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

Defining Business 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***

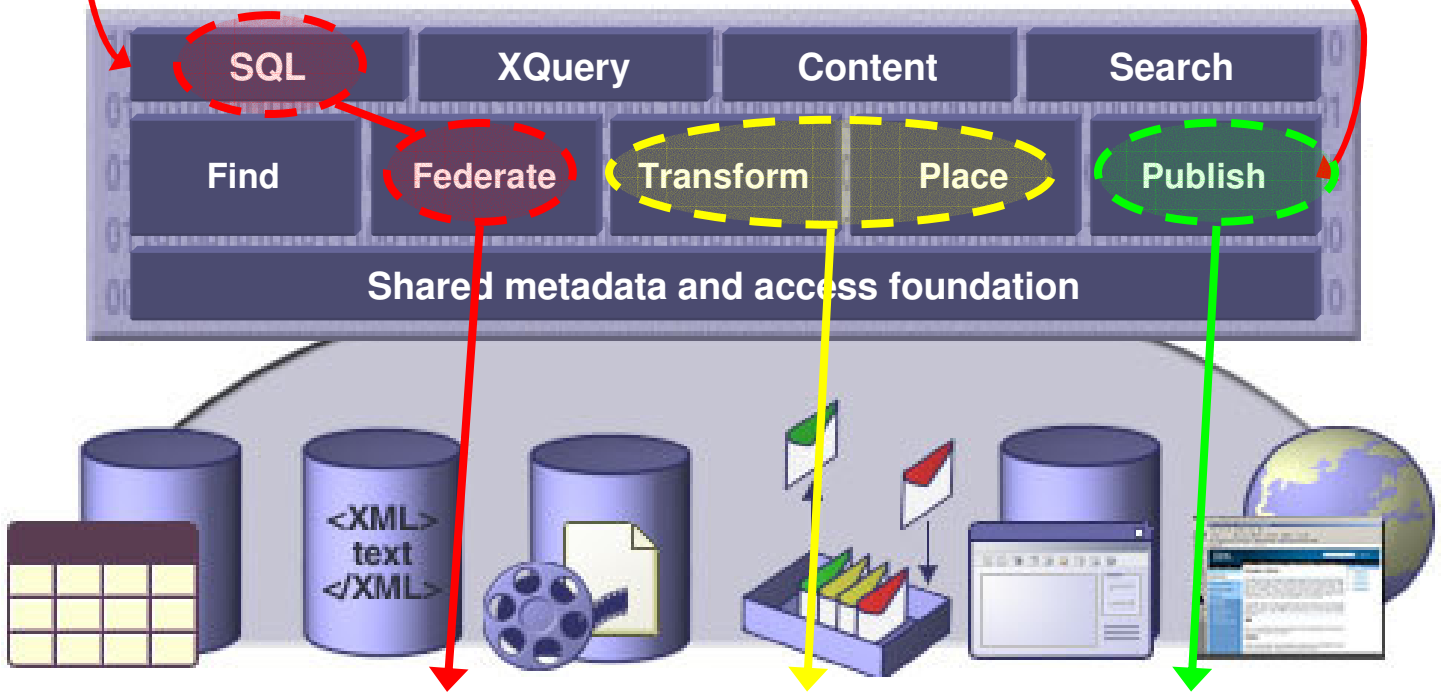
Can users react in real-time to the most recent information? → ***People, Information***

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

IBM Information Integration Vision

Any Data

- ***Multiple access paradigms***
- ***Multiple integration disciplines***



zSeries II Offering :

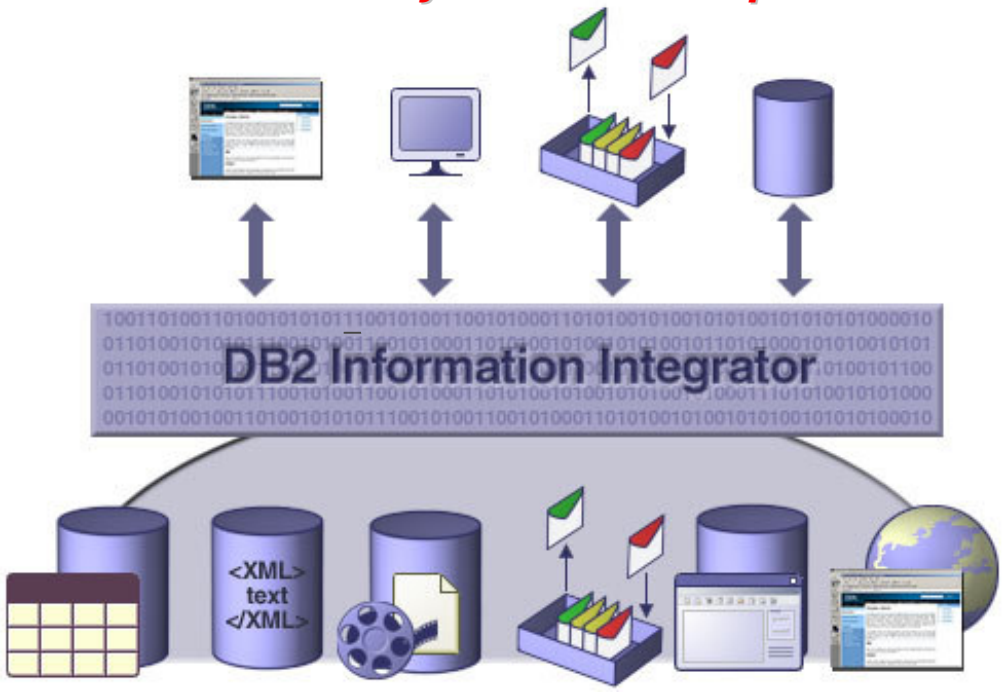
Federated Data

Replicated Data

Event Publishing

IBM WebSphere Information Integration Software

Integrating diverse business information across and beyond the enterprise



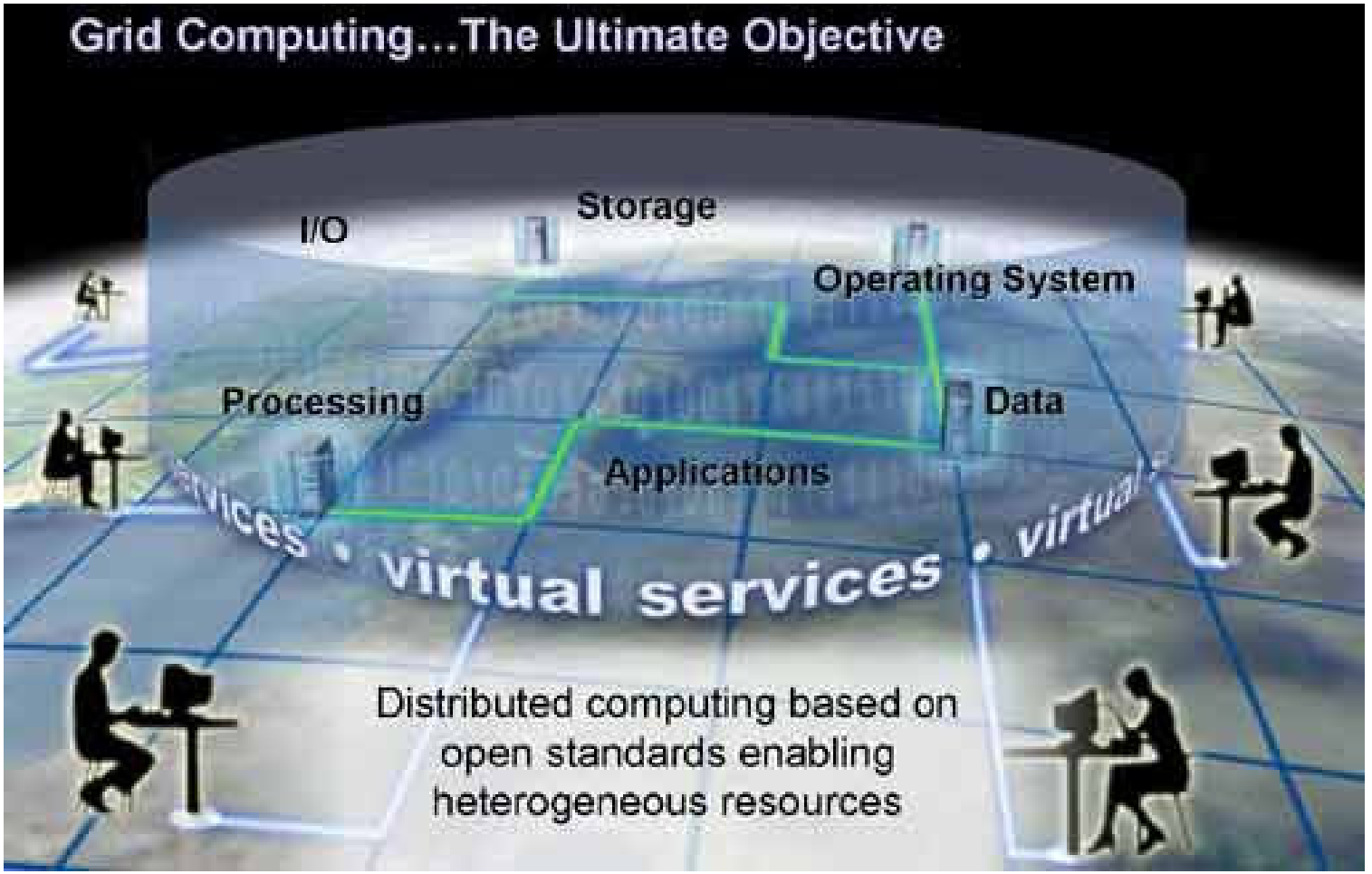
Information Integration :

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

- Data federation
  - Extensible read/write access across diverse data and content sources
  - Database programming model (SQL)
  - Content programming model (OO API)
- Data placement
  - Caching and replication over heterogeneous information
- Data transformation
  - SQL, XML, Web services
  - Advanced search and mining
  - Metadata management
- Part of a complete integration solution
  - XML publishing, consumption, and interchange
  - WebSphere business integration
  - Open platform based on industry standards

IBM Strategy: Heterogeneous Information GRIDs

Getting Access to Information regardless of where it resides …

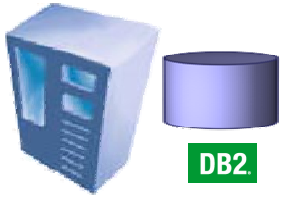


… leveraging existing Assets of an open on demand Infrastructure

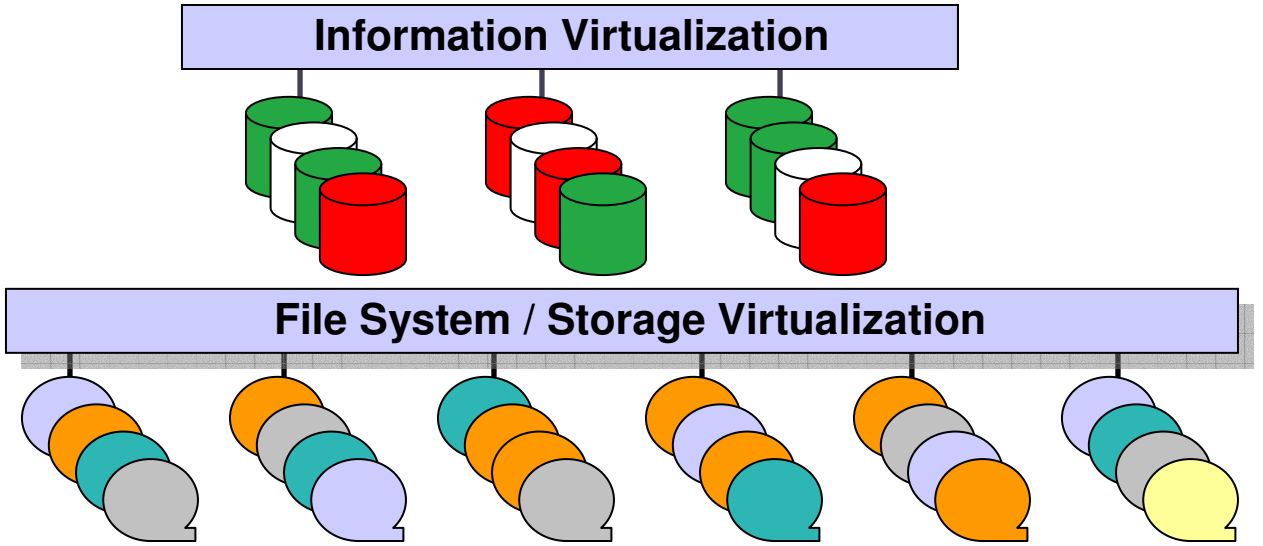
Using Websphere Information Integrator as GRID Enabler

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

Federation instead of Centralization



WebSphere Information Integrator



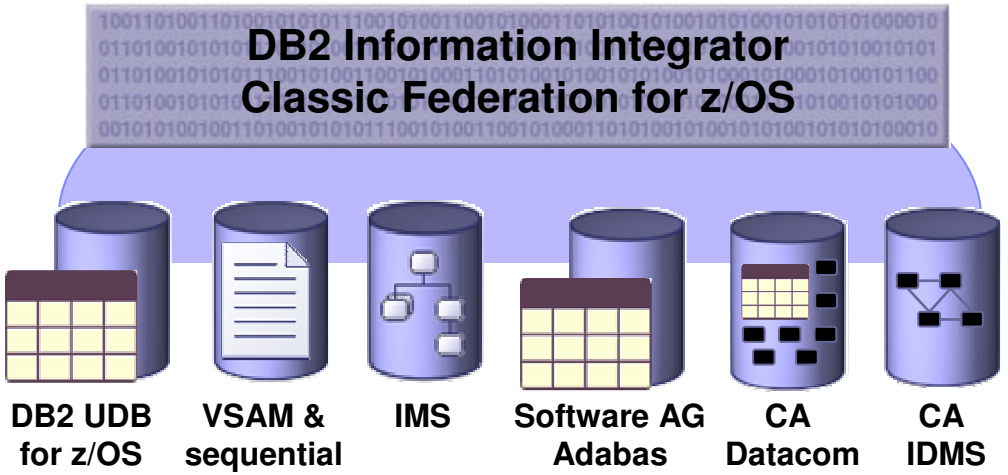
Federated Data Server

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

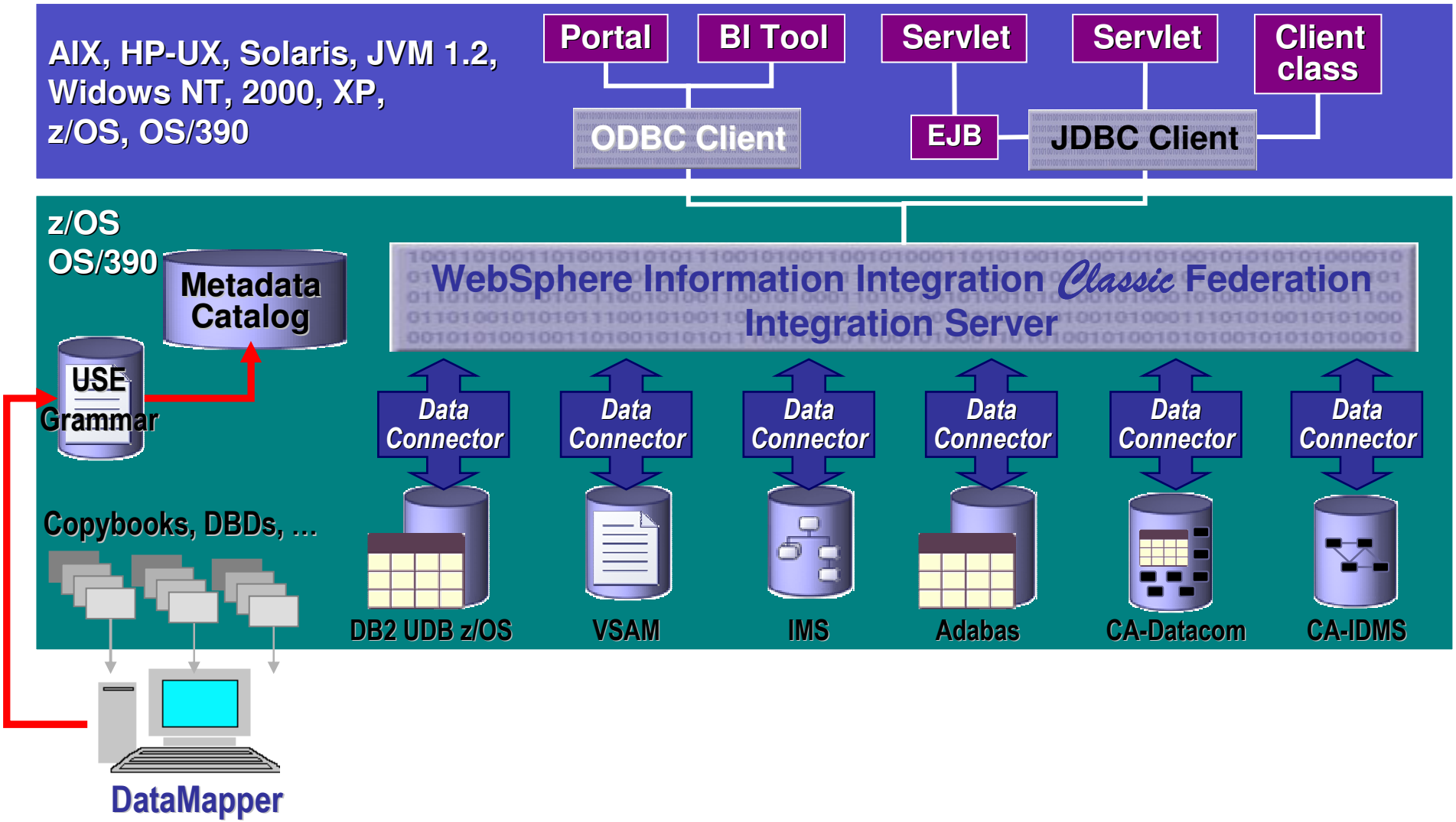
WebSphere Information Integrator Classic Federation for z/OS

- 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

- WebSphere 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 Integratror *Classic* Federation

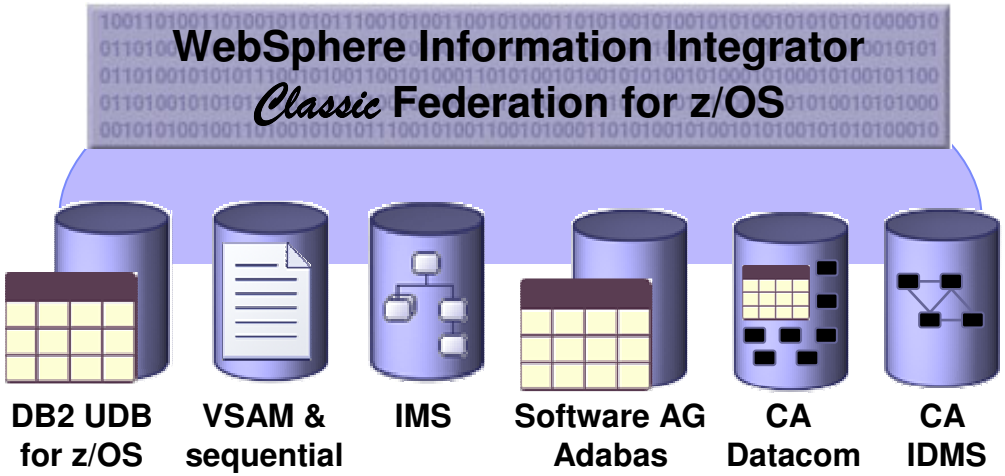


WebSphere II Classic Federation: Standard SQL 92 Support

- SELECT/INSERT/UPDATE/DELETE all supported
- Standard SQL error handling
  - SQL error and response codes returned as part of result
- Single and two phase commit
  - Commit - Rollback - Autocommit supported for all data sources
  - Rolling delivery of two phase commit: DB2, IMS and CA-Datacom available now
- Stored Procedure “Call” leverages existing programs
  - Reuse mainframe algorithms
  - Invoke IMS transactions
- DBCS Support
  - Client-based conversion
  - IMS, VSAM, DB2, CA-IDMS, sequential in v8.2
    - Adabas mixed mode available, full graphic data type support planned
    - CA-Datacom planned

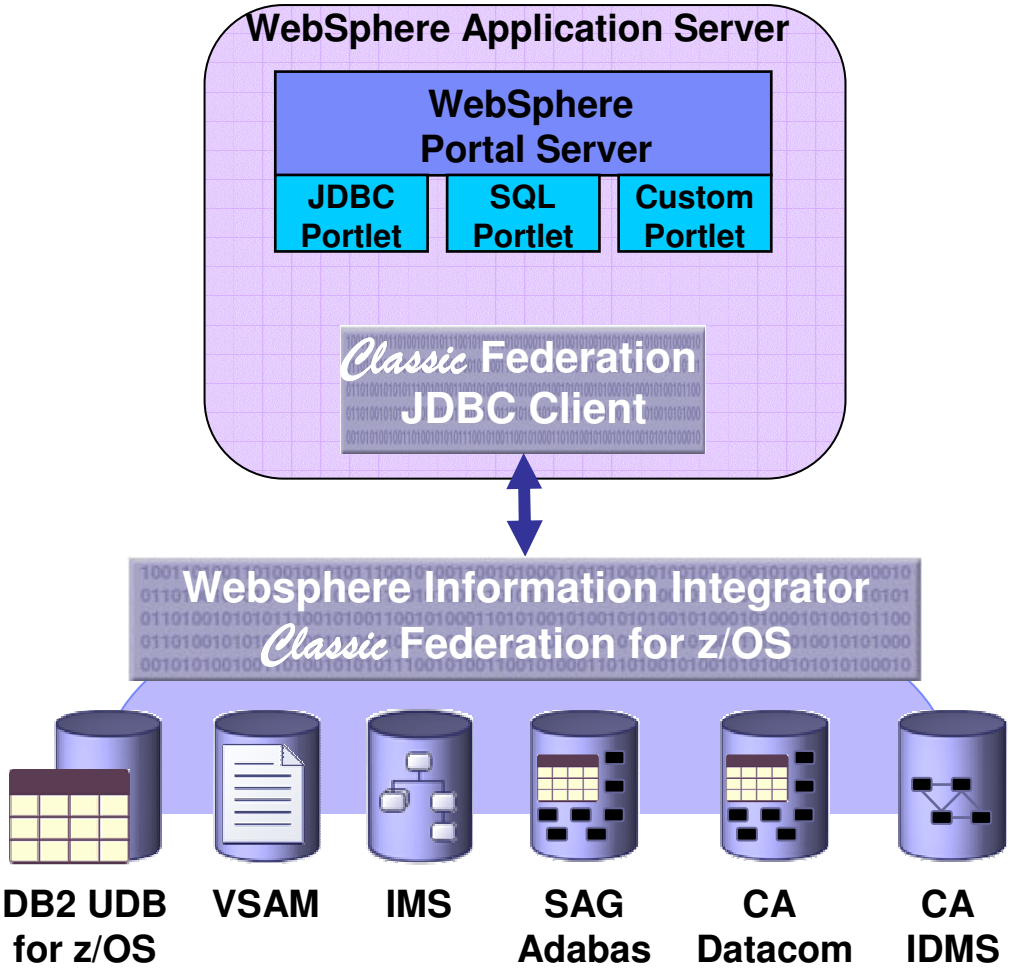
Platform Support

- Operating Systems
  - Server:
    - z/OS 1.4
  - Clients:
    - AIX 5.1.0, HP-UX 11.01, Solaris 2.7
    - Windows* (NT, 2000, XP or Server 2003)
- Communications
  - TCP/IP or MQ Series v5 or higher
- Databases
  - DB2 UDB for z/OS 6.1 or 7.1
  - IMS/DB 7.1
  - CA-IDMS 13 or 14
  - CA-Datacom 10
  - Adabas 7.1



* Data Mapper requires NT, 2000 or XP

WebSphere and II Classic Federation

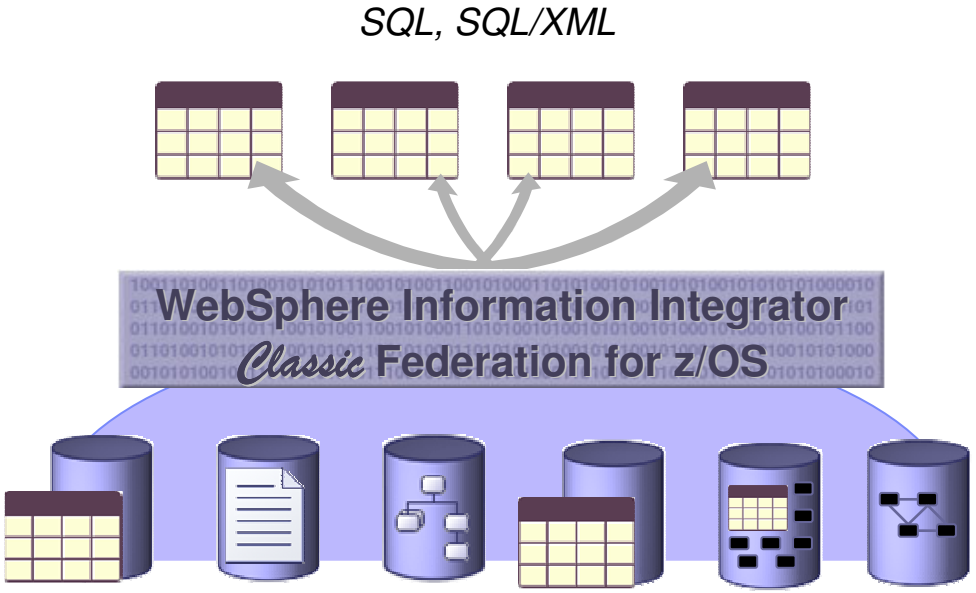


- Integrate mainframe data directly with Web applications, Portals, EAI platforms
- Supports JDBC 2.1 Core APIs Plus
  - Scrollable/Updatable ResultSets
  - Commit – Rollback – AutoCommit - XA
  - Stored Procedure to mainframe programs
  - Parameter Markers
  - Metadata commands for tables, columns,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

Integrating Classic Legacy Data with WS II Classic Federation

Access various kinds of legacy data sources

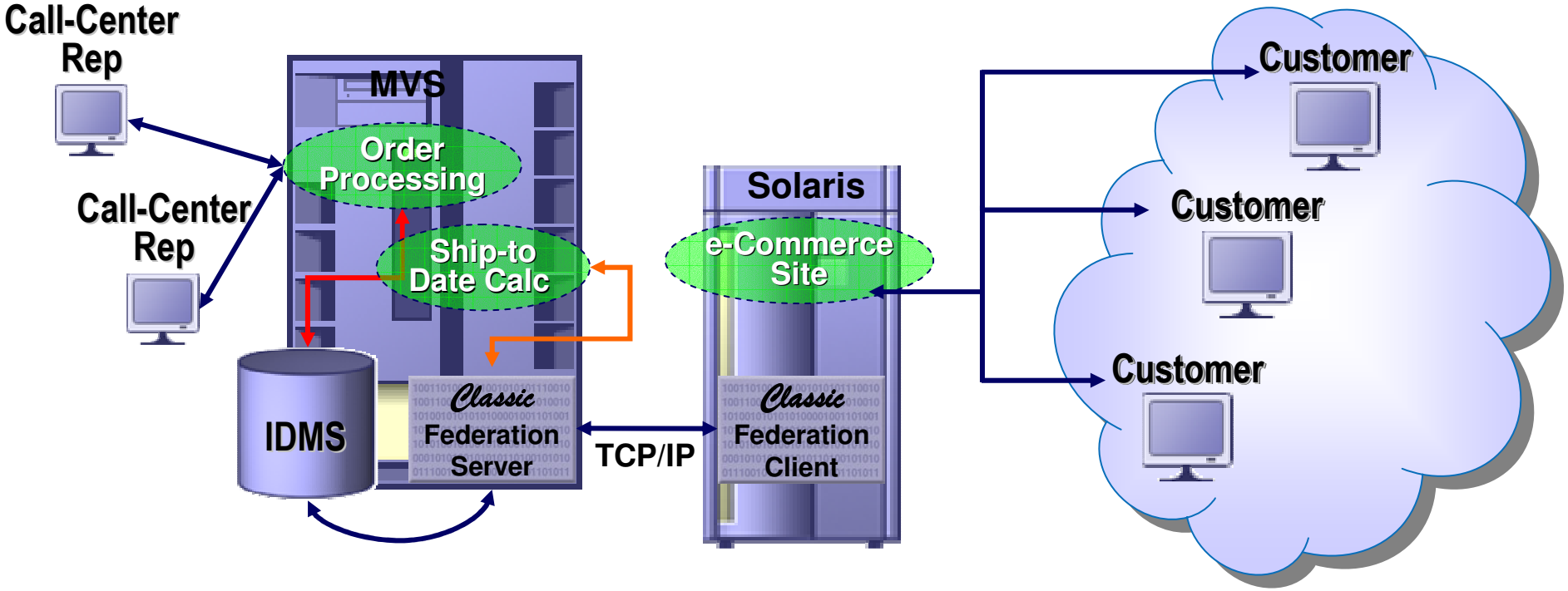
- Native SQL/JDBC Interface not available
- Programming skills falling short
- Mapping of non-relational legacy Data into relational Format
- Access Legacy Data through JDBC Interface
- Legacy Sources can be mapped into WebSphere II Nicknames



Integration in Action – European Catalog House

Seamlessly share order processing data and algorithms between:

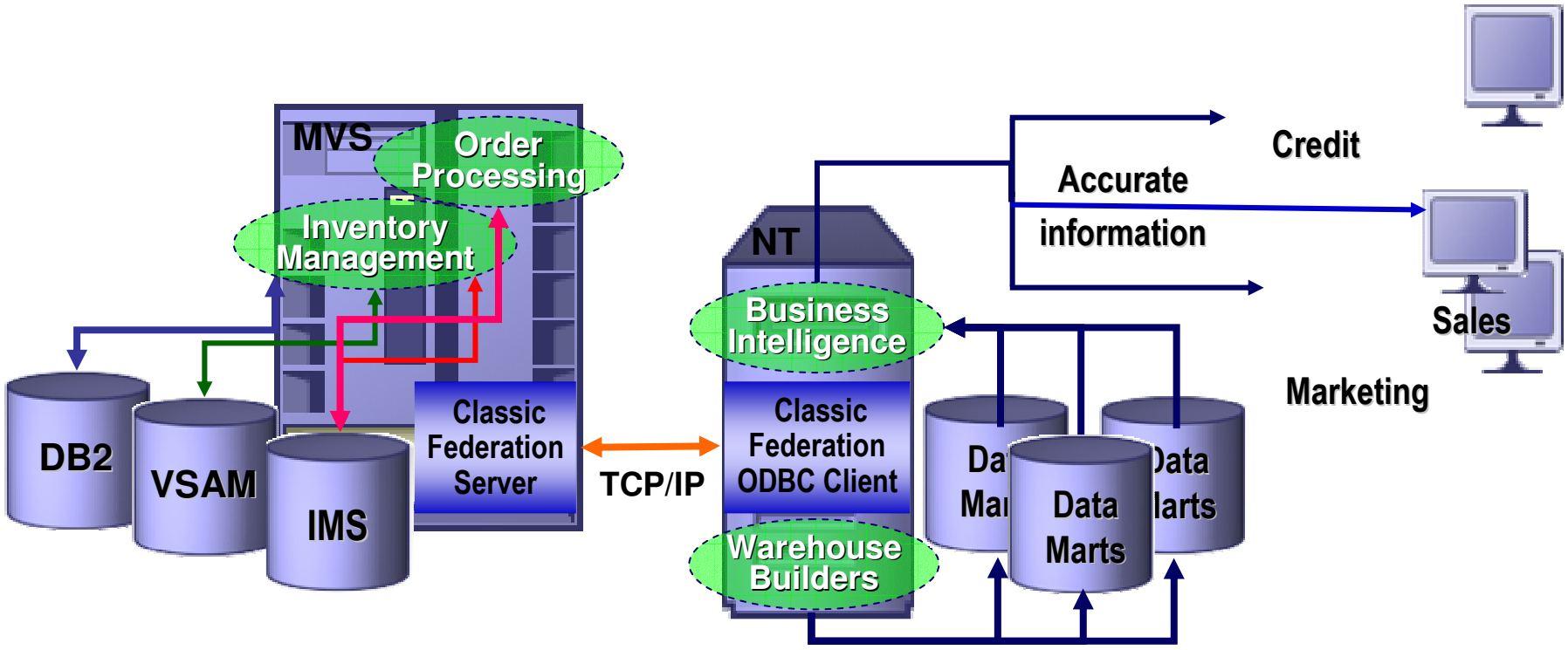
- Legacy call-center systems
- New e-commerce applications
- No mainframe skills required for e-commerce site development



Integration in Action – Recreation Vehicle Manufacturer

Provide accurate inventory and sales data to data warehouse

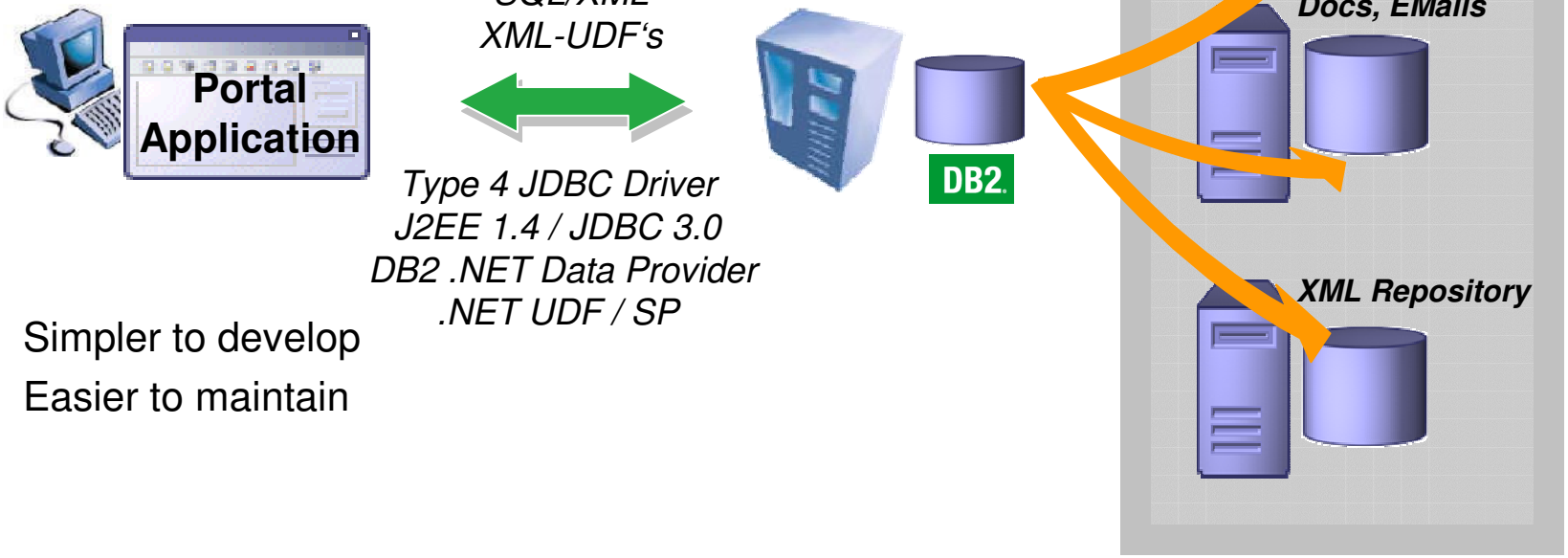
- Dynamically connect data warehouse tool with mainframe data
- Cut development time in half
- Accelerate product delivery with warehouse "pull" of new shipment data



Speeding Portal Development

Business Scenario

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

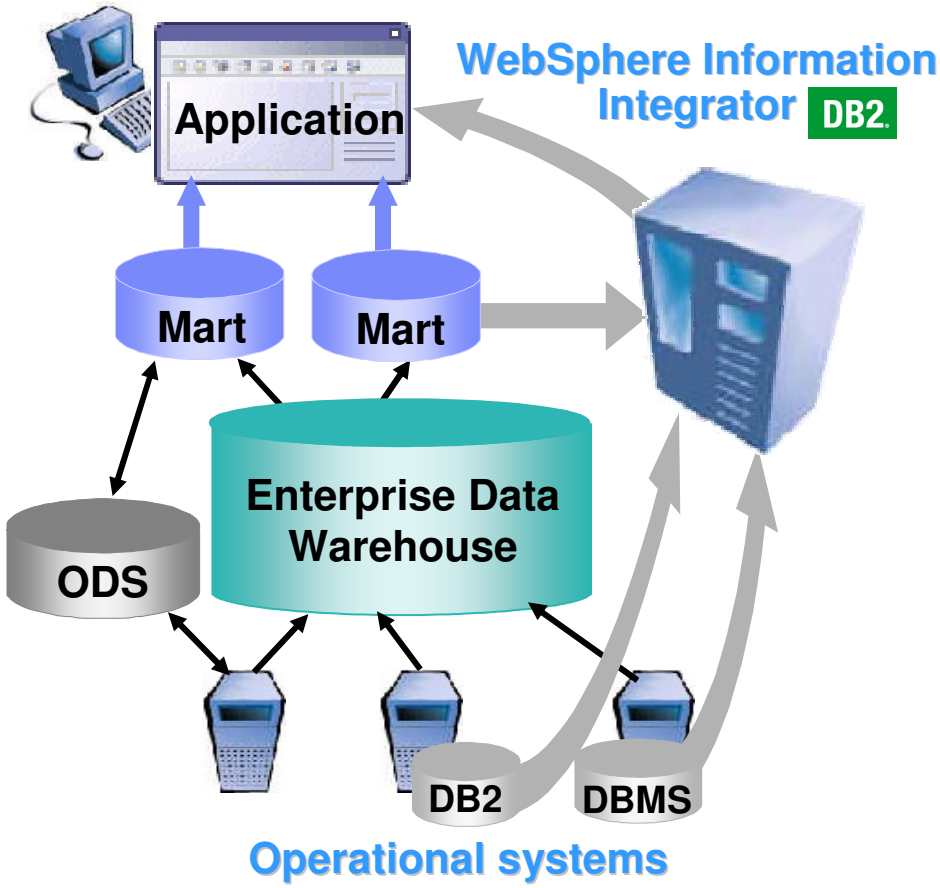


- Simpler to develop
- Easier to maintain

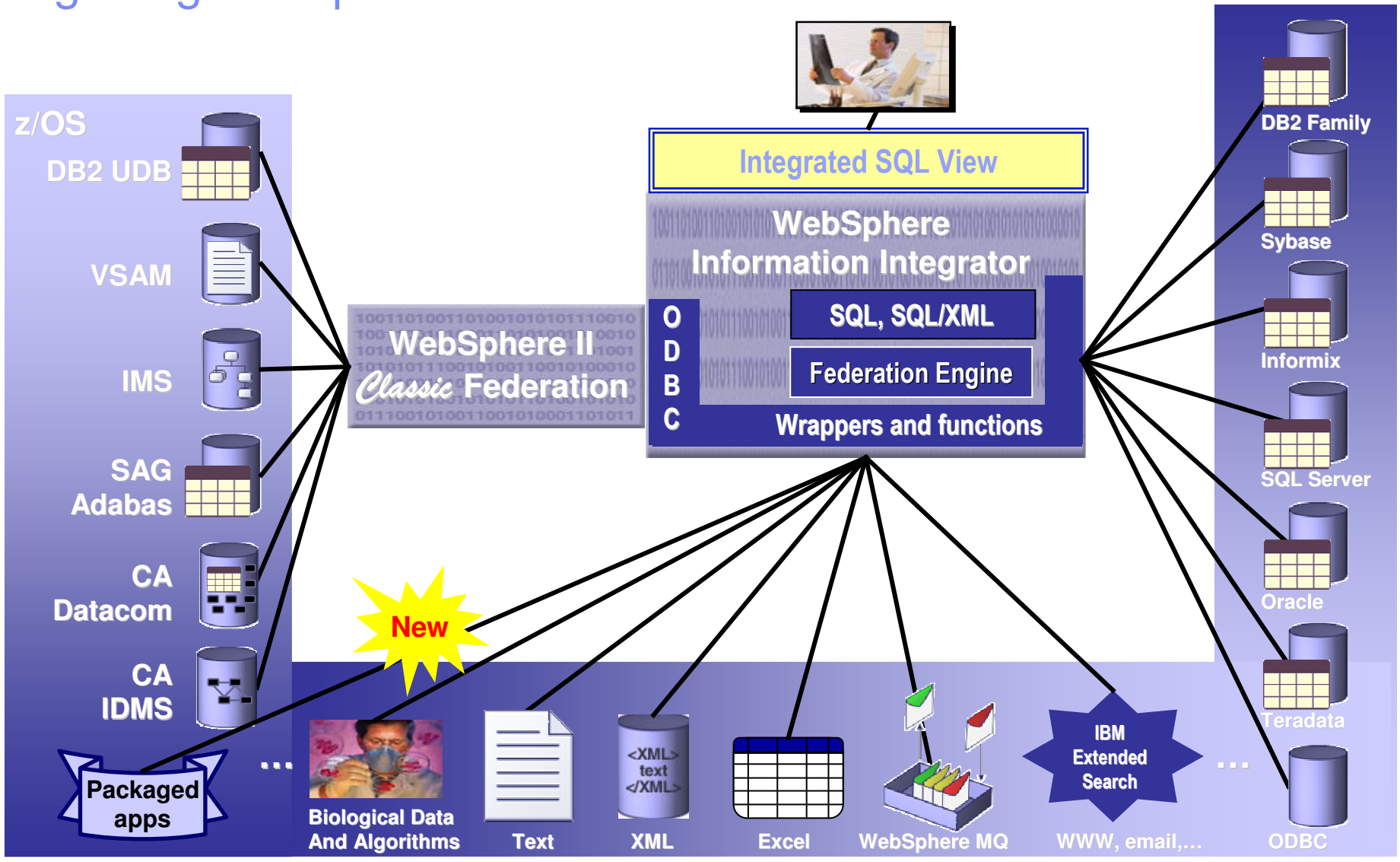
Extending Data Warehouses with Real-Time Data

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



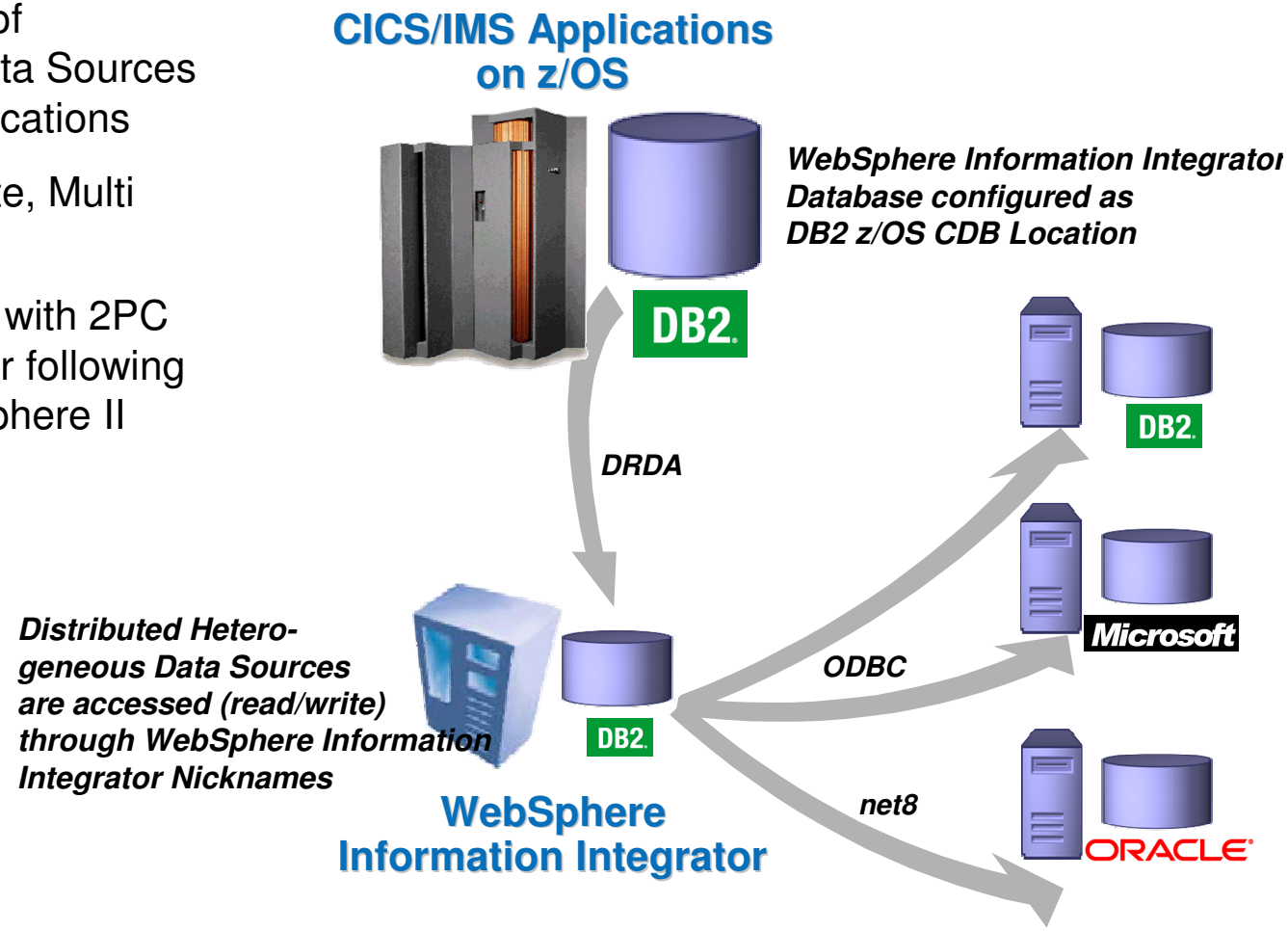
Integrating Enterprise Data



Access Distributed Heterogeneous Data from CICS Applications

Business Scenario

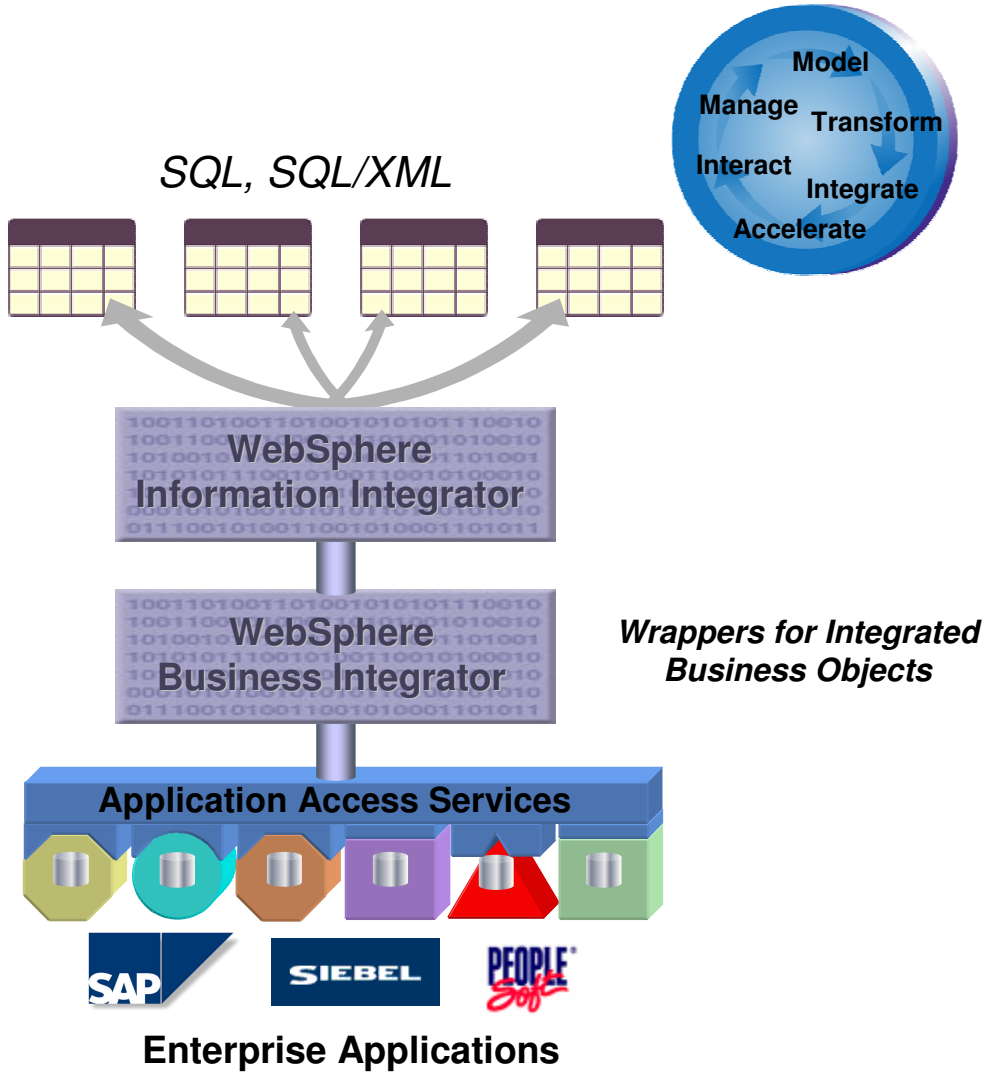
- Access all kinds of WebSphere II Data Sources from legacy Applications
- Single Site Update, Multi Site Read
- Distributed UOW with 2PC support in plan for following version of WebSphere II



Information and Process Integration with WebShpere

Business Scenario

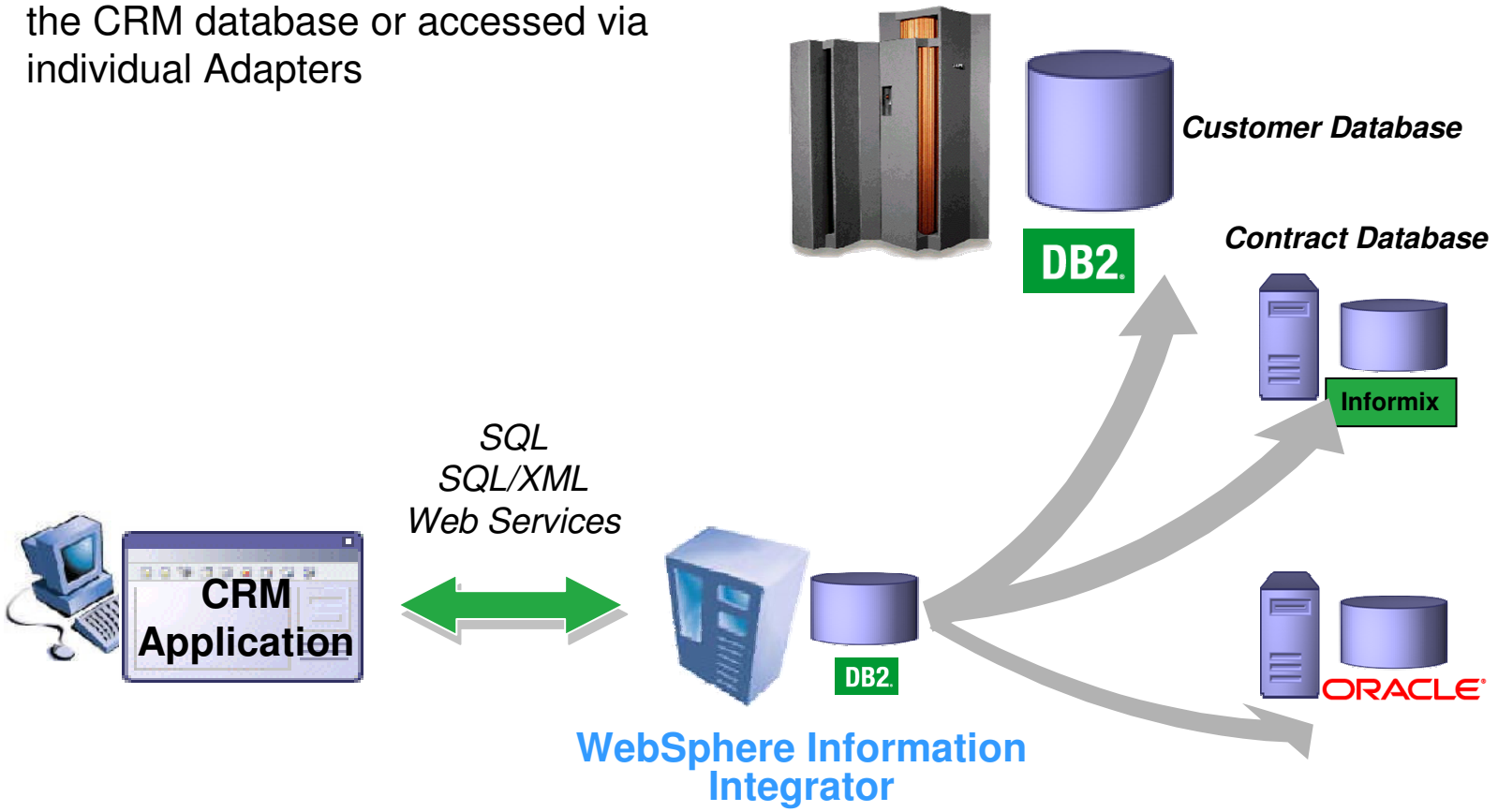
- Enterprise Applications provide APIs for Business Object/Component Retrieval
- Enterprise Business Components can be mapped into relational Format using WebSphere components :
  - ✓ Information Integration
  - ✓ Business Adapters
- Business Objects can be joined with other relational / non-relational Information



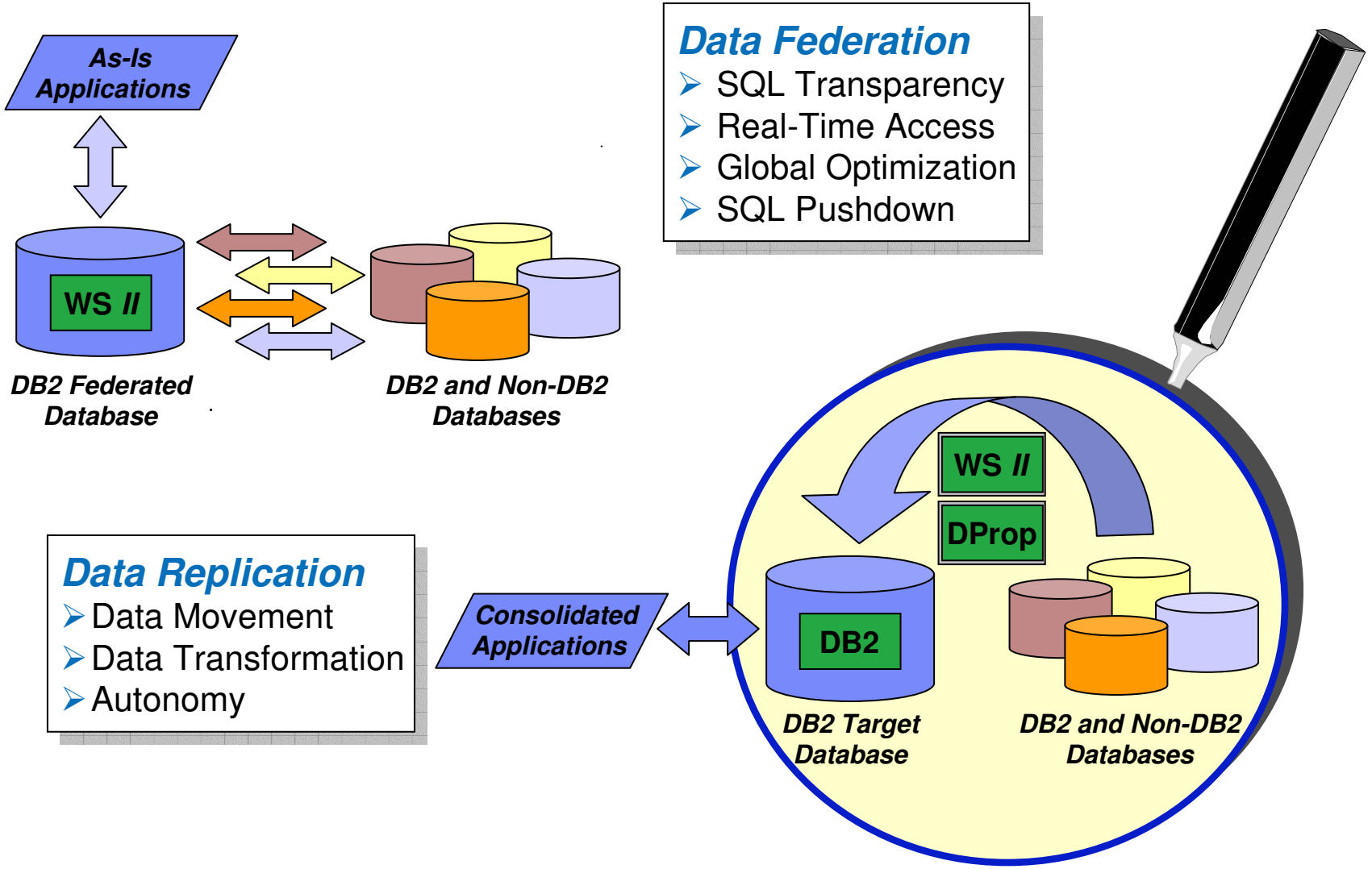
Enabling Customer 360° Analysis

CRM Systems offer 360o Analysis

- Without federated Information, all Customer Data has to be moved into the CRM database or accessed via individual Adapters



Federate or Replicate – That is the Question !!

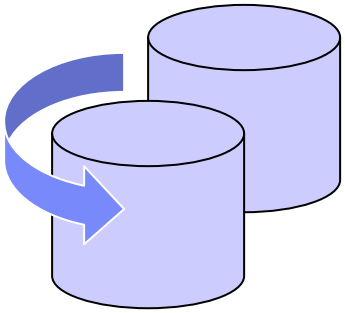


Replicated Data

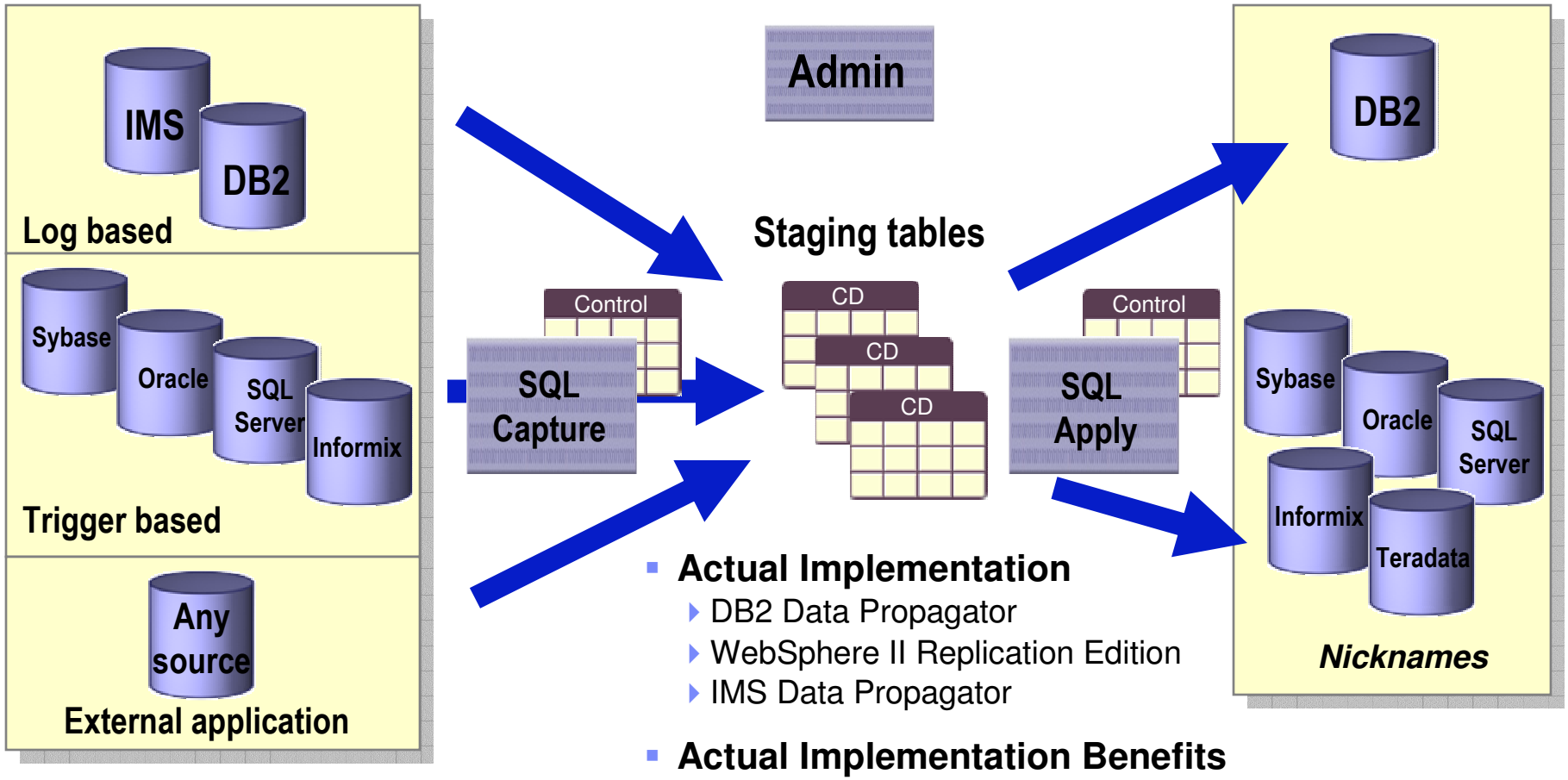
- ***SQL Replication***
- *Q Replication*

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

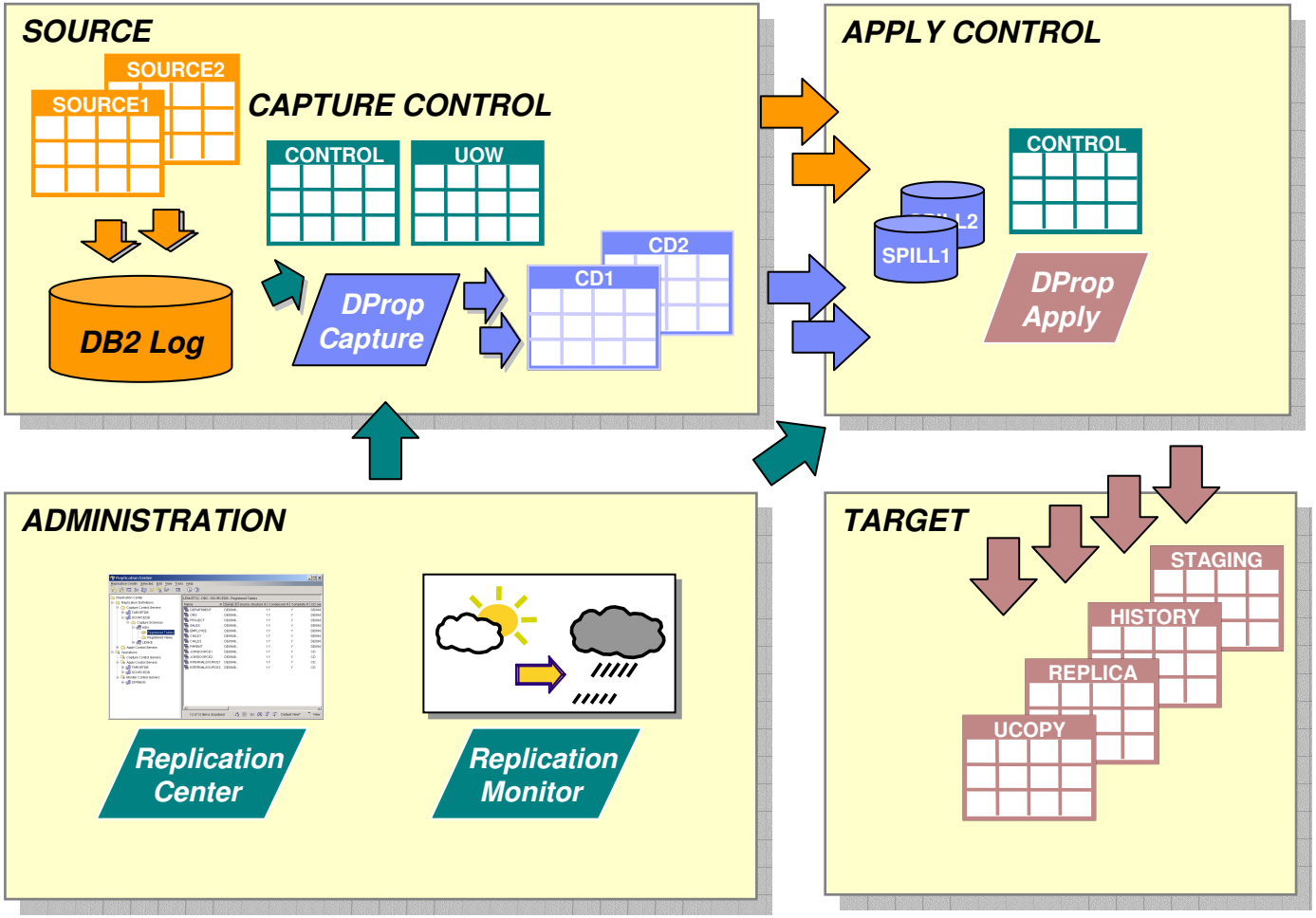


SQL Replication Architecture

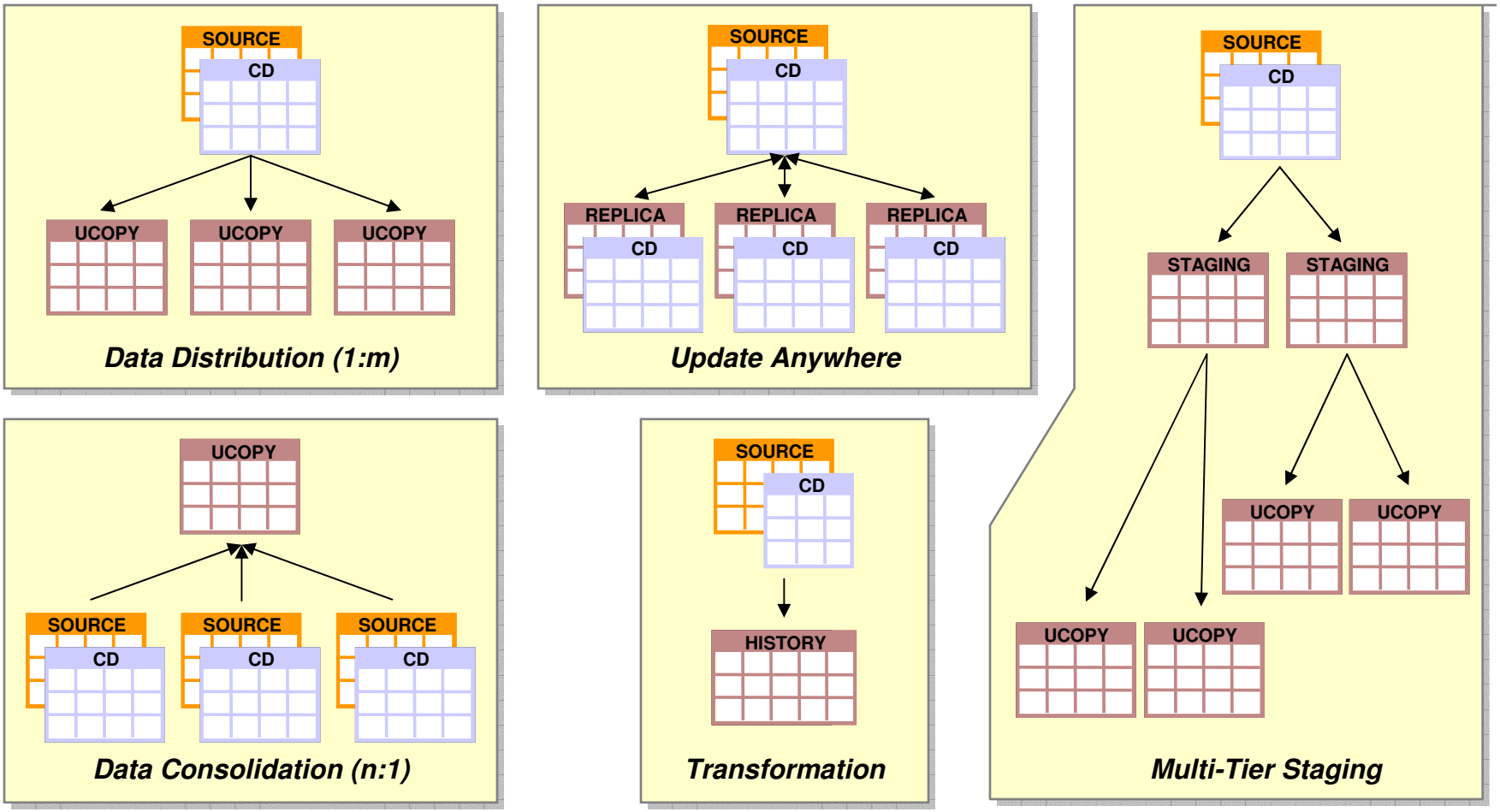


- **Actual Implementation**
  - DB2 Data Propagator
  - WebSphere II Replication Edition
  - IMS Data Propagator
- **Actual Implementation Benefits**
  - Extremely flexible and resilient
  - Very easy to set up transformations
  - Scales well to reach multiple targets
  - Homogeneous & Heterogeneous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

SQL Replication Component Overview

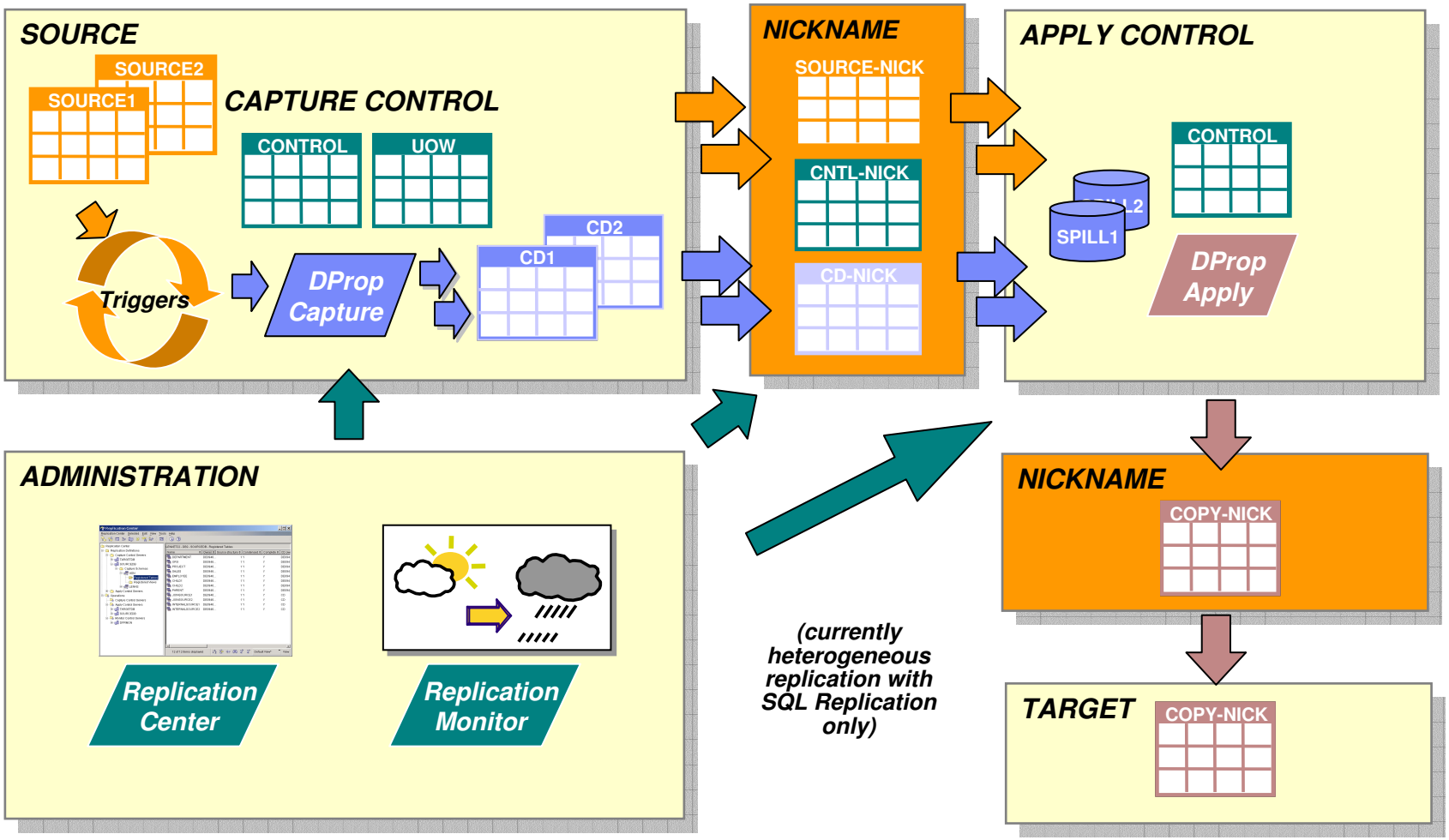


Sample SQL Replication Scenarios



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

DB2 Data Replication from Federated SOURCES



Federated Sources & Targets

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

Replicated Data

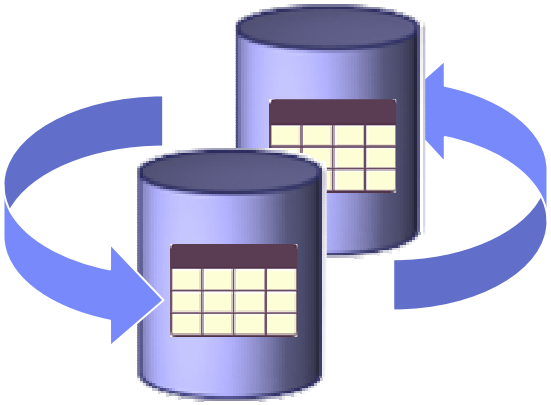
- *SQL Replication*
- ***Q Replication***

Why Create Another Replication Architecture?

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

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

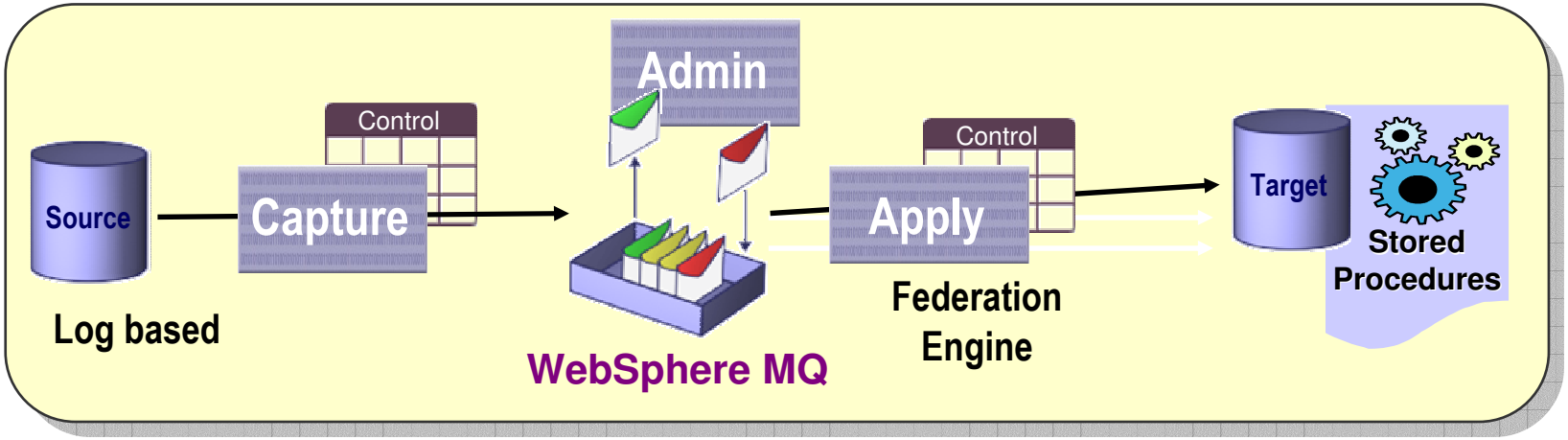
- **Capability**
  - Significantly improve multi-directional Replication Support



- **Manageability**
  - Reduce the number of Replication Objects to be defined and managed
  - Ease the Definition Process with new Replication Center Wizards

Q Replication Architecture

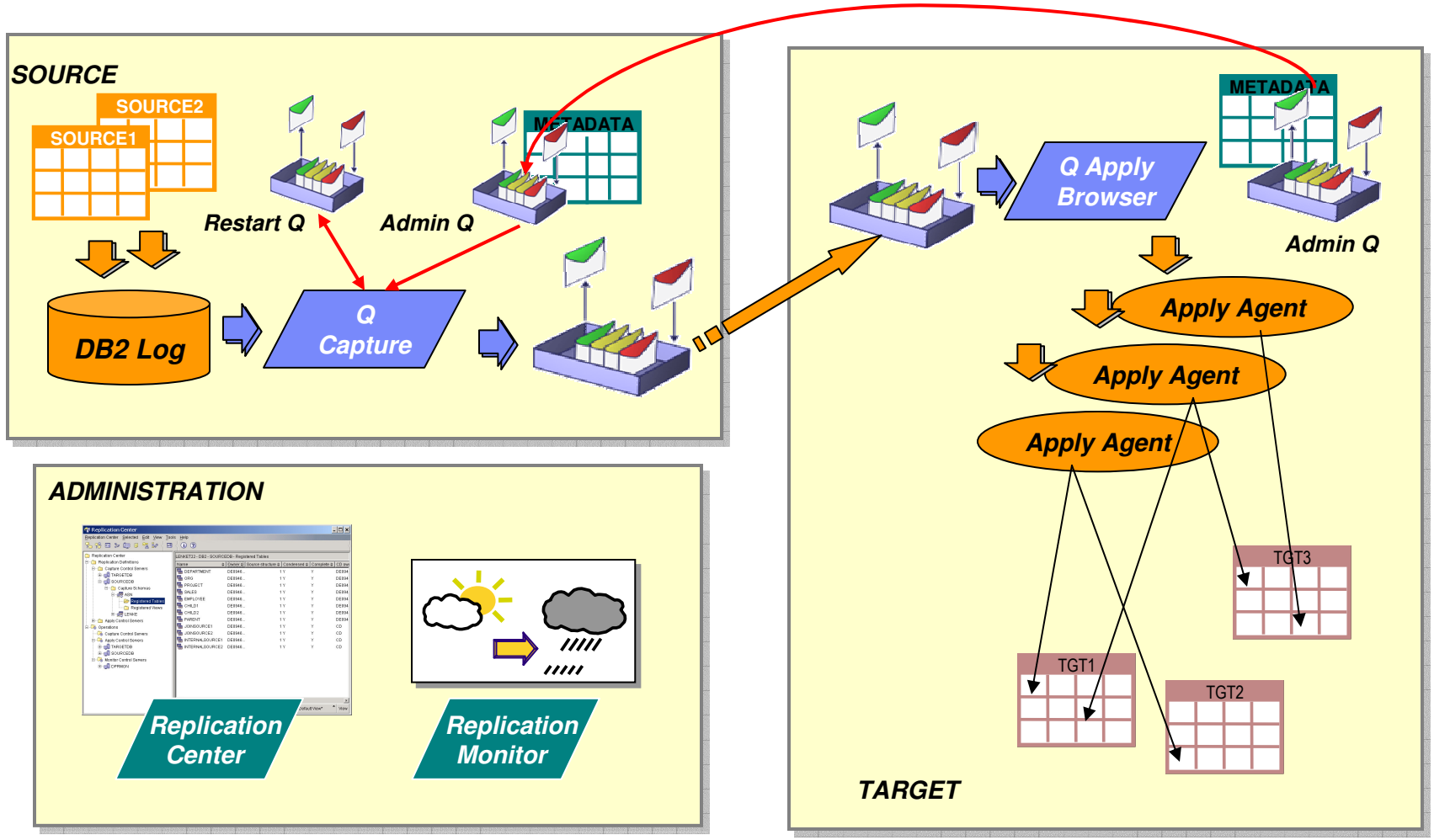
WebSphere Information Integrator introduces new replication architecture for delivering extremely low latency replication



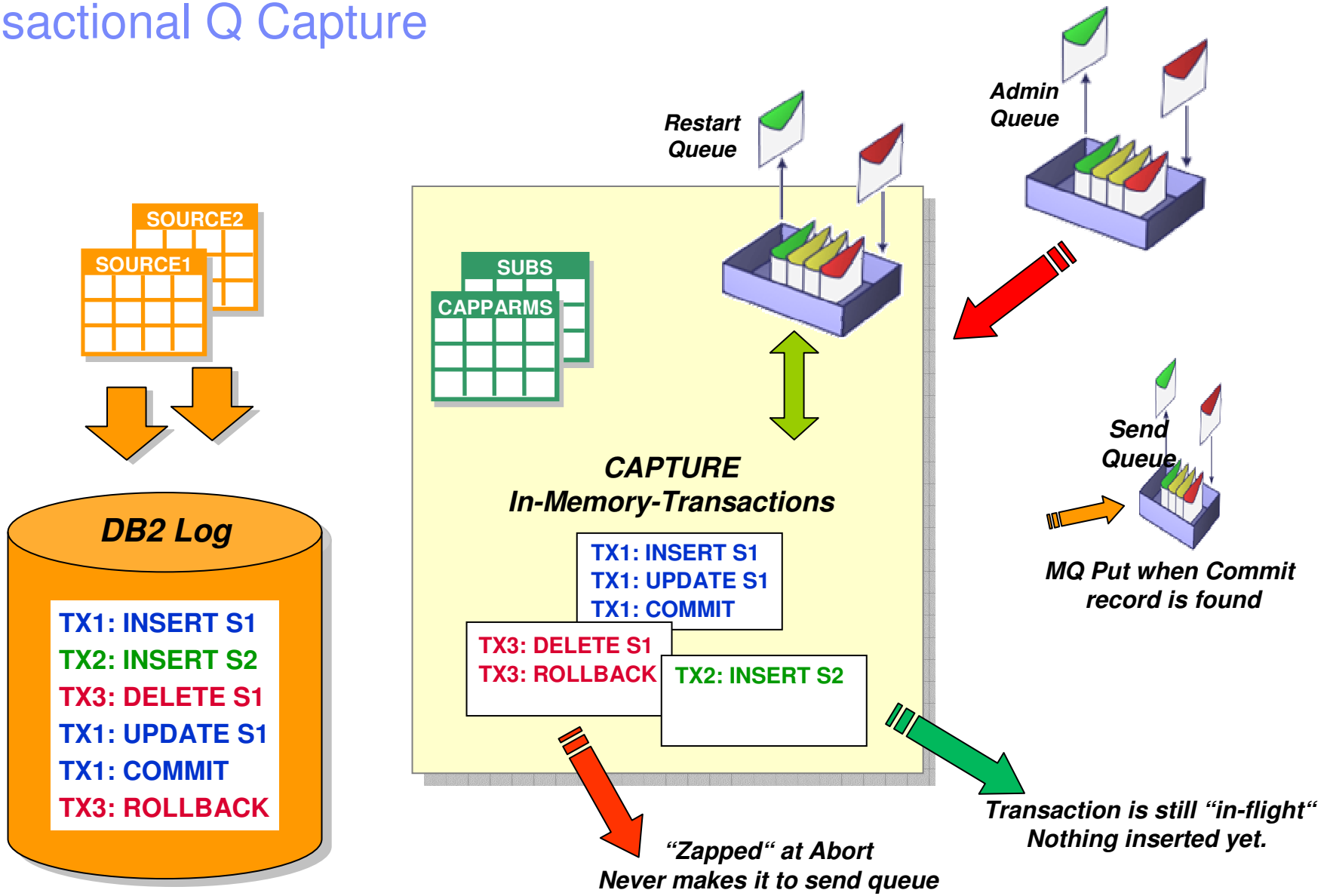
- Each message represents a transaction
- Highly parallel apply process
  - Non dependent transactions re-parallelized at the target
- Differentiated conflict detection and resolution
- Integrated infrastructure for replication and publishing
- DB2 to DB2 today

- 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 – Q Subscription Process



Transactional Q Capture



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

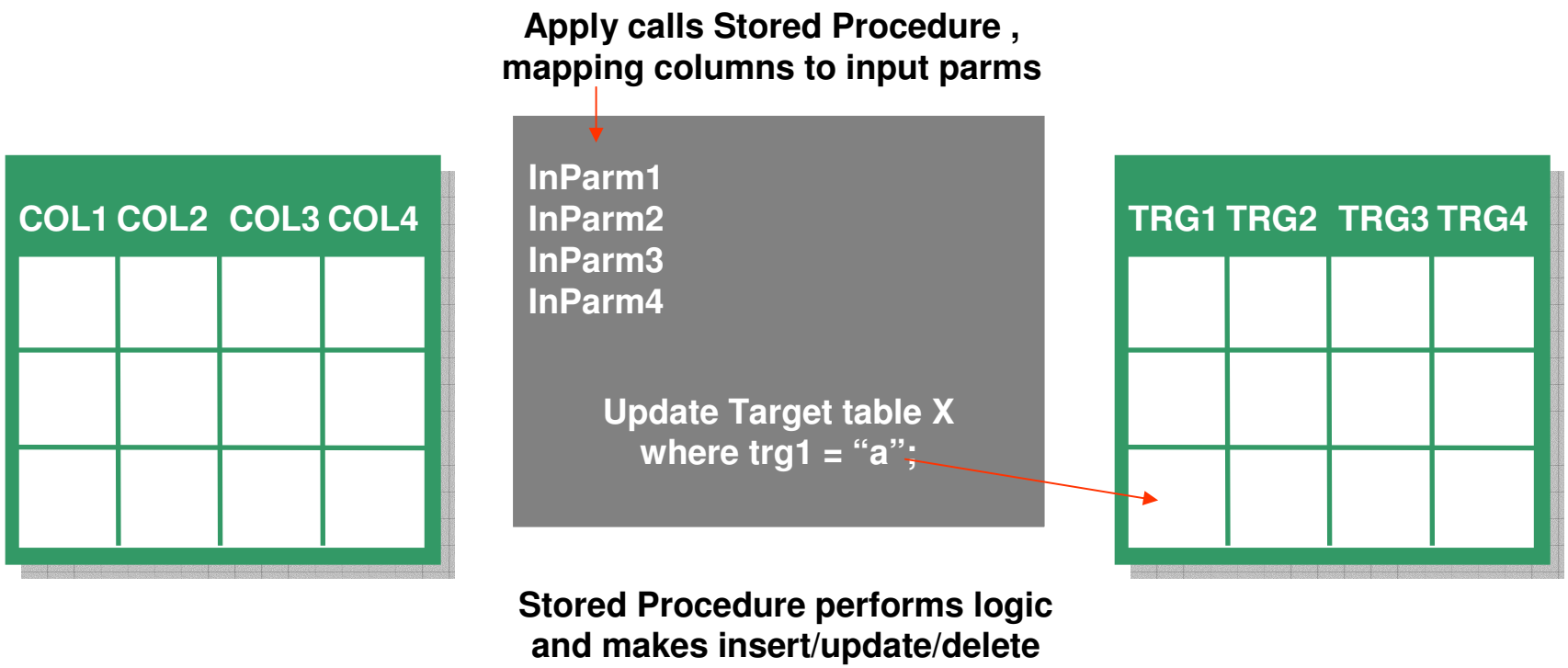
```
WHERE :LOCATION ='EAST' AND :SALES > (SELECT SUM(expense)
FROM STORES WHERE stores.deptno = :DEPTNO)
```

Q Replication - Transformations

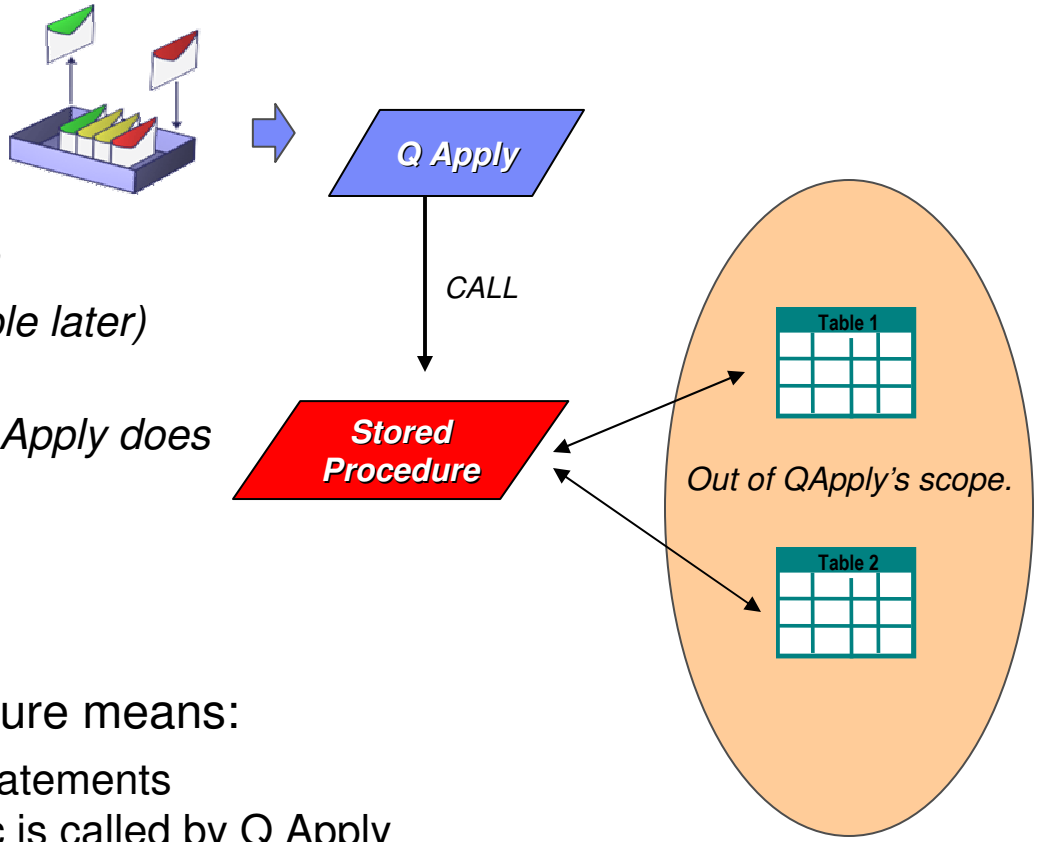
- Transformations achieved through:
  - Triggers on the target table
  - Stored Procedures called by Apply at the row level

. . .

  - Publish Event to user application



Stored Procedures Apply

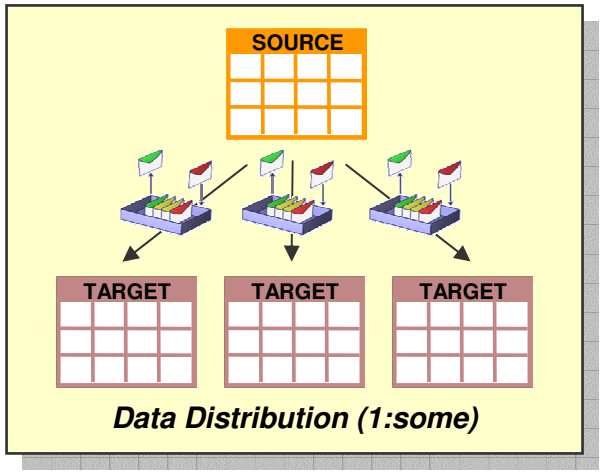


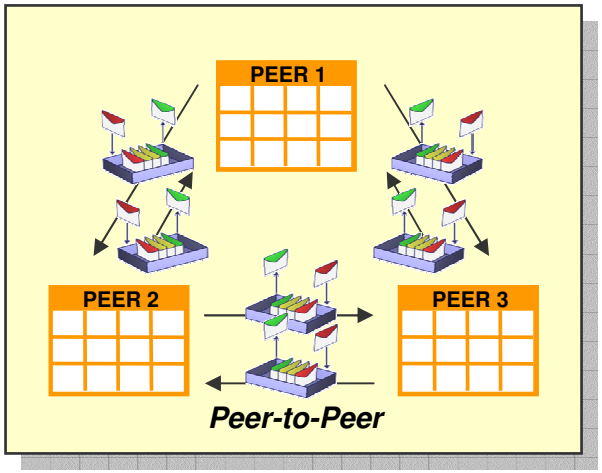
Apply via Stored Proc, why?

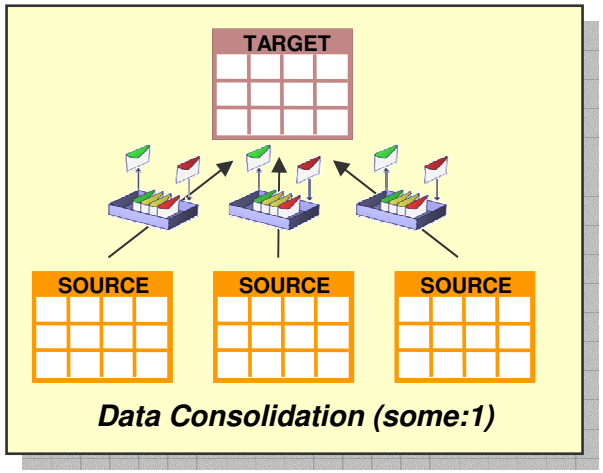
- *Data Transformation (example later)*
- *Data Aggregation*
- *Anything different to what Q Apply does*

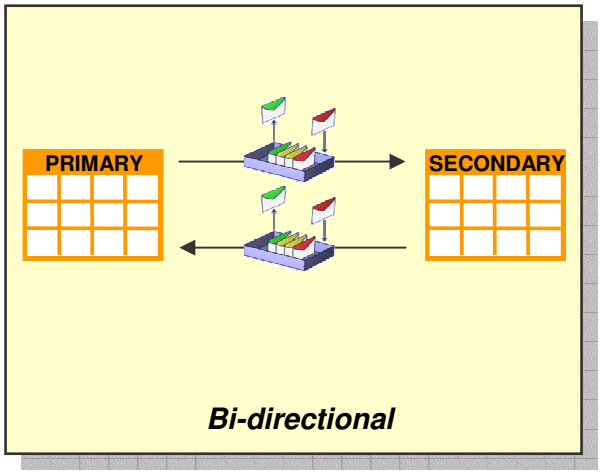
- Applying to a Stored Procedure means:
  - Instead of executing I/U/D statements
    a db2 registered Stored Proc is called by Q Apply
  - Row operation and column values are passed as parameters
    one row is one Stored Proc call
  - Q Apply has no information at all about any target table

Sample Q-Replication Scenarios









Key Scenarios:

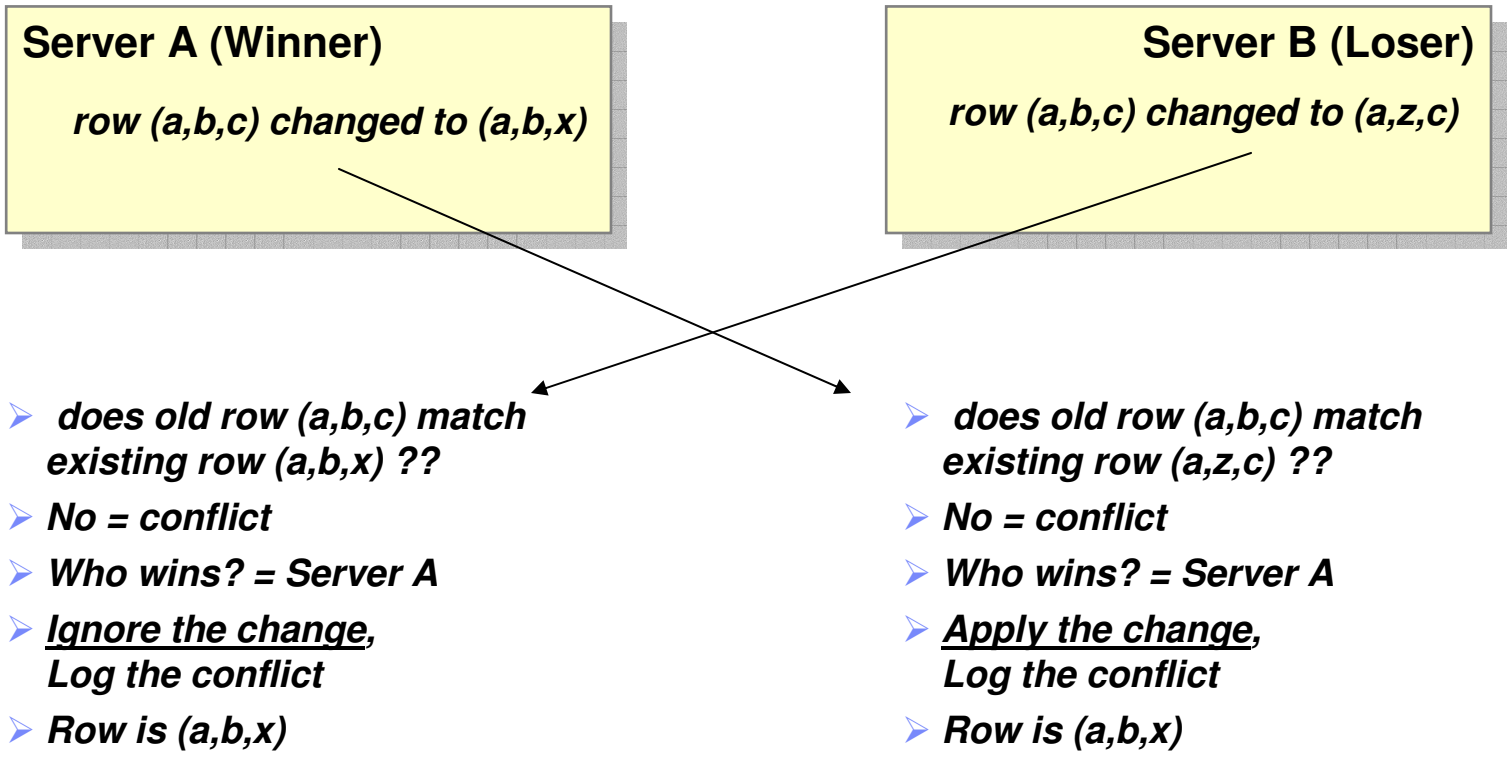
- Low-Latency Replication
- Geographically dispersed Applications with distributed Databases
- Bi-directional Replication with Conflict Checking, Handling, and Notification
- Software-based Hot-Standby
- Cross DB2-Family

Conflict Detection and Resolution

- Enables multi-directional replication that may result in conflicts
- Important for
  - “Active” standby systems
  - Workload balancing
- Value based conflict resolution
  - 2 participating nodes
  - Minimal overhead
- Version based conflict resolution
  - 2 or more participating nodes (practical limit around 6)
  - Requires extra columns and triggers
  - Most robust conflict detection and resolution

Value Based Conflict Detection

- Do the current row values at the apply target match the old row (before values) carried over from the source update?
- Designated site wins.

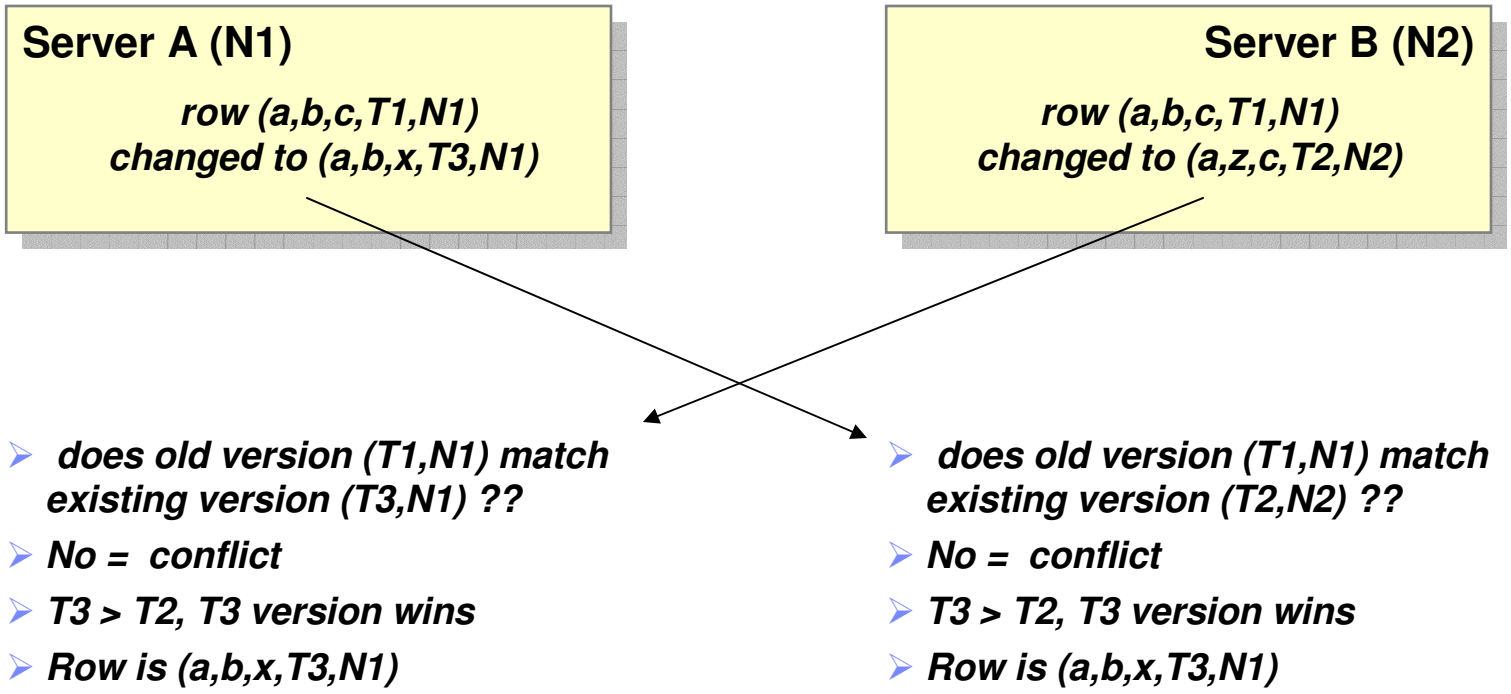


Value Based Conflict Detection and Resolution

- Disadvantage: Does not detect all possible forms of conflict
  - Does not detect insert/insert+delete conflicts
  - Is not offered for more than 2 participating database nodes
- Advantage: Requires less overhead
  - No extra columns or triggers
  - No effect to source updating applications
  - Problematic conflict cases may not be applicable to user applications
  - Can supplement with reconciliation utility (Tdiff/Trepair)
  - Might be appropriate for planned outage/failover/DR

Version Based Conflict Detection

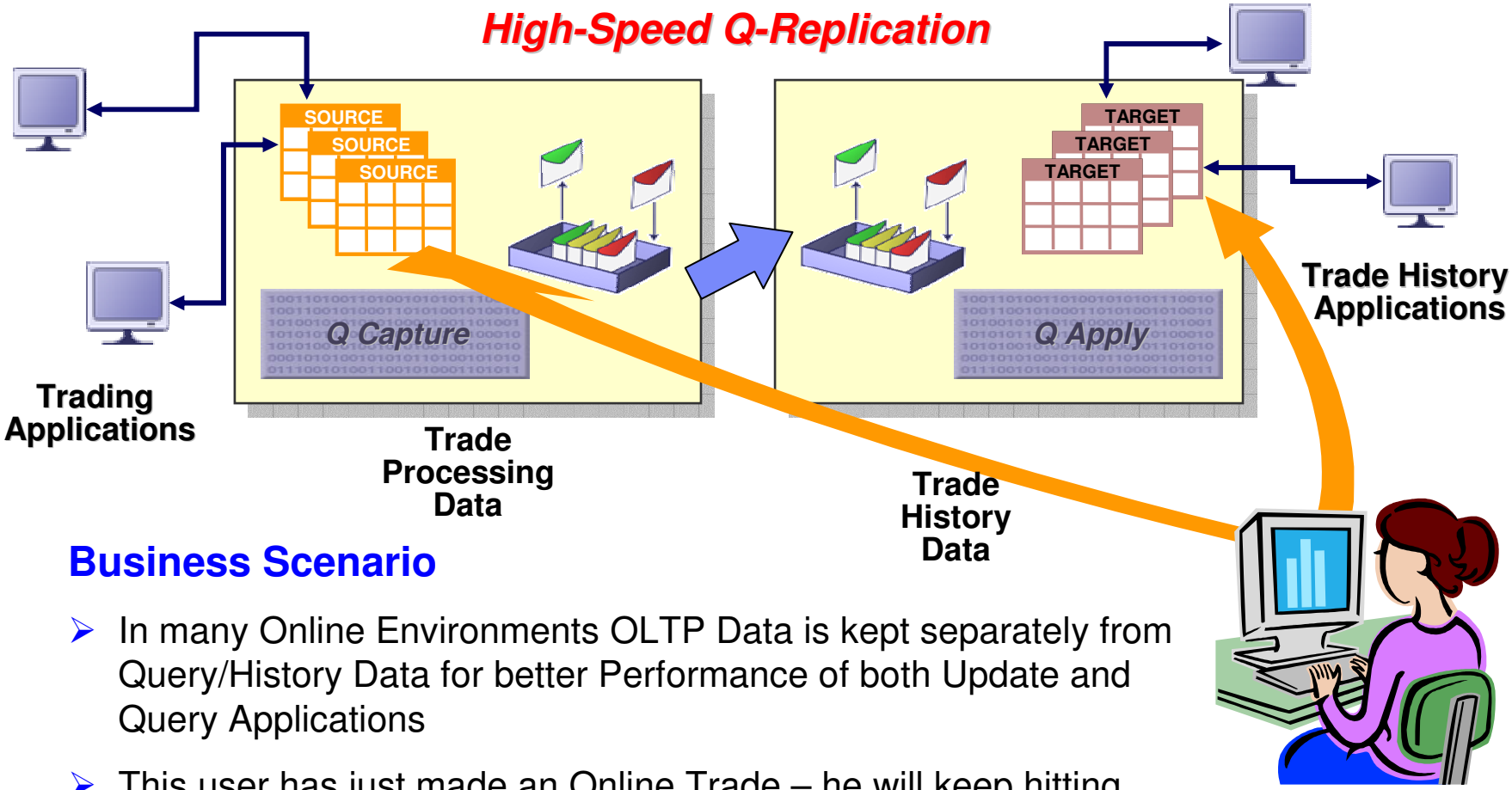
- All rows are augmented with a “Version” = timestamp Tx and smallint Nx, indicating when and by which server the row was last updated
- Do the current values of Tx and Nx at the apply target match the old values of Tx and Nx carried over from the source update?
- Most current timestamp Tx wins.



Version Based Conflict Detection and Resolution

- Advantage: Detects all possible forms of conflict
  - Based on time zone adjusted timestamps
  - Theoretically can support any number of participating database nodes
  - Practical limit for V8.2 is about 6
- Disadvantage: Requires more overhead
  - Requires versioning columns on tables (admin adds)
  - Requires triggers to maintain versioning columns (admin builds)
  - Does impact source updating applications
  - All participating nodes must be connected to all other participating nodes
  - Time zones can vary , but machine clocks should be well synchronized
  - Clocks that are off will slow down the Apply – data from the future will never be applied

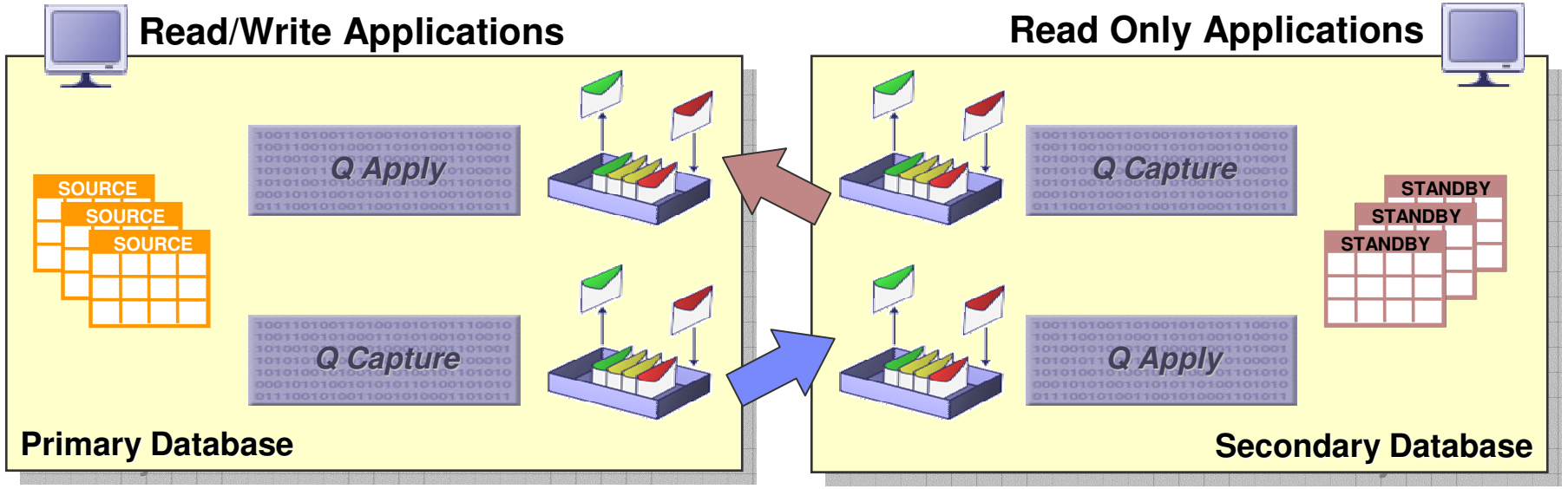
Feeding Trade-History Database with Q-Replication



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

Summary – Q-Replication Overview

- **Q-Capture**
  - Captures Changes from DB2 Log into Message Queue(s)
  - Websphere MQ replaces the use of Staging Tables (CD Tables)
  - Each Message represents a Transaction
  - Very Compact Internal Message Format
- **Q-Apply**
  - Highly-Parallel Apply Processing
  - Conflict Detection, Resolution and Documentation
- **Websphere MQ**
  - Robust, Secure, and High-Performance Messaging Infrastructure
  - Available on all commercially relevant Platforms

IBM's Information Replication Architectures

- SQL-Replication (Formaly and still known as DataPropagator)
  - Log-based asynchronous Change Capture
  - Versatile Replication Architecture for Data Replication between all Members of the DB2 Family and beyond (in Combination with WebSphere Information Integrator)
  - Relational Staging Concept
  - Successfully used for Years by huge User Community

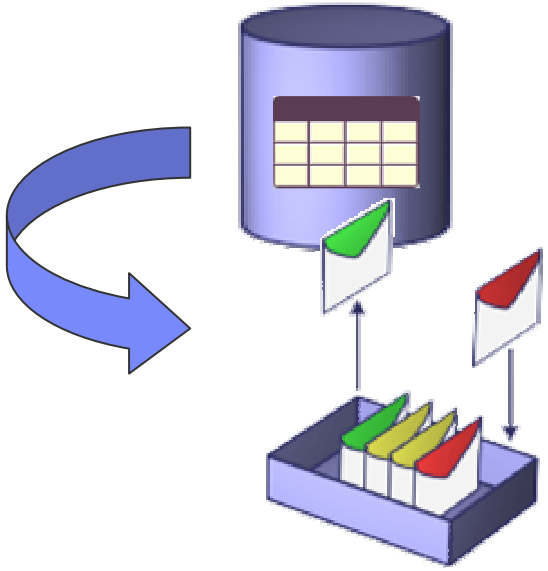
- Q-Replication / Event Publishing
  - Log-based asynchronous Change Capture
  - Captured Transactions immediately sent via Message Queues
  - High-Volume, Low-Latency Architecture

- HADR (DB2 UDB for Linux, Unix, Windows only)
  - Replicates entire Database by Log Buffer Shipping (over IP)
  - Initially no Read-Access at Secondary Site

Event Publishing

Why Publish Data?

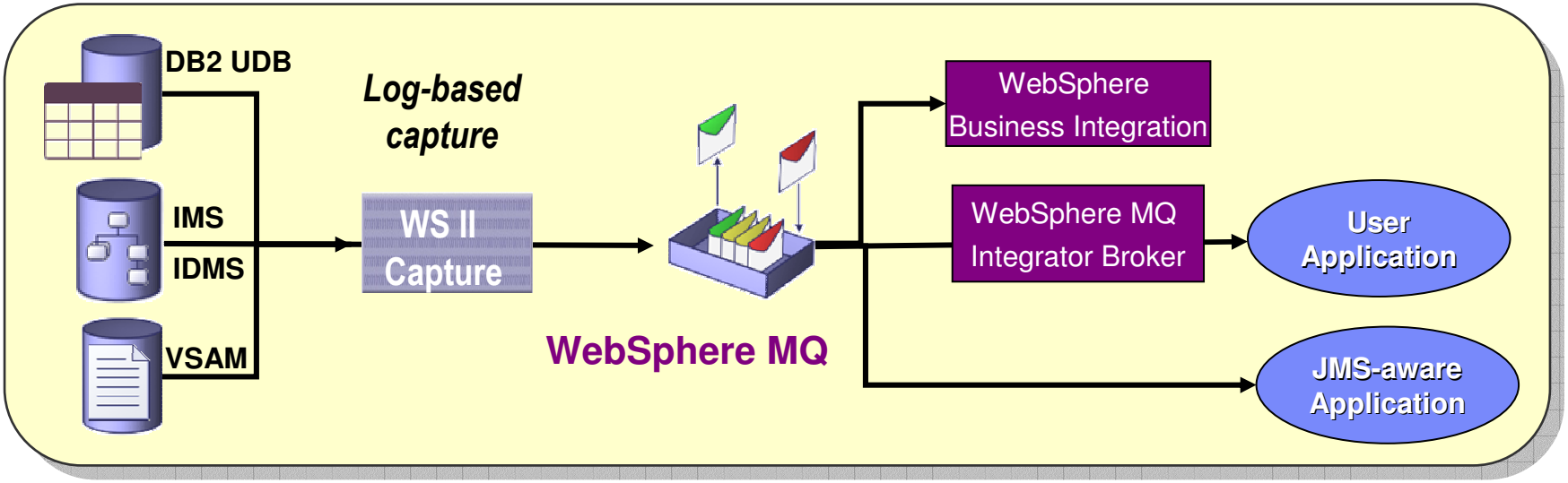
- **Database to Application Messaging**
  - Drive downstream Applications or APIs based on the Transactional Data of the changed Database Events
- **Event Notification**
  - Stream changed Data Information to Web Interfaces
  - Stream only particular Events of Interest (filter Data)
- **Data Warehouse / 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

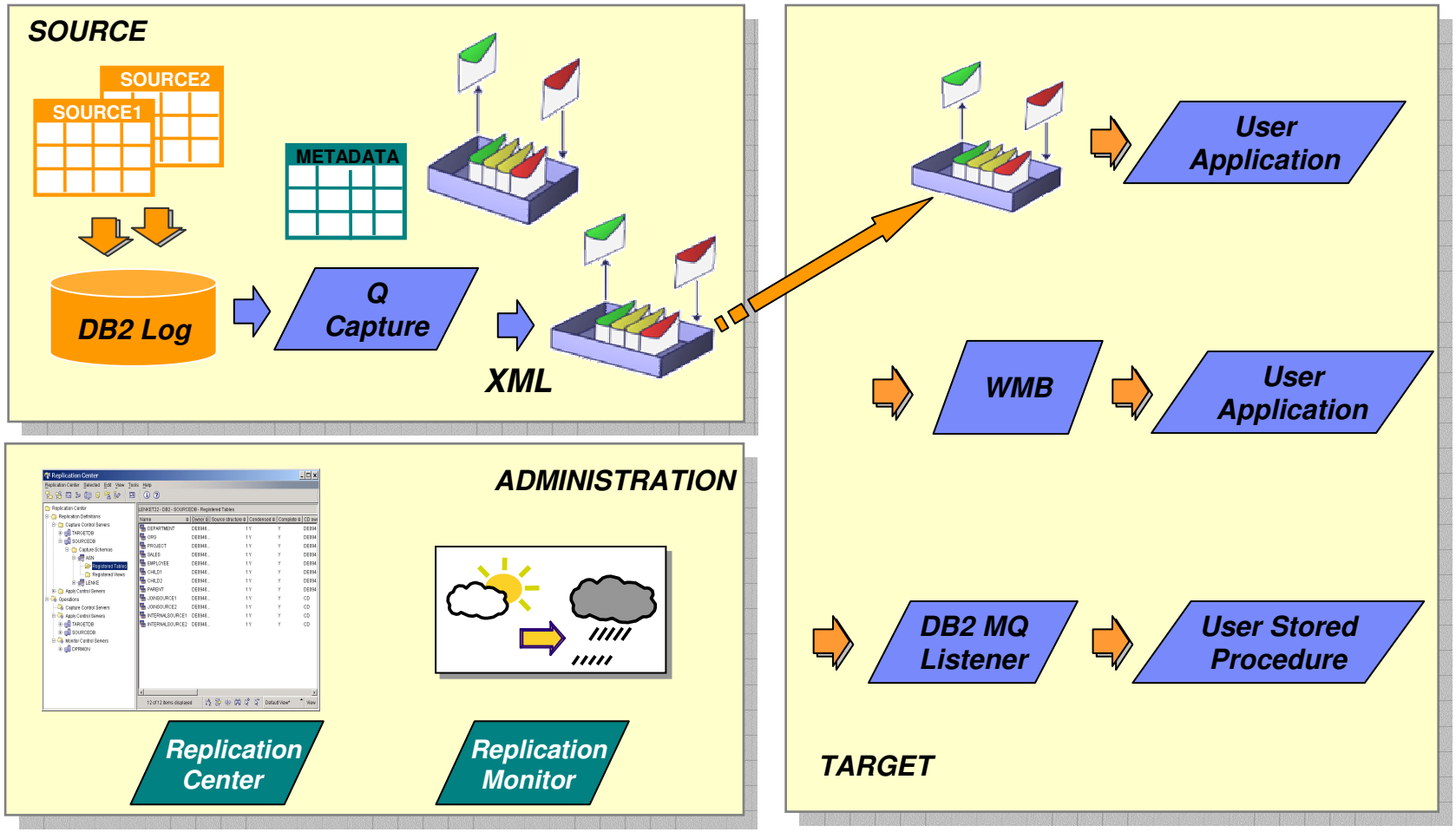
Capture database changes as XML messages and publish them to WebSphere MQ

- **Function**
  - Publish events to a message queue
  - XML self-describing format (UTF-8)
  - Wizard-driven configuration
- **Usage**
  - Application to application messaging
  - Event streaming
  - Source for ETL tool



Classic Data is REACTIVATED !

DB2 Implementation: Process Flow



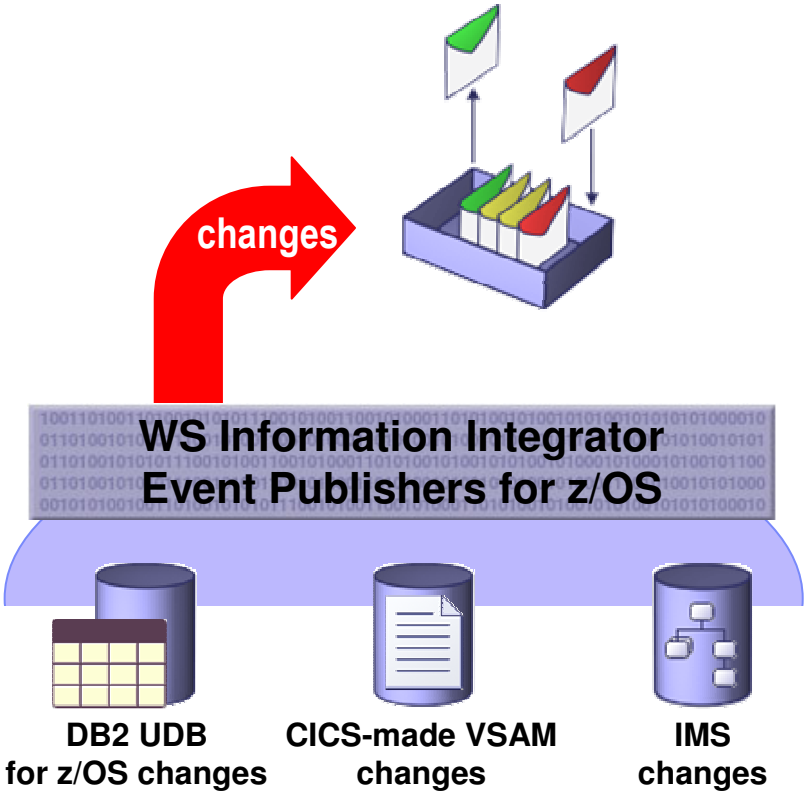
Event Publishing - Publication Options

- **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

WebSphere Information Integrator Event Publishers for z/OS

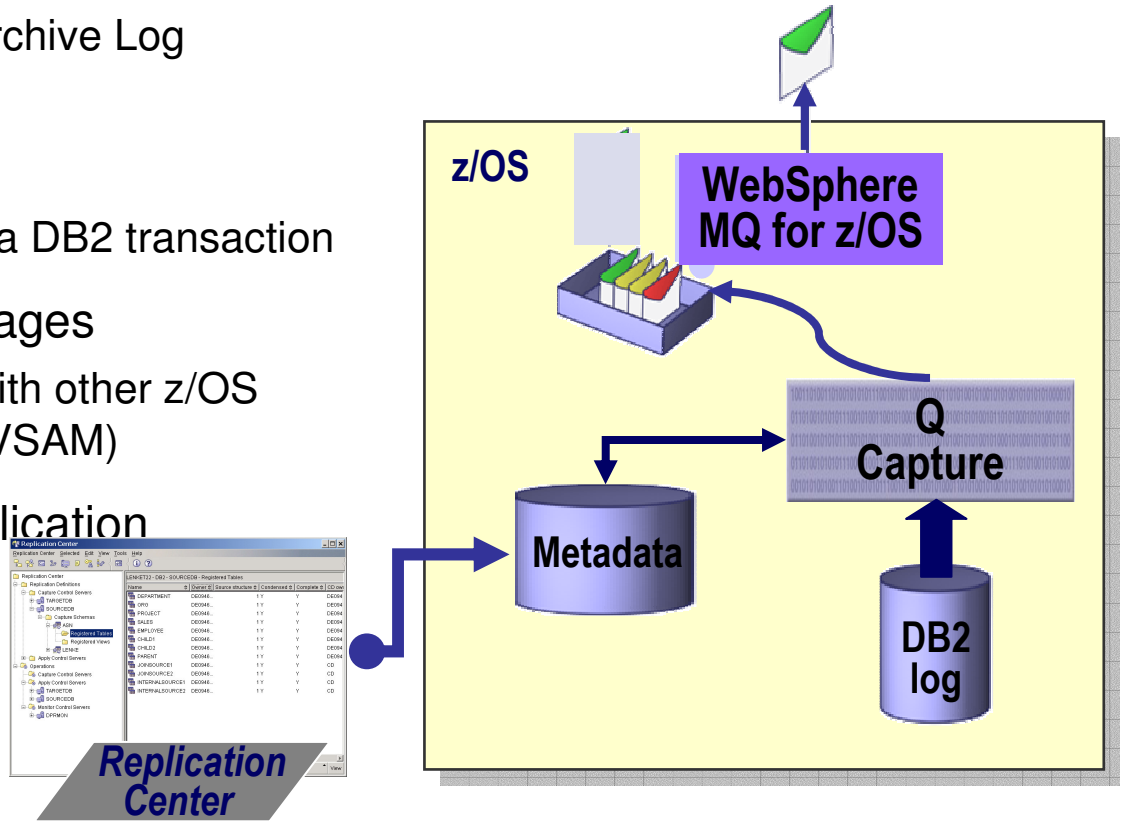
- Real time DB2, IMS, VSAM and CA-IDMS changed-data capture and publishing
- Publish to WebSphere MQ
- Relational XML format
- WebSphere listener application/tool
  - Picks up message(s)
  - Takes action
- Two Event Publisher infrastructures:
  - DB2 Universal Database for z/OS
    - Based on WebSphere II Q-replication
  - IMS, VSAM and CA-IDMS*
    - Based on WebSphere II Classic Federation



Capture, Externalize (XML) and Deliver to MQ

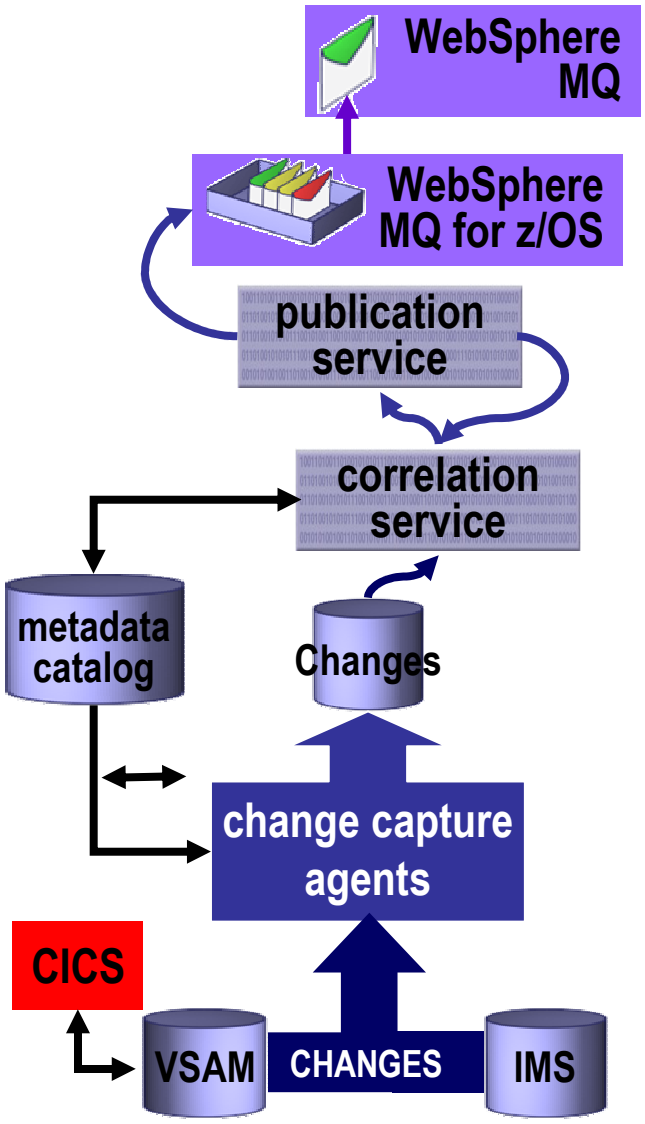
DB2 UDB for z/OS Implementation

- Leverages Q Replication infrastructure
  - Replication Center metadata management
  - Capture and Publish with no Apply
- Log-based changed-data capture
  - Log Buffer - Active Log - Archive Log with seamless transition
- Transaction aware
  - Each message represents a DB2 transaction
- Publish XML format messages
  - XML format is consistent with other z/OS event publishers (IMS and VSAM)
- Upgradeable to full Q Replication

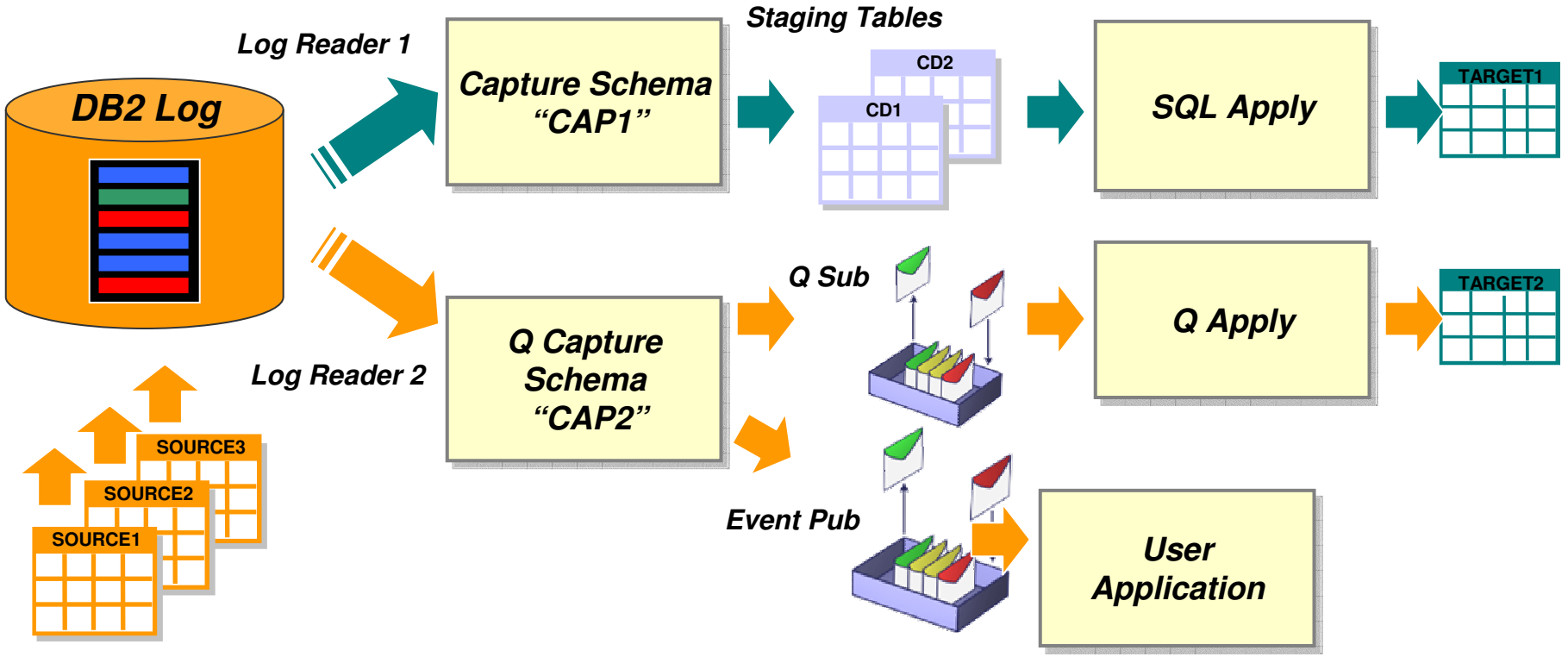


Classic Event Publisher: IMS & VSAM Implementation

- Change capture agents intercept changes
  - Active log stream capture or log file access
  - Changes are forwarded to the correlation service
- Correlation service
  - Sorts data by unit-of-work identifiers
  - At end of unit-of-work
    - Rollback - flush all data for this unit-of-work
    - Commit - reformat data into relational XML messages
      - push data to the publication service
  - Metadata catalog holds mapping between IMS and VSAM changes and relational table/column definitions that will be published
- Publication service
  - Manages publication to WebSphere MQ persistent queue
  - Initiates recovery data update with the correlation service
  - Queue info is defined in configuration files



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

WebSphere II Event Publisher & Business Intelligence

Feeding Changed Data to :

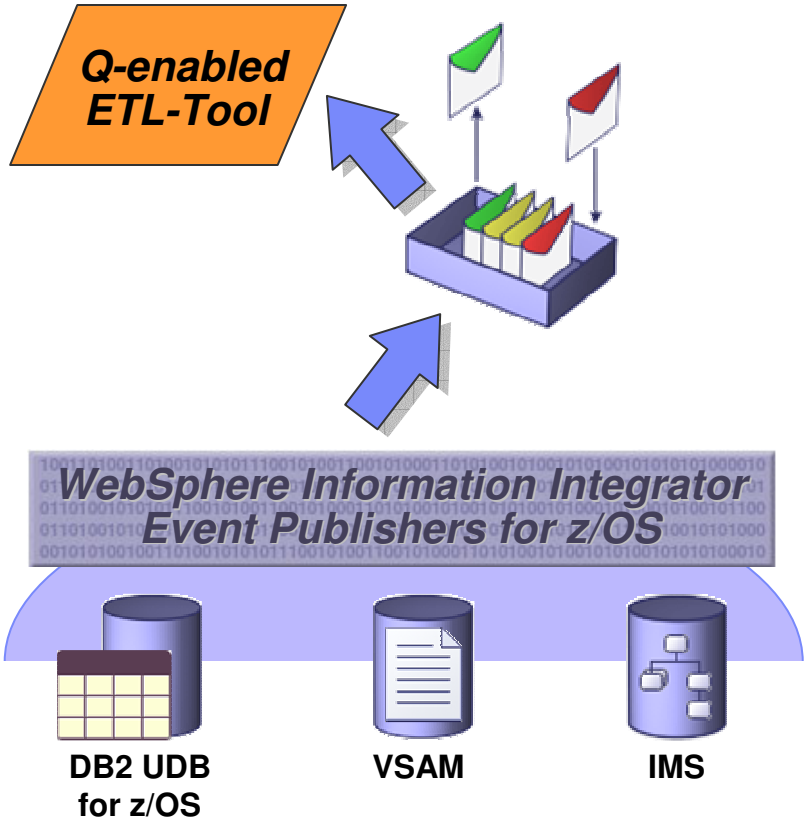
- Data Warehouse
- Datamart
- Operational Data Store (ODS)

Optimize Resource Utilization

- Minimize Bandwidth Requirements
- Maximize Data Currency

Complements with WebSphere II Federation

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



WebSphere II Event Publisher & Business Integration

Data "Events" drive business integration

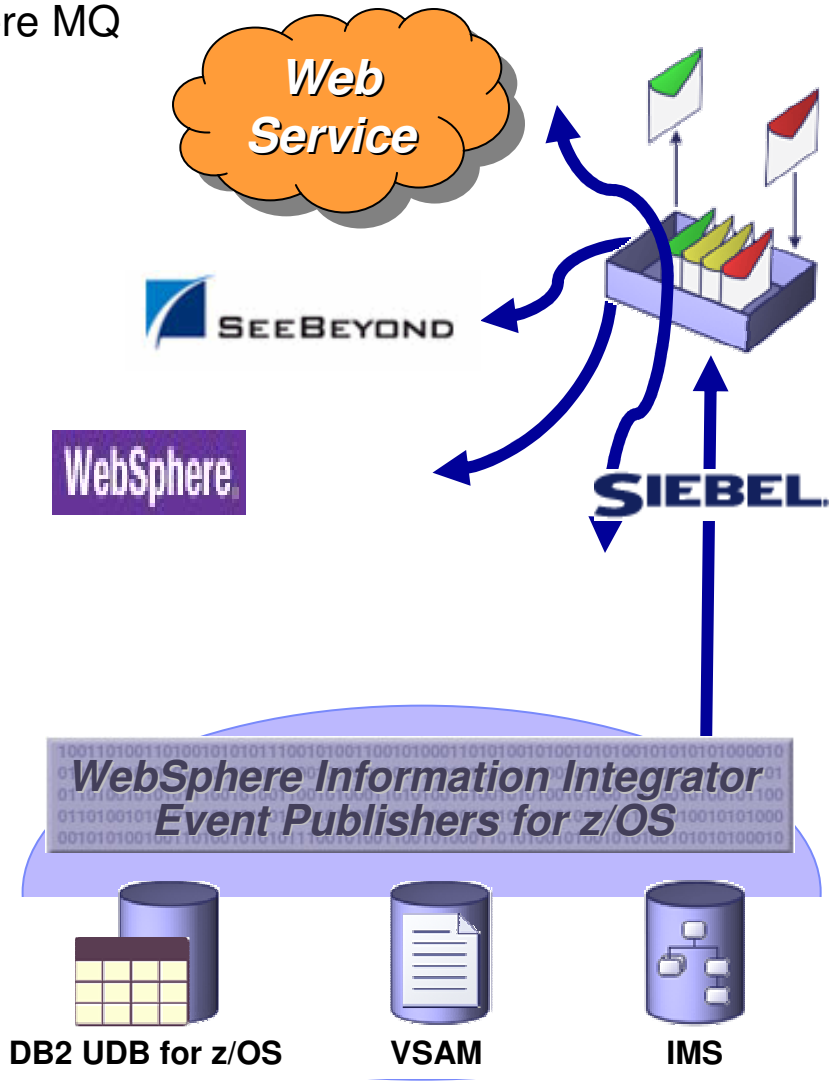
- Seamless integration with EAI via WebSphere MQ

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

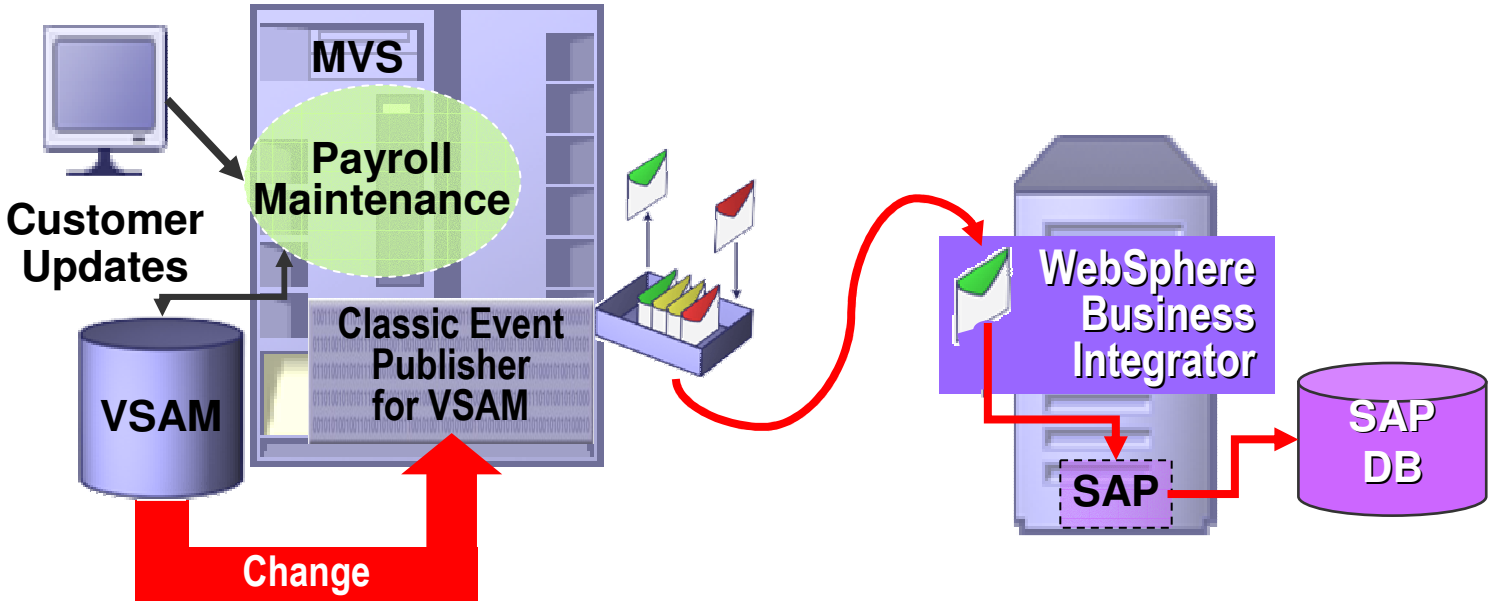
- Synchronize mainframe updates with:
  - ✓ CRM
  - ✓ ERP
  - ✓ HR, etc.



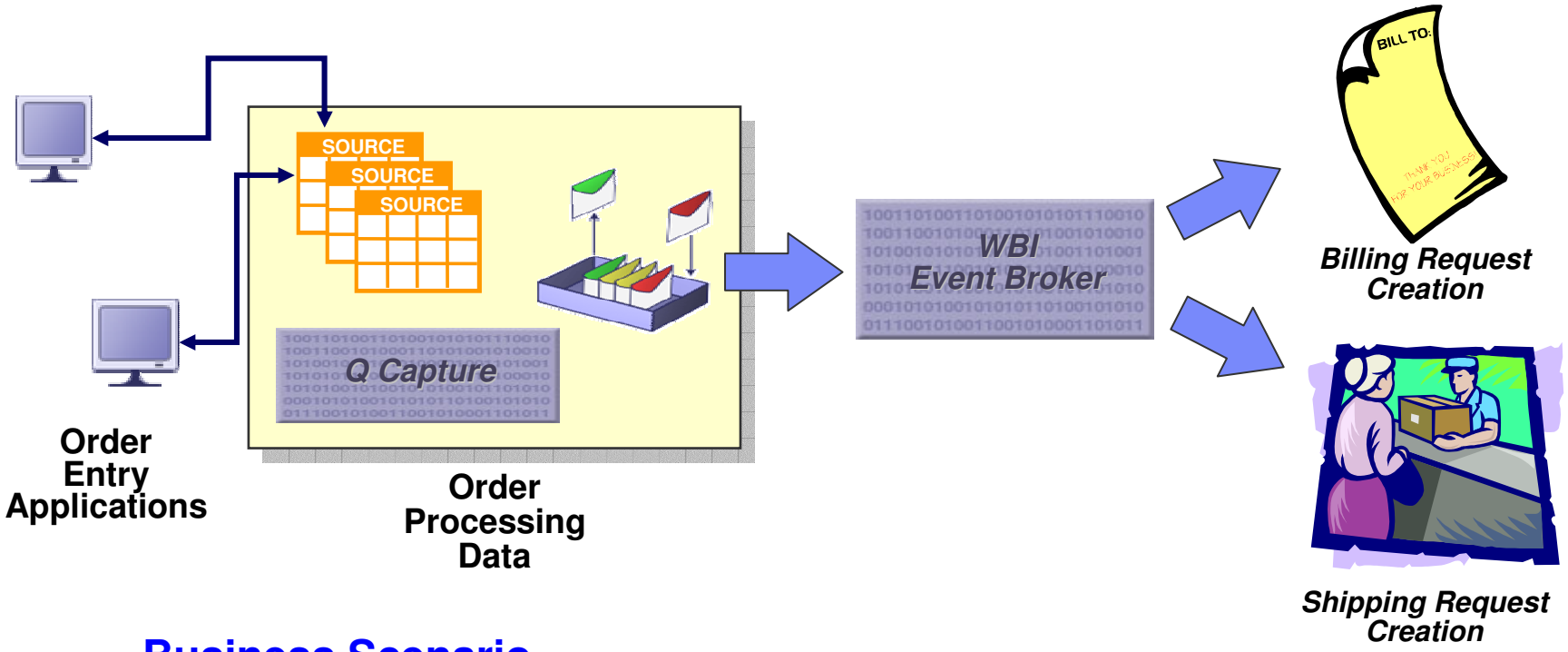
Sample Application

- 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

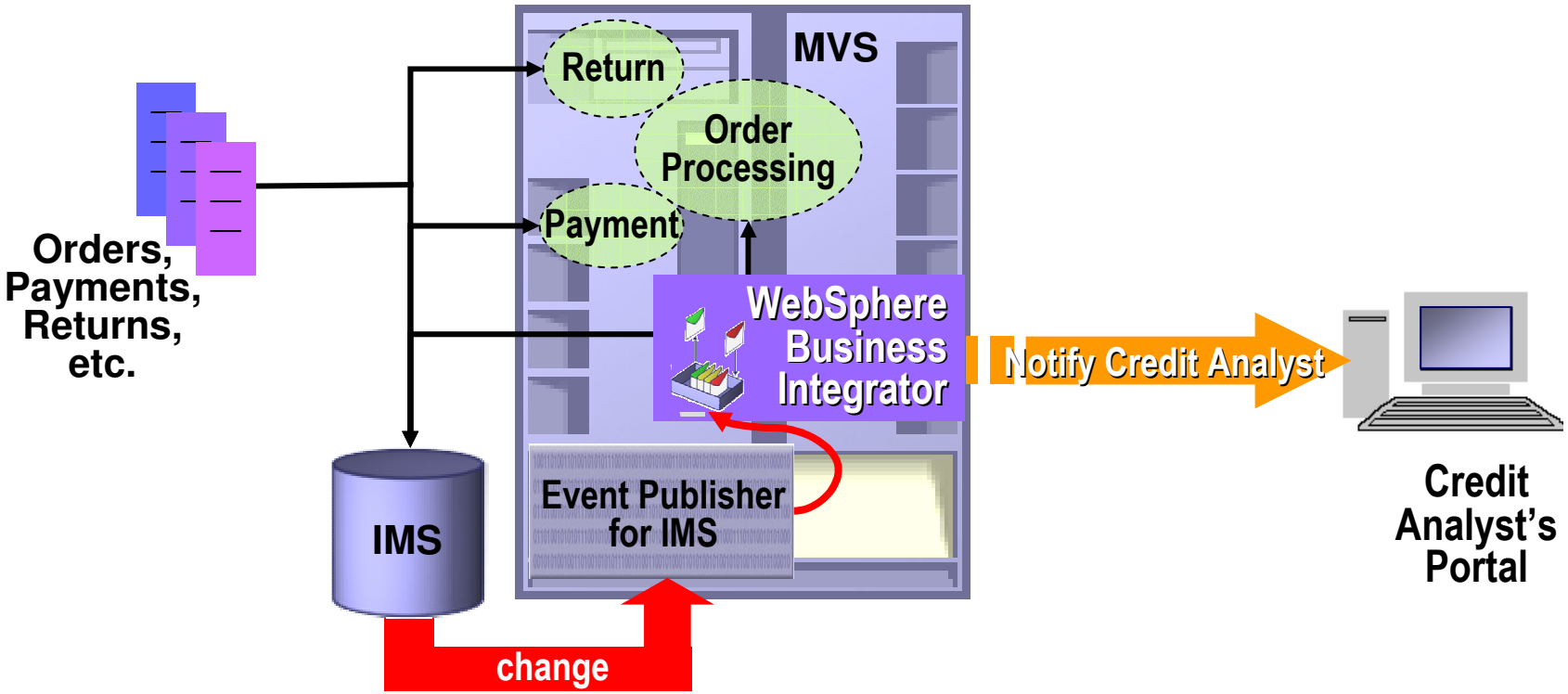


Business Scenario

- 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

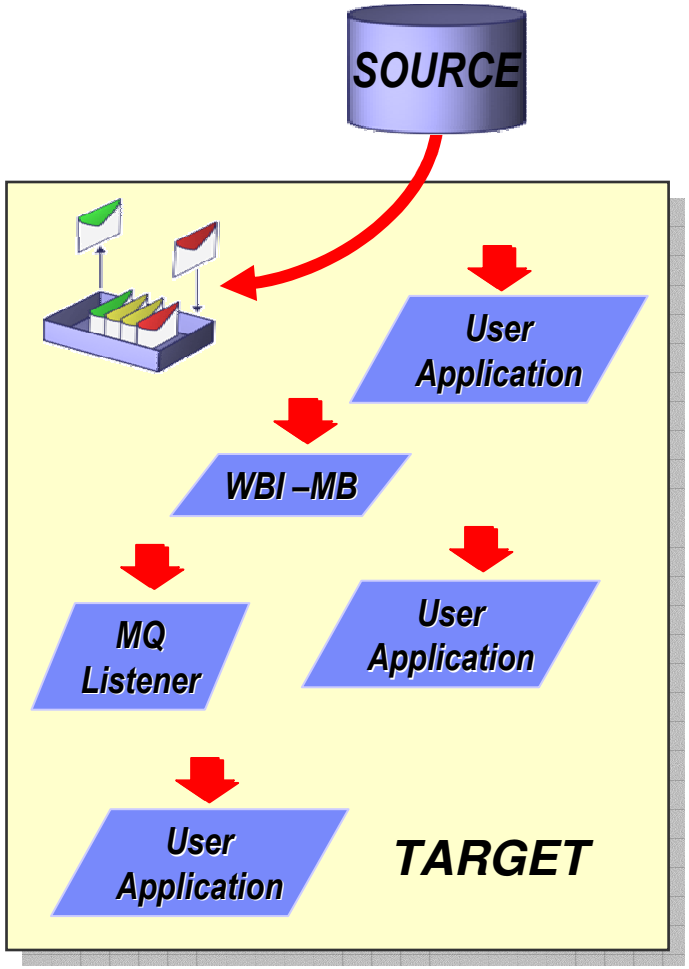
- Event Notification
  - Receivable balances approaching credit limit pushed to a credit analyst
    - Threshold is independent of order processing and accounting applications
    - No “hard-hooks” in OLTP applications necessary



Why data events versus application events?

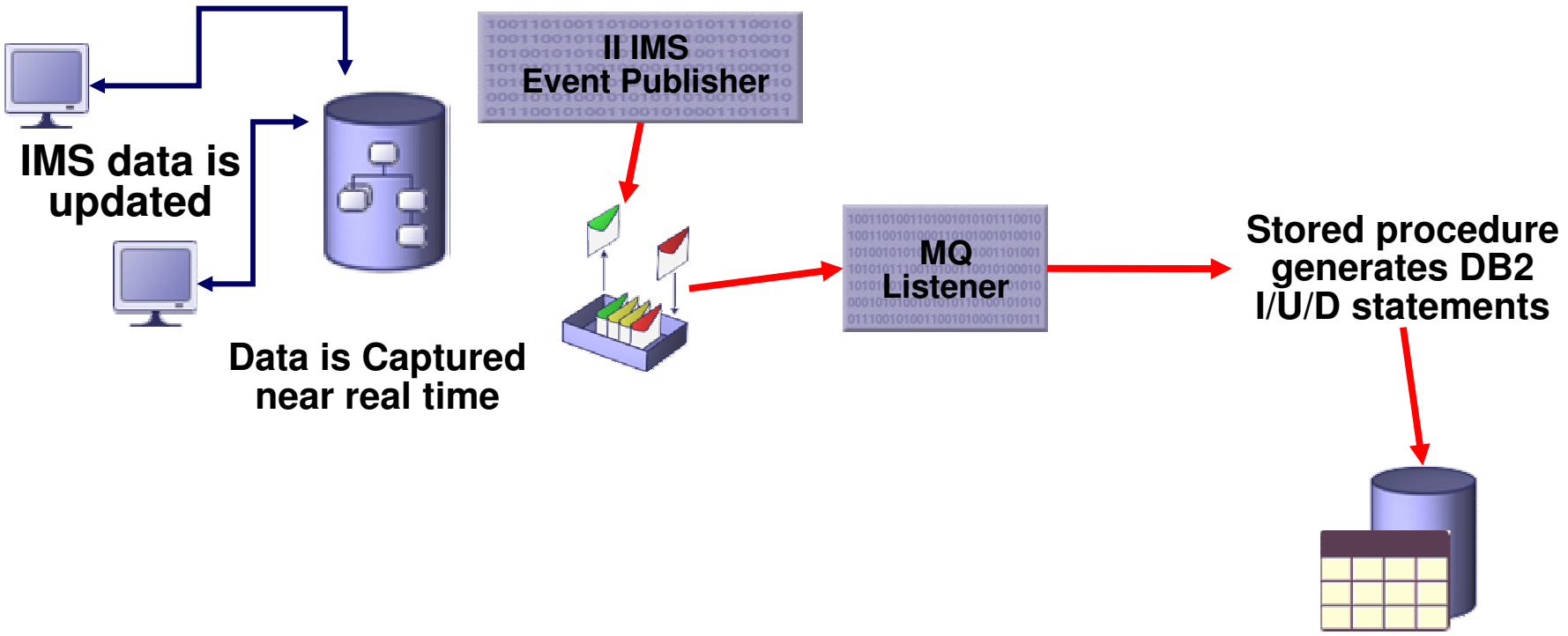
Integration is independent of the source applications

- 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



Summary

Using Classic II Event Publishing for Replication



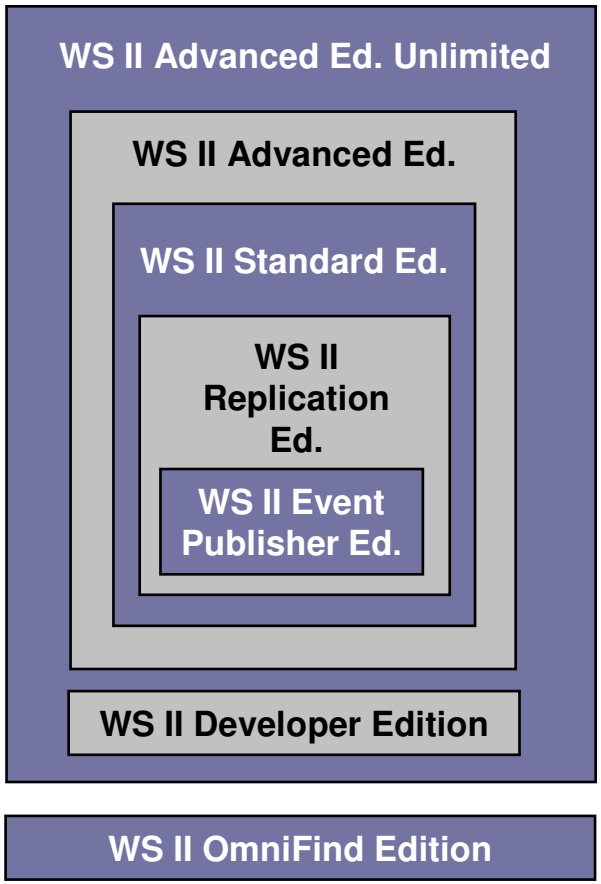
- It is our direction to support replication of Classic data sources using the highly parallel Q Apply of WebSphere II Replication.
- As an interim suggestion, SQL stored procedures can be used to apply the data captured through the II Classic Event Publisher

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)

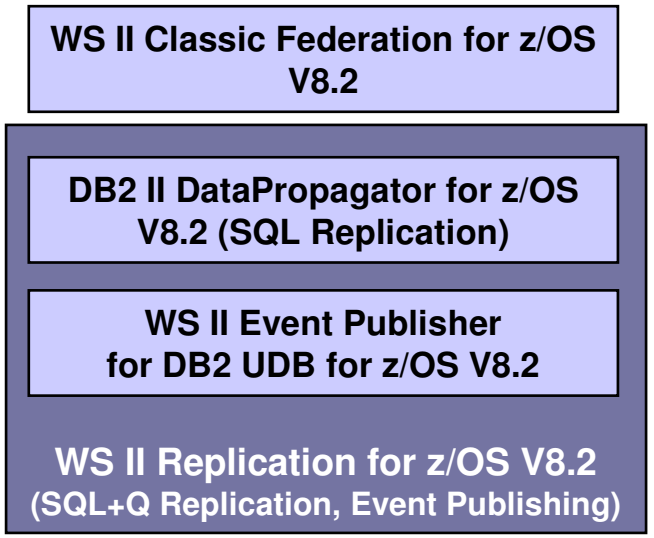


PID 5724-C74

PID 5724-C74 (same as for WebSphere II)

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

Mainframe (IBM eServer zSeries)



PID 5697-I82

PID 5655-I60

PID 5655-M36

PID 5655-L88

PID 5655-M38

PID 5655-M35

PID 5655-N56

- Value Unit pricing Model

Replication and Event Publishing Products : z/OS

DB2 DataPropagator for z/OS V8.2

- SQL Replication Architecture (DProp Capture and Apply)
  - Available for DB2 UDB z/OS V7 and V8

WebSphere Information Integrator Replication for z/OS V8.2

- Q Replication and SQL Replication Architecture
  - Includes Event Publisher
  - Available for DB2 UDB z/OS V7 and V8
  - Websphere MQ prerequisite when using Q Replication

WebSphere Information Integrator Event Publisher for DB2 z/OS V8.2

- Event Publisher for DB2 (Q Capture)
  - Websphere MQ prerequisite

WebSphere Information Integrator Classic Event Publisher for IMS V8.2

- Event Publisher for IMS
  - Websphere MQ prerequisite

WebSphere Information Integrator Classic Event Publisher for VSAM V8.2

- Event Publisher for CICS/VSAM
  - Websphere MQ prerequisite

Replication and Event Publishing Products : Linux, Unix, Windows

DB2 UDB V8.2 for Linux, Unix, Windows includes DB2 DataPropagator

- SQL Replication Architecture
- SQL Capture and SQL Apply (for all DB2 UDB V8 Editions incl. Partitioning Feature)
- DB2 Sources and Targets. Informix IDS Sources and Targets supported through Federation Capability

WebSphere Information Integrator Replication Edition V8.2

- SQL Architecture: DB2 & multi-vendor Sources and Targets
- Q Architecture: DB2 Sources and Targets
- Note that Websphere MQ is bundled with this Product

WebSphere Information Integrator Event Publisher Edition V8.2

- Q Architecture: DB2 LUW Sources
- Note that Websphere MQ is bundled with this Product
- Data Changes pubished through Message Queues in external XML Format