Université du Mainframe 2005

Extraction, intégration, diffusion de données
en environnement hétérogène

IBM Software et IBM System & Technology Groups

Agenda

- **Stratégie IBM « Information Integration »**
  - Yannick Barel
- **L'offre WebSphere Information Integration**
  - Eric Derbanne
- **L'offre Websphere Data Integration Solution (ex Ascential)**
  - Steven Haddad
- **Scenarios d'utilisation et de mise en œuvre**
  - Bob DeCori

Université du Mainframe 2005

Information Integration :

« Leveraging your mainframe asset »

Yannick Barel
Mgr, WW Information Integration Customer Enablement
IBM - Silicon Valley Lab
yannick.barel@us.ibm.com

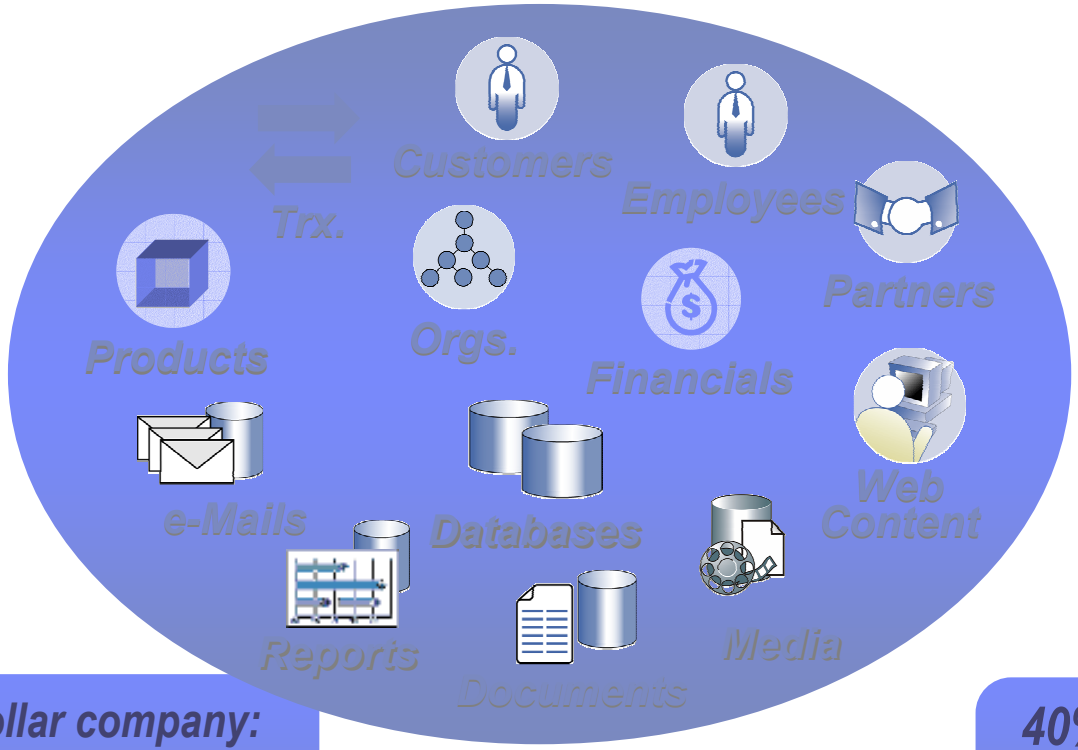
IBM Software et IBM System & Technology Groups

Market Insights

60% + of CEOs: Need to do a better job capturing and understanding information rapidly in order to make swift business decisions.

Only 1/3rd of CFOs believe that the information is easy to use, tailored, cost effective or integrated.

85% of information is unstructured.



30-50% of application design time is spent on copy management.

42% of transactions are still paper-based.

30% of people's time: searching for relevant information.

The average billion dollar company: 48 disparate financial systems 2.7 ERP systems

79% of companies have more than two repositories and 25% have more than 15

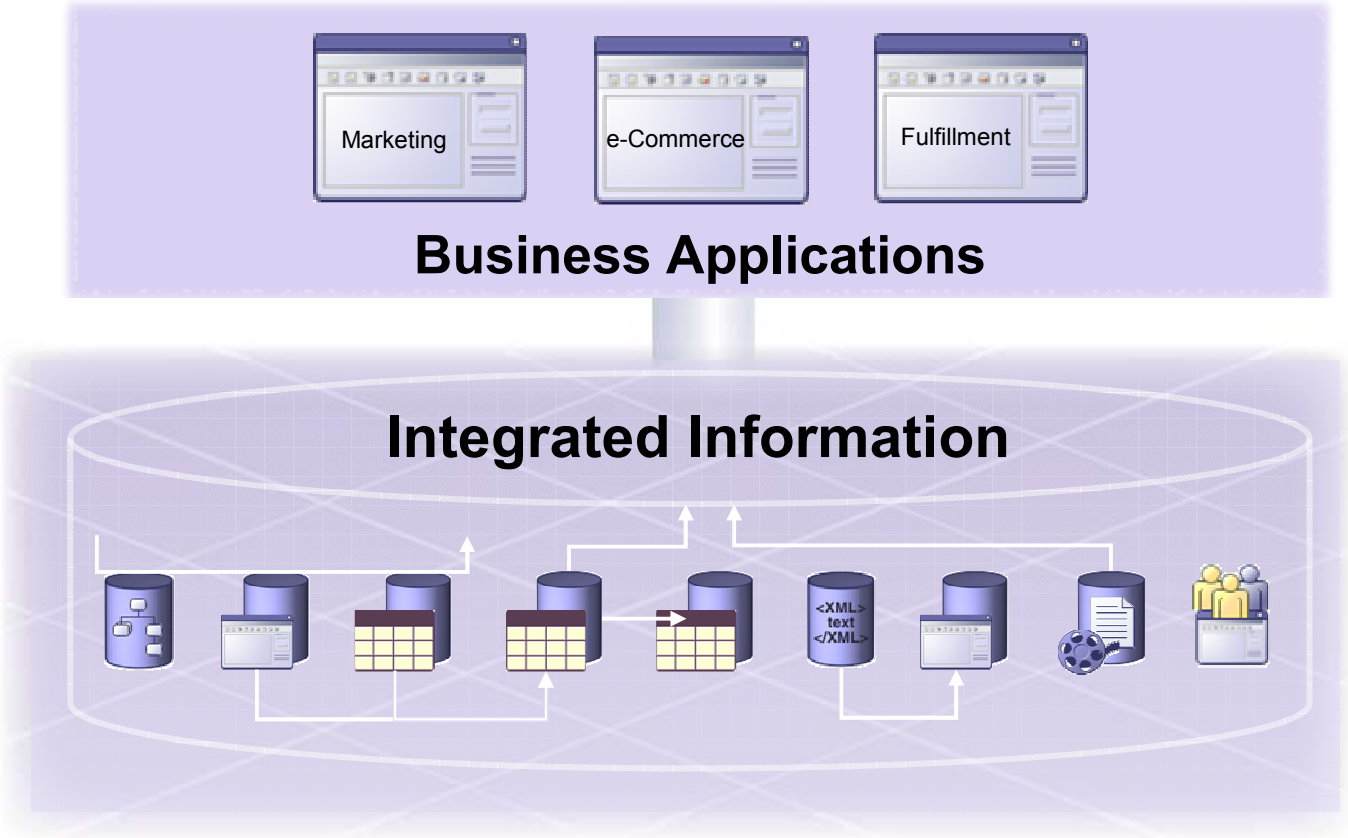
40% of IT budgets may be spent on integration.

Sources: IBM & Industry Studies, Customer Interviews, Forrester

IBM Software et IBM System & Technology Groups

IBM Information Integration Vision

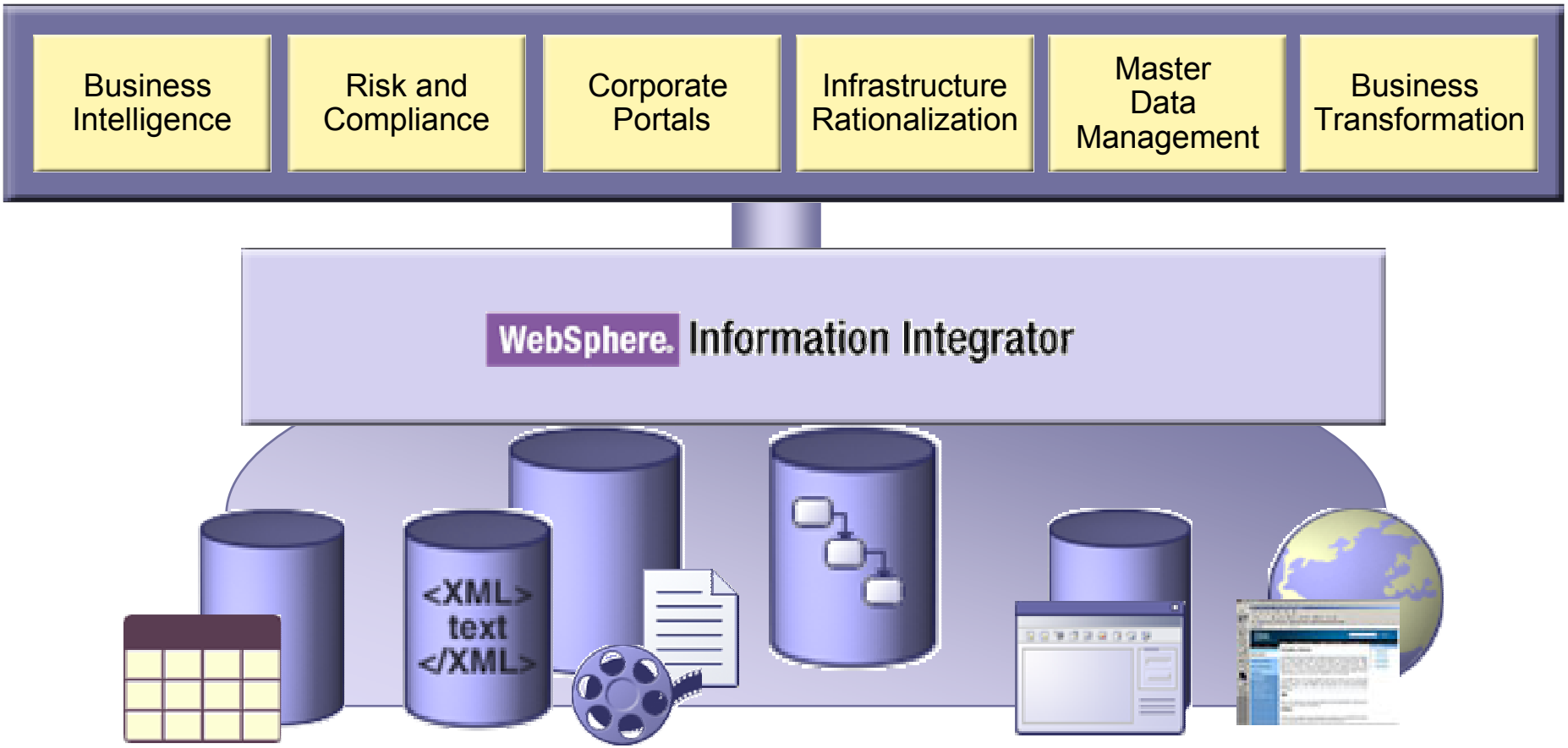
Delivering accurate, consistent, timely, and coherent business information



Get the right information, in the form you want, whenever you need it.
Isolate applications from information complexity.

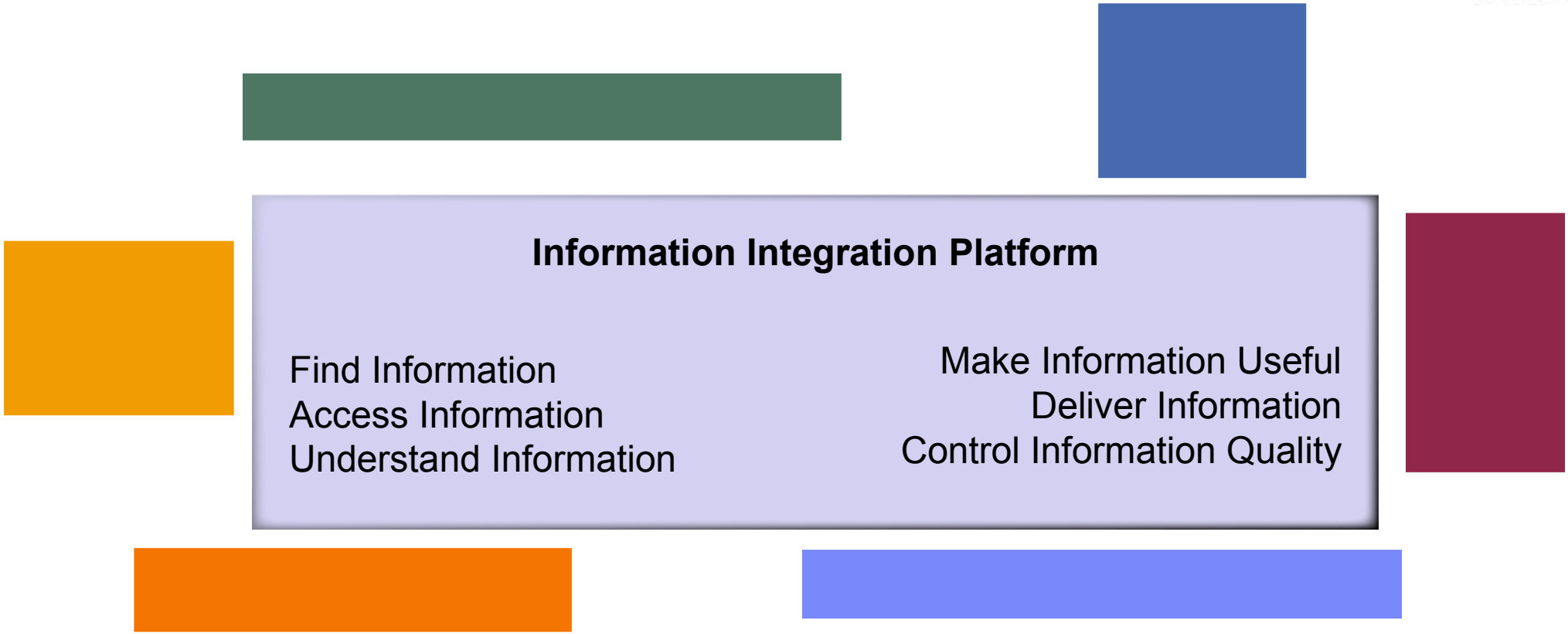
WebSphere Information Integration Platform

A strategic information integration platform to help enterprises become on demand businesses



IBM Software et IBM System & Technology Groups

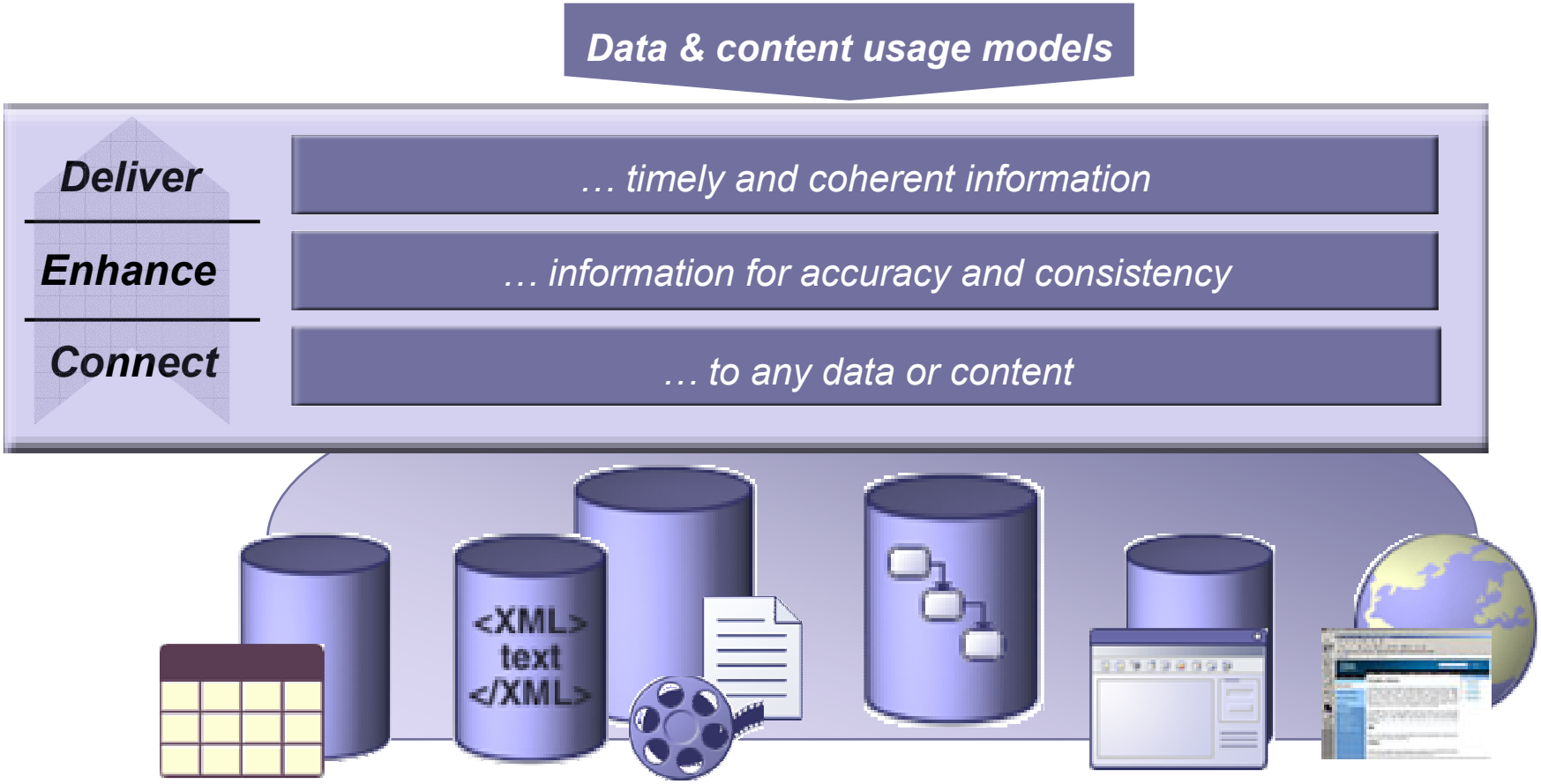
Value of a Platform Approach



- **A platform that could address all of these problems could address multiple business initiatives**
- **It could provide leverage and reuse across projects and ensure consistency**
- **It could enable governance and control over information**

Comprehensive Information Integration Services

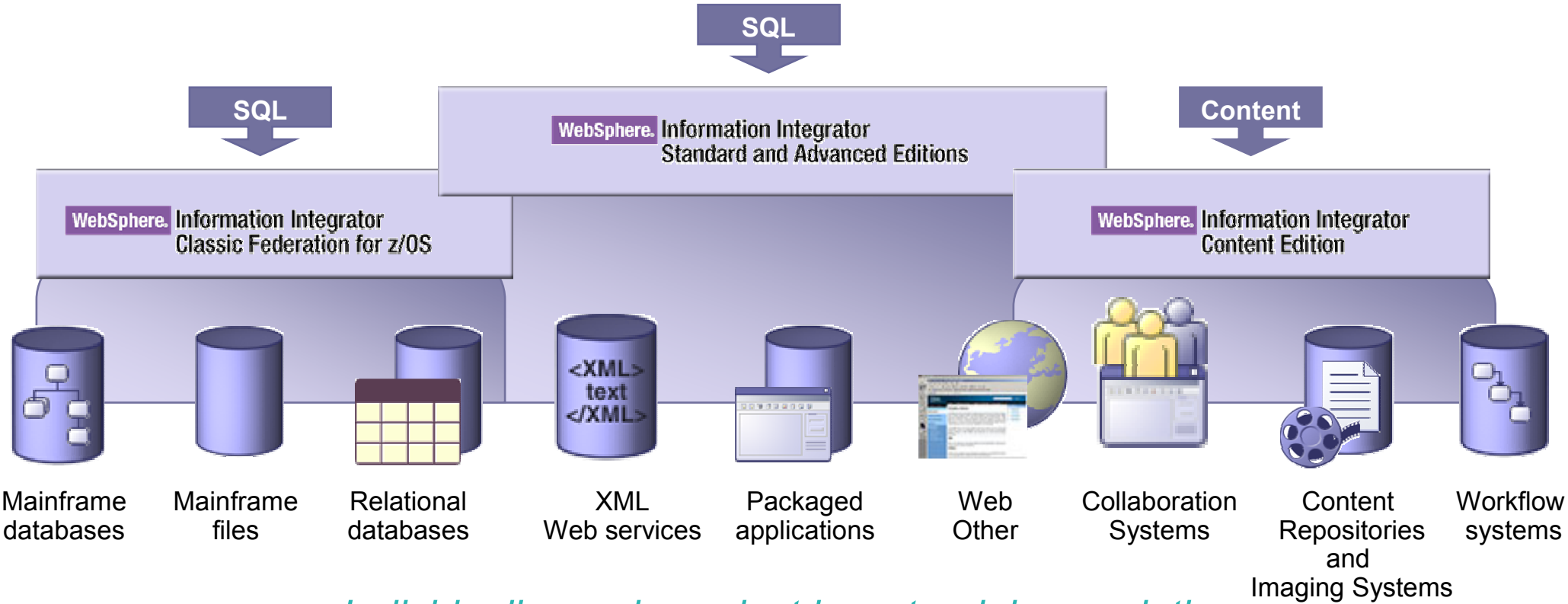
Delivering accurate, consistent, timely, and coherent business information



Federation

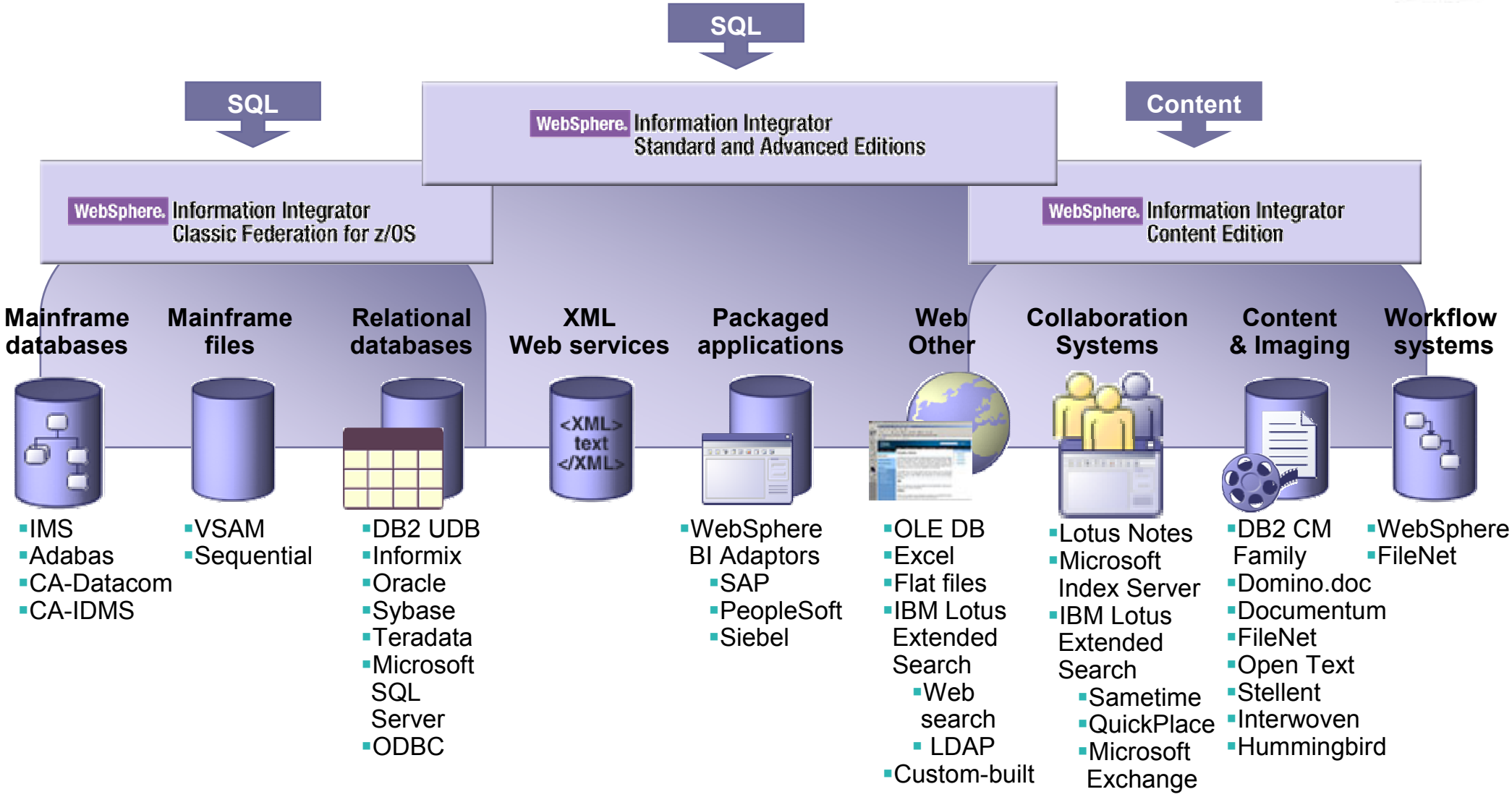
Access diverse and distributed information as if it were in one system

Single sign on – Unified views – Common language – Web services or Java API Query and update – Optimized access



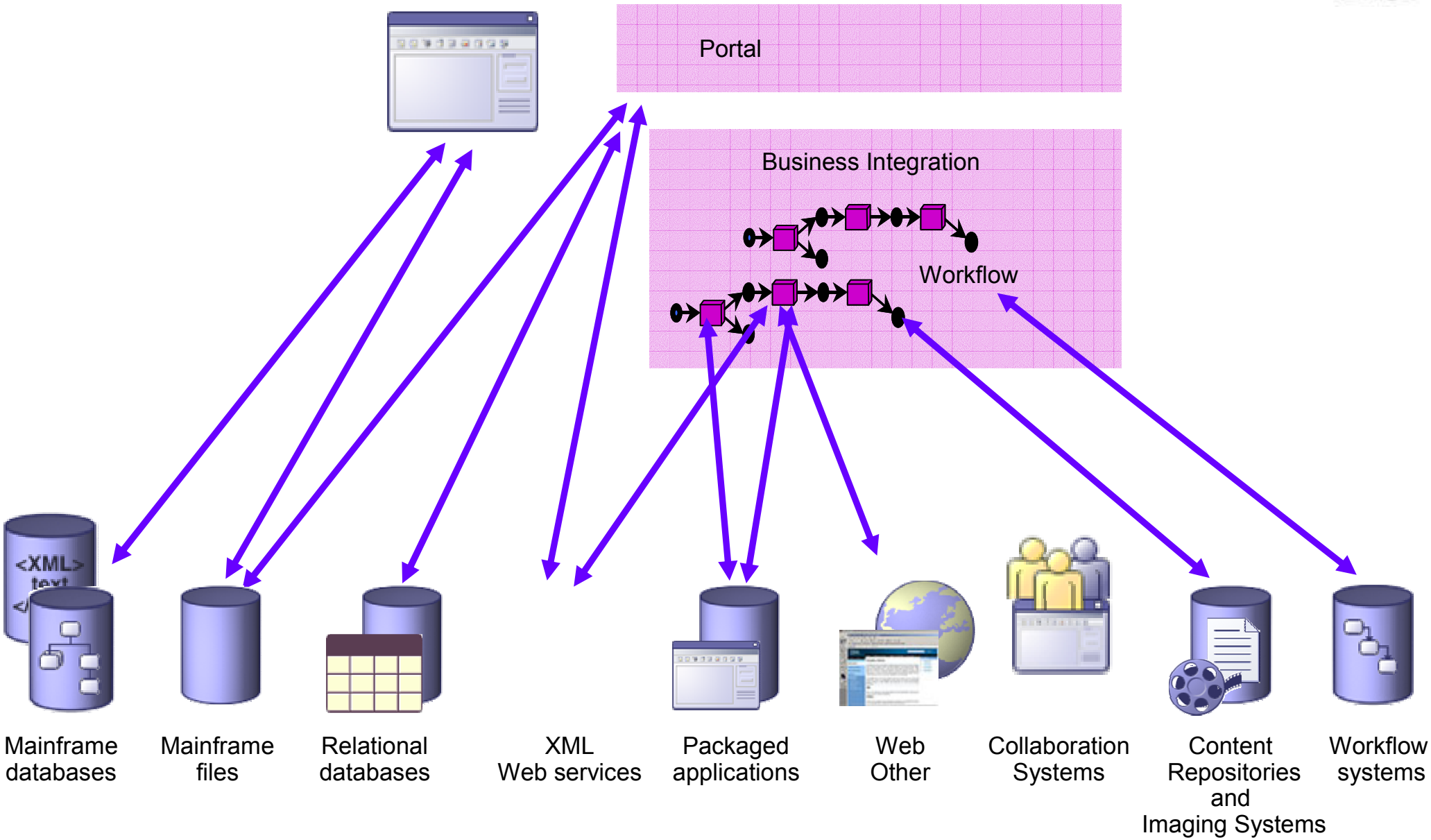
Individually, each product is a standalone solution

Federated Sources



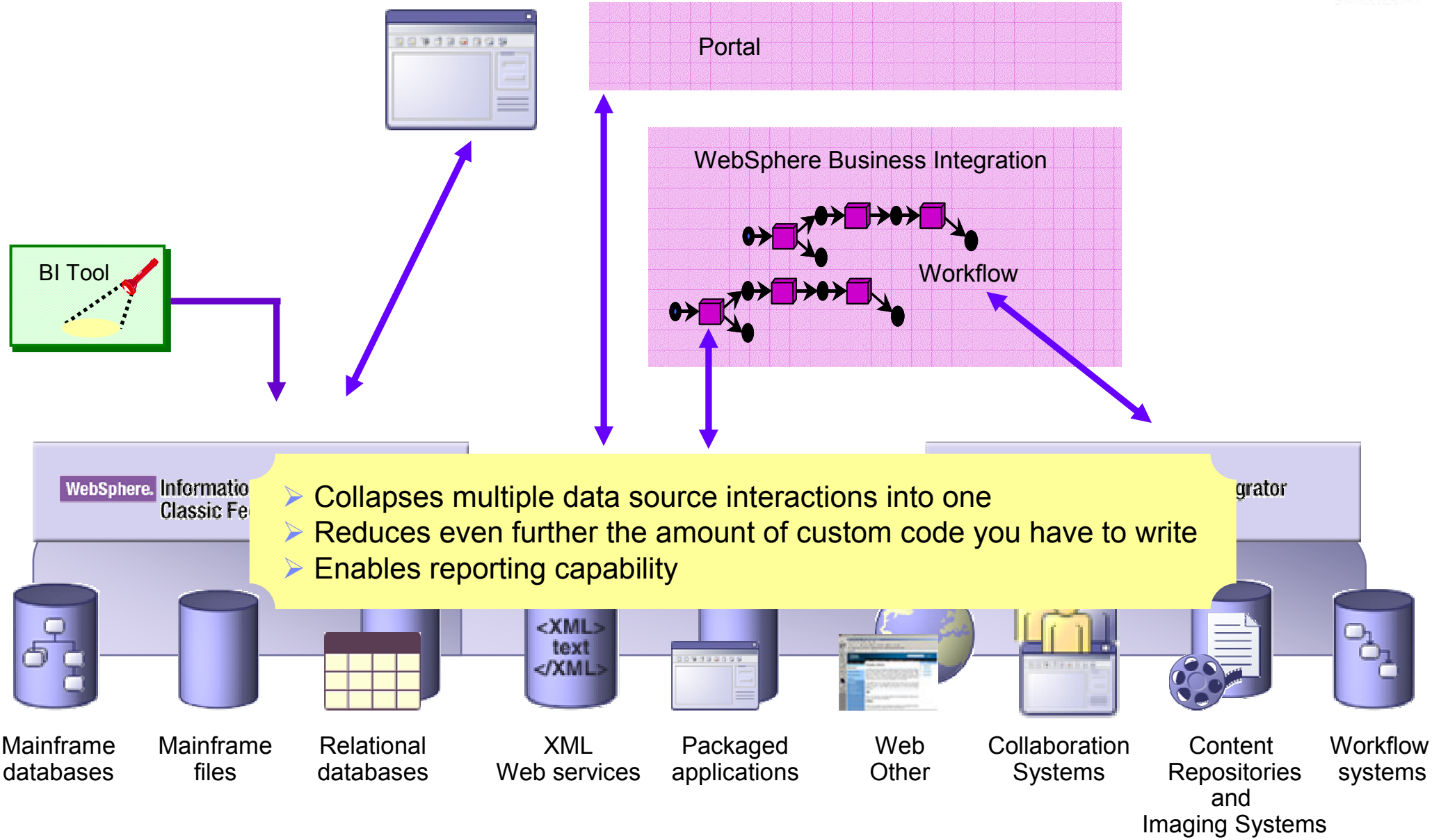
Plus partner tools and custom-built connectors extend access to more sources

Enhance Business Intelligence and Business Integration



IBM Software et IBM System & Technology Groups

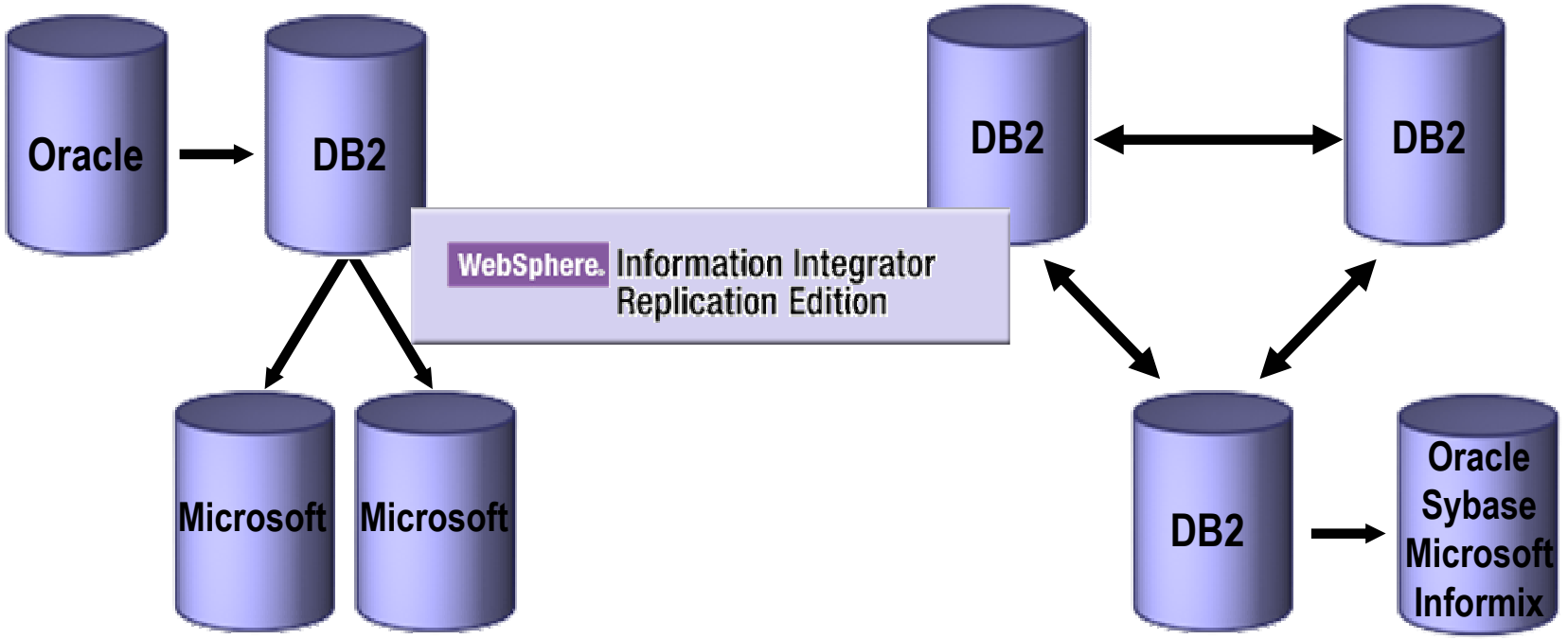
Enhance Business Intelligence and Business Integration



- Collapses multiple data source interactions into one
- Reduces even further the amount of custom code you have to write
- Enables reporting capability

Replication

Manage placement for performance, currency, and availability



SQL Replication

- Flexible scheduling, transformation, distribution
- Typically used for business intelligence, distribution and consolidation, application integration

Q Replication

- Low latency, high volume replication
- Conflict detection and resolution (DB2 UDB)
- Typically used for high availability, workload distribution, application integration

Event Publishing

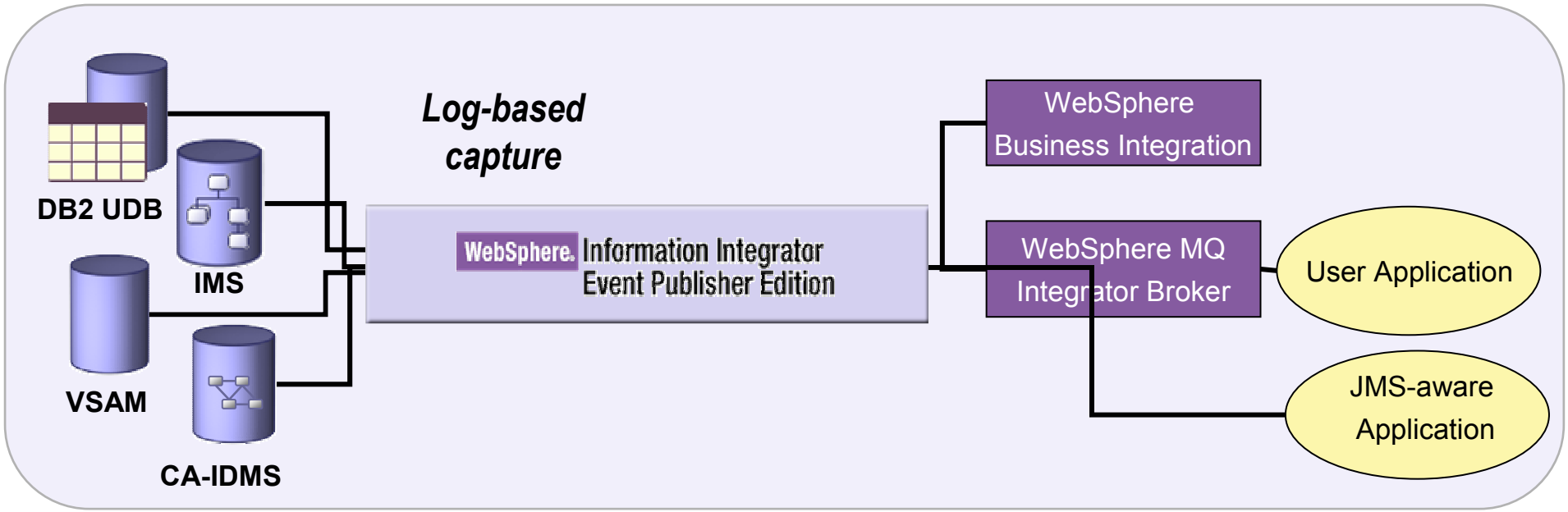
Capture database changes as XML messages and publish them to WebSphere MQ

Function

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

Usage

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



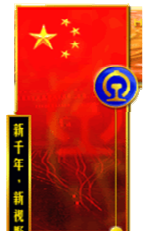
Speeding Business Integration

- **Use standard SQL skills and tools**
  - Leverage existing assets
- **Unified Views, e.g. Customer Data Integration**
  - Reduce hand-written code and development time by half
- **Increasing Portal Adoption and Satisfaction**
  - Correlate disparate information, Reduce portal maintenance
- **Business Performance Management**
  - Correlate real-time events with historical views or planned targets
- **Data-driven Application Integration**
  - Replicate data between application databases
  - Support consolidation and distribution between HQ and branch/store loca
  - Publish events from data to drive processes or link applications

Select Customers by Top Industries

1) Financial Services
2) Manufacturing
3) Insurance
4) Retail
5) Life Sciences
6) Service
7) Imaging/Publishing
8) Telecommunications
9) Automotive
10) Government

Merrill Lynch
citistreet
招商银行 CHINA MERCHANTS BANK
Pioneer
tsmc
CATERPILLAR
泰康人寿 TAIKANG LIFE
ITERGO
FONDIARIA
RadioShack
GALERIA KAUFHOF
Aventis
CYTOGEN
CHIRON
west
ADP
HYATT
Canon
QuadGraphics
速博 sparq
T··Mobile
DAIMLERCHRYSLER
mazda
Kawasaki
Let the good times roll.



Chinese Ministry of Railways

MINNESOTA DEPARTMENT OF AGRICULTURE
FROM THE FARM TO YOUR FAMILY

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

Technical RedBooks

For a complete list of Redbooks, visit the following Web site:
http://www.redbooks.ibm.com/
To find Redbooks related to WebSphere Information Integrator use some of the following keywords in the Redbooks Search field:

- Information Integration
- Federation
- Replication

Redbooks	URL
DB2 II Performance Monitoring, Tuning and Capacity Planning Guide	http://www.redbooks.ibm.com/abstracts/sg247073.html?Open *Redbook*, published 04 November 2004
Patterns: Information Aggregation and Data Integration with DB2 Information Integrator	http://www.redbooks.ibm.com/abstracts/sg247101.html?Open *Redbook*, published 20 September 2004
WebSphere Portal Server and DB2 Information Integrator: A Synergistic Solution	http://www.redbooks.ibm.com/abstracts/sg246433.html?Open *Redbook*, published 09 March 2004
XML for DB2 Information Integration	http://www.redbooks.ibm.com/abstracts/sg246994.html?Open *Redbook*, published 15 July 2004
Data Federation with IBM DB2 Information Integrator V8.1	http://www.redbooks.ibm.com/abstracts/sg247052.html?Open *Redbook*, published 16 October 2003
Getting Started on Integrating Your Information	http://www.redbooks.ibm.com/abstracts/sg246892.html?Open *Redbook*, published 11 February 2003
A Practical Guide to DB2 UDB Data Replication V8	http://www.redbooks.ibm.com/abstracts/sg246828.html?Open *Redbook*, published 27 December 2002, last updated 22 July 2003

Université du Mainframe 2005

IBM WebSphere Information Integration

L'offre IBM « Information Integration » sur z/OS

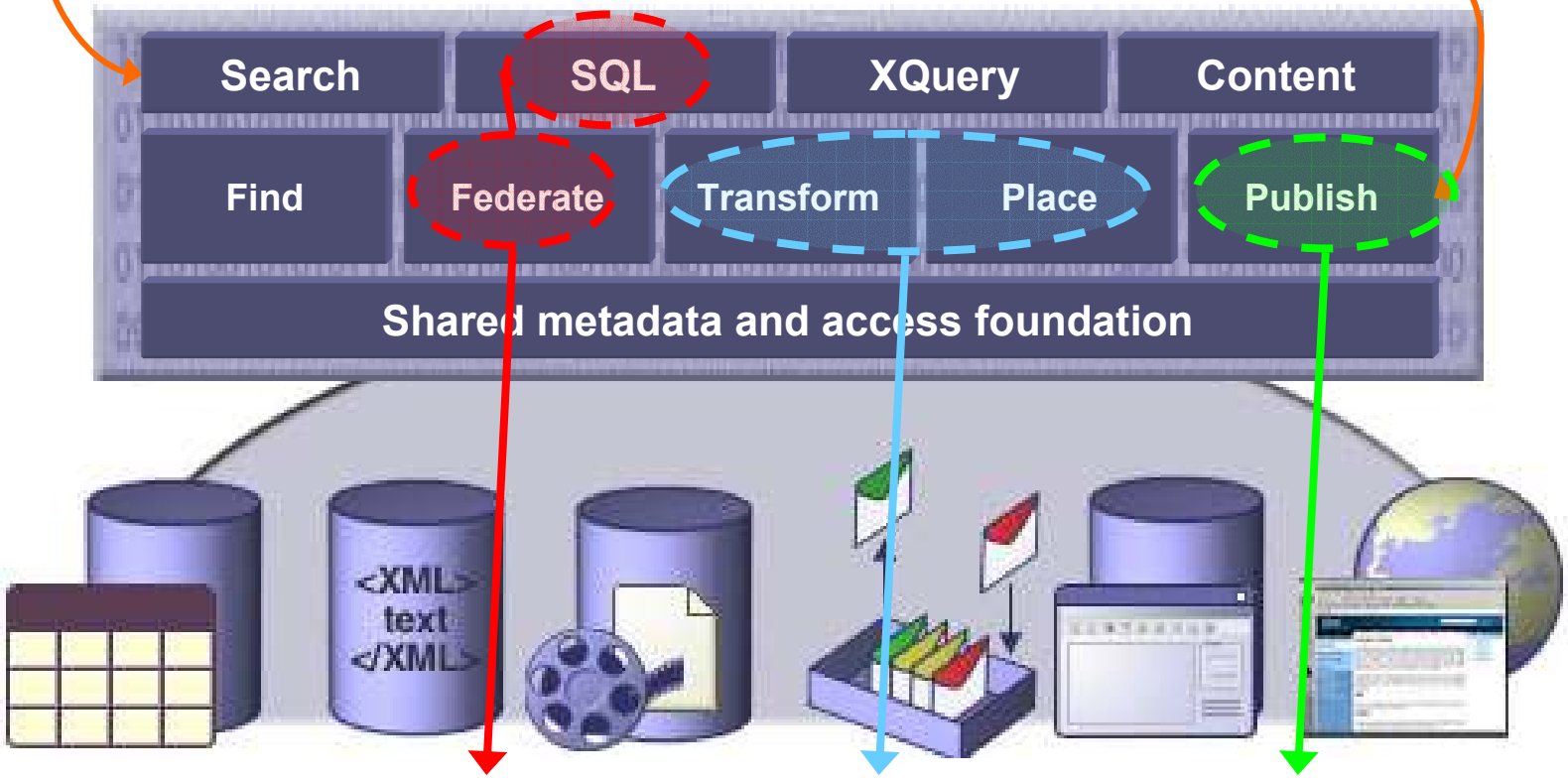
Eric Derbanne
IBM France Software Group
eric.derbanne@fr.ibm.com

IBM Software et IBM System & Technology Groups

IBM Information Integration Vision

Any Data

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



zSeries II Offering :

IBM Software et IBM System & Technology Groups

Université du Mainframe 2005

1. Serveur de Données Fédérées

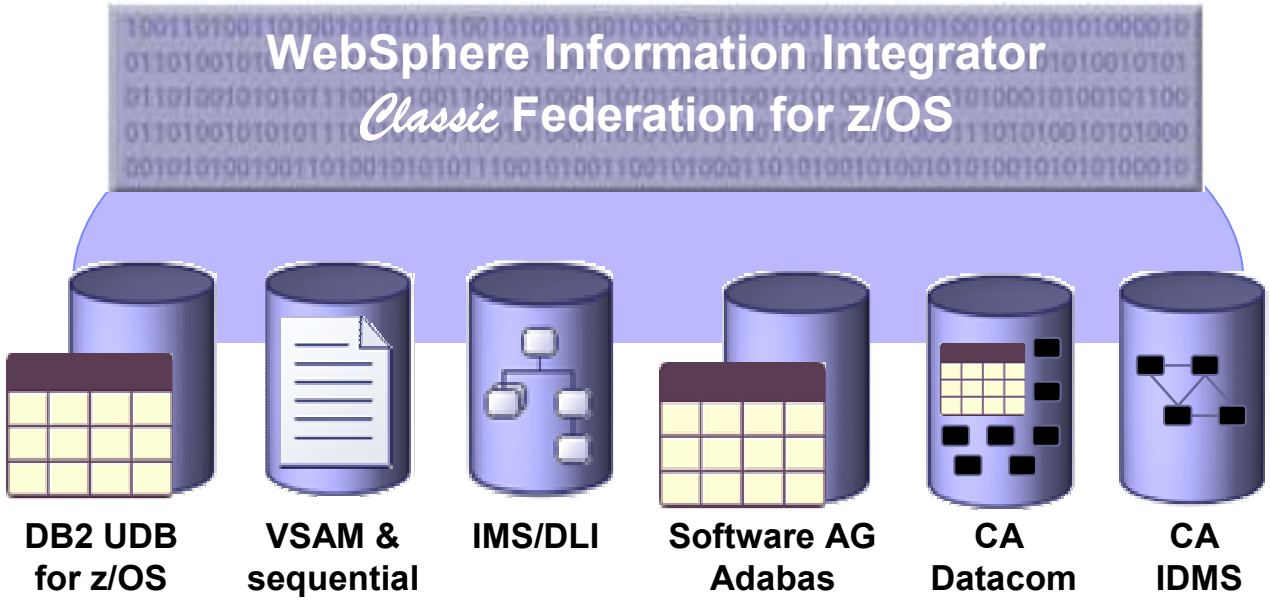
- *WebSphere Information Integrator Classic Federation – z/OS Platforms*
- *WebSphere Information Integrator - LUW Platforms*

IBM Software et IBM System & Technology Groups

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



IBM Software et IBM System & Technology Groups

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

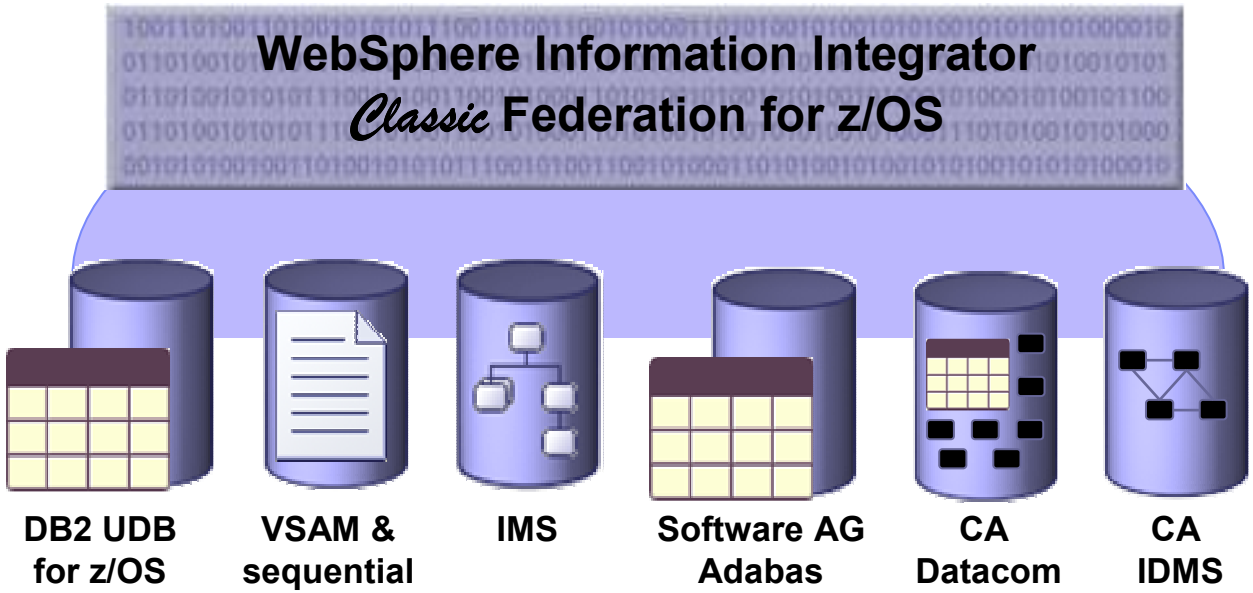
- **Single and two phase commit**
  - Commit - Rollback - Autocommit supported for all data sources
  - Rolling delivery of two phase commit: DB2, IMS and CA-Datacom available now

- **Stored Procedure "Call" leverages existing programs**
  - Reuse mainframe algorithms, allows direct target data source programming if needed
  - Invoke IMS transactions
  - Invoke CICS 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

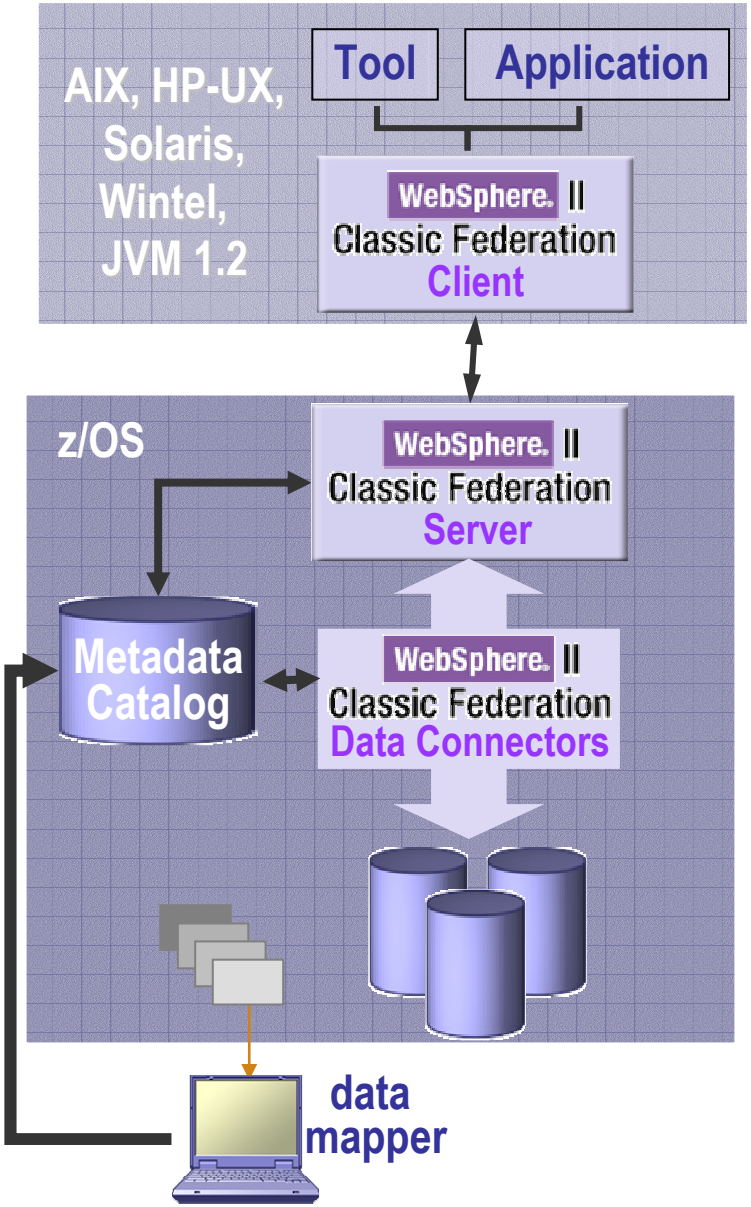
- **Operating Systems**
  - Server:
    - z/OS 1.4
  - Clients:
    - AIX 5.1.0, HP-UX 11.01, Solaris 2.7
    - Windows* (NT, 2000, XP or Server 2003)
- **Communications**
  - TCP/IP or MQ Series v5 or higher
- **Databases**
  - DB2 UDB for z/OS 6.1 or 7.1
  - IMS/DB 7.1
  - CA-IDMS 13 or 14
  - CA-Datacom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



* Data Mapper requires NT, 2000 or XP

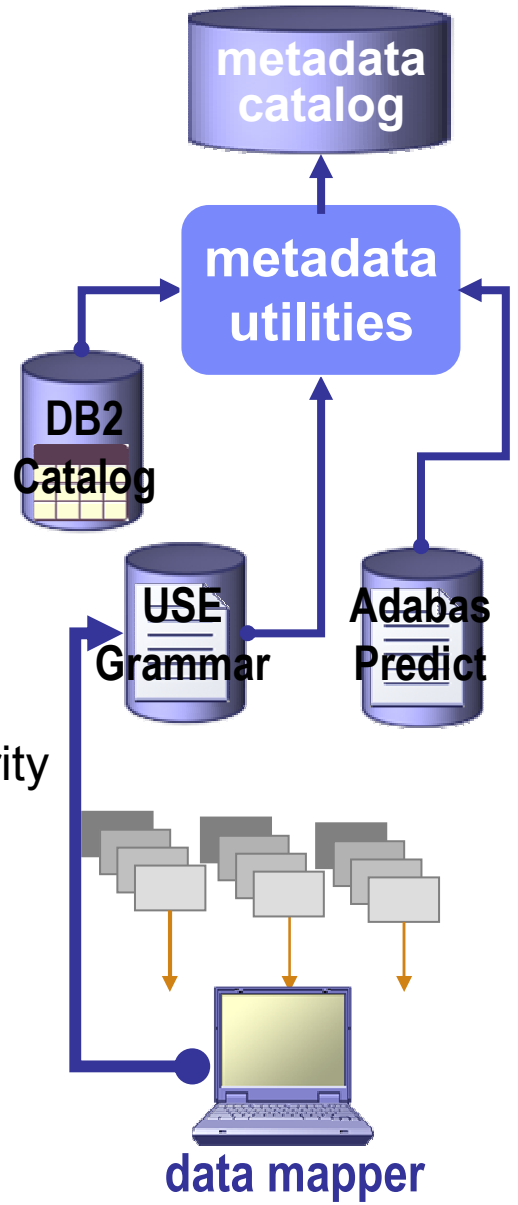
WebSphere II Classic Federation Implementation

- **Create relational description of mainframe data sources by mapping the physical data definitions to logical tables and views**
- **Mainframe Server and components act as a relational database engine**
- **JDBC and/or ODBC drivers provide standardized interface for tools and applications**

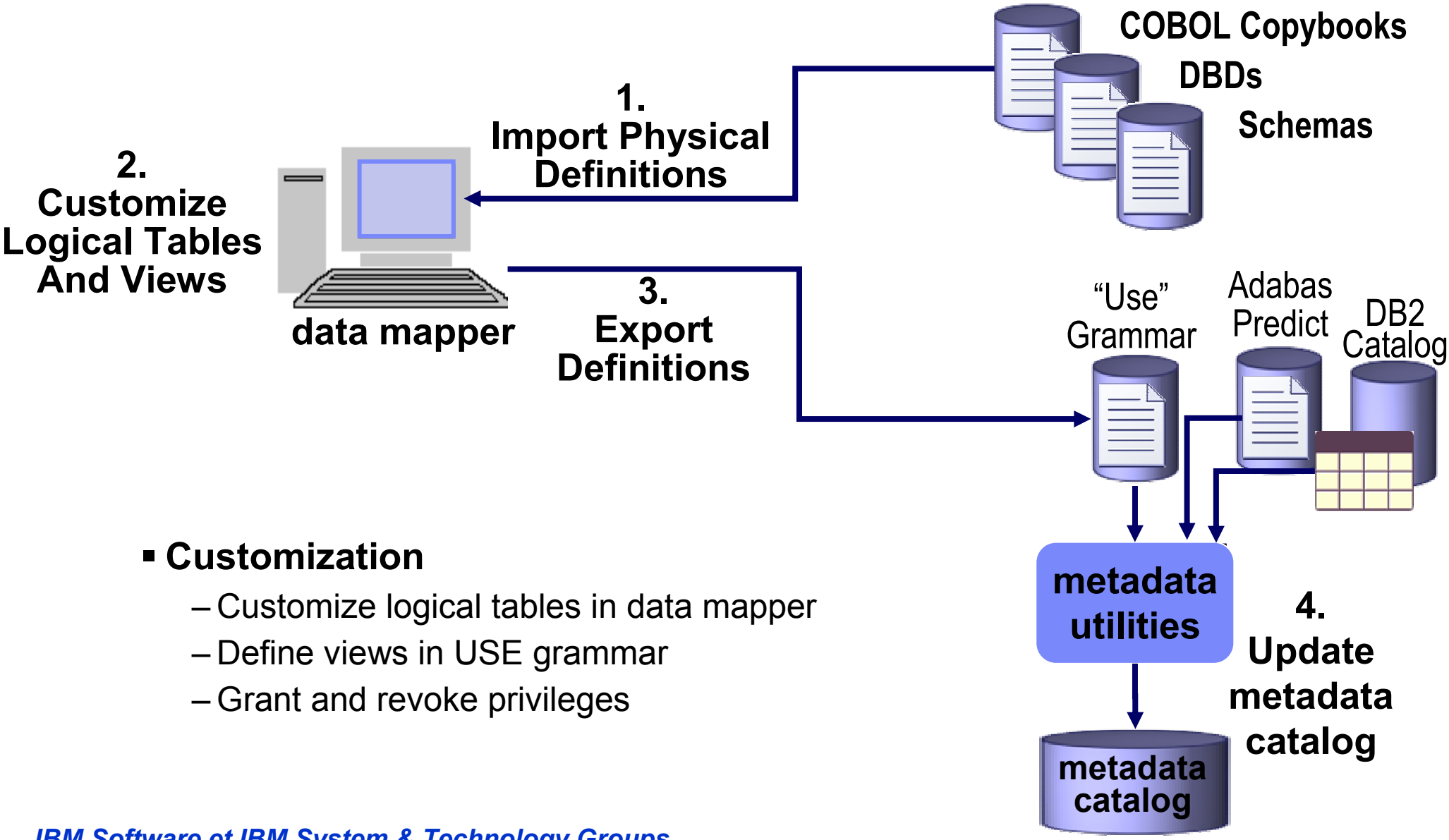


Metadata Management

- **Metadata defines business-oriented relational mappings**
  - Import existing copybooks, IDMS schemas, IMS DBDs, etc.
  - Generate logical relational reference table definitions
  - GUI to customize logical tables to business requirements
- **Simulated RDBMS catalog and more**
  - RDBMS-like catalog support: systables, syscolumns, etc.
  - Query-able tables for non-relational metadata
- **Some metadata-driven features**
  - Automatic translation of legacy data types
  - Handles legacy constructs like recurring data and redefines
  - Complex tables can span segments, records, etc.
  - Metadata-driven filtering using WHERE clauses
  - Enhances security via schema mapping, views, & DB2-like security
- **Metadata Utilities**
  - Create and update metadata catalog entries
  - Verify metadata against physical (e.g. VSAM index checks)
- **Data mapper**
  - Metadata customization and visual administration



Metadata Management Workflow



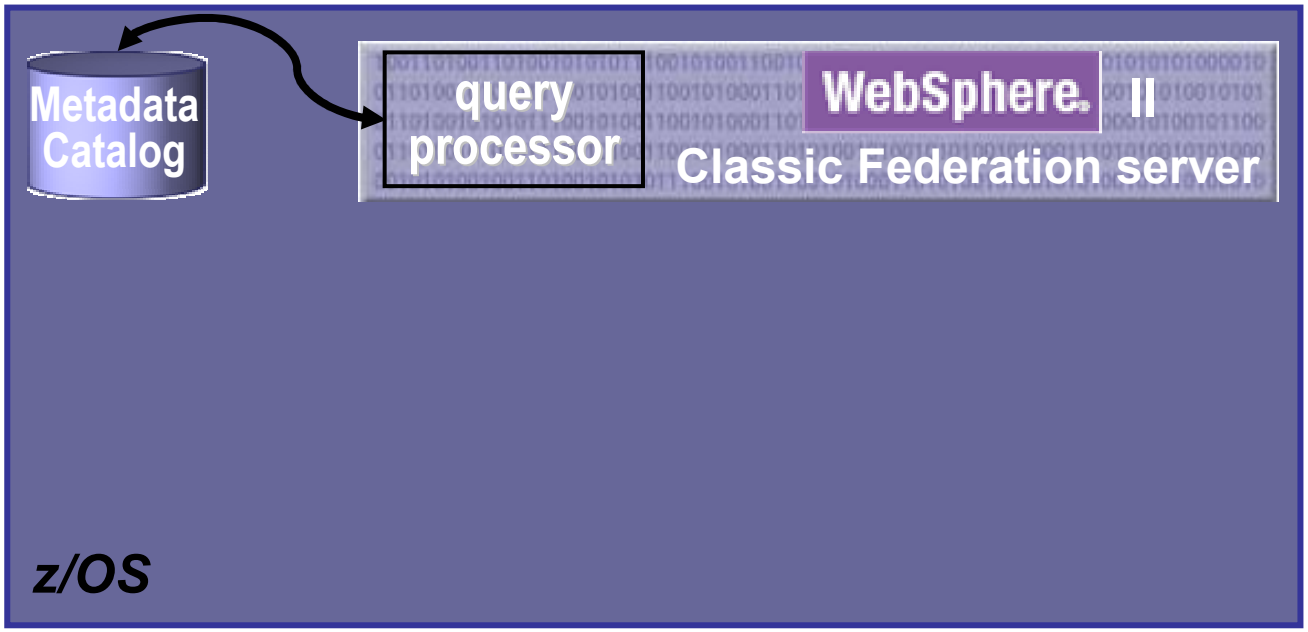
- **Customization**
  - Customize logical tables in data mapper
  - Define views in USE grammar
  - Grant and revoke privileges

Component Overview

AIX, HP-UX, Solaris, JVM 1.4,
Widows NT, 2000, XP,
UNIX System Services

Portal | BI Tool | Servlet | Servlet | Client class

EJB



z/OS-based data server

- Runs in its own address space
- Responds to catalog queries
- Manages mainframe resources
- Distributes WebSphere II Classic Federation SQL to data connectors
- Assembles WebSphere II Classic Federation result sets

In-depth look – Query Processor

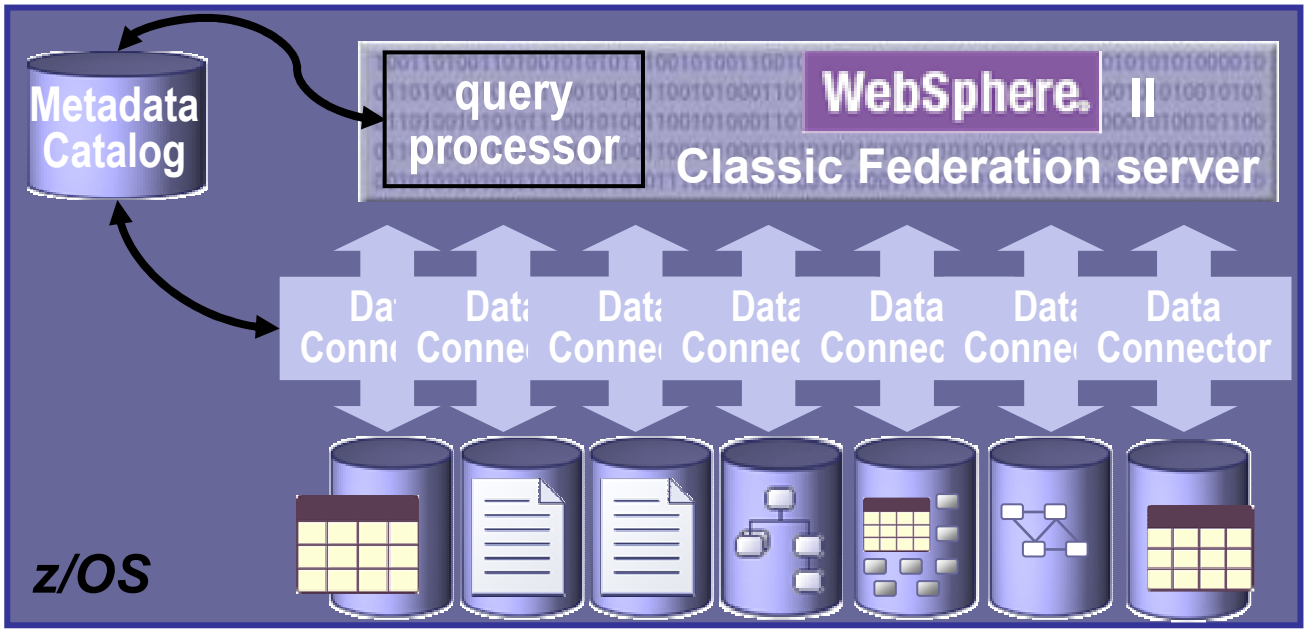
- Based on SQL, determines the type or types of files and/or databases to be accessed, and passes control to the appropriate database connectors. If required, rewrites the SQL query into one or more native file or database commands.
- Optimize native data access calls based on the contents of the SQL statements and the known characteristics of the file or database. The guiding principles of query optimization are :
  - **least amount of data should be read.**
  - **all database access aids should be utilized.**
- Ensure that the criteria information needed to filter the data to satisfy the SQL qualification, is **pushed down** to the database connectors
- Construct a standard relational result set. For non-relational data stores this involves restructuring the data, as it is accessed, into columns and rows and translating data types. When segments contain embedded recurring data items (OCCURS and OCCURS DEPENDING ON constructs.), perform **normalization** of the data.
- Perform post-processing of result sets as needed (e.g., ORDER BY or GROUP BY).
- Process catalog queries using the metadata catalog
- Process INSERT, UPDATE and DELETE with commit/rollback suppor

Component Overview

AIX, HP-UX, Solaris, JVM 1.4,
Widows NT, 2000, XP,
UNIX System Services

Portal | BI Tool | Servlet | Servlet | Client class

EJB



Data connectors

- Database-specific read/write engines
- Determine how to access databases and files
- Leverage native access aids: keys, indexed, etc.
- Reentrant for small footprint

In-depth look – IMS Connector

- Requires a mapping between the physical record layouts and one or more logical relational tables, honoring the hierarchical relationships between the segments / secondary indexes.
- Datamapper ensures this by using IMS DBD's and COBOL copybooks to generate "logical relational table definitions" , which is stored in the metadata catalog.
- DRA/ DBCTL is the recommended interface into IMS, as it supports dynamic scheduling of multiple PSBs that is needed for a multi-user environment. This is the same method used by CICS to interface with IMS DB databases. Even IMS/DC shops have DBCTL, even though they might not use it.
- The IMS connector acts as the coordination controller with IMS DB and is responsible for scheduling PSBs, issuing DL/I calls, unscheduling PSBs and resolving indoubt units-of-recovery. IMS DB connections and DL/I call completion are monitored using a DRA Control Exit running in the data server's address space.address space.
- SSA generation is highly optimized based on the contents of the WHERE clause.
- Joins are optimized in two ways: **by mapping paths in the IMS hierarchy**, and **using appropriate JOIN attributes in the SQL.**
- **Path calls**: When accessing a path in the hierarchy, IMS DB "path calls" will be issued if the WHERE clause references a search field located in the bottom-most "leaf" segment in the hierarchy.

In-depth look – Adabas Connector

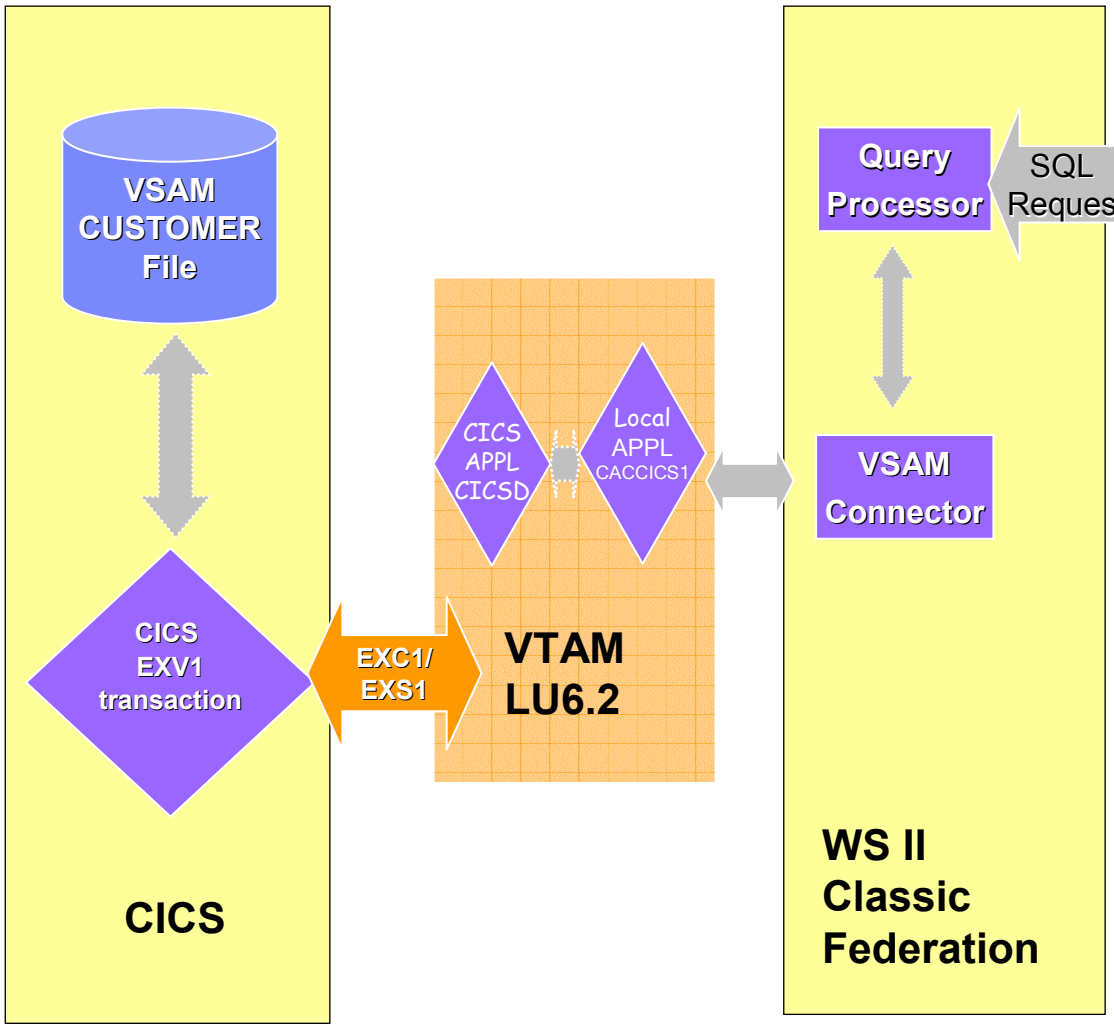
- Uses the direct call interface to access Software AG Adabas data
- Connects to the Adabas nucleus to service the component of the SQL statement that is passed to it by the data server's query processor
- A distinct connection is established for each logical table in the query. Users do not share connections to the Adabas nucleus. Once a query has been processed, the connections to the nucleus on behalf of that user are released
- Descriptor, super descriptor and sub-descriptor information can all be provided as part of the Adabas related metadata. This information is then automatically used by the Adabas connector to optimize access to the Adabas data
- Does support complex Adabas constructs including Periodic Groups (PE) and Multiple-Value fields (MU) as well as descriptors, superdescriptors and subdescriptors in indexing
- Does support complex SQL processing including cross-database joins.
- Does NOT depend on a CICS transaction environment or Adabas access
- Does access Adabas from its own address space with integrity so that there is no impact on existing online transaction processing environments.

In-depth look – IDMS Connector

- Requires a mapping between the physical record layouts and one or more logical relational tables, honoring the relationships between the records, ie, SET's.
- Datamapper ensures this by using CA-IDMS' schema and subschema information to generate "logical relational table definitions" , which is stored in the metadata catalog.
- Uses the native DML interface to access IDMS data. Determines the DML call(s) to issue based on the SQL statement's WHERE clause, and the metadata information for the logical table (as generated above).
- Establishes a separate run unit for each IDMS "branch" of the execution tree created by the Query Processor from the SQL statement.

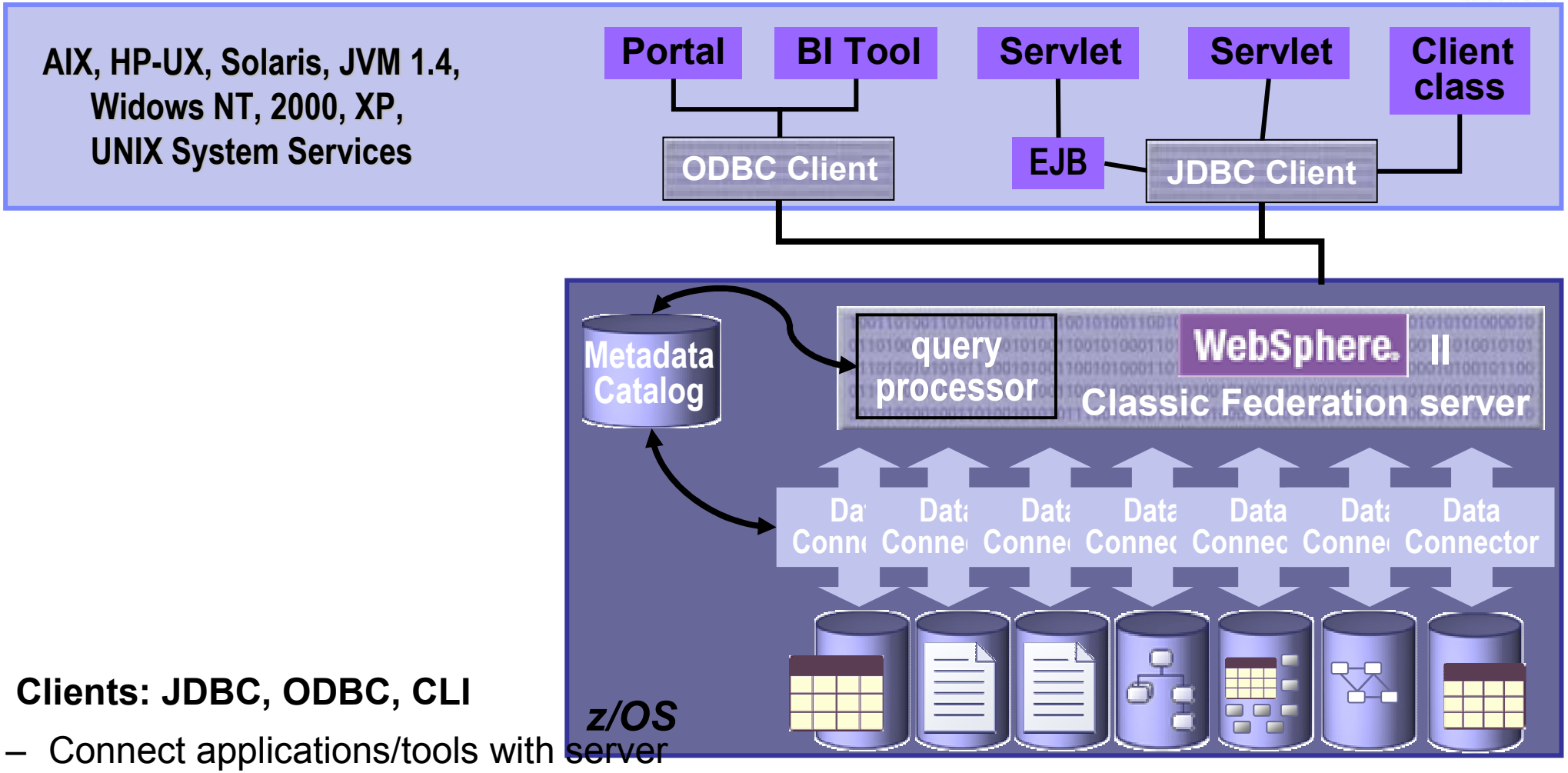
- **Unique feature**
  - Supports access to multiple IDMS databases from a single integration server. It can access multiple databases within a single IDMS CV or across multiple IDMS CVs. This empowers an organization to JOIN data across business units, divisions and/or other independent operations when each is using it's own dedicated IDMS databases.

In-depth look – VSAM Connector



- Uses the supplied CICS transaction EXV1 to access VSAM data – leverages powerful CICS VSAM I/O.
- Some VTAM & CICS expertise needed to configure – challenge in some customer environments
- Mapping required between the physical record layouts and one or more logical relational tables
- Breakout of one physical file into multiple logical table usually based on REDEFINES & OCCURS constructs
- Superior performance, and scalability obtained by tuning the VTAM & CICS definitions
- Supports transactions (currently single phase commit only)

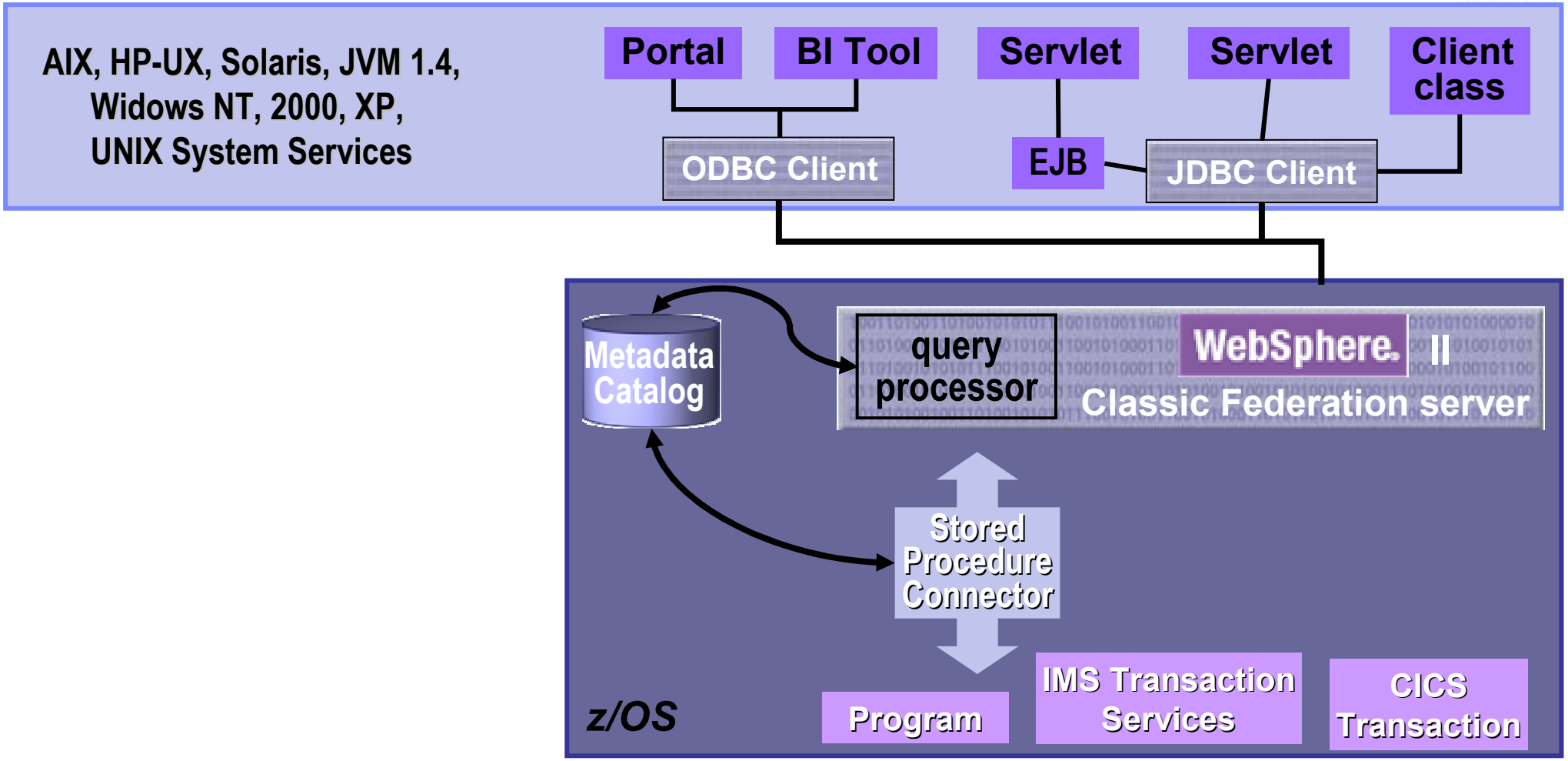
Component Overview



Clients: JDBC, ODBC, CLI

- Connect applications/tools with server
- Multi-threaded, standardized interfaces
- TCP/IP & MQ communication protocols
- Service catalog queries for both products
- Service SQL data access requests

Component Overview



Stored Procedure Connector

- Invokes mainframe program running within data server address space
- Returns: Single result as parameters OR Multi-row result set

Two broad categories of usage

- **e-Business**
  - Deliver mainframe data to
    - Self-service portals (real-time account details)
    - e-commerce solutions (real-time inventory)
    - Employee portals (real-time claims detail)
  - Web developers become productive with no mainframe skills
  - Eliminates data latency business issues caused by copied data

- **Business intelligence**
  - Integrates seamlessly with
    - Reporting and analytical tools, e.g. Business Objects
    - Portals, e.g. WebSphere Portal
    - ETL, e.g. Ascential DataStage

Will be covered later at end of session

Integrating Enterprise Data

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

WS II *Classic* Federation

Integrated SQL View
WebSphere Information Integrator
SQL, SQL/XML
Federation Engine
Wrappers and functions
ODBC

DB2 Family
Sybase
Informix
SQL Server
Oracle
Teradata
ODBC

New

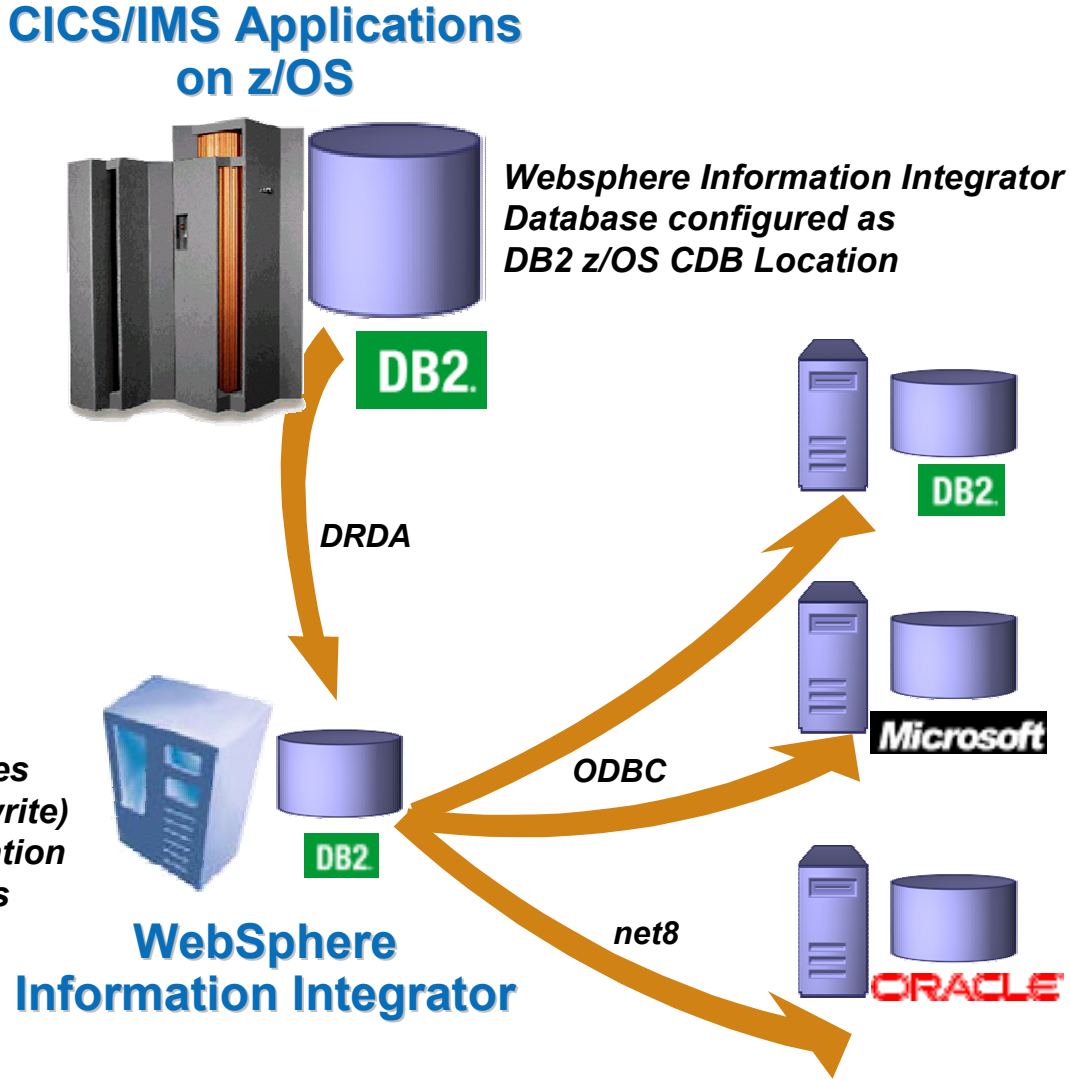
Packaged apps
Biological Data And Algorithms
Text
XML
Excel
WebSphere MQ
IBM Extended Search
WWW, email,...

IBM Software et IBM System & Technology Groups

Access Distributed Heterogeneous Data from CICS Applications

Business Scenario

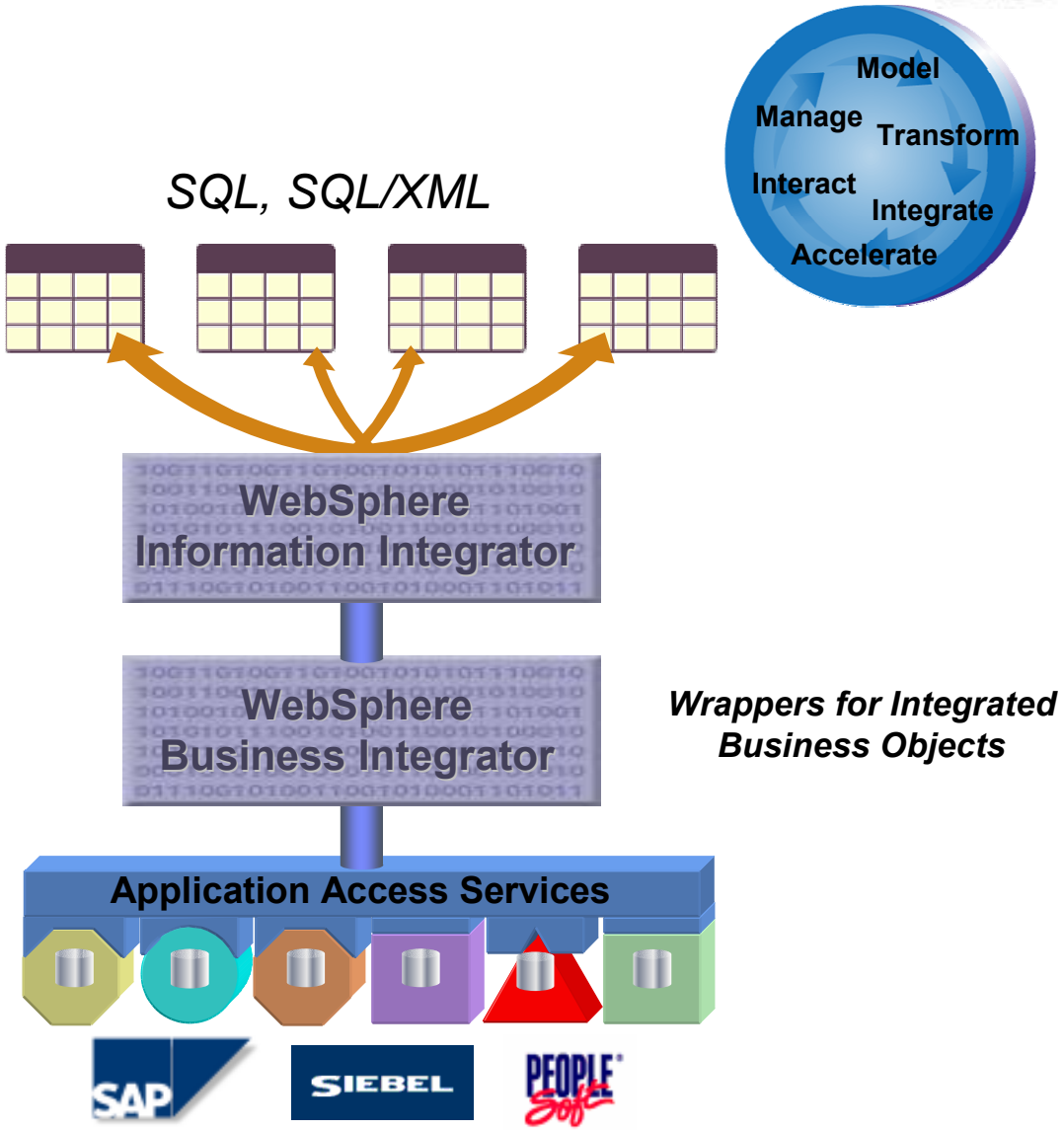
- Access all kinds of DB2 II Data Sources from legacy Applications
- Single Site Update, Multi Site Read
- Distributed UOW with 2PC support in plan for following version of WebSphere II



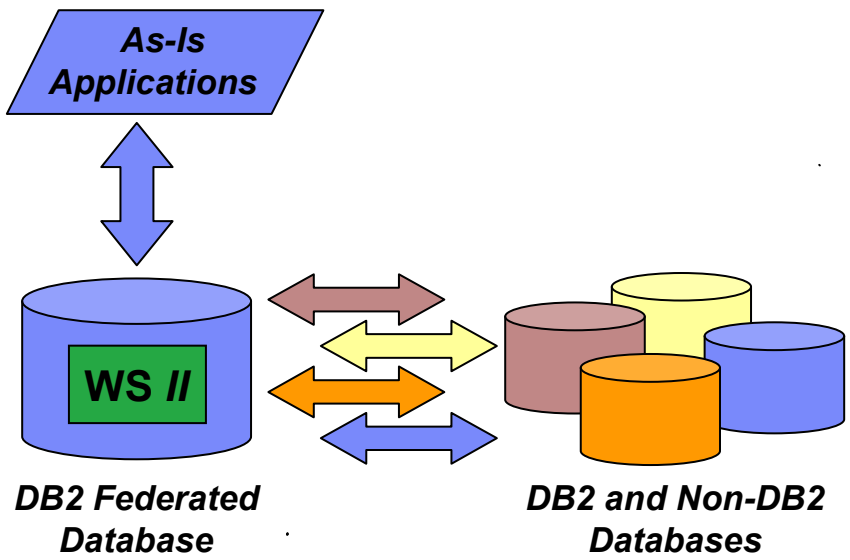
Information and Process Integration with WebShpere

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

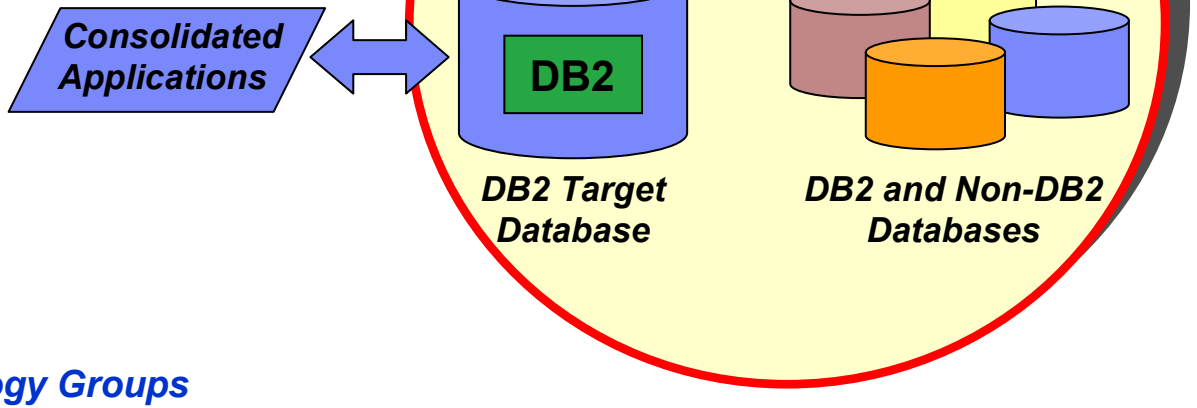


Federate or Replicate – That is the Question !!



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

Data Replication
- Data Movement
- Data Transformation
- Autonomy



Université du Mainframe 2005

1. Serveur de Données Fédérées

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

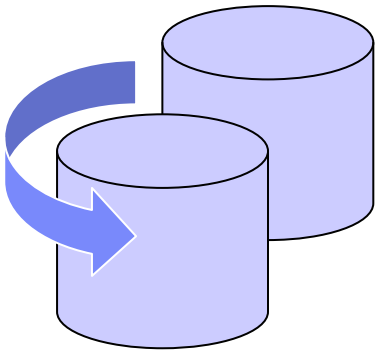
2. Données Répliquées

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

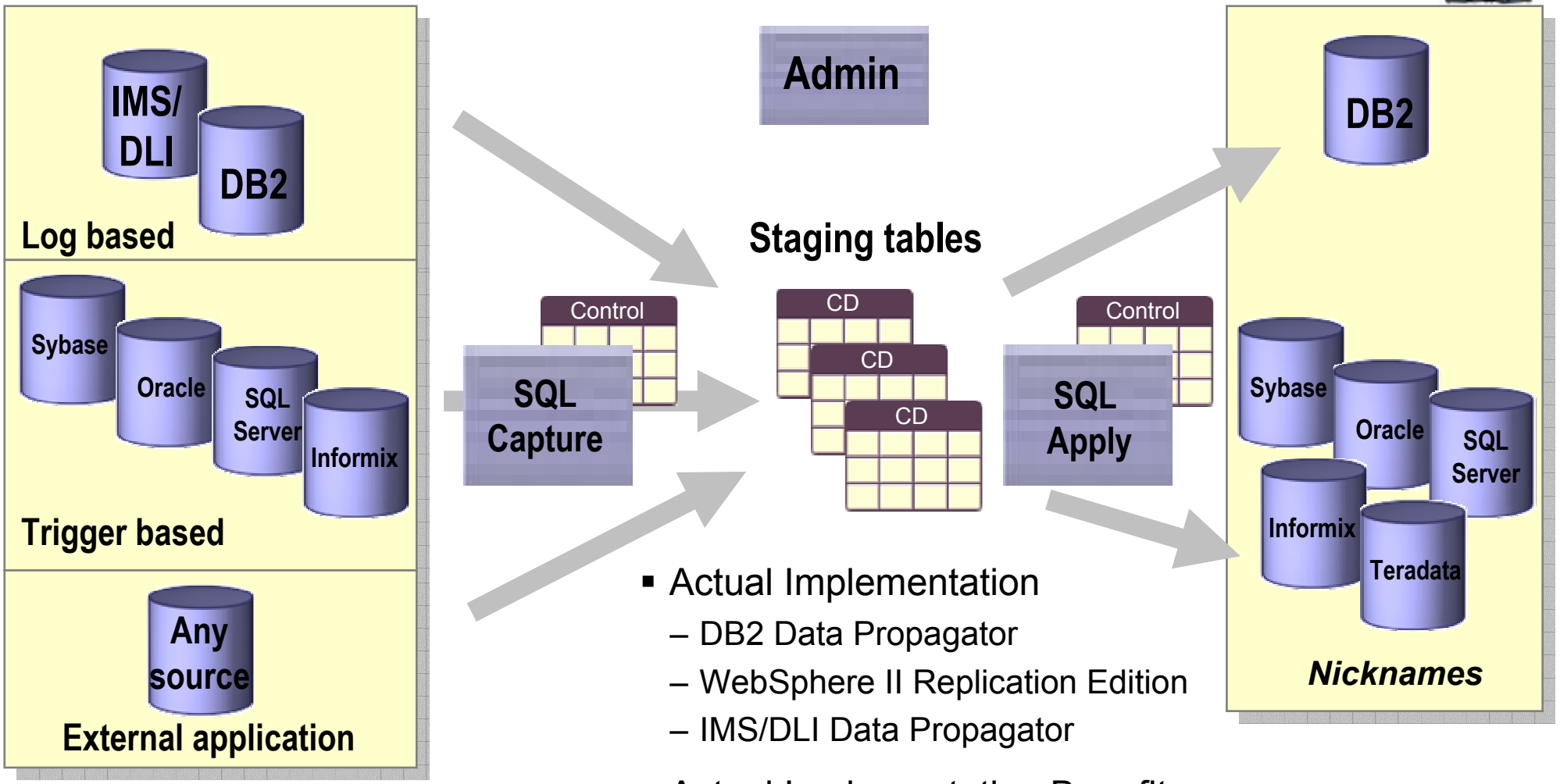
IBM Software et IBM System & Technology Groups

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

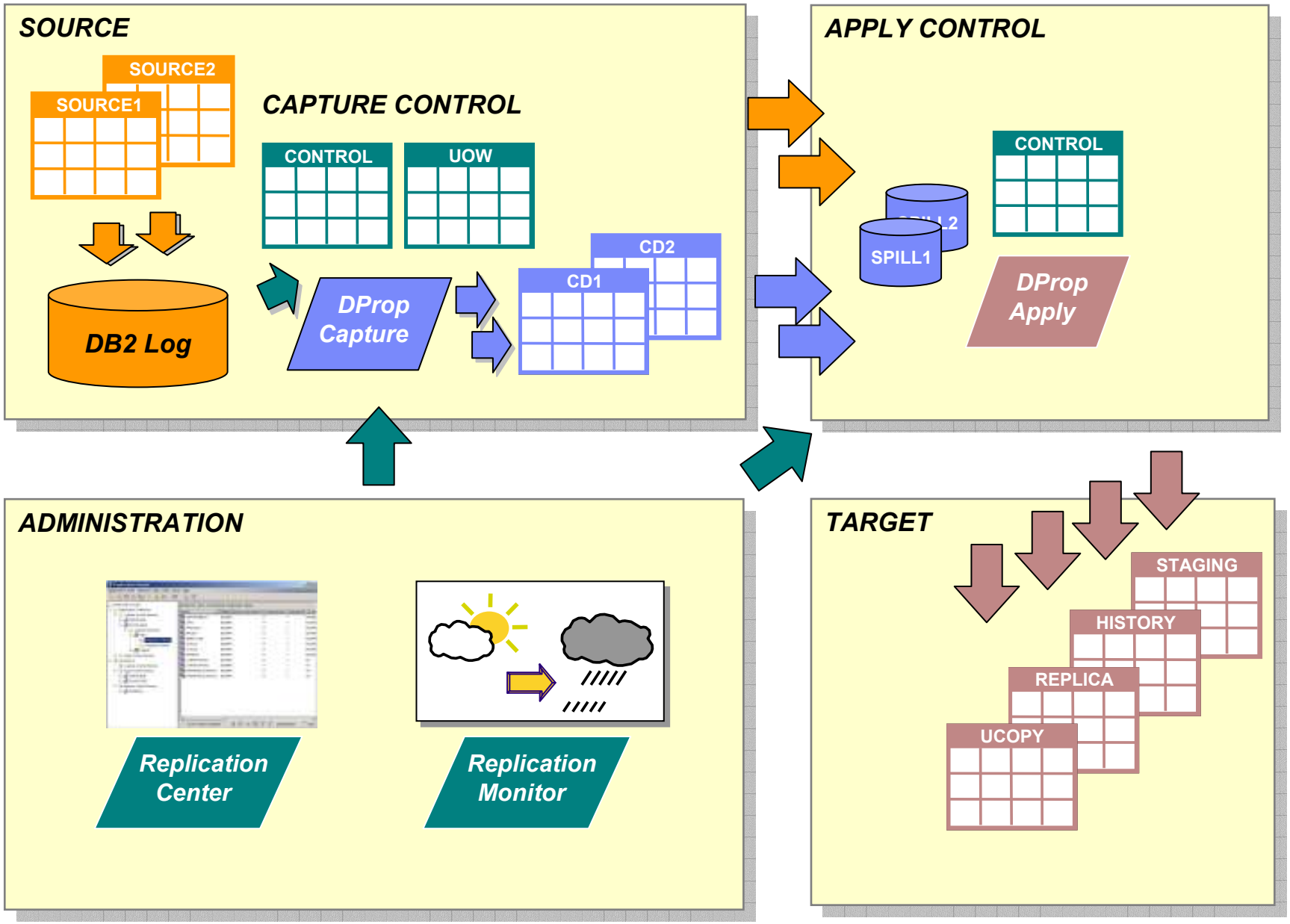


SQL Replication Architecture



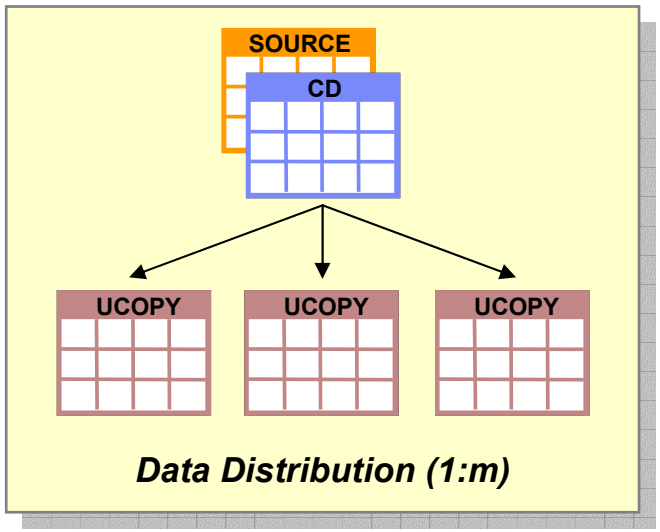
- Actual Implementation
  - DB2 Data Propagator
  - WebSphere II Replication Edition
  - IMS/DLI Data Propagator
- Actual Implementation Benefits
  - Extremely flexible and resilient
  - Very easy to set up transformations
  - Scales well to reach multiple targets
  - Homogeneous & Heterogeneous Sources

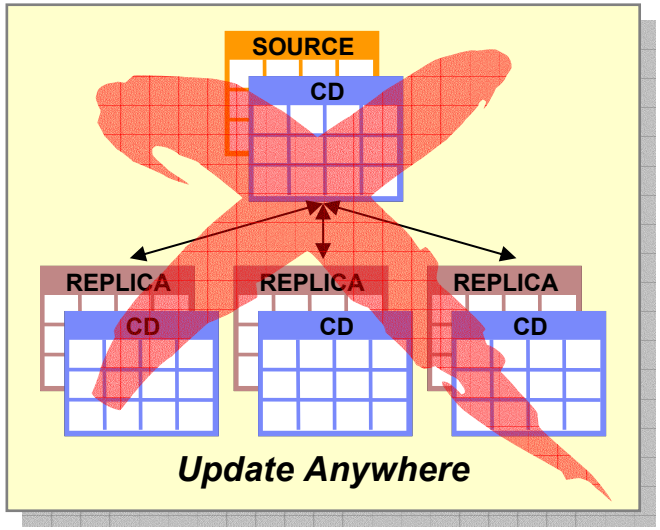
SQL Replication Component Overview

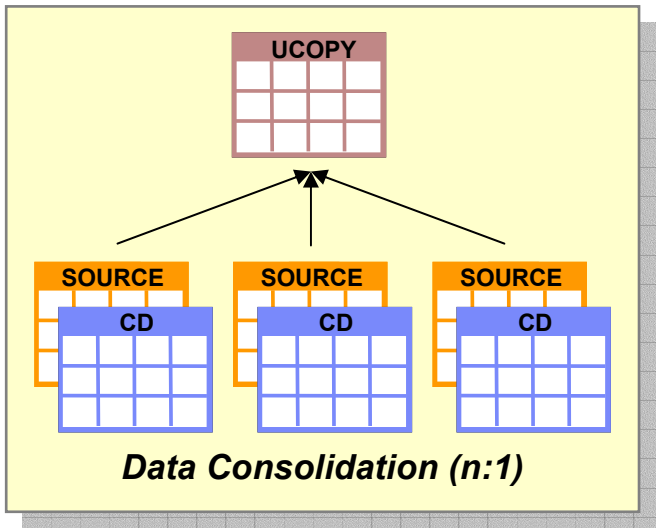


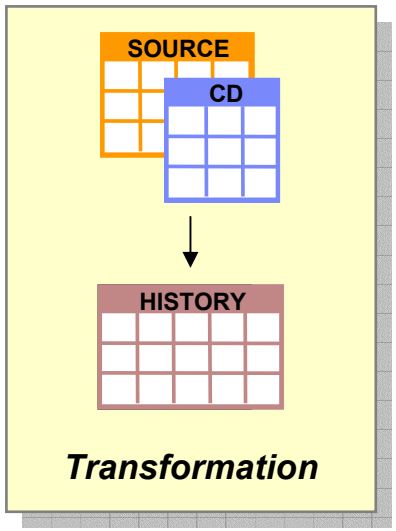
IBM Software et IBM System & Technology Groups

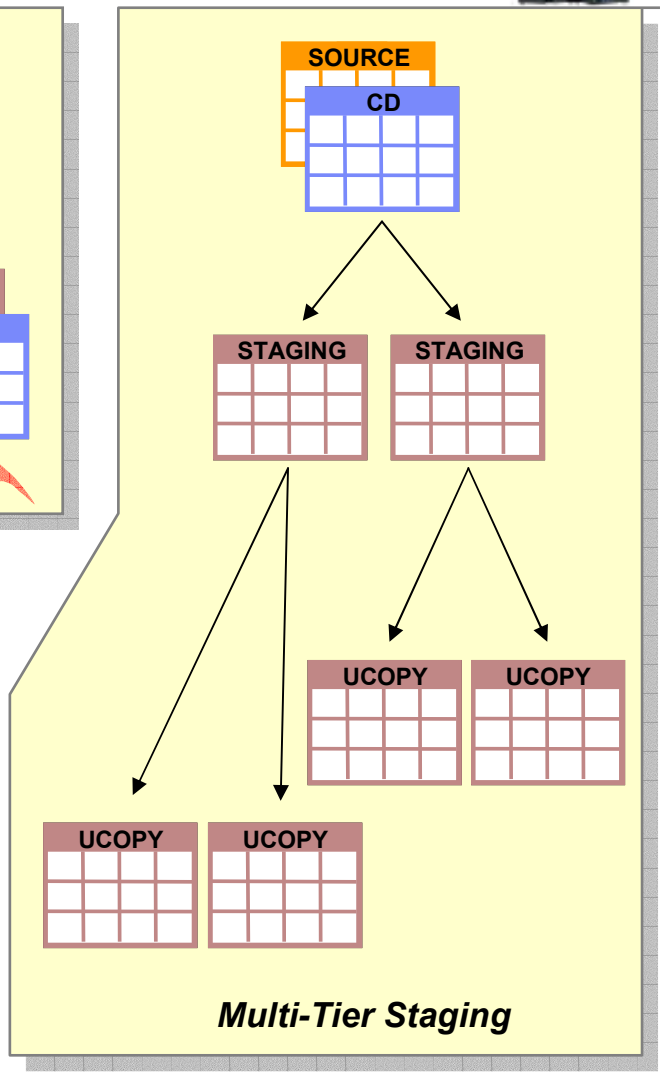
Sample SQL Replication Scenarios







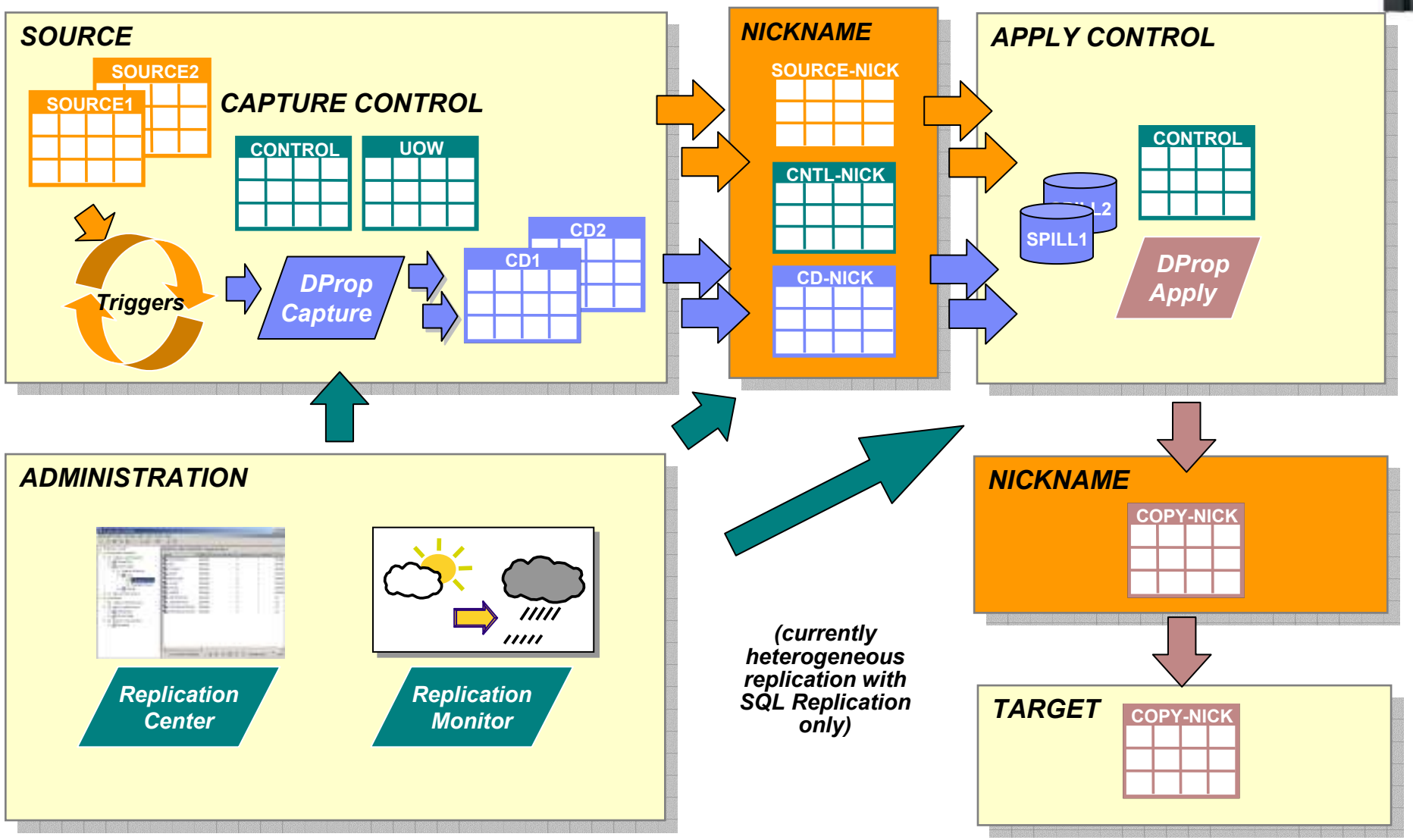




- Subsets
- SQL Transformations
- Updateable Predicates
- Updateable Primary Keys

IBM Software et IBM System & Technology Groups

DB2 Data Replication TO / FROM Federated SOURCES



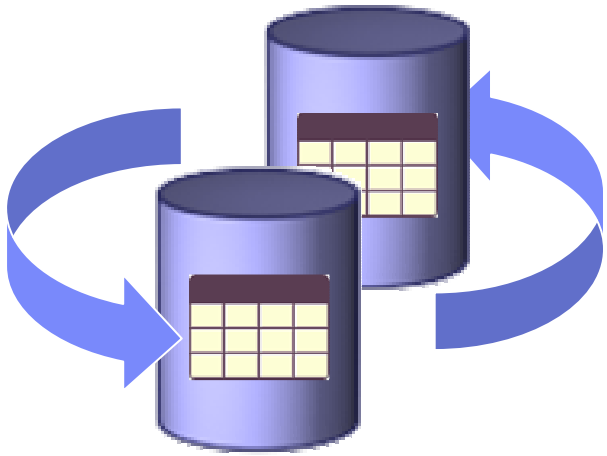
Federated Sources & Targets

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

Why Create Another Replication Architecture?

- Performance
  - Combine high Throughput with low Latency

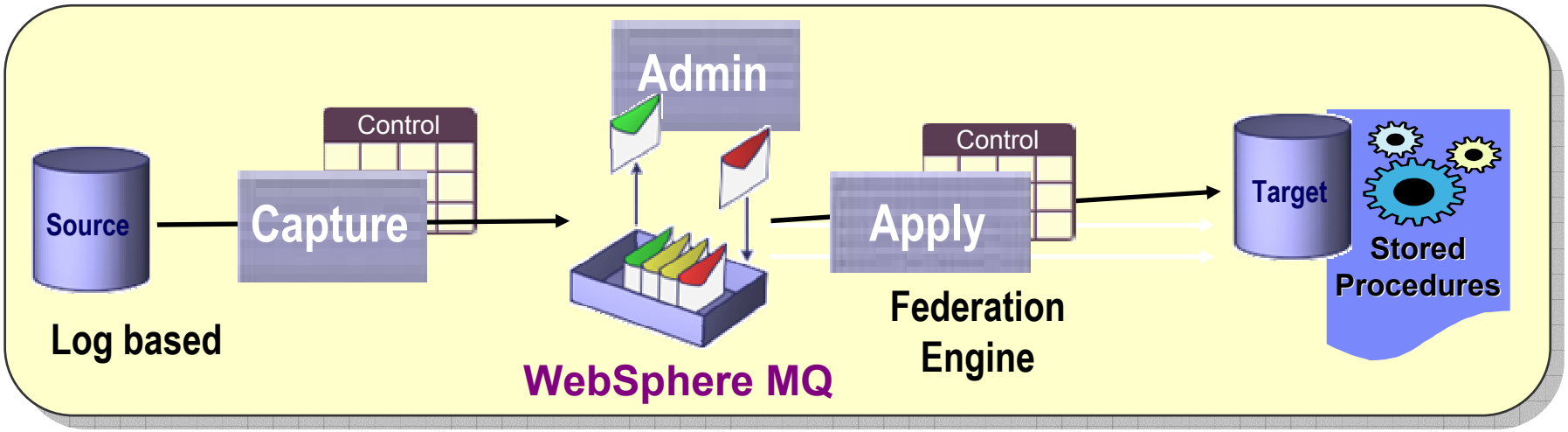
- New Function
  - Event Publishing from DB2 and Classic Sources
- Capability
  - Significantly improve multi-directional Replication Support



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

Q Replication Architecture

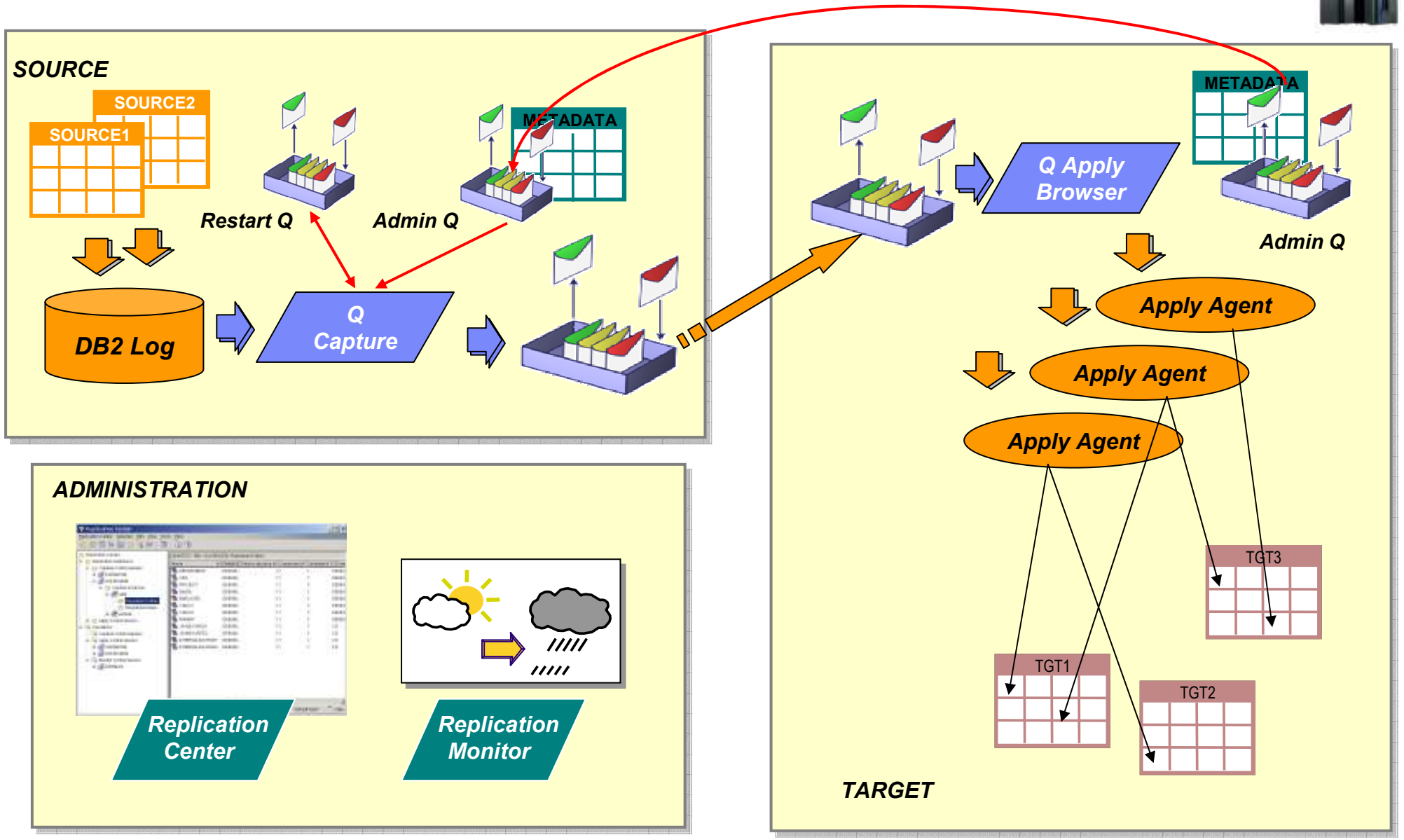
Websphere Information Integrator introduces new replication architecture for delivering extremely low latency replication



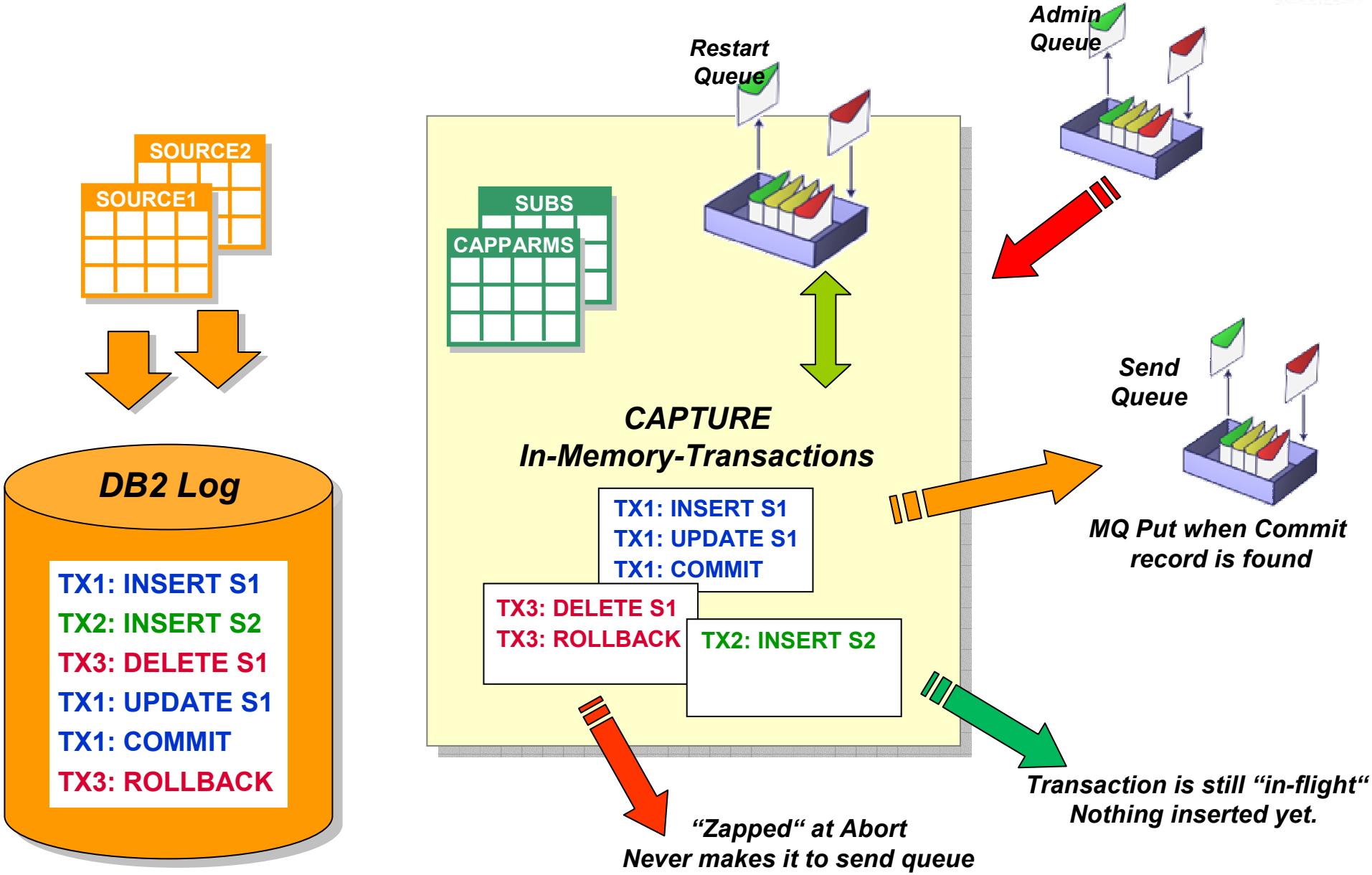
- **Each message represents a transaction**
- **Highly parallel apply process**
  - Non dependent transactions re-parallelized at the target
- **Differentiated conflict detection and resolution**
- **Integrated infrastructure for replication and publishing**
- **DB2 to DB2 today**

- **Data Integrity**
  - Persistent messaging with WebsphereMQ
  - Detects missing messages
- **Data transformation**
  - Triggers on the target table
  - Stored Procs called by Apply at the row level
  - Publish Event to user application
- **DB2 to DB2 today**
  - Staged availability of heterogeneous support

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
```

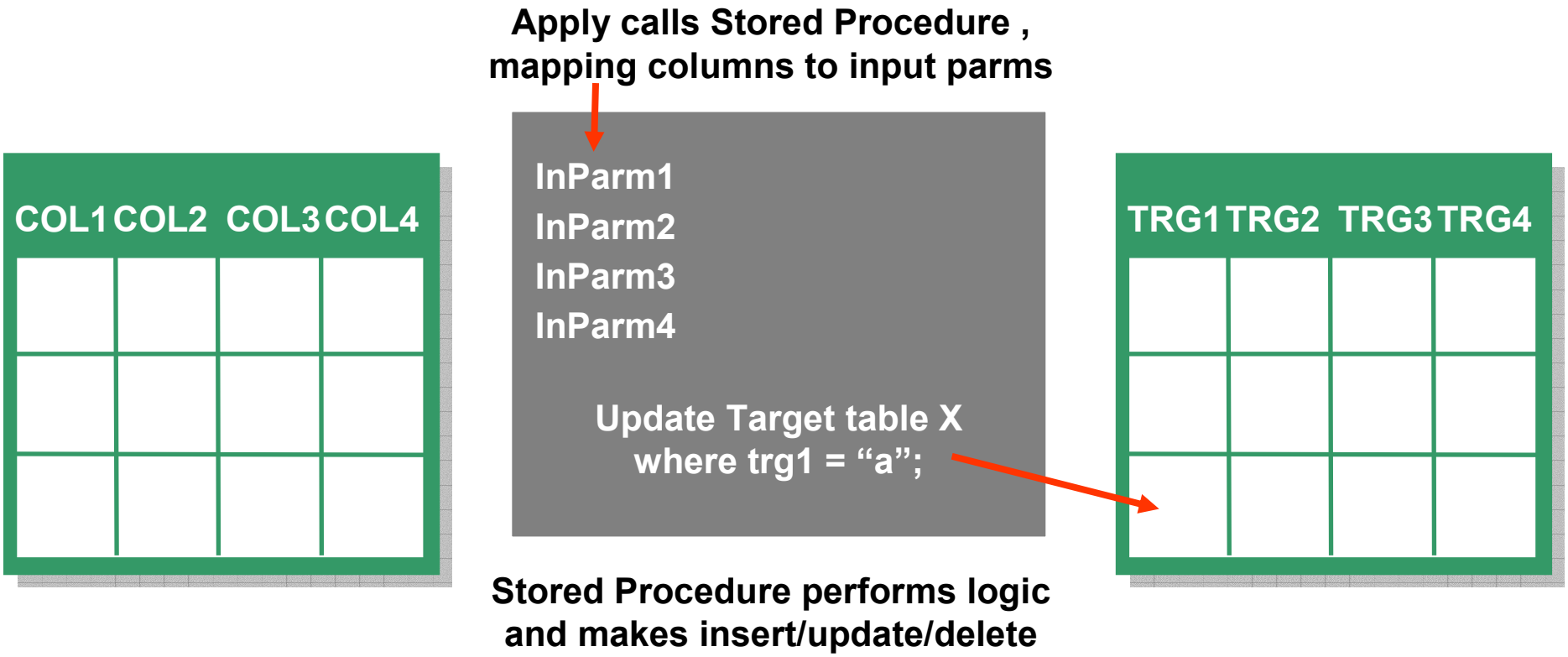
  - Based on values in data in other tables

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

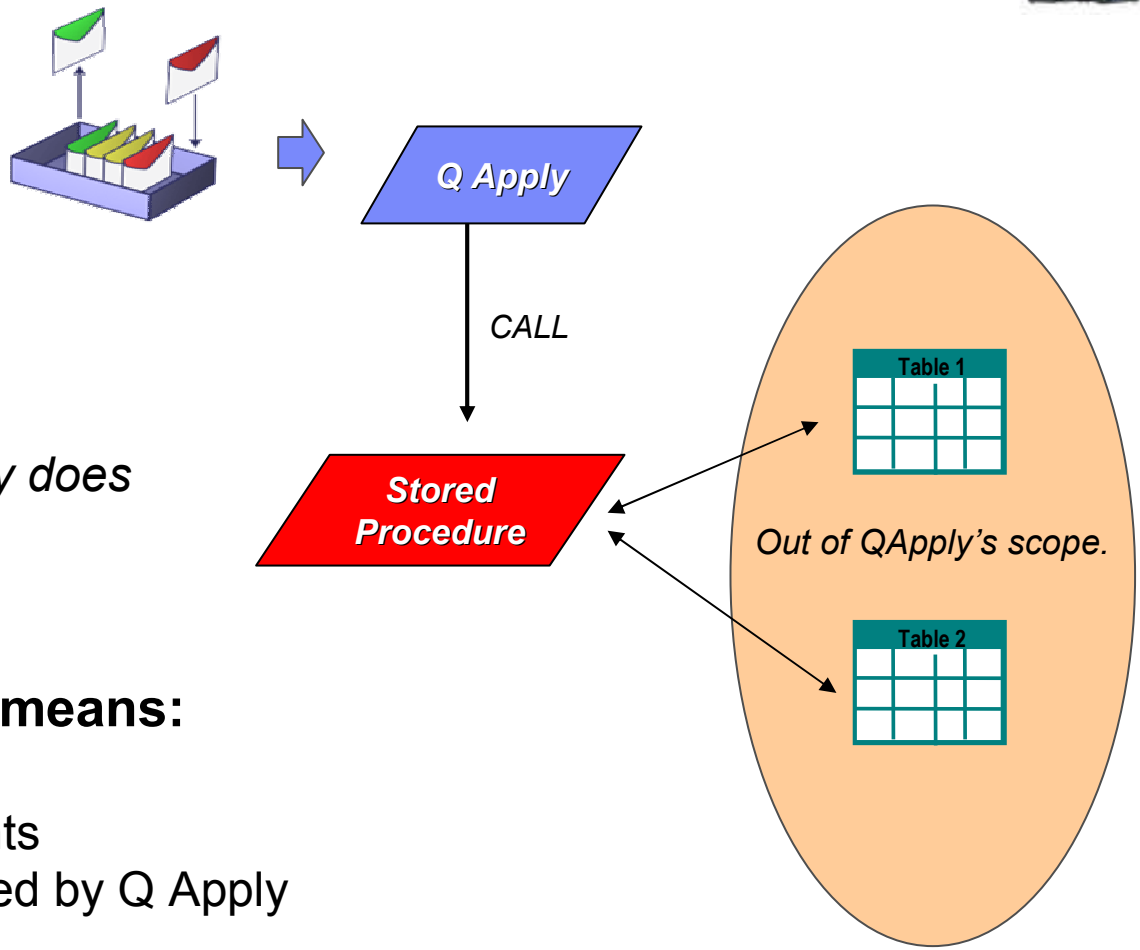
Q Replication - Transformations

- **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

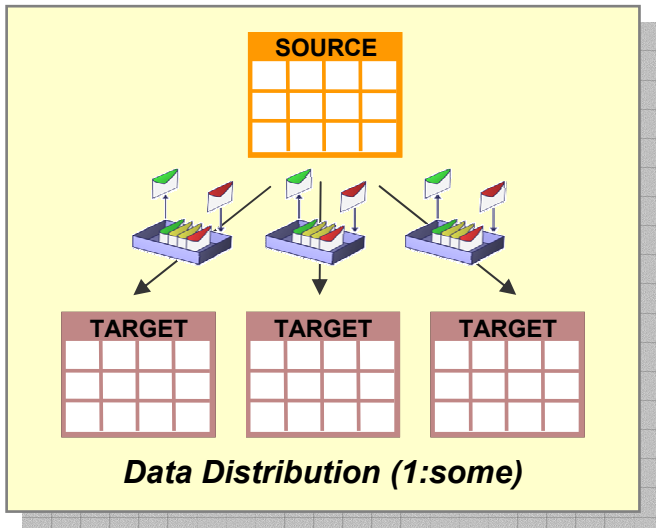


Apply via Stored Proc, why?

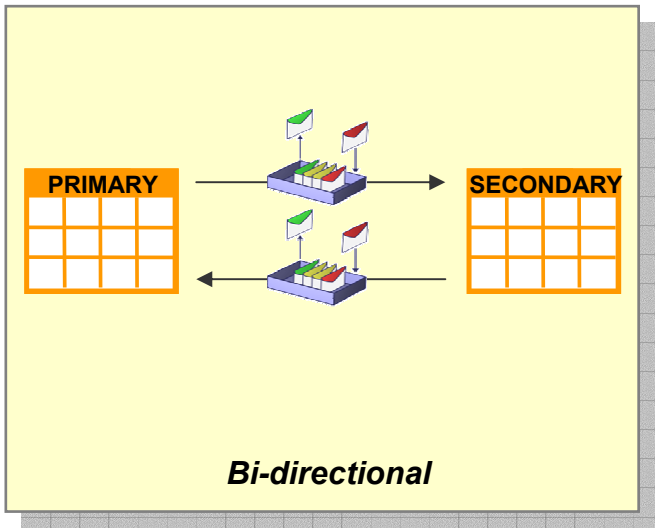
- *Data Transformation*
- *Data Aggregation*
- *Anything different to what Q Apply does*

- **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

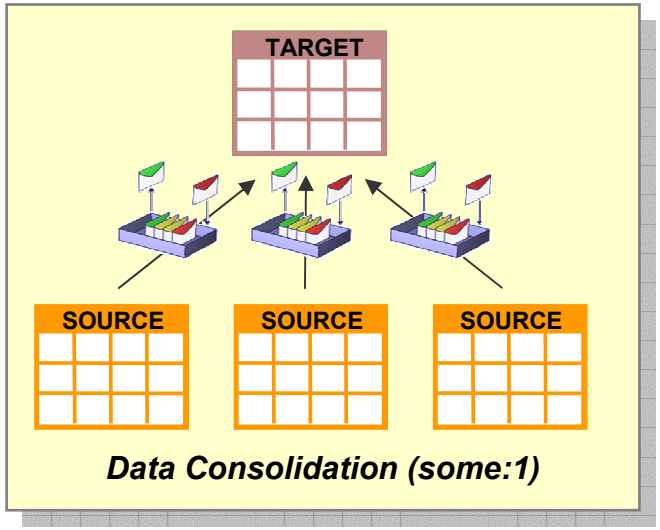
Sample Q-Replication Scenarios



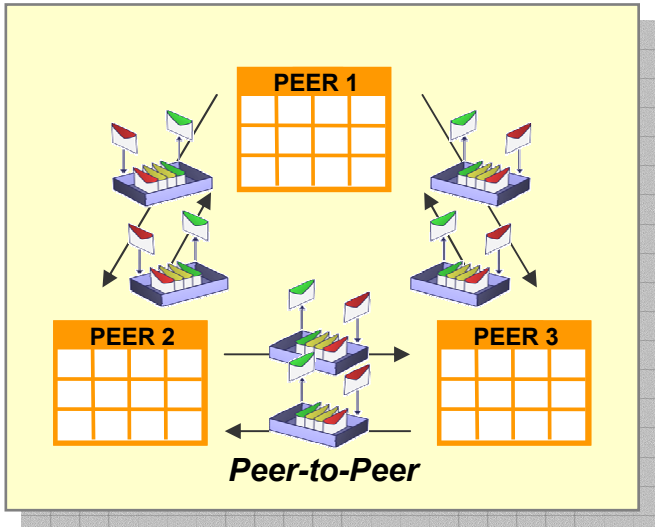
Data Distribution (1:some)



Bi-directional



Data Consolidation (some:1)

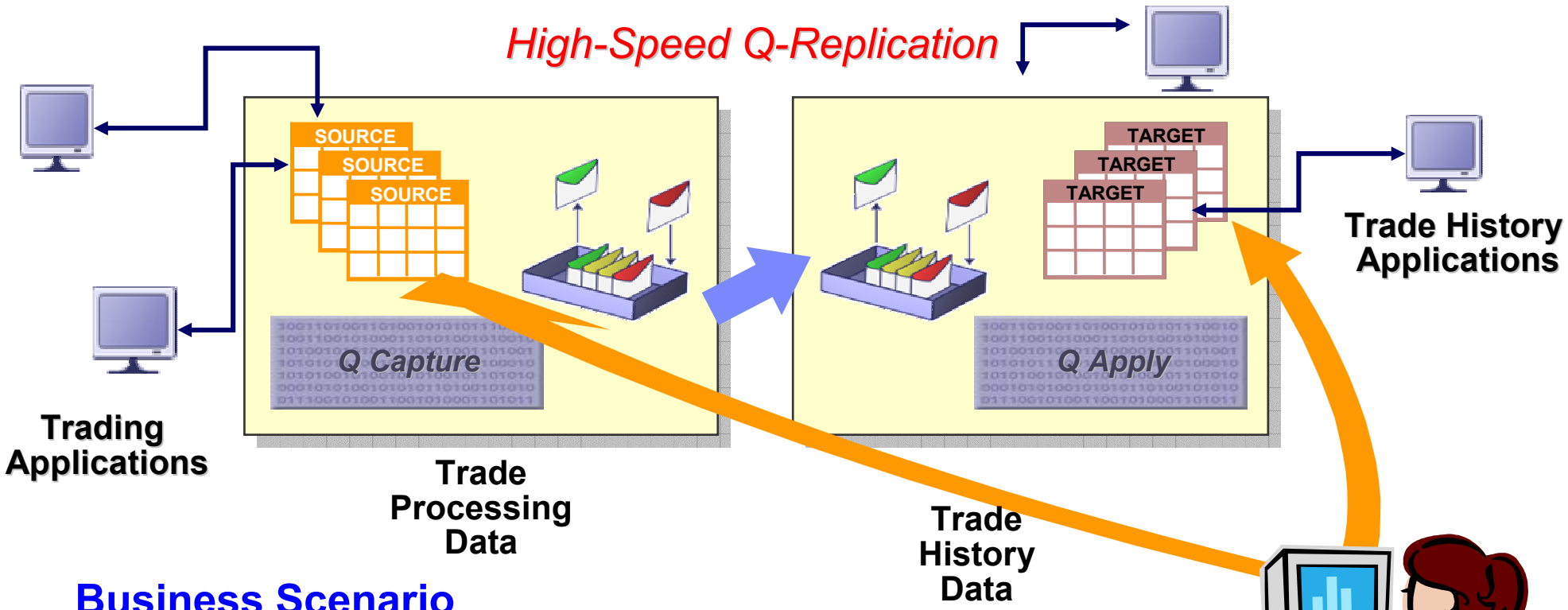


Peer-to-Peer

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

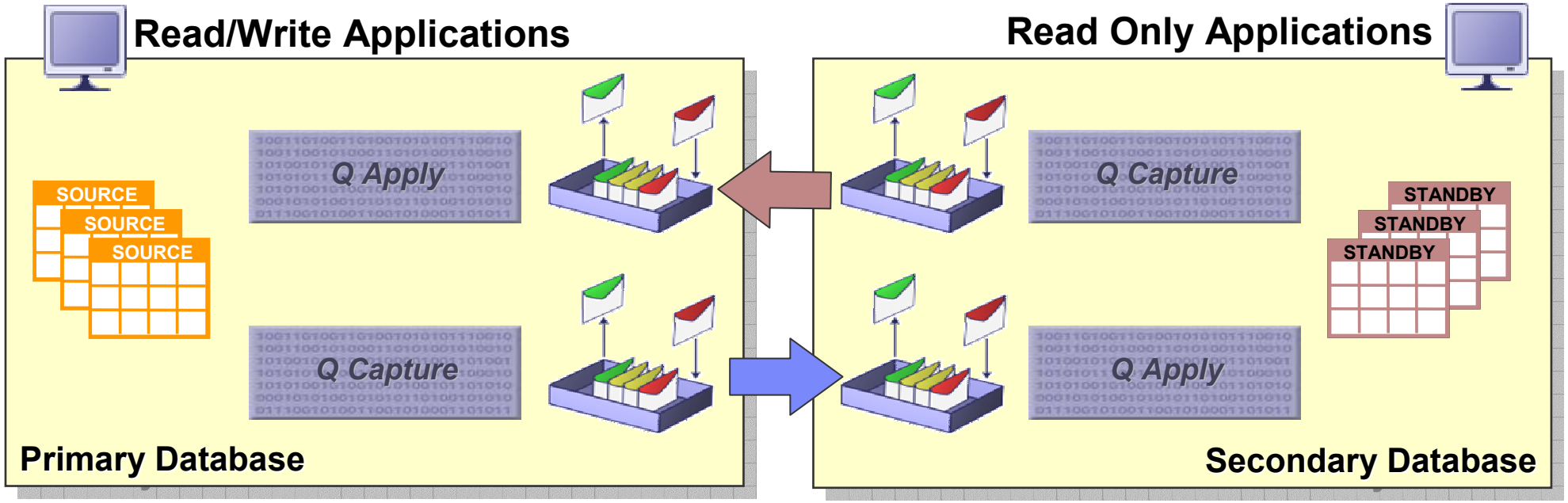
Feeding Trade-History Database with Q-Replication



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 transitionally 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

Q Replication has higher throughput and lower latency

- **Preliminary Lab measurements have show very high throughput**

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

Université du Mainframe 2005

1. Serveur de Données Fédérées

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

2. Données Répliquées

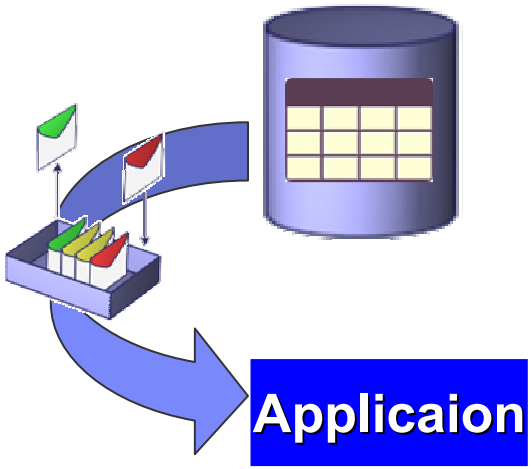
- *SQL-Replication*
- *Q-Replication*

3. Publication d' Événement (« Event Publishig »)

IBM Software et IBM System & Technology Groups

Why Publish Data?

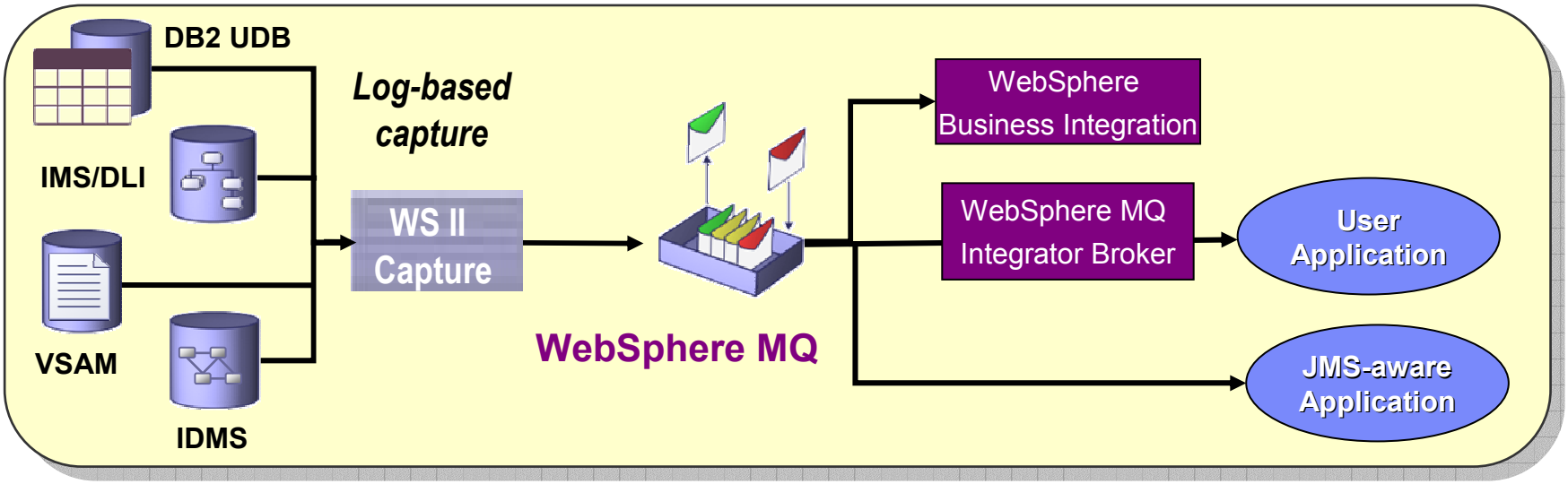
- **Database to Application Messaging**
  - Drive downstream Applications or APIs based on the Transactional Data of the changed Database Events
- **Meet Auditing Requirements**
  - Capture and store information regarding what changes were made to critical business data and by whom
- **Event Notification**
  - Stream changed Data Information to Web Interfaces
  - Stream only particular Events of Interest (filter Data)
- **MQ provides guaranteed delivery**
  - Avoids the need for 2-Phase-Commit (2PC)
  - Works even when the Target is not available
- **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

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

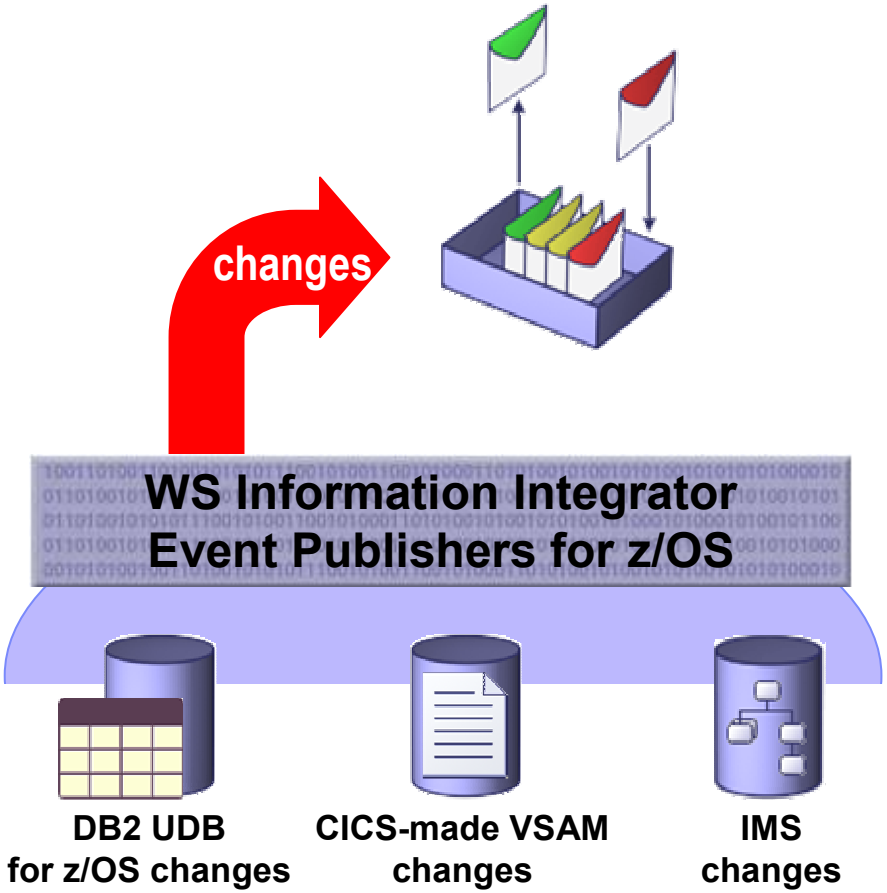
- **Function**
  - Publish events to a message queue
  - XML self-describing format
  - Wizard-driven configuration
- **Usage**
  - Application to application messaging
  - Event streaming
  - Source for ETL tool



Classic Data is REACTIVATED !

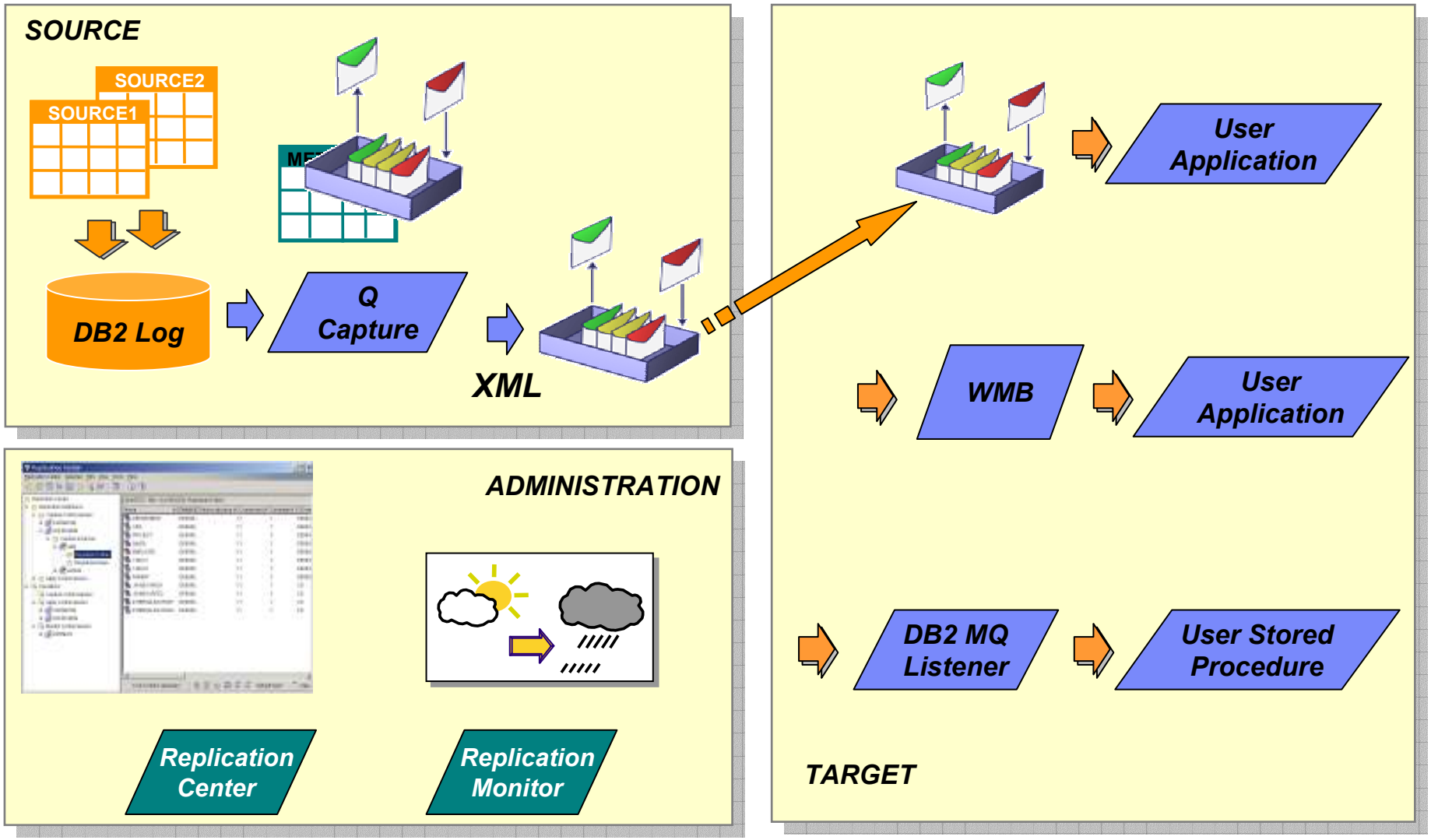
WebSphere Information Integrator Event Publishers for z/OS

- **Real time DB2, IMS, VSAM and CA-IDMS changed-data capture and publishing**
- **Publish to WebSphere MQ**
- **Relational XML format**
  - Each message represents a DB2 transaction
- **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
  - XML format is consistent with other event publishers (IMS, VSAM, IDMS, Distributed)



Capture, Externalize (XML) and Deliver to MQ

DB2 Implementation: Process Flow

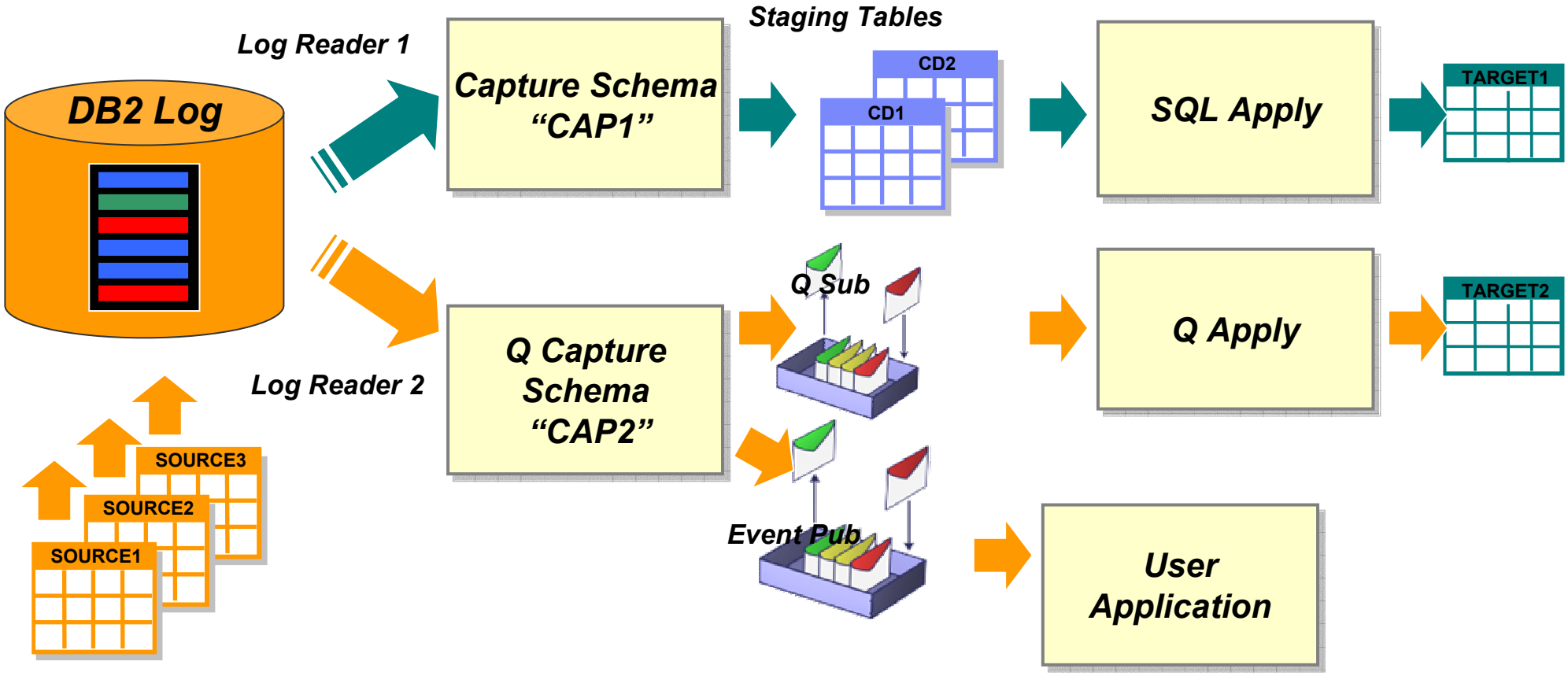


Event Publishing - Publication Options

- **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

Combining SQL and Q Replication with Event Publishing

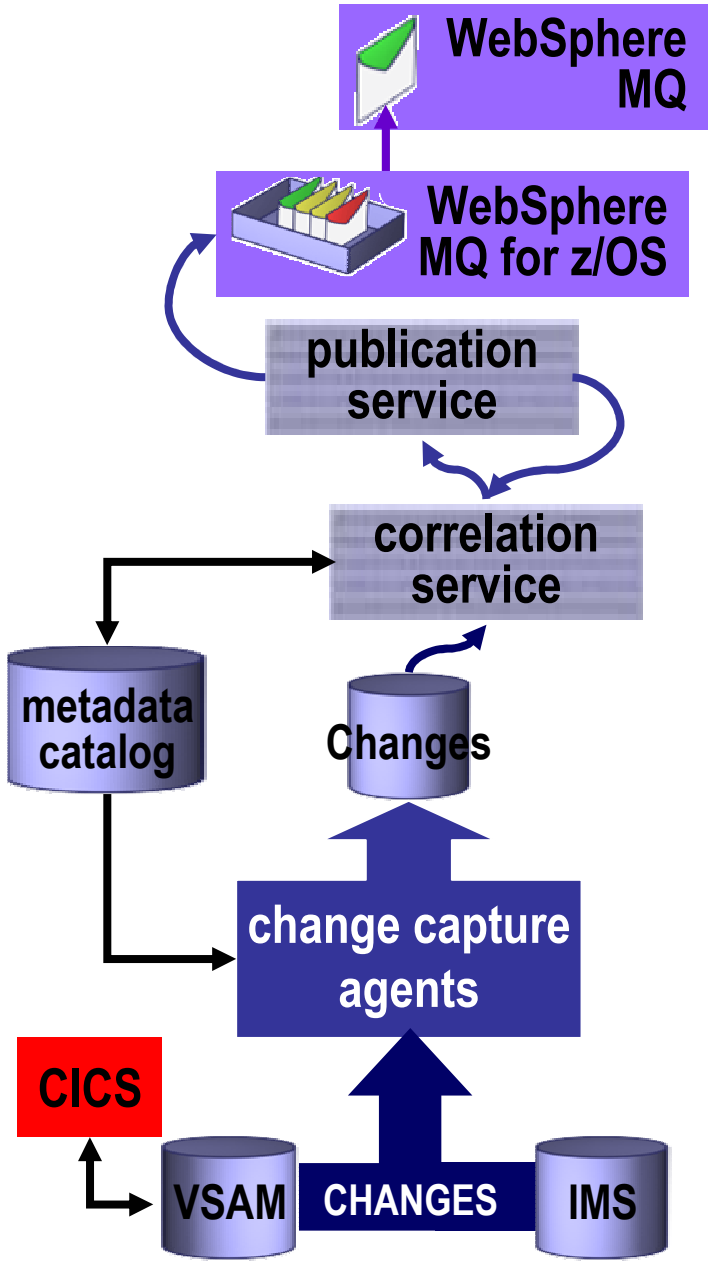


SQL Replication and Q Replication can co-exist

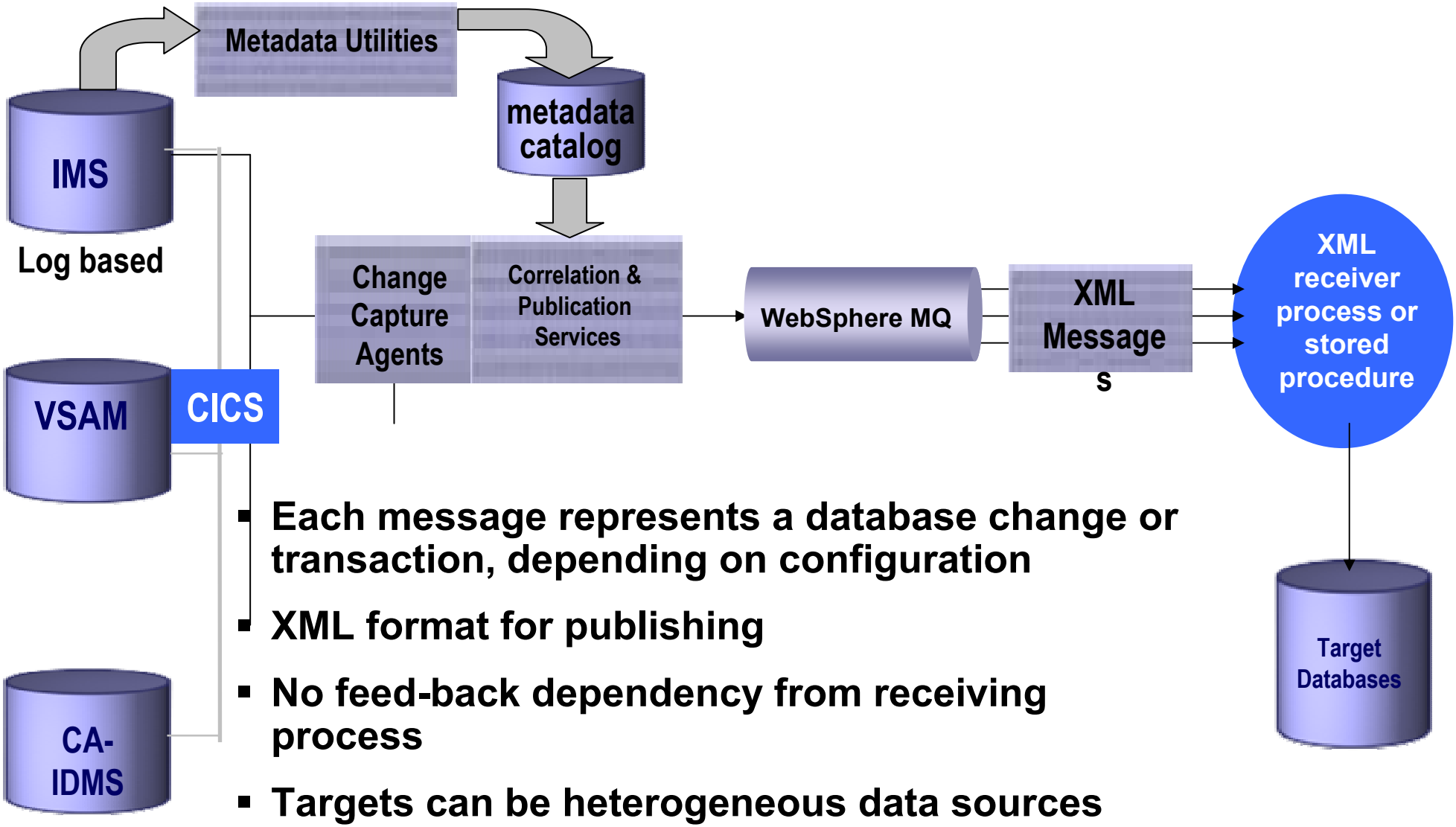
- Managed at source by using multiple capture schemas
- One Q Capture can handle both Publications and Subscriptions

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



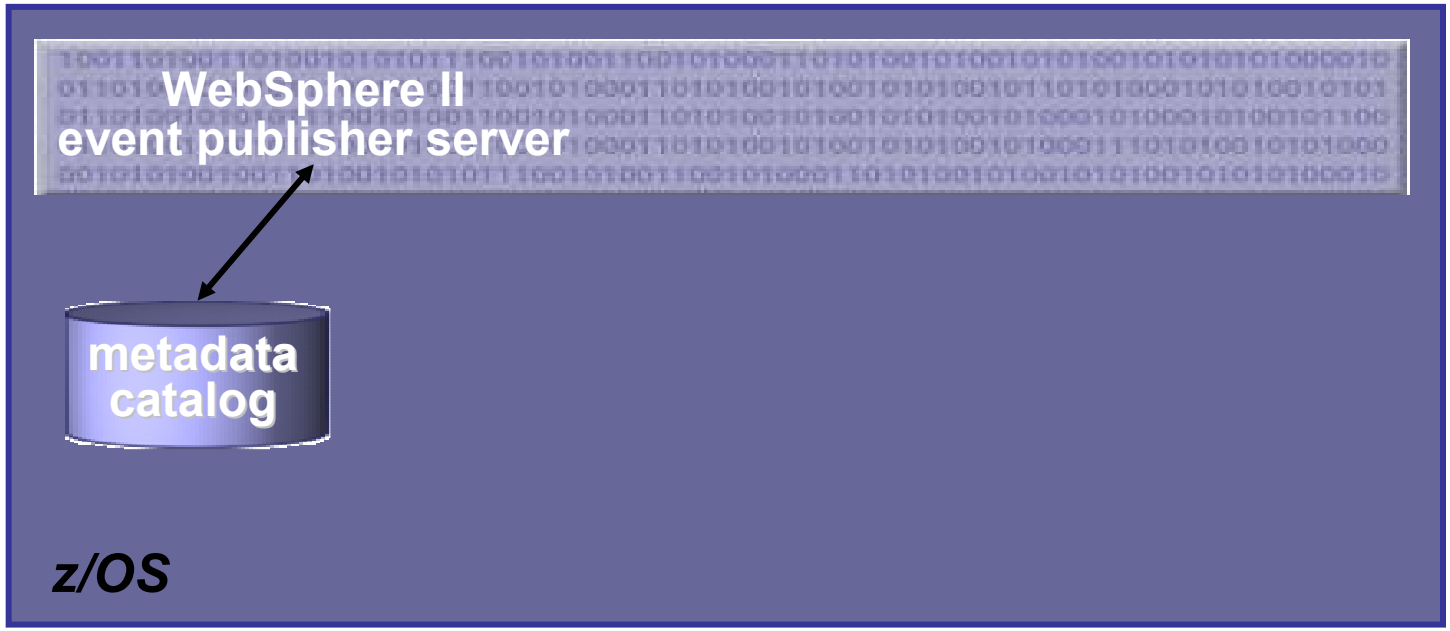
Classic Event Publishers - Architecture



- Each message represents a database change or transaction, depending on configuration
- XML format for publishing
- No feed-back dependency from receiving process
- Targets can be heterogeneous data sources

WebSphere II Classic Event Publishers

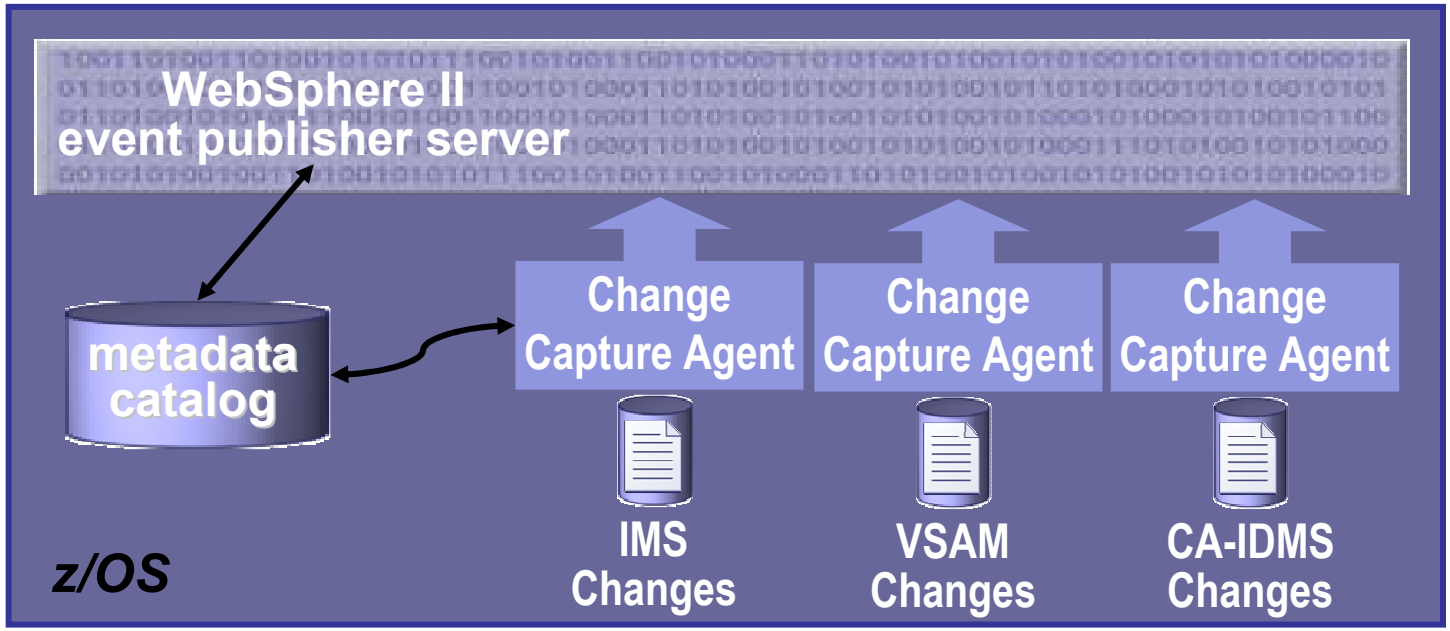
Operational Elements – server and catalog



Meta Data Catalog

- Holds mapping of non-relational data to logical tables
- Designates data to be captured and published
- z/OS-based data server
- Runs in its own address space
- Manages mainframe resources

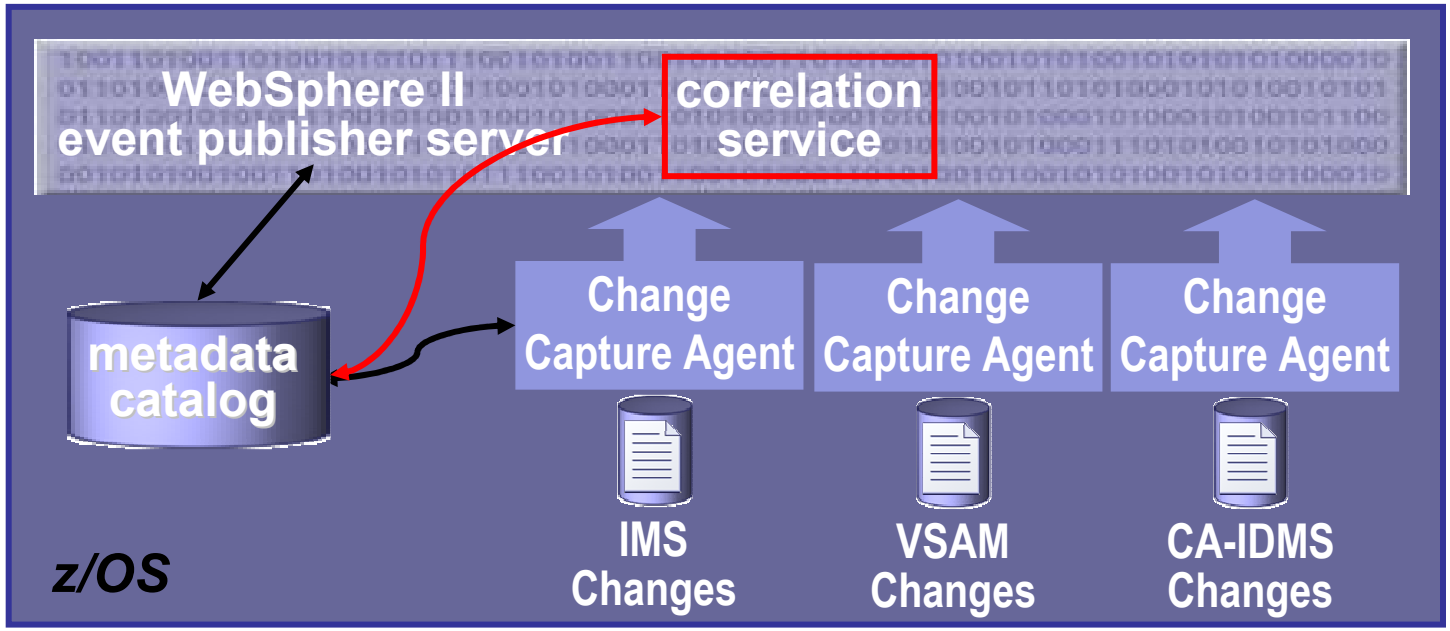
WebSphere II Classic Event Publishers Operational Elements – the CCA's



Change Capture Agents

- Database/file specific changed-data capture agents
- Filter changes based on metadata flag*
- Forward changes to correlation service
- Active (log exit) and Recovery (log file)

WebSphere II Classic Event Publishers
Operational Elements – Correlation Service

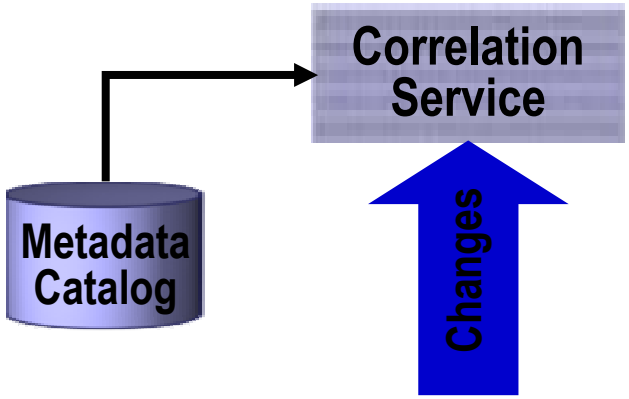


Correlation Service

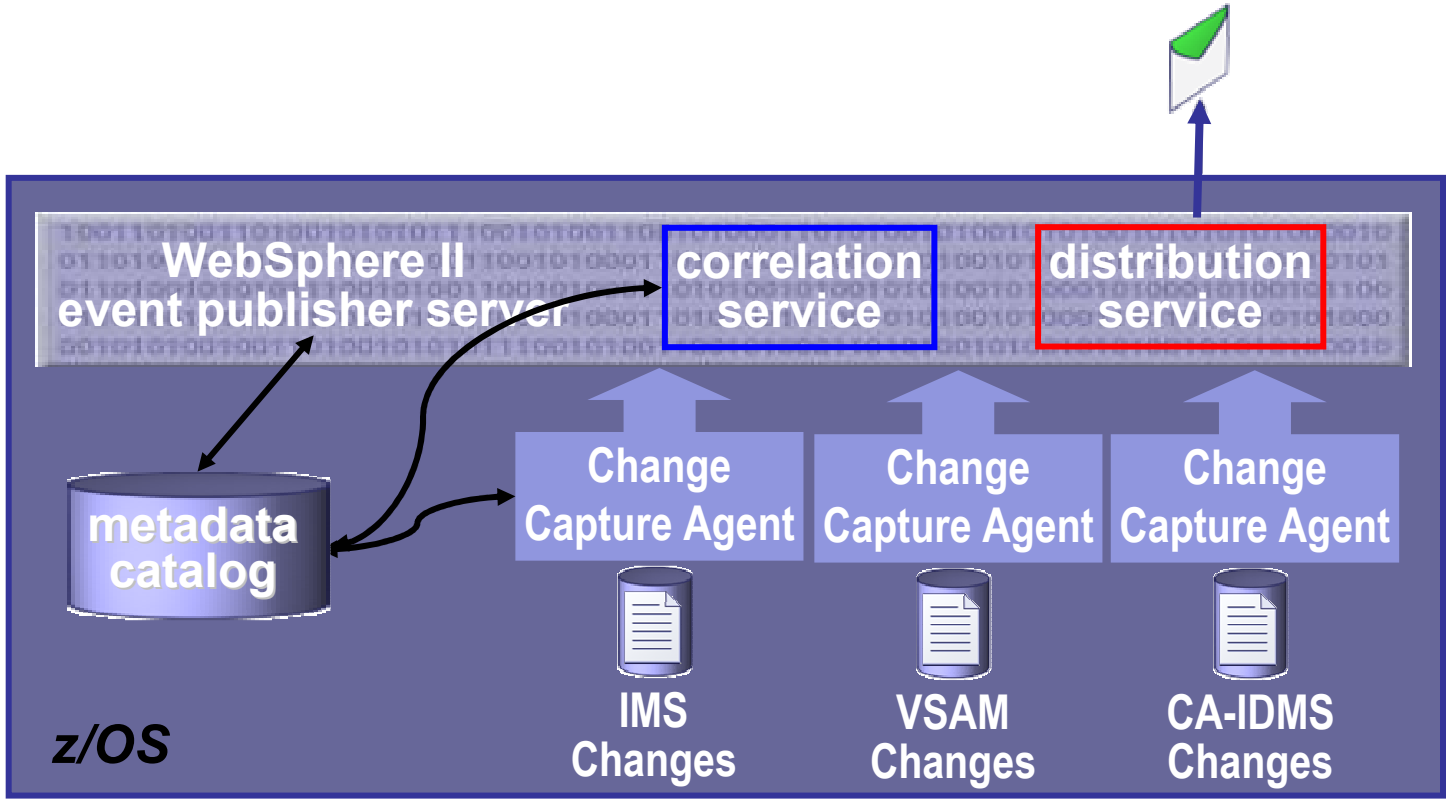
- Filter changes when necessary
- Format data items into relational data types
- Delivers committed changes to the publication service
- Handle recovery as needed

In-depth look - Correlation Service

- **Manage data publishing by unit-of-work**
  - Sort 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 internal SQLDA messages
      - send to Distribution Services
- **Metadata Catalog holds mapping for source DBMS tables to relational table/column projections that will be captured**



WebSphere II Classic Event Publishers Operational Elements – Distribution Service

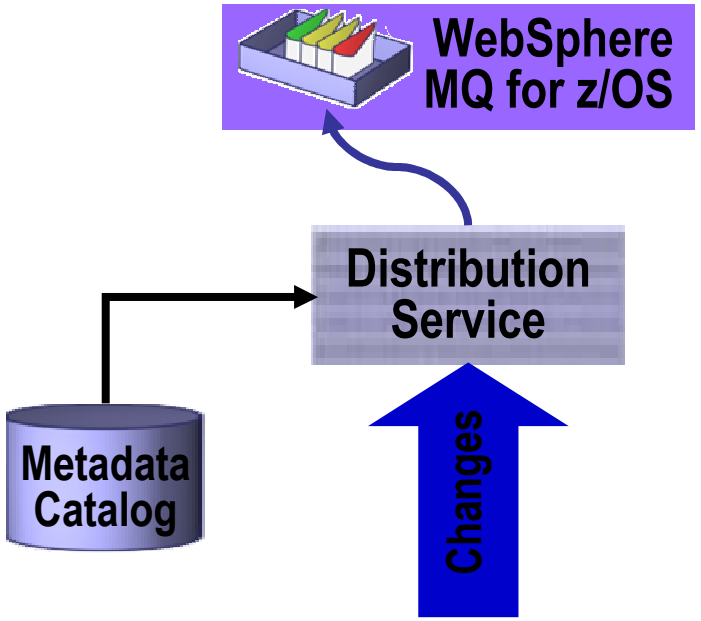


Distribution Service

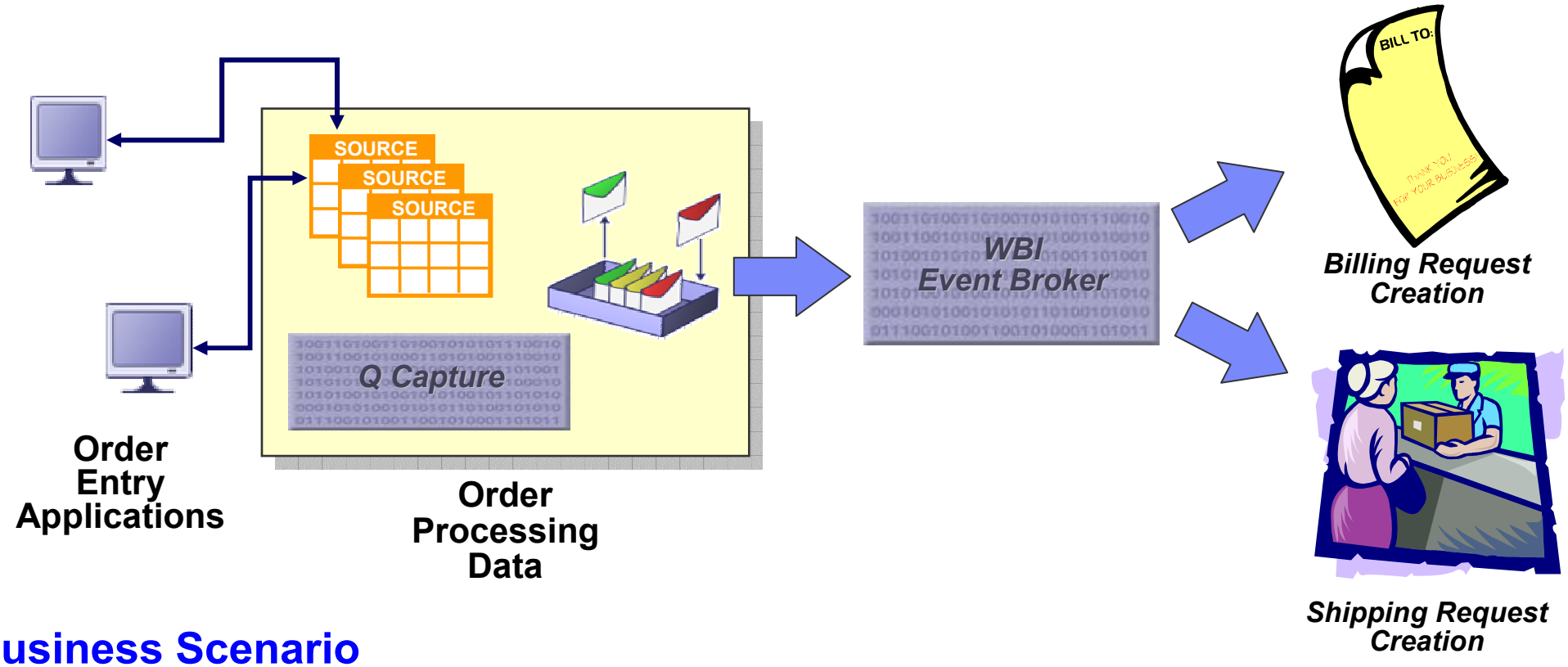
- Creates XML message
- Publishes changes to WebSphere MQ for z/OS queue
- Confirms delivery for the correlation service

In-depth look - Distribution Service

- **Manage distribution of data changes by unit-of-work**
  - WebSphere MQ only in ver 8.2
- **Handles SQLDA transformations to final presentation format(s)**
  - XML only in ver 8.2
- **Publishing metadata is defined in configuration files**
- **Metadata catalog used for validation**



Order Processing – Exploiting II Event Publishing

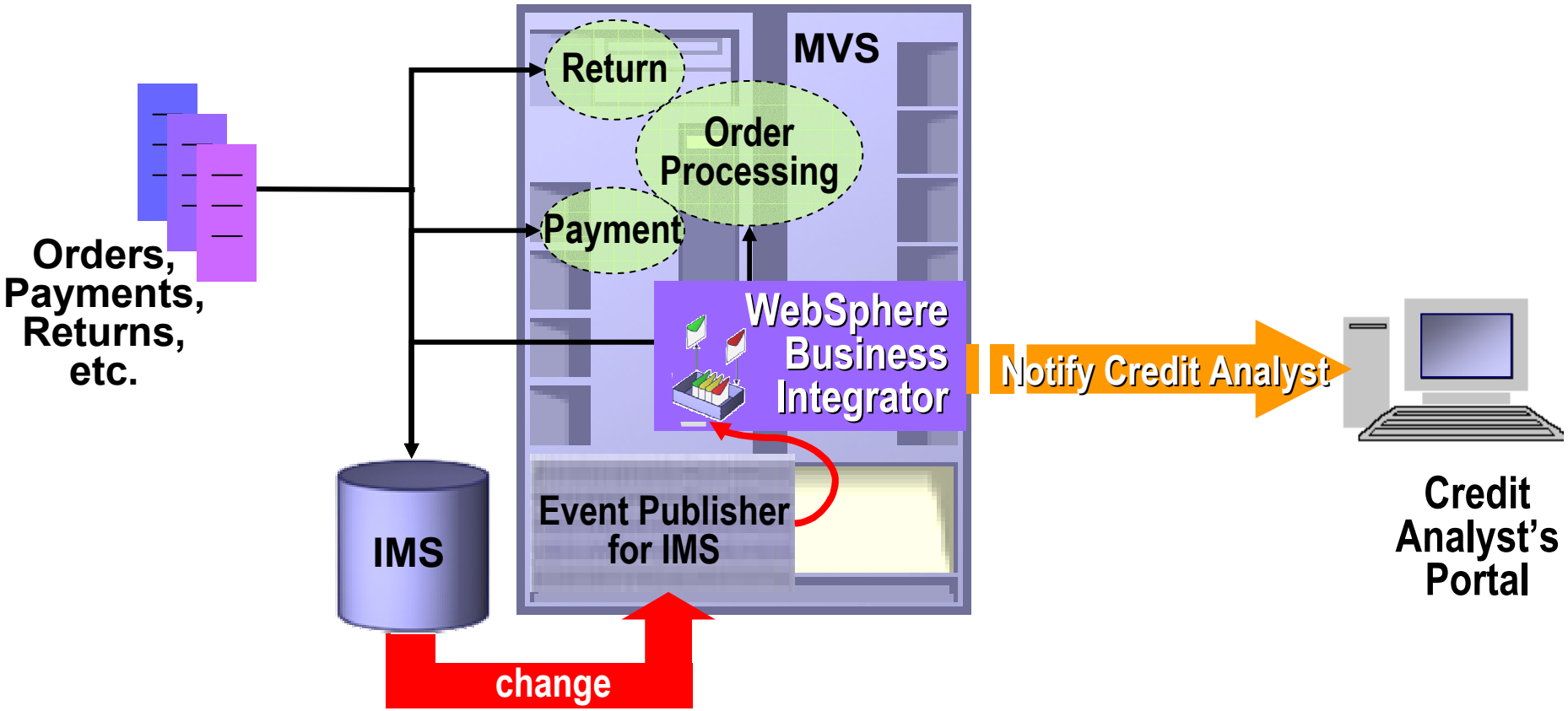


Business Scenario

- 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*



IBM Software et IBM System & Technology Groups

Why data events versus application events?

- Integration is independent of the source applications

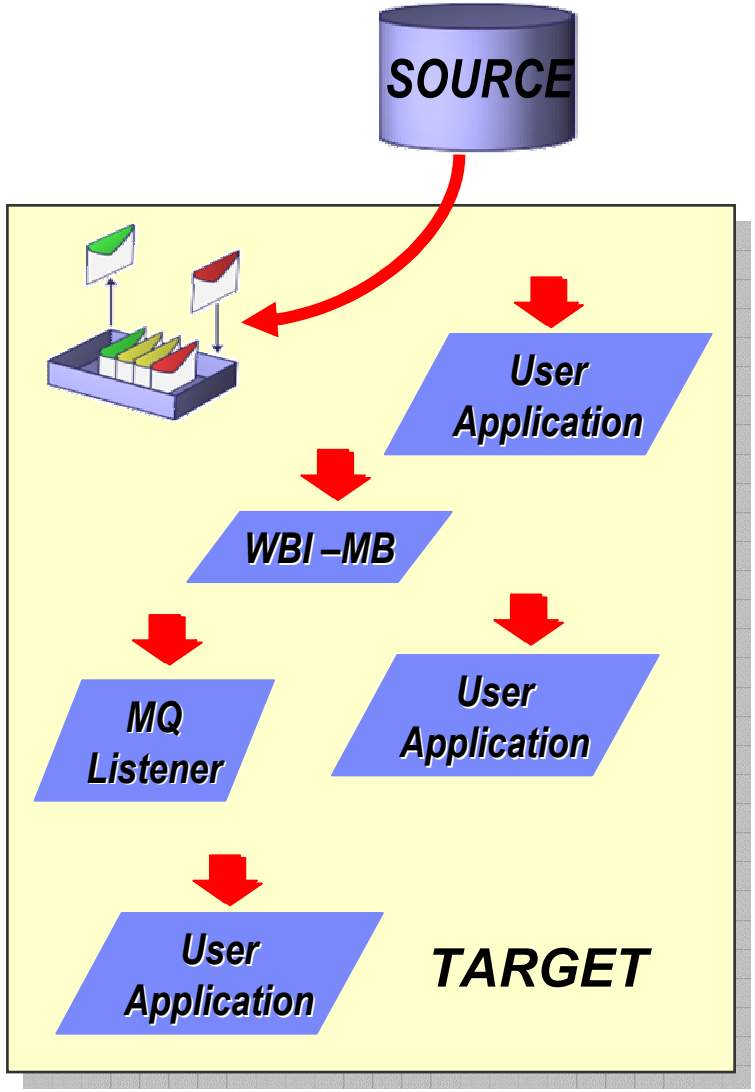
  –Applications grow and evolve with
  " minimal impact on the integration "

  –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



Université du Mainframe 2005

1. Serveur de Données Fédérées

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

2. Données Répliquées

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

3. Publication d' Événement (« Event Publishig »)

IBM Software et IBM System & Technology Groups

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)

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

IBM Software et IBM System & Technology Groups

Université du Mainframe 2005

IBM WebSphere Data Integration Solution

L'offre IBM « Ascential » sur z/OS

Steven Haddad
Ascential, an IBM company
steven.haddad@fr.ibm.com

IBM Software et IBM System & Technology Groups

Université du Mainframe 2005

WS Information Integrator Classic

1. *Usage Scenarios*
2. *Integration scenarios with DataStage*

Bob DeCori
WS II Solutions & Deployment
rdecori@us.ibm.com

IBM Software et IBM System & Technology Groups

Agenda

- **WS II Classic Federation Usage Case Studies**
  - Broad categories
  - In-depth discussion

[redacted]
WebSphere Information Integrator

- WebSphere II Classic integration scenarios with DataStage

Two broad categories of usage

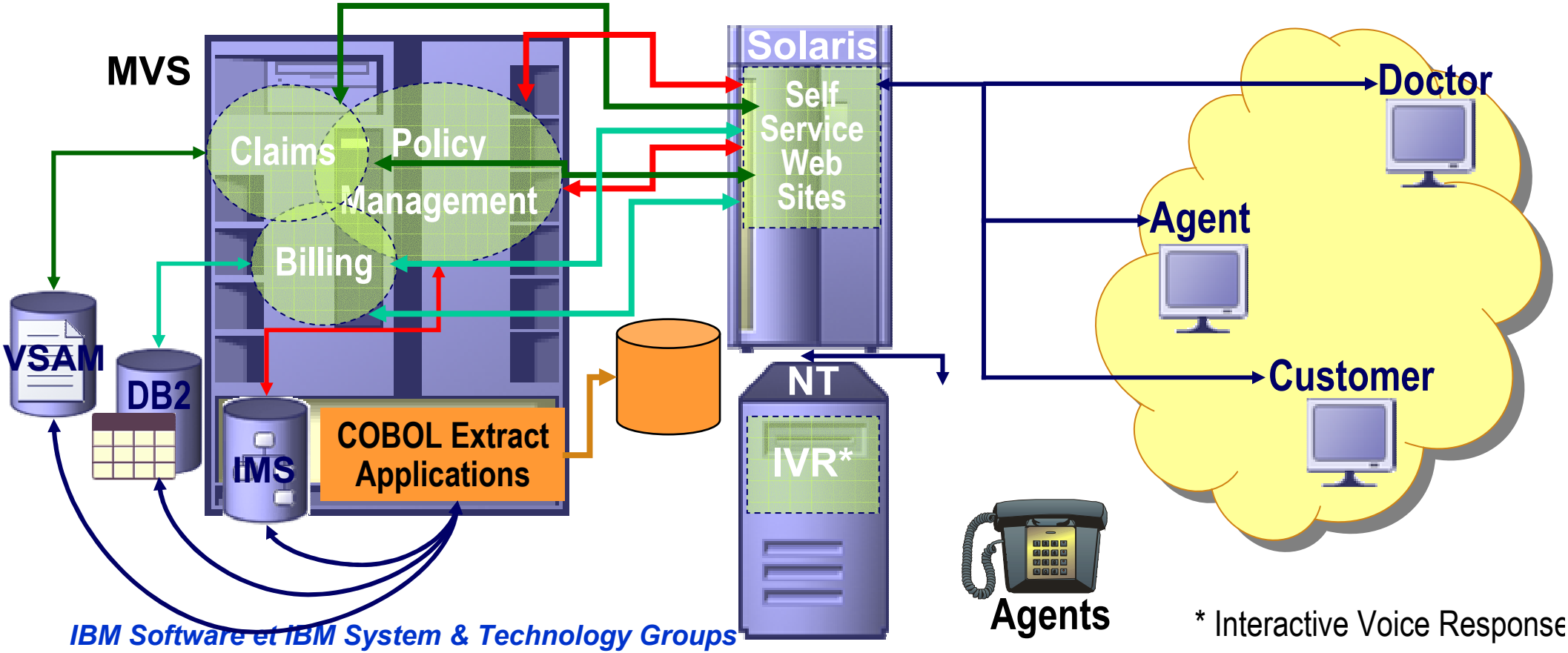
- **e-Business**
  - Deliver mainframe data to
    - Self-service portals (real-time account details)
    - e-commerce solutions (real-time inventory)
    - Employee portals (real-time claims detail)
  - Web developers become productive with no mainframe skills
  - Eliminates data latency business issues caused by copied data

- **Business intelligence**
  - Integrates seamlessly with
    - Reporting and analytical tools, e.g. Business Objects
    - Portals, e.g. WebSphere Portal
    - ETL, e.g. Ascential DataStage

Self-service Application for Insurance Carrier

The Pain associated with "traditional" implementations

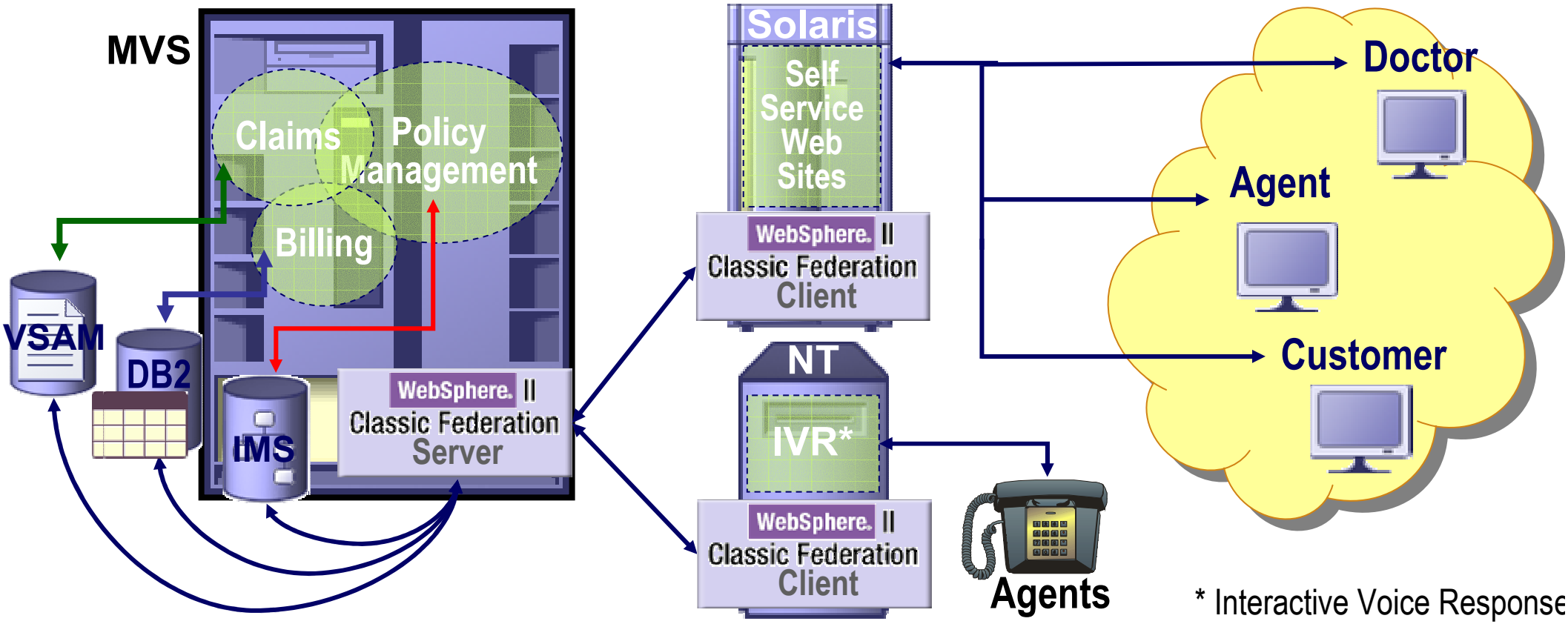
- **Option a : copy data to non-mainframe environments**
  - Estimated cost $2M
  - Data refreshed every 30 hours or so
- **Option b : integrate the IMS transactions**
  - Estimated cost 10,000 man-hours per application



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



IBM Software et IBM System & Technology Groups

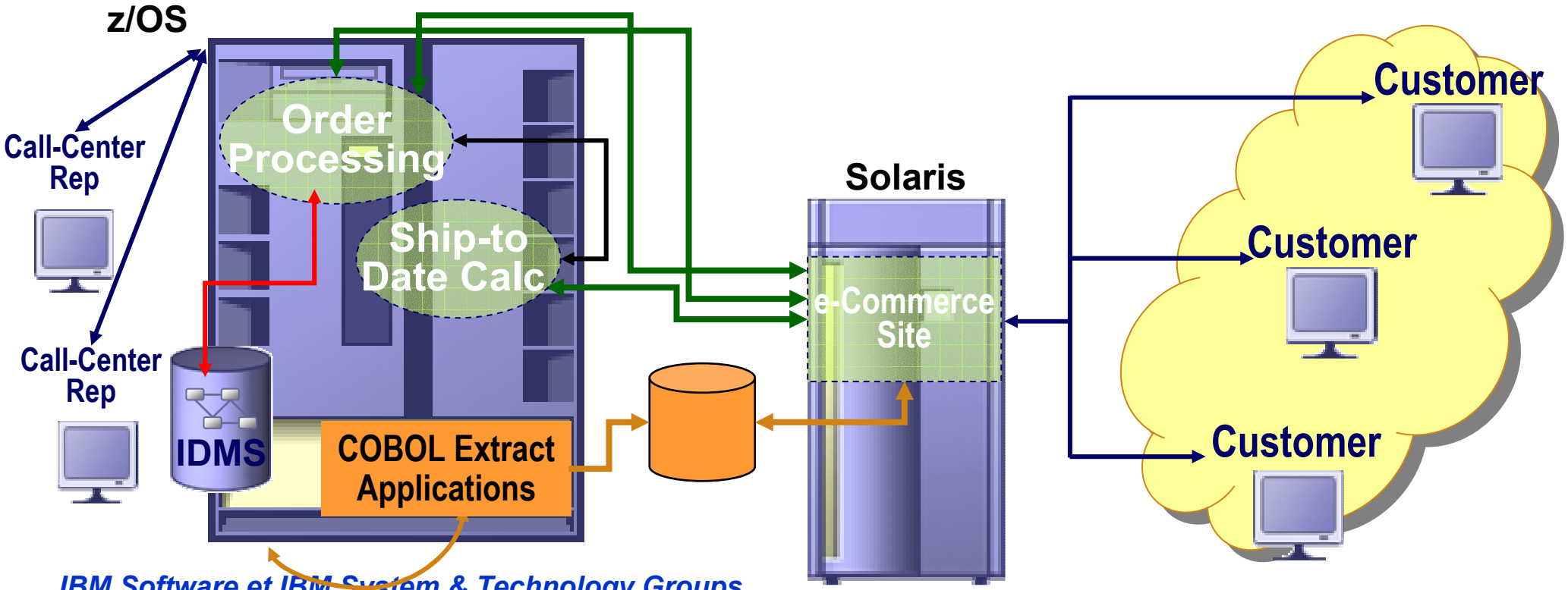
Self-service Application at large insurance carrier

Description	**A large US insurance carrier is using WSIICF to provide up-to-the-minute policy, claims and account information to their customers and agents through its web site as well as IVR (automated phone response system).**
Architecture	**The insurance carrier's operational data resides in IMS running on z/OS. Nortel's IVR system runs on Windows server, using IICF ODBC driver to provide access to account and claim information to users of the automated phone system. WebSphere Application Server runs on Solaris using IICF JDBC driver providing the same access to the users of the website.**
Benefits	**Eliminates the significant expense associated with ETL or invoking CICS programs (IICF came in less than 25% of the cost of either process). Current architecture provides availability of current data to the customer service systems, thus providing significant customer satisfaction benefits in addition to costs savings mentioned previously.**

E-commerce website for European-based catalog retailer

The pain associated with "traditional" implementations

- **Option a: copy data to non-mainframe environments**
  - Out-of-date inventory = potential "out-of-stock" sales
  - Dissatisfied buyers stop shopping here
- **Option b: integrate the CICS transactions**
  - Java Web developers have no mainframe skills
  - e-Commerce is "shopping" NOT "order processing"

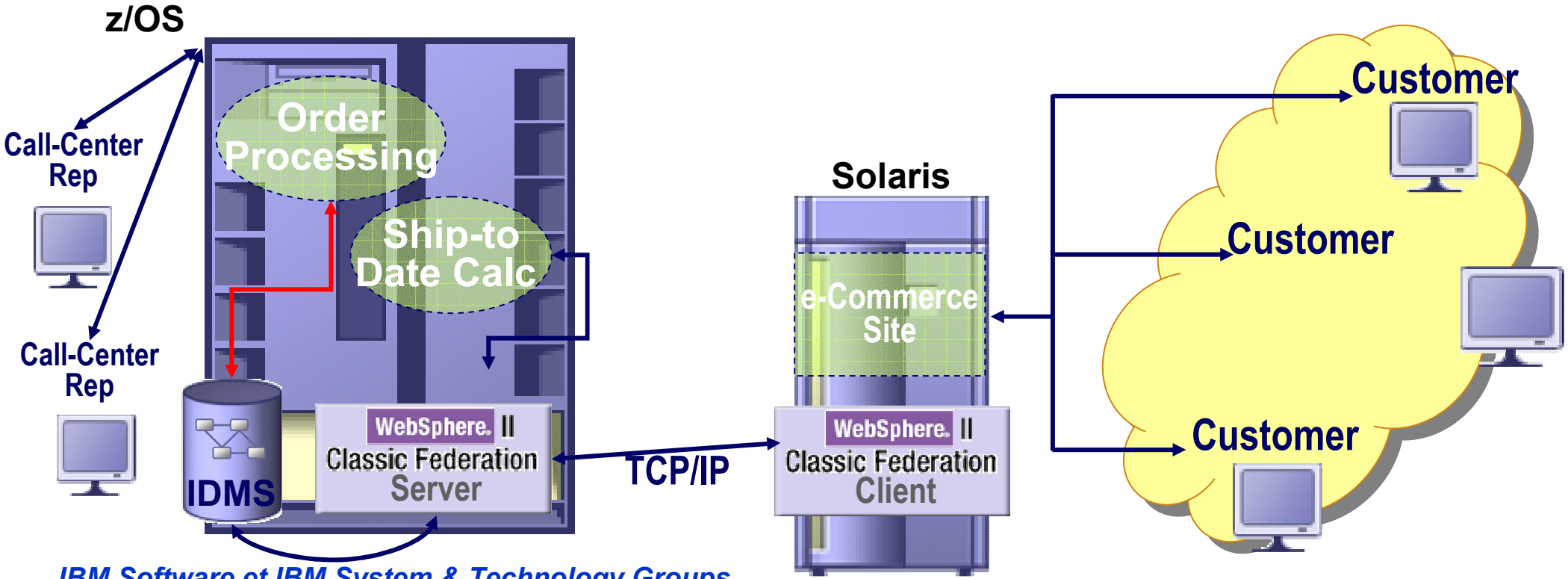


IBM Software et IBM System & Technology Groups

E-commerce website for European-based catalog retailer

IBM Solution - *Single-source mission critical data*

- **Seamlessly share order processing data and business logic**
  - No impact on call-centers
  - New WebSphere e-commerce applications share critical data
  - Leverage “common” procedures such as ship-to-date calc or pricing
  - WebSphere Studio development independent of mainframe skills



IBM Software et IBM System & Technology Groups

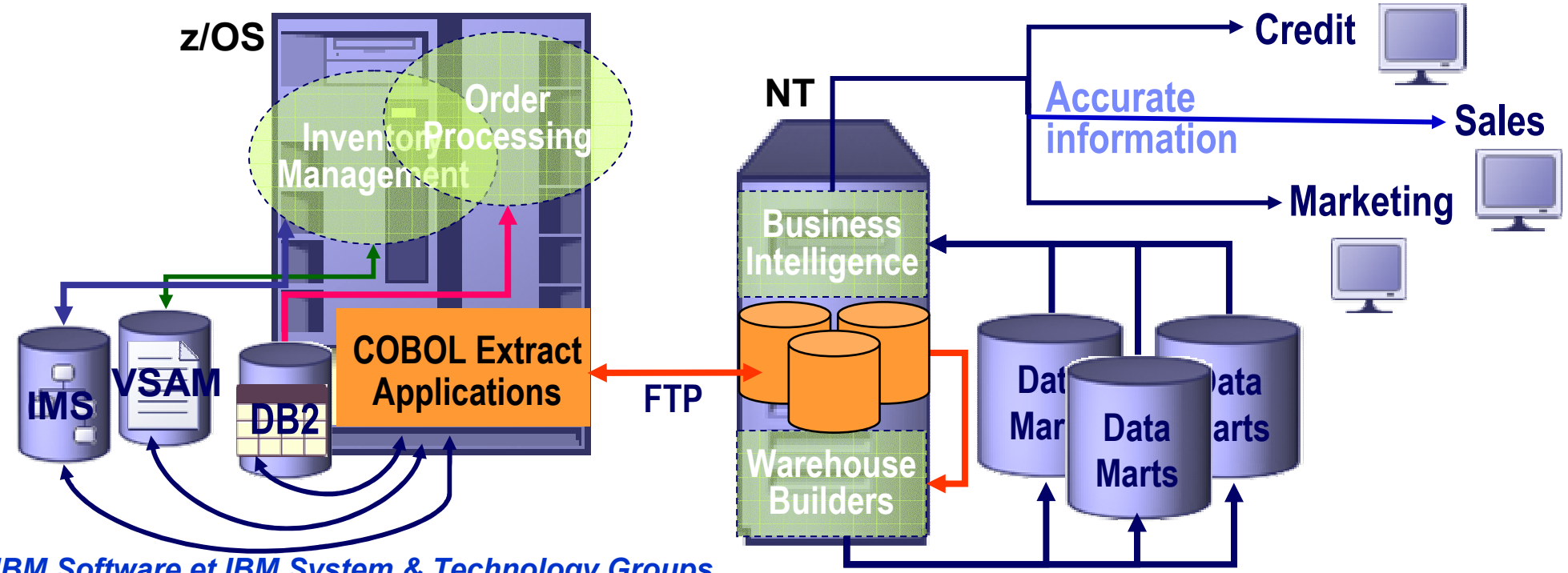
E-commerce website for European-based catalog retailer

Description	**One of the largest European-based catalog retailers is using WSIICF to provide a superior shopping experience for their web customers, without impacting the traditional sales channels that use CICS transactions on 3270 terminals in its call centers.**
Architecture	**The retailer's operational data resides in IDMS running on z/OS. The e-commerce site is running on Websphere App Server on Solaris, utilizing WSIICF JDBC driver to provide transactional SQL access for the J2EE applications to the IDMS operational data.**
Benefits	**Using WSIICF enabled this catalog retailer to make the shift from a order-entry to a shopping cart paradigm. While growing their customer base by launching a slick e-commerce site, they were able to keep their existing sales channels that depended on the CICS 3270 transaction based call centers. This was possible as the impact on the existing CICS system was minimal.**

BI solution for international manufacturer of recreational vehicles

The pain associated with "traditional" implementations

- **Disjoint Process**
  - Build and maintain mainframe "extract" process
  - Build and maintain distributed data transform & load
- **Management challenges lead to increasing costs**
  - Multiple skill sets required:
    Mainframe programming & data warehouse design/build
  - Coordinating multiple components and development teams

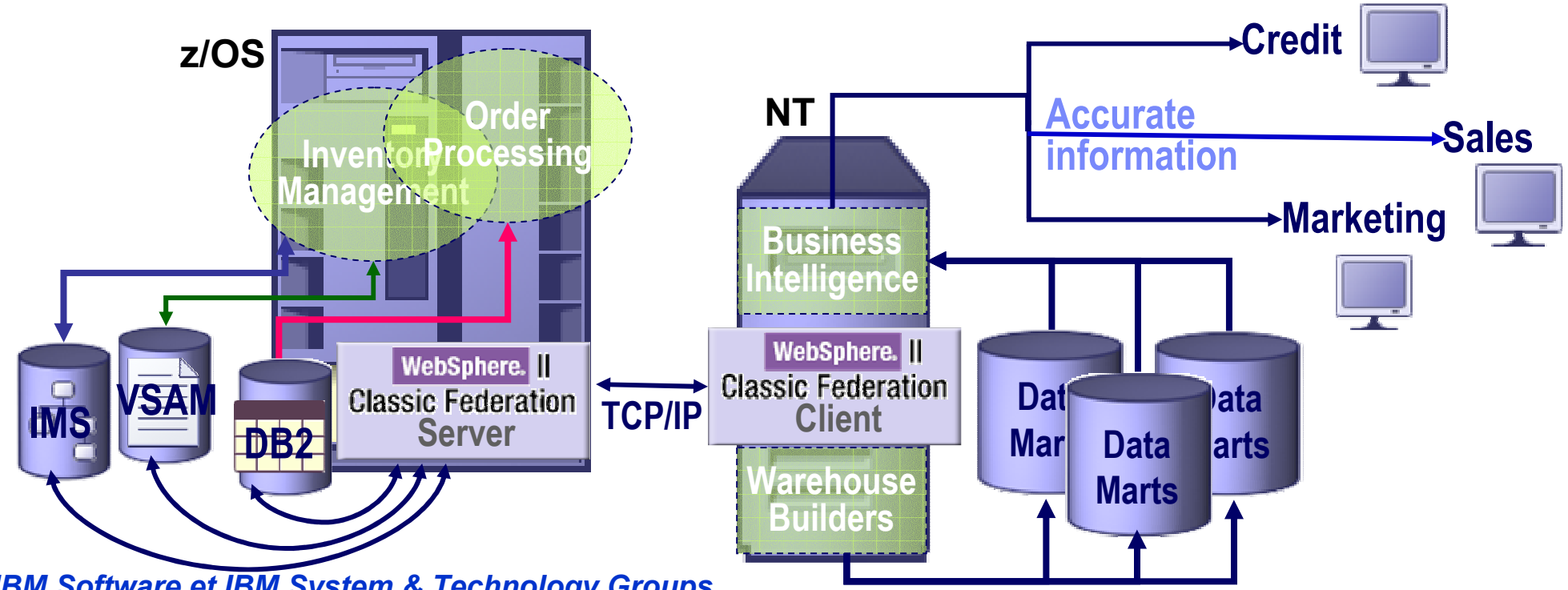


IBM Software et IBM System & Technology Groups

BI solution for international manufacturer of recreational vehicles

IBM solution -- *feed operational data to ETL via SQL*

- **Dynamically connect data warehouse tool with mainframe data**
  - No dependence on mainframe development
    - Dramatically simplified management: One team owns it all
  - One consistent process leverages "power" of ETL tooling
    - Development time "cut in half"
  - Empowers additional uses
    - Dynamic query by business intelligence tools extends the warehouse

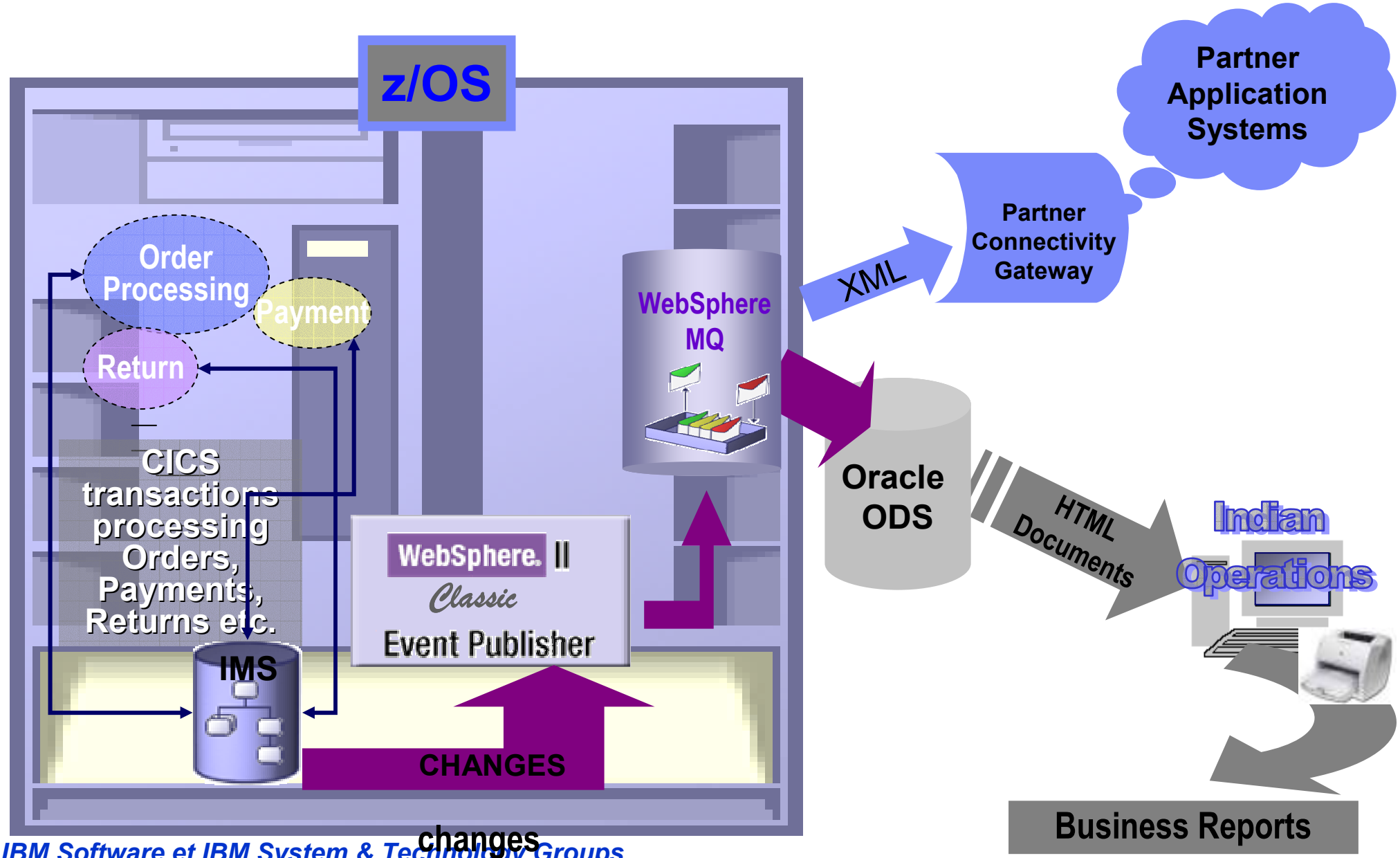


IBM Software et IBM System & Technology Groups

BI solution for international manufacturer of recreational vehicles

Description	**An international manufacturer and distributor of recreational vehicles is using WSIICF to populate its datawarehouse which is then used to drive their business intelligence platform, powering key inventory and sales decisions.**
Architecture	**The manufacturer's operational data resides in IMS, VSAM & DB2 running on z/OS. The ETL tool running on Windows uses WSIICF ODBC driver to pull relevant mainframe data as if it were all one relational database.**
Benefits	**By eliminating the dependency on batch COBOL programs (which typically need some lead time), the manufacturer is now able to react instantly to market conditions by using the always current data in the datawarehouse to drive business decisions. This has allowed them to fight back the competition and regain market share in the parts business. Conservative management reviews indicate deploying WSIICF has cut development times in half, with corresponding cost savings..**

Event Monitoring
using Classic Event Publisher at large technology reseller

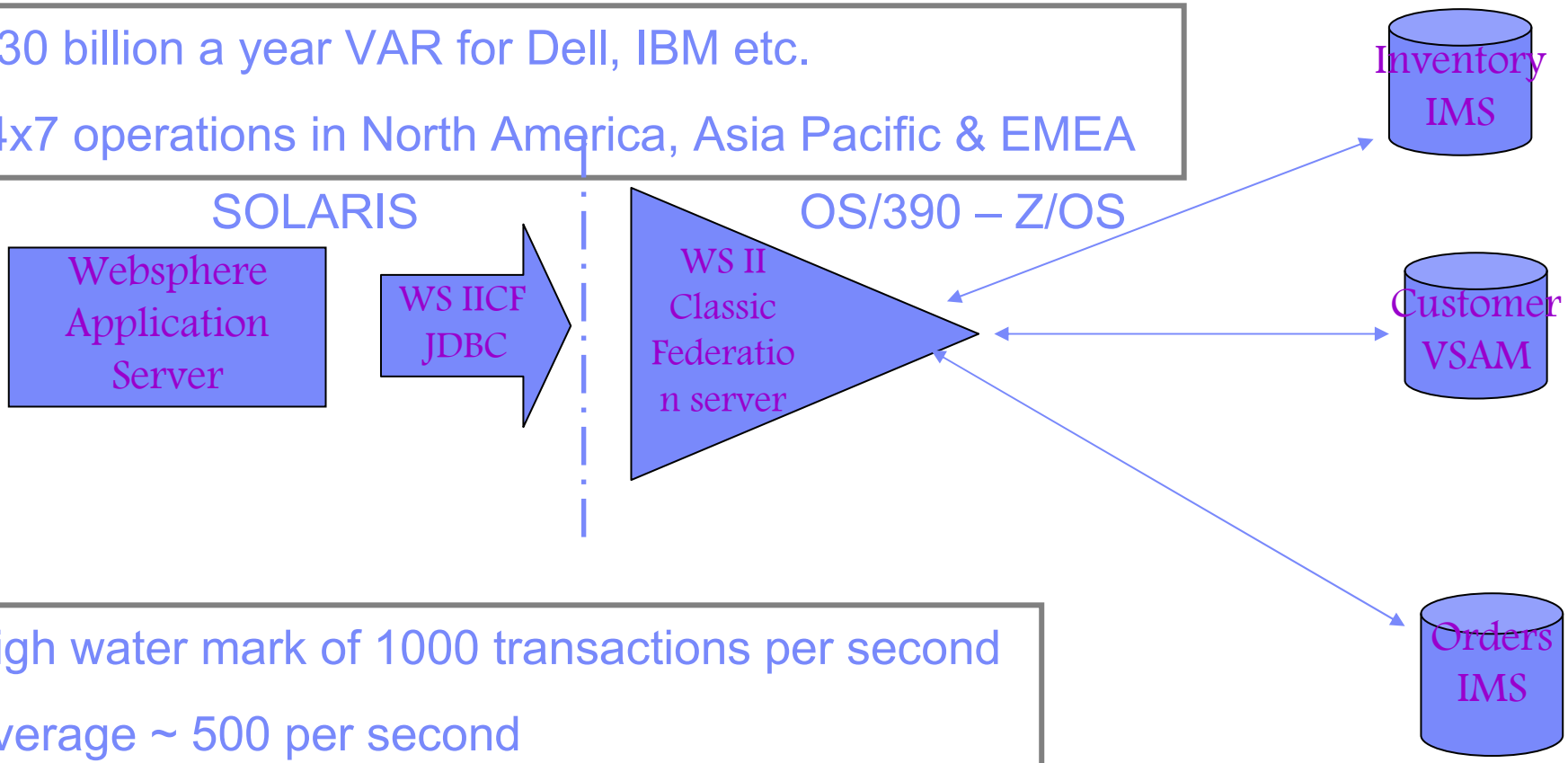


Event Monitoring using Classic Event Publisher at large technology reseller

- ✓ **Currently capturing changes from US, AP and CA**
- ✓ **Volume is about 1 million transactions per day**
- ✓ **Translates to about 3 million database changes per day.**
- ✓ **Will soon be adding EMEA and LA, which is estimated to increase the volumes to 4.5 million changes per day.**
- ✓ **Changes are populating an ODS on Oracle to provide data for operational reports.**
- ✓ **Enables the Partner Connectivity Gateway (PCG) so that partner companies can integrate their applications with Ingram's inventory system**

E-Commerce and infrastructure applications at international technology reseller

- $30 billion a year VAR for Dell, IBM etc.
- 24x7 operations in North America, Asia Pacific & EMEA



- High water mark of 1000 transactions per second
- Average ~ 500 per second
- Impact on existing online transactions acceptable

E-Commerce and infrastructure applications at international technology reseller

Description	**Large international technology reseller is using WSIICF to provide access to mainframe non-relational data for their e-commerce and infrastructure (shipping, reporting and licensing) applications. In production for over 2 years.**
Architecture	**The reseller's operational data resides in VSAM, IMS and DB2 running on z/OS. There are two production WSIICF servers – one for North America, and the other serving EMEA and Asia, each is serving 100 simultaneous clients at any given point of time, on average with a response time of 100-200 ms.**
Benefits	**The real time federation, along with IICF's robust and scalable behavior makes all these applications possible. Alternatives like transaction wrappers and ETL was not considered practical because of the expense & lack of scalability. Other key factors include insignificant impact on mainframe resources (monitored using Omegamon), and simplicity of "map once, deploy whenever" approach.**

Case management application at US State Govt.

Description	**The Department of Social Services at an US State Government has deployed a .Net based application to provide case workers out in the field the facility to open, review and update the cases they are working on from the convenience of a PC, while office based workers continue to use the existing Natural applications from their terminals ("green screen"). The backend database (Adabas 7.3) is not changed in any way.**
Architecture	**The case information is stored in Adabas files. The case & customer information is accessed by a .Net application via IBM WebSphere Information Integrator Classic Federation (WSIICF) ODBC driver. The application allows the user to add, modify and view the case data using various web pages and writes it back to Adabas as required. There is a new version of this app, based on WebSphere App Server 5.1 that is in test currently.**
Benefits	**With WebSphere II Classic Federation, this US State Govt. department was able to leverage existing assets in a new interface. Solution benefits include elimination of ETL resources, elimination of the need for a call center to service calls from field case workers, and reduction of case management errors by real time delivery relevant information.**

Standards compliance application at large German bank

Description	**A large German public sector bank (nearly $150B in assets) is leveraging its existing assets to roll out compliance with IFRS & Basel II.**
Architecture	**The master information is stored in disparate databases and different formats, ie, Adabas and sequential files. The master information is accessed by WBI Message Broker (WBI MB) application via IBM WebSphere Information Integrator Classic Federation (WSIICF) ODBC driver. Then, the WBI MB application modifies and transforms the data and writes it back to Adabas and new DB systems, like Microsoft SQL Server, DB2 and SAP IFRS solution.**
Benefits	**With WebSphere II Classic Federation, this German bank was able to integrate existing assets in new business context. Solution benefits include reduced requirements for ETL resources, increased reliability, robust performance and real time delivery of complex information.**

Convergent Billing System at US telecom giant

Description	**An US telecom giant is building an integrated customer management and convergent billing solution on the HP-UX platform, using Singl.eView, a leading telecom sector solution.**
Architecture	**A significant part of the data needed by this application resides in Adabas 7.4 running on z/OS 1.4. WebSphere II runs on the same HP-UX box as Singl.eView, providing a federated view of Oracle tables and Adabas files using the WebSphere II Classic Federation ODBC/CLI driver with its ODBC wrapper.**
Benefits	**Provides availability of current data to the billing system, thus reducing billing errors. Also eliminates the additional expense associated with ETL and batch programs.**

Web-based reporting at large financial institution

Description	**An US payroll processing giant is using WSIICF to provide access to their operational database running on CA-Datacom on z/OS to their Tax and Finance reporting infrastructure running on a Solaris WebSphere Application Server environment.**
Architecture	**A significant part of the data needed by the reporting infrastructure resides in CA-Datacom running on z/OS 1.3. WebSphere Application Server runs on Solaris, providing an app server environment for Java servlets and EJB's that generate reports, using the WebSphere II Classic Federation JDBC driver for legacy (Datacom) access.**
Benefits	**Eliminates the significant expense associated with ETL and batch programs, a process which was not satisfying their business requirements anyway, since a latency of more than two hours led to loss of revenue. Current architecture provides availability of current data to the reporting system (currency of data for the reports being a key component of their business), thus providing significant revenue savings.**

Enterprise connectivity of mainframe data to .Net apps at large credit card processing institution

Description	**One of the largest credit card processing firms in the world is using WSIICF to provide access to their .Net based apps, their platform of choice in the distributed environment. One of their first apps to roll out is the corporate dashboard app.**
Architecture	**The credit card processor's operational data resides in VSAM, IMS, DB2 and QSAM files running on z/OS. The corporate dashboard app needs real time data federation across the heterogeneous data sources mentioned above, providing decision support information for corporate executives.**
Benefits	**The real time federation provided by IICF makes this app possible. Alternatives like transaction wrappers and ETL was not considered practical because of the complexity and wide diversity of the data. Other key factors include lack of any significant impact on mainframe resources (monitored using Omegamon), and a staggered security implementation in which mainframe security credentials are hidden from the distributed platform users and developers.**

E-Commerce and infrastructure applications at international technology reseller

Description	**Large international technology reseller is using WSIICF to provide access to mainframe non-relational data for their e-commerce and infrastructure (shipping, reporting and licensing) applications. In production for over 2 years.**
Architecture	**The reseller's operational data resides in VSAM, IMS and DB2 running on z/OS. There are two production WSIICF servers – one for North America, and the other serving EMEA and Asia, each is serving 100 simultaneous clients at any given point of time, on average with a response time of 100-200 ms.**
Benefits	**The real time federation, along with IICF's robust and scalable behavior makes all these applications possible. Alternatives like transaction wrappers and ETL was not considered practical because of the expense & lack of scalability. Other key factors include insignificant impact on mainframe resources (monitored using Omegamon), and simplicity of "map once, deploy whenever" approach.**

E-Commerce and infrastructure applications at giant food chain

Description	**Giant food chain is using WSIICF to develop new enterprise-wide applications on the Weblogic platform, integrating their diverse data sources (result of many acquisitions over the years)**
Architecture	**The giant food chain has its operational data in VSAM, IMS, DB2 and IDMS running on z/OS. They made a strategic decision to keep this operational data in place, and develop new enterprise apps on BEA's Weblogic platform.**
Benefits	**IICF's standard JDBC driver enabled this giant to migrate to BEA's Weblogic as their development platform of choice, allowing them to leverage existing Weblogic/ J2EE expertise without having to worry about the backend legacy data structures (IMS, IDMS etc.) IICF's robust and scalable behavior is also a plus. Alternatives like transaction wrappers and ETL was not considered practical because of the expense & lack of scalability. Other key factors include insignificant impact on mainframe resources (monitored using Omegamon), and simplicity of "map once, deploy whenever" approach.**

Two-way replication at insurance company

Description	**An insurance company based in Wisconsin is creating an ODS on Oracle to deal with the disparate IMS operational systems resulting from acquisitions in Texas & New York.**
Architecture	**The insurance company has its operational data in several IMS systems on z/OS. To standardize their data for use by new applications, they are creating an Operational Data Store in Oracle/ Solaris, using Informatica's Powerstage product to synchronize the two. Even though Informatica now owns PowerExchange (the old Striva) for IMS access, it has no transactional capabilities. Hence, Informatica brought IICF into the picture to fulfill this role, which it did successfully.**
Benefits	**IICF's standard ODBC interface (which ensured painless integration into a competitive product) coupled with full transactional IMS access ensured the success of this complex integration process, which allows the customer to develop new applications on Oracle, hiding the complexity & diversity of backend data from the application developers.**

Reporting infrastructure at major North American transportation organization

Description	**One of the largest transportation organizations in North America replaced the incumbent VSAM access product with a better performing IICF, which incidentally also has smaller footprint and the latest ODBC/ JDBC interface**
Architecture	**The transportation organization has its operational data in VSAM/ CICS. They had the i-Way product to provide VSAM connectivity to their reporting system (based on Crystal Reports), but found it to be too unstable and resource-intensive. After a very comprehensive PoC (IICF proved itself to be not only more stable & light on resources, but also superior in performance), IICF replaced the incumbent as their mainframe connectivity solution.**
Benefits	**IICF's superior robustness and scalability, coupled with superior performance allowed the customer to save time and money being spent on maintaining a temperamental solution that was also eating up a lot of scarce mainframe resources.**

Agenda

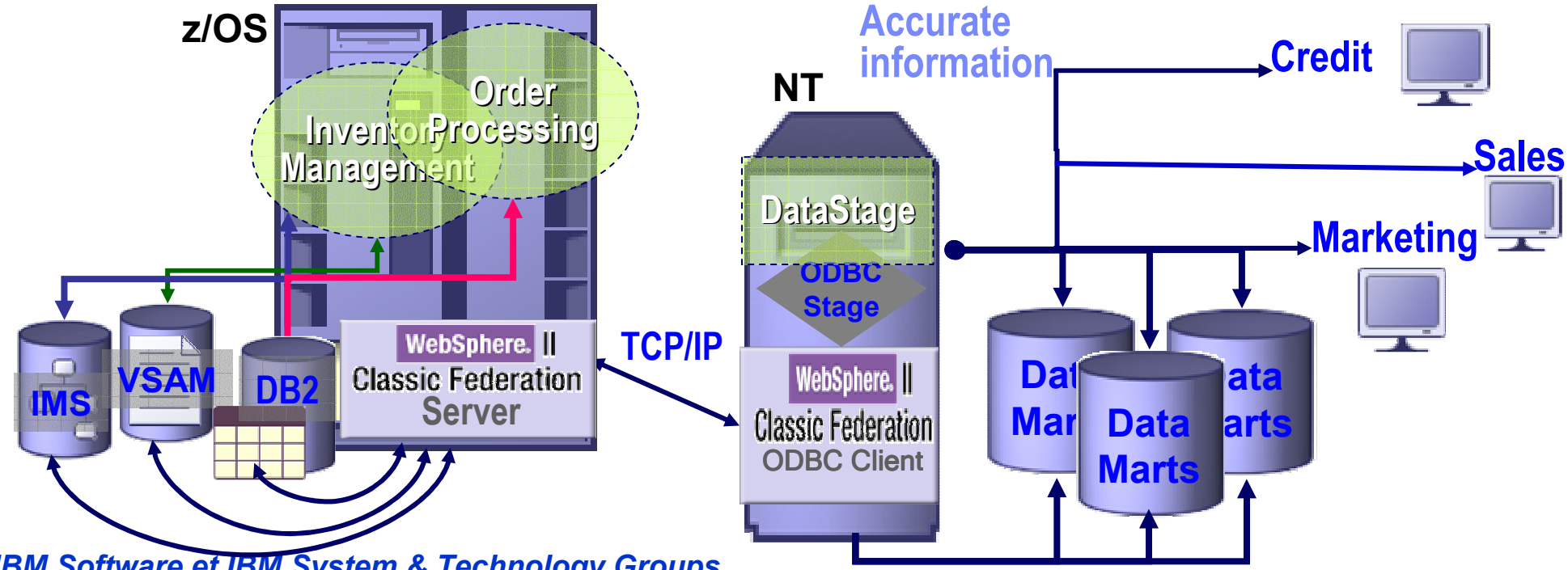
- WS II Classic Federation Usage Case Studies
  - Broad categories
  - In-depth discussion

[redacted]
WebSphere Information Integrator

- **WebSphere II Classic integration scenarios with DataStage**

Example of operational data feed to ETL via SQL using Classic Federation

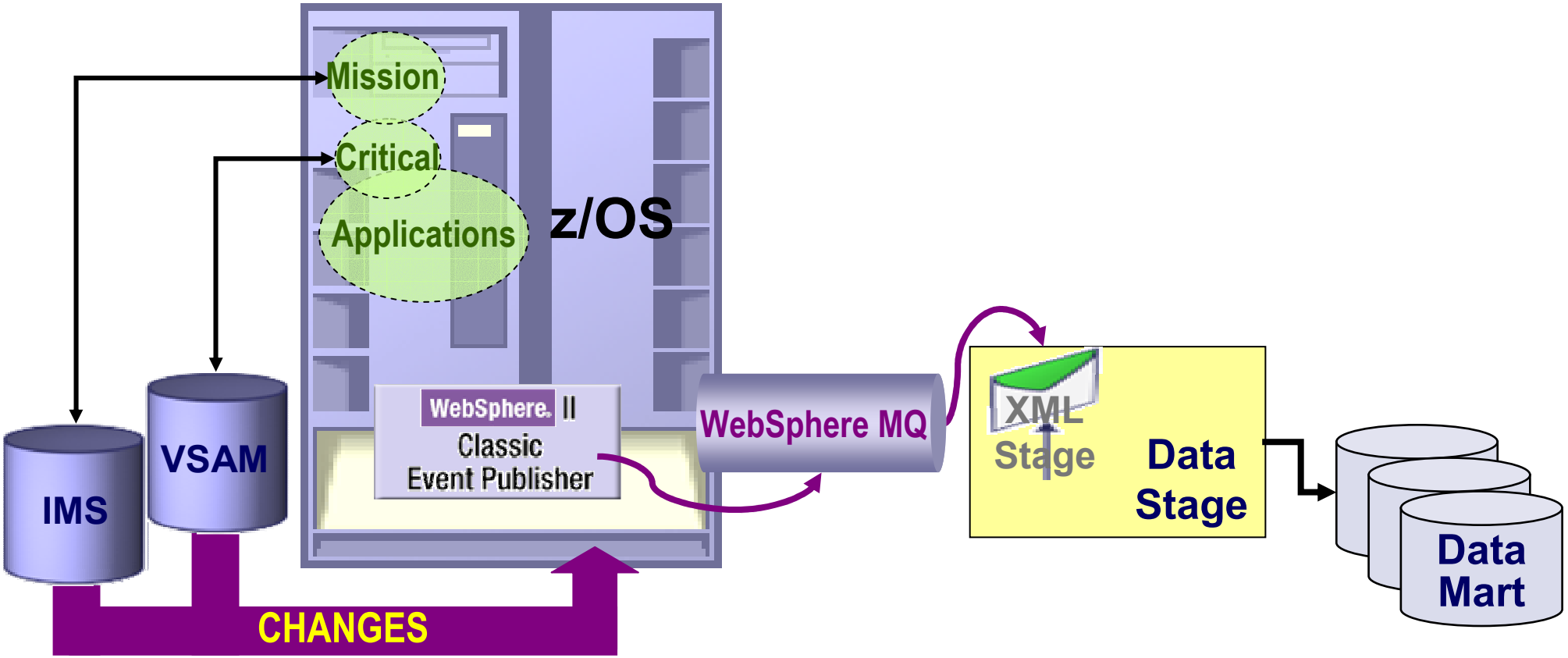
- **Dynamically connect data warehouse tool with mainframe data**
  - No dependence on mainframe development
    - Dramatically simplifies management: One team owns it all
  - One consistent process leverages "power" of ETL tooling
    - Development time "cut in half"
  - Empowers additional uses
    - Dynamic query by business intelligence tools extends the warehouse



IBM Software et IBM System & Technology Groups

Example of Replication using Classic Event Publisher

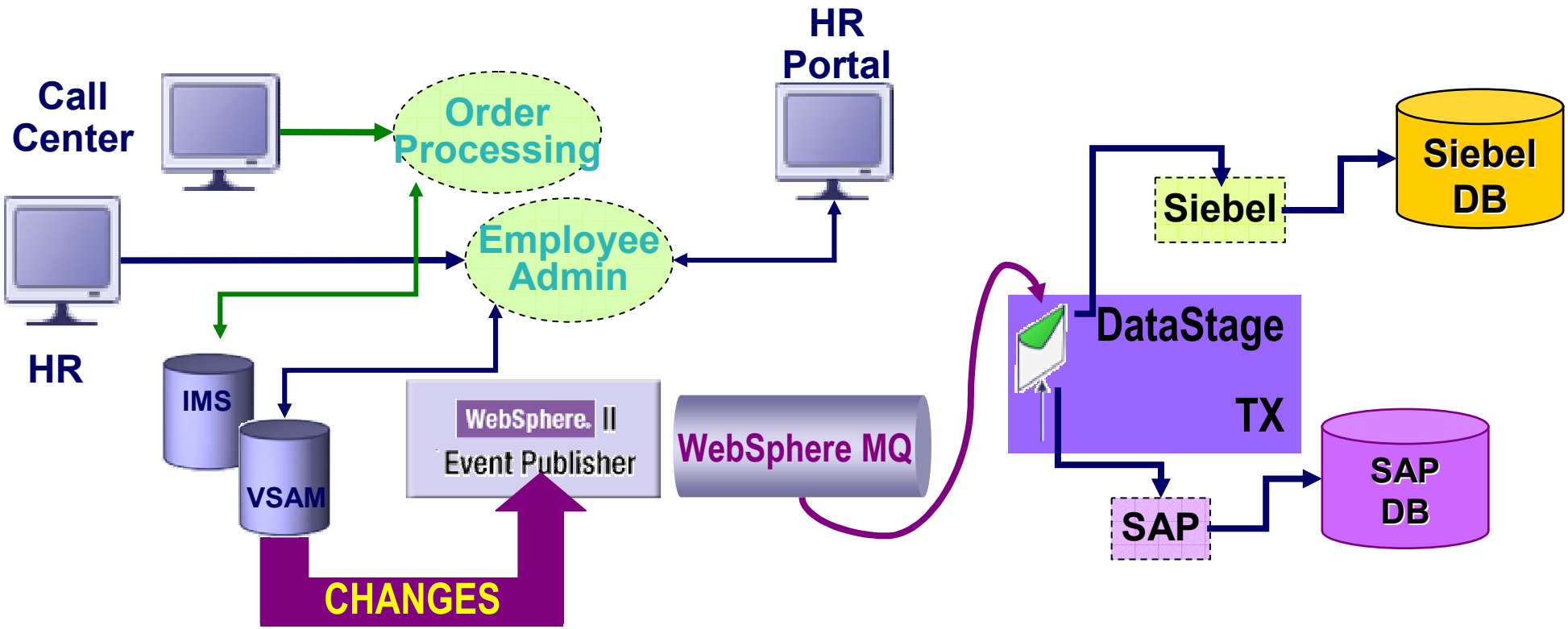
- **Dynamic, changed-data feed**
  - Maximize data currency while minimizing & stabilizing bandwidth utilization
- **Reliable and recoverable**
  - Recoverability built-in and WebSphere MQ assures high performance delivery



IBM Software et IBM System & Technology Groups

Example of Synchronization with CRM and ERP using Classic Event Publisher

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



Université du Mainframe 2005

Thank you!

WS II Solutions & Deployment : Bob DeCori
email: rdecori@us.ibm.com