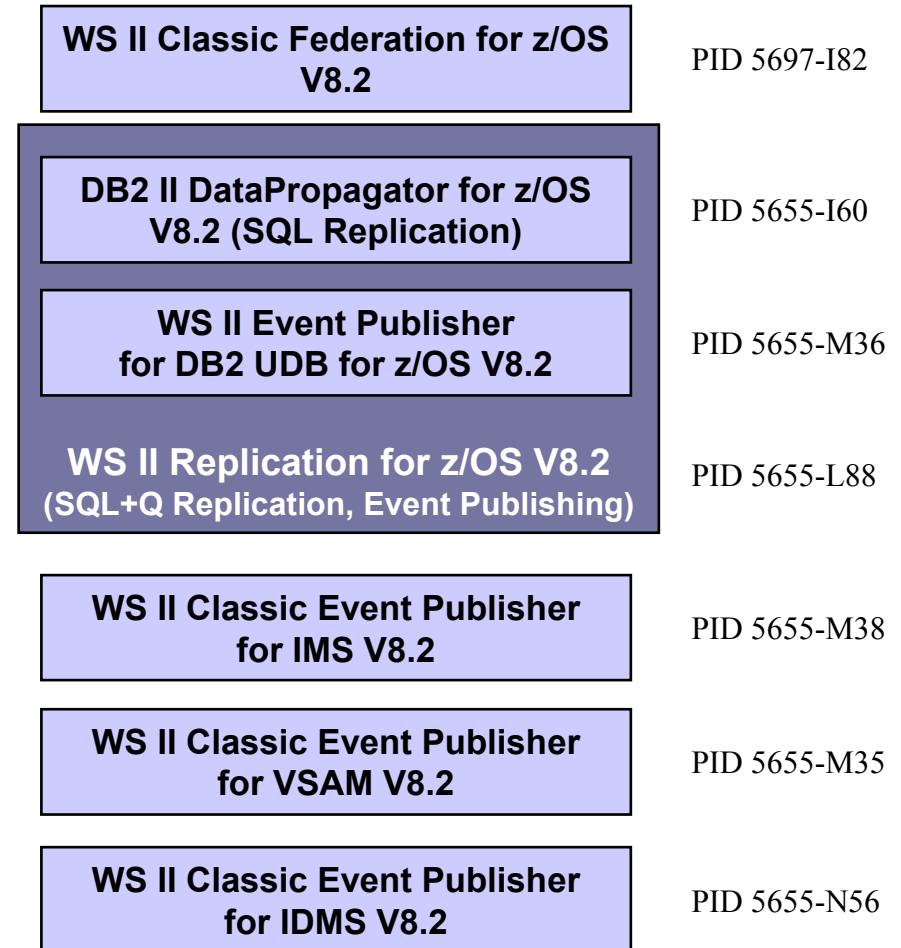
WebSphere Information Integrator Packaging and PIDs

Distributed (Linux, UNIX, Windows)



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

Mainframe (IBM eServer zSeries)



- Value Unit pricing Model