Key Considerations for Mainframe Data Integration -- One size does not fit all

Karen Durward
Information Integration Product Management

TAKE BACK **CONTROL**

Agenda

- Information on demand and data delivery
  - Concepts
  - Information Integration challenges
  - The Integration Landscape
  - Process-Centric Integration versus Data-Centric Integration

- Applying the Integration Patterns
  - Sample Implementations
  - The Role of Replication
  - 3rd Generation Data Warehousing

- IBM Information Server
  - A platform approach
  - Technology
  - Sneak Peek: Information Server for System z

- Summary
  - Where do you go from here?

TAKE BACK CONTROL

Information on Demand
… and Data Delivery

Integration – Some Definitions

- Process-Centric
  - EAI (Enterprise Application Integration)
    - Process-centric, automate workflows and process dependencies
    - Leverage process assets across applications, to create new applications
- Data-Centric
  - Federation
    - Data and content centric "pull"
    - Virtualizes access to information sources: structured, unstructured, content
  - ETL (Extract, Transform, Load)
    - Data placement and transformation solutions
    - Typically scheduled, but moving to more real-time approaches
  - Data Event Publishing
    - Data centric "push"
    - Driver behind other data movement and integration patterns
  - Replication
    - Creating and synchronizing like-to-like copies of data
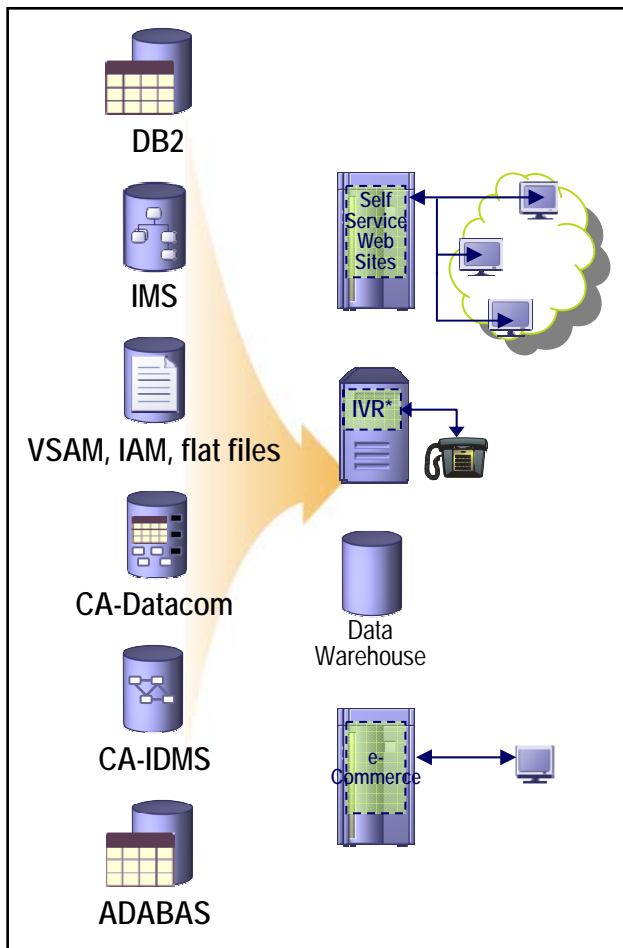   - Focus is on low latency, reliability and multi-directional capabilities

Data Integration Challenges You Face Every Day

- New applications on diverse platforms drive integration demands
  How do you reuse critical "z" operational data on these platforms?

- Mergers & acquisitions and business intelligence drive consolidation
  How do you manage complex transformations to rationalize diversity?

- Reliance on proprietary skills can result in workload backlogs.
  Can you leverage standards on the mainframe?

- Performance, geography and the politics of "data ownership" drive you to create more and more copies of your data.
  Can you ensure timely, accurate and consistent synchronization?

- "Data and content volumes are expected to increase by a factor of 10 during the next five years"   Gartner, November 2005
  How will you deal with these growing volumes?

TAKE BACK CONTROL

Mission Critical "z" Data Can Be Difficult
Over 60% of operational data is sourced from System z

DB2

IMS

VSAM, IAM, flat files

CA-Datacom

CA-IDMS

ADABAS

Self Service Web Sites

IVR*

Data Warehouse

e-Commerce

Why is system z data delivery an issue?

- Proprietary databases – multiple, complex APIs
- Billions of lines of mainframe code – must leverage
- Integration with modern initiatives like data warehousing and SOA are not native to this data

What are you doing to meet these challenges?

- *Hand-coding* – COBOL/PLI programs for every data need
  - Bottleneck for new initiatives … slows everything down
  - Dependence on unique skills … proprietary API experts

- *The One-Size-Fits-All Approach:*
  - ▸ Integrate the transactions
    - … but this can be
    - overly complex
    - maintenance intensive due to hard wired integration
    - ill suited to the functional demands of new workloads

  - ▸ A unique "copy" for every requirement
    - … but this can lead to:
    - Latency errors
    - Synchronization challenges
    - Quality issues

TAKE BACK CONTROL

Business Integration Market Landscape

- Process &/or Information
  - EAI "primarily" focuses on application assets
    …secondarily data
  - ETL differentiates on transformation
  - Replication focuses on placement
  - Federation is pure data access
  - Publishing is an enabler to the rest

- Emphasis is on complementary uses
  - Update only data warehousing using ETL
    with Event Publishing
  - Empower SOA-based data access
    via Federation
  - Use Published "data events" to drive
    EAI workflow
  - Leverage Replication for guaranteed delivery
    of high volume changes to ETL
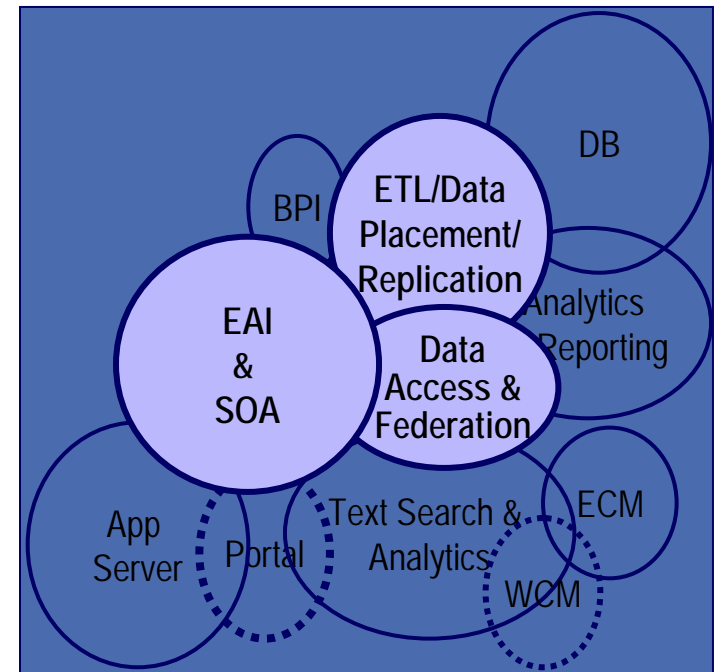
The Information-Centric Patterns
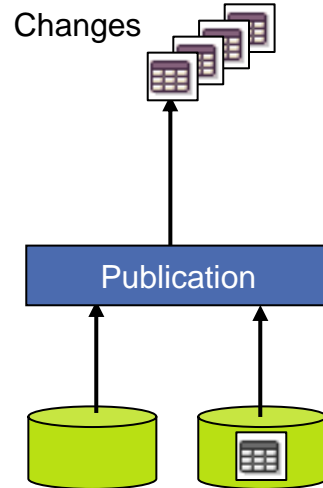
Virtualize

All data participates equally

Application,
BI Tool, …

Warehouse
Data Mart

SQL

Federation
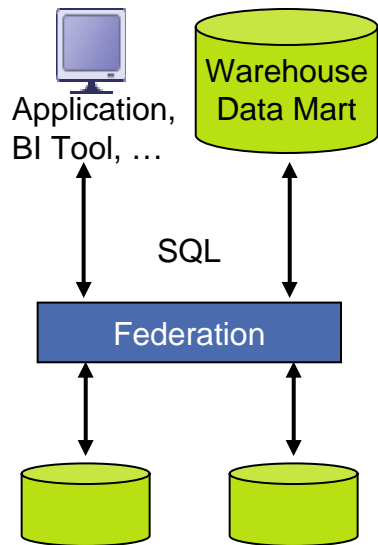
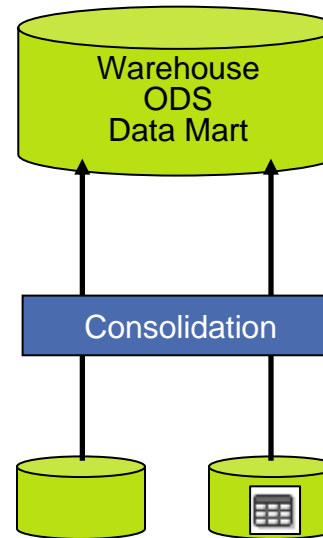Seamlessly integrate with:
– e-commerce
– Self service

Move

Data changes drive action

Changes

Publication

Proactively deliver data to:
– Drive integration
– Enable change only updating to:
  • Reduce batch window dependencies
  • Reduce data latency

Transformation rich copies

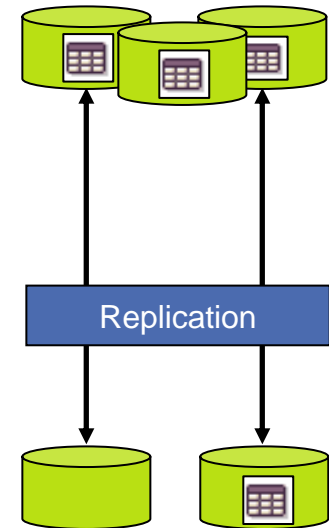Warehouse
ODS
Data Mart

Consolidation

Repackage & repurpose
– Data Warehousing
– Operational Data Stores
– MDM
– CDI

Synchronize

Like-to-like copies

Replication

Create & synchronize copies
– Distributed operations
– High availability data
– Local copies for performance
– DBMS migrations

TAKE BACK CONTROL

Data Delivery Requirements
… Applying the Patterns

TAKE BACK **CONTROL**

Choosing an Integration Pattern
A multi-dimensional self service environment

Insurance carrier wants to reduce costs and improve customer and agent satisfaction by implementing self-service environments…BOTH Interactive-voice-response and Web sites

- Initial Environment
  - Call center representatives take calls from agents who interface with clients
  - Mainframe systems consisting of over 1,000 transactions handle:
    - Accounting, Policy management, Claims and more
  - Technology is
    - IMS transactions
    - IMS, VSAM and DB2 data
    - Very complex after decades of application evolution

- Requirements
  - Minimal disruption on existing call-center
  - Staged implementations showing value every 3 – 6 months
  - Performance and accuracy are essential to credibility
  - Unskilled users (agents and customers) mandate an easy-to-use interface
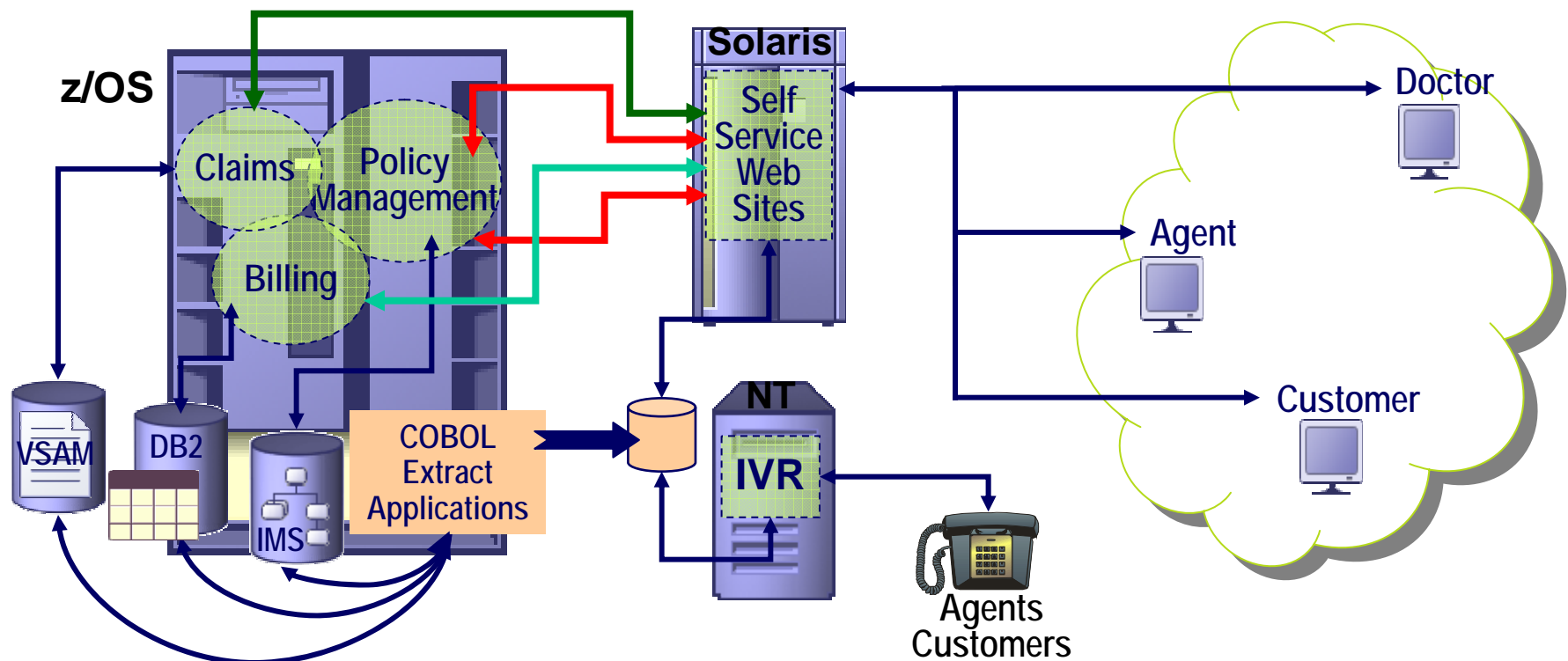
TAKE BACK CONTROL

A Multi-Dimensional Self-Service Environment
EAI and ETL are traditional choices

Complexity and volume made them too costly and too time consuming
- Copy the data to Oracle ……..… $2M for hardware and software
- Connect through transactions….   10,000 man hours per system with little reusability

A Multi-Dimensional Self-Service Environment
Federation delivered in a fraction of the time and cost

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
- No disruption to operational systems
- Minimal overhead in OLTP environment

Choosing an Integration Pattern
You may need more than one

Catalog retailer is building an e-commerce site

- Initial Environment
  - Call center representatives take orders
  - Mainframe order-processing system used by call center reps
    - CICS transactions
    - CA-IDMS data
   - Highly-optimized for use by skilled employees

- Requirements
  - Minimal disruption on existing call-center order processing
  - Must support "shopping", not just order taking
  - E-commerce site must be optimized for unskilled customer usage
  - Performance and accuracy are essential to business continuity and reputation

"Best" solution may leverage multiple patterns
ETL, Federation and EAI

- **ETL**: merge operational product data with image content
  - Product information is copied to "local" RDBMS for performance
  - Changes are infrequent and can be scheduled
  - Value-add image data added as part of the data transformation
- **Federation**: critical "real time" data is dynamically accessed
  - There can only be one version of the "truth" for inventory
- **EAI**: critical processes are shared
  - Leverage the CICS transactions for pricing, ship-to-date, etc
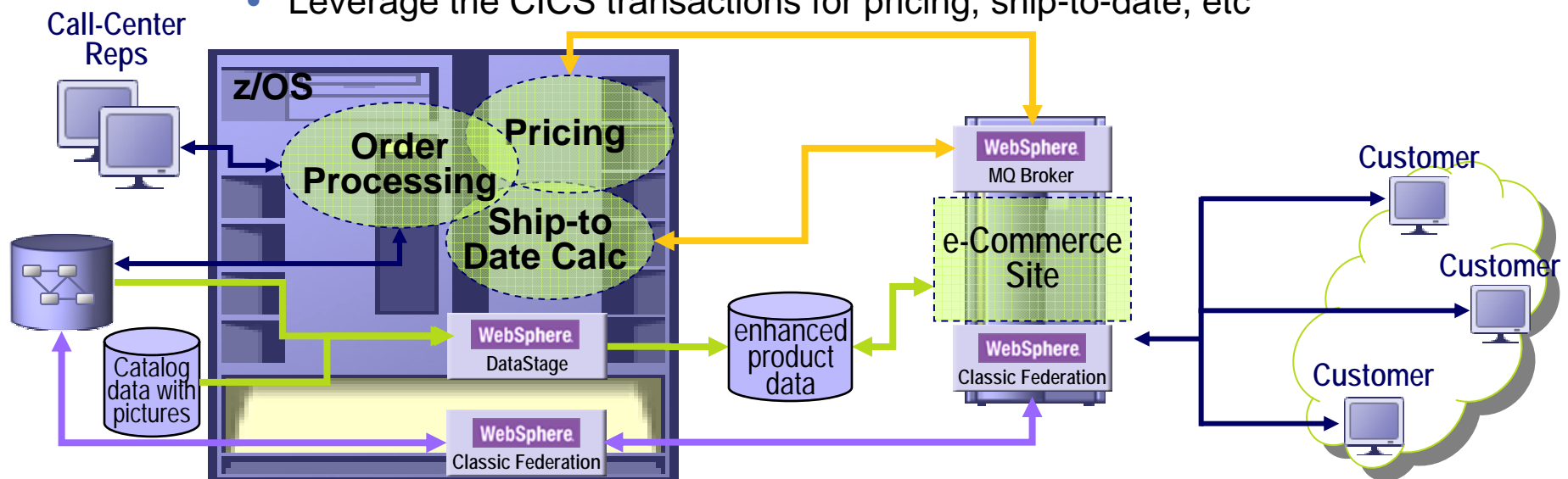
Choosing an Integration Pattern
Federation & Event Publishing as enabling technologies

Global technology distributor wants to implement a Web e-commerce environment, reduce inventory costs and improve decision making

- Initial Environment
  - Two mainframe-centric transaction processing centers
    - Ensures 24 by 7 operations for their Global business
  - Technology environments
    - Mainframe CICS environment servicing ~50 million transactions/week
    - IMS databases hold most of their data

- Requirements
  - Implement their Next-Gen transaction platform on WebSphere
    - Staying with z/OS & IMS data for performance, reliability and security
  - Improve interfaces with their suppliers
    - Implement "real time" interfaces for a just-in-time inventory environment
  - Leverage reduced cost of an "off shore" data center for non-transactional workloads
    - Implement an ODS to support business reporting and decision processing

Federation & Event Publishing Behind …
Transaction processing, ETL and EAI

- Federation enables Web-based applications
  - Java-based access to IMS data with transactional performance

- Changes delivered to India data center's ODS via ETL
  - Continuous, 24 by 7 delivery:
    - Ensure up-to-the-second data is available
    - Eliminate dependence on batch cycle
  - Steady, predictable usage of communication resources

- Changes delivered to Partner Gateway (EAI)
  - Enables Just-in-Time inventory management
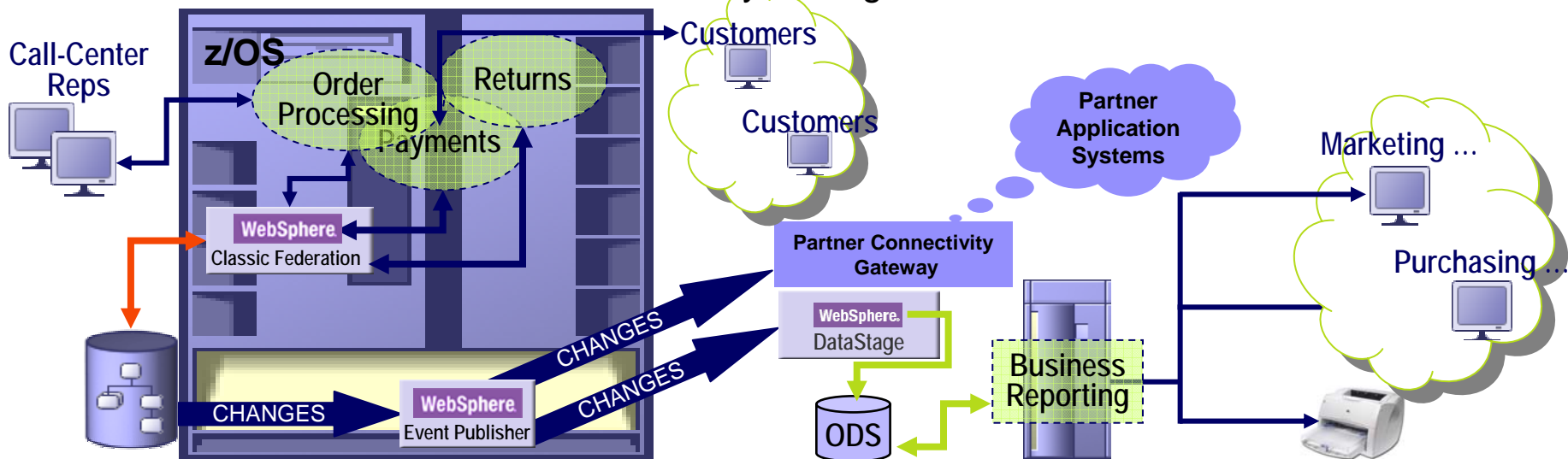
Choosing a Data Movement Solution
Replication or ETL?

Financial Services firm requires 24 by 7 operational environment for financial trading applications

- Starting Environment
  – Two mainframe-centric data centers approximately 1,500 miles apart
  – HADR environment close to primary data center
  – Technology environments
    • Mainframe CICS environment servicing 1,000s of transactions per second
    • DB2 databases hold most of their data

- Requirements
  – Provide secondary operational platform for rapid switch-over
  – Improve TCO profile by leveraging secondary platform for "real" work

Replication is "best" because:
  – *Low latency is critical*
  – *No transformations are required*
  – *Easy to switch source & target during a failure*

TAKE BACK **CONTROL**

On-line Trading Environment
High Availability Replication
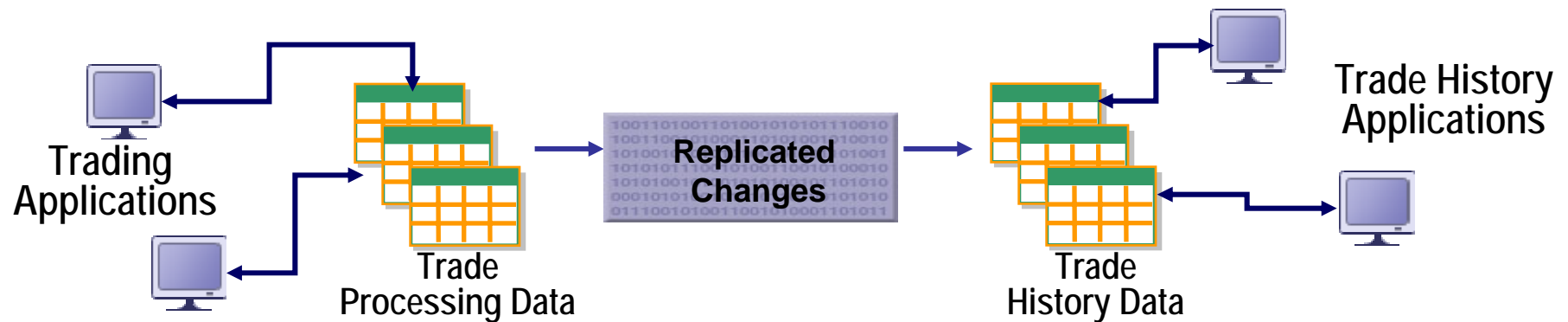
- Replication synchronizes data between the primary & secondary platforms
  - 5-10 million transactions daily replicated to 2ndary platform w/ <1 second delay
  - All query processing routed to 2ndary platform to improve primary throughput

- On primary system failure:
  - Secondary system becomes primary

- On primary recovery:
  - Secondary system replicates to primary until "caught up"
  - Trading users redirected back to primary

TAKE BACK **CONTROL**

Choosing a Data Movement Solution
Replication or ETL?

Rapidly expanding retailer wants to push price changes from central system to hundreds of stores to improve accuracy and competitive position

- Starting Environment
  - Headquarters mainframe data center with DB2 pricing data
  - Store inventories and departments vary … not all prices apply to every store
  - Communications bandwidth varies from 256k to 1Meg

- Requirements
  - Push price changes out as needed to rapidly react to changing demand
  - Ensure independence of each store's updates
  - Anticipate need to return some data back to HQ

Replication is "best" because:
- *Automatic recovery with guaranteed delivery*
- *Minimal transformations are required*
- *Bidirectional capabilities inherent in Replication*

TAKE BACK **CONTROL**

Rapid, Reliable Data Distribution
Leveraging Multi-faceted Replication
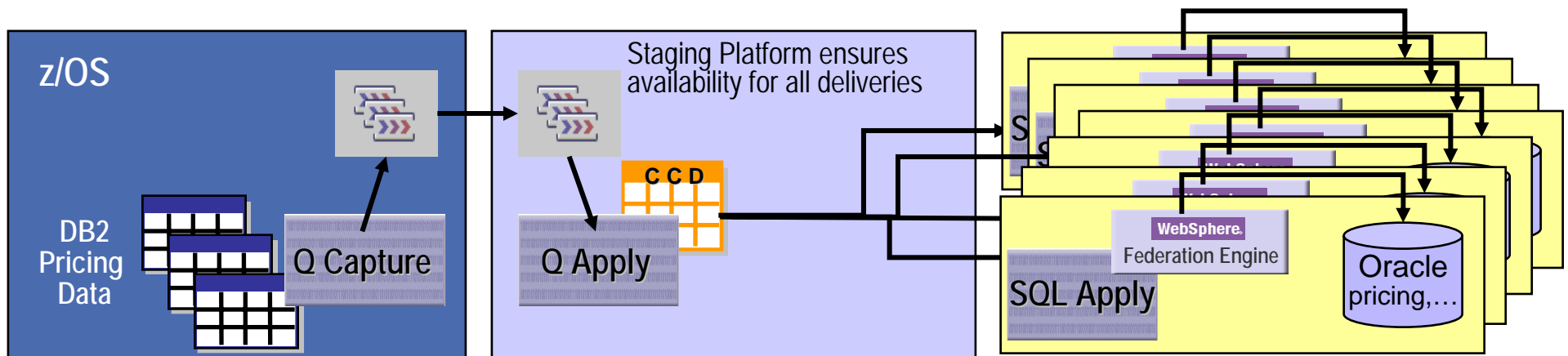
- **Customer Pain points**
  - Hundreds of thousands of products, with variations from store to store demands automation
  - Guaranteed delivery is essential … customer satisfaction requires consistency
  - Adding hundreds more stores in next few years

- **Value Proposition**
  - Capture & guaranteed delivery of "pricing updates"
  - Each store: (1) receives only what it needs (2) is fully recoverable (3) is independently updated
  - Flexible targeting enables changes
  - Enables store-to-HQ replication as needed

TAKE BACK **CONTROL**

What About SOA and Mainframe Data?
Financial Industry Case Study

Enable IT to rapidly deliver new business processes and applications that are dependent on mainframe data. Create an environment that keeps operational costs low, reuses existing investments and skills

- Starting Environment
  - Primarily DB2 for z/OS
  - Multiple portals and user applications

- Requirements
  - Quickly deliver solutions to clients requiring current, authoritative data with high service levels
  - Keep operational costs low
  - Reuse both existing IT assets and skills

Solution
- *Java Stored Procedures, published as shared services*
- *Provides access to DB2 queries and stored procedures via Web services*
- *Over 50% cost savings by re-engineering and converting traditional CICS/COBOL APIs to Java stored procedures as Web services*

TAKE BACK **CONTROL**

Service-Enabling Heterogeneous Data
Financial Industry Regulatory Compliance

Improve risk management across all member institutions and meet deadline compliance with Basel II.

- Starting Environment
  - Information in 23 different retail systems from over 2500 branch offices
  - Multiple portals and user applications

- Requirements
  - Enhanced risk management and increased efficiency of data collection for Basel II required data
  - Ability to view data in operational systems spread across the enterprise including third party information without disrupting retail systems

TAKE BACK CONTROL

Service-Enabling Heterogeneous Data
Finance: Regulatory Compliance

Risk measurement application

DB2 OLAP

Solution – Basel II Exception Monitoring
- *ETL packaged as a service provides a single point for access and control of risk-bearing information*
- *Spans many different mainframe systems*
- *Federation combines data stored in different technologies & formats*

Web Service Interface
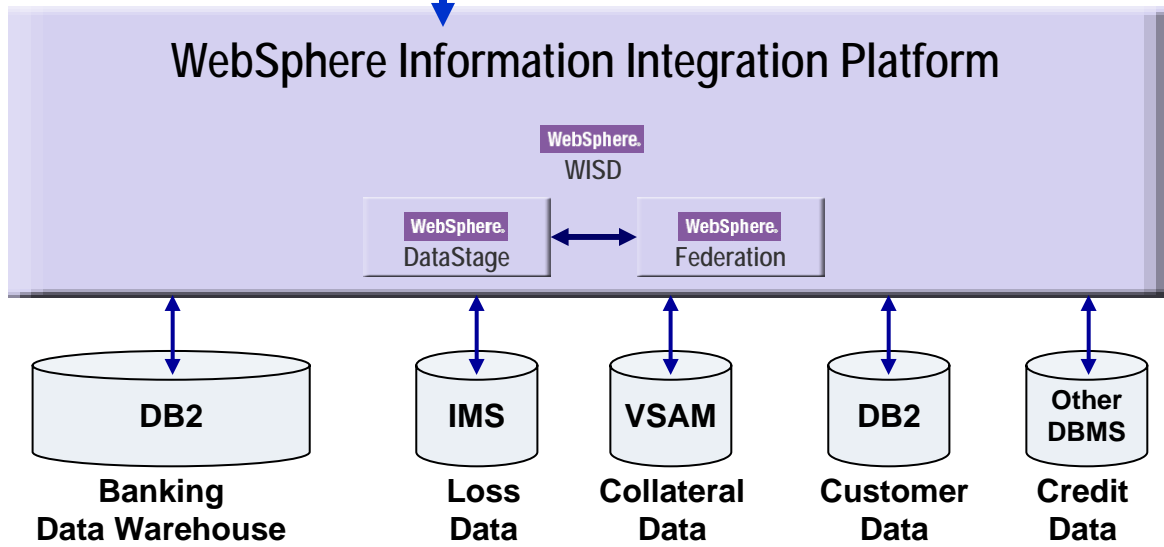
WebSphere Information Integration Platform

WebSphere.
WISD

WebSphere. DataStage ←→ **WebSphere.** Federation

DB2	IMS	VSAM	DB2	Other DBMS
Banking Data Warehouse	**Loss Data**	**Collateral Data**	**Customer Data**	**Credit Data**

Managing the Dynamic Data Warehouse Content
ETL Solution Accelerates with Automation & Reuse
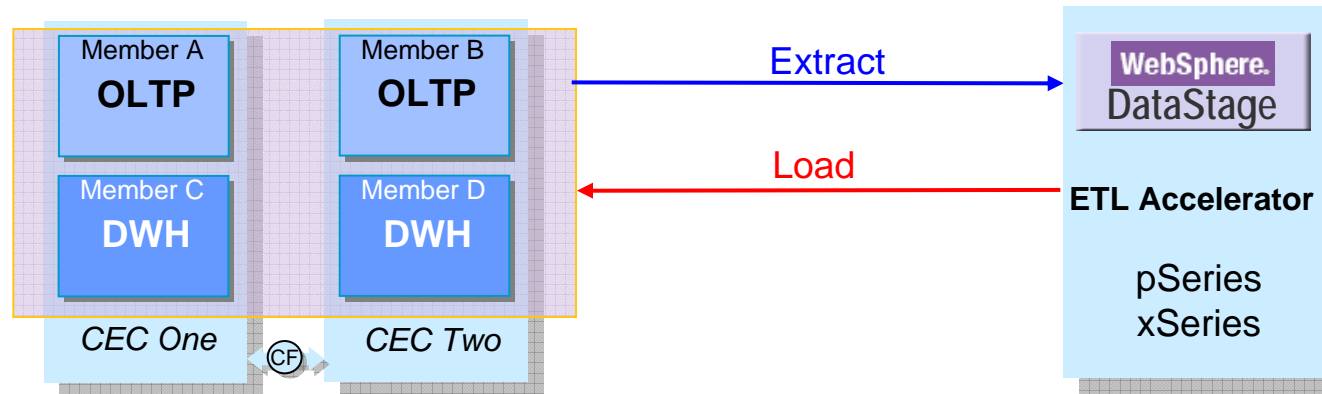
ETL is done using an ETL Accelerator, **WebSphere DataStage.**

- *E* – the data is extracted from the OLTP information sources

- *T* – 100s of embedded transforms & custom built transforms are applied

- *L* – Aggregated and transformed data is loaded into the warehouse tables

Managing the Dynamic Data Warehouse Content
Non-Relational Mainframe Data Must Participate
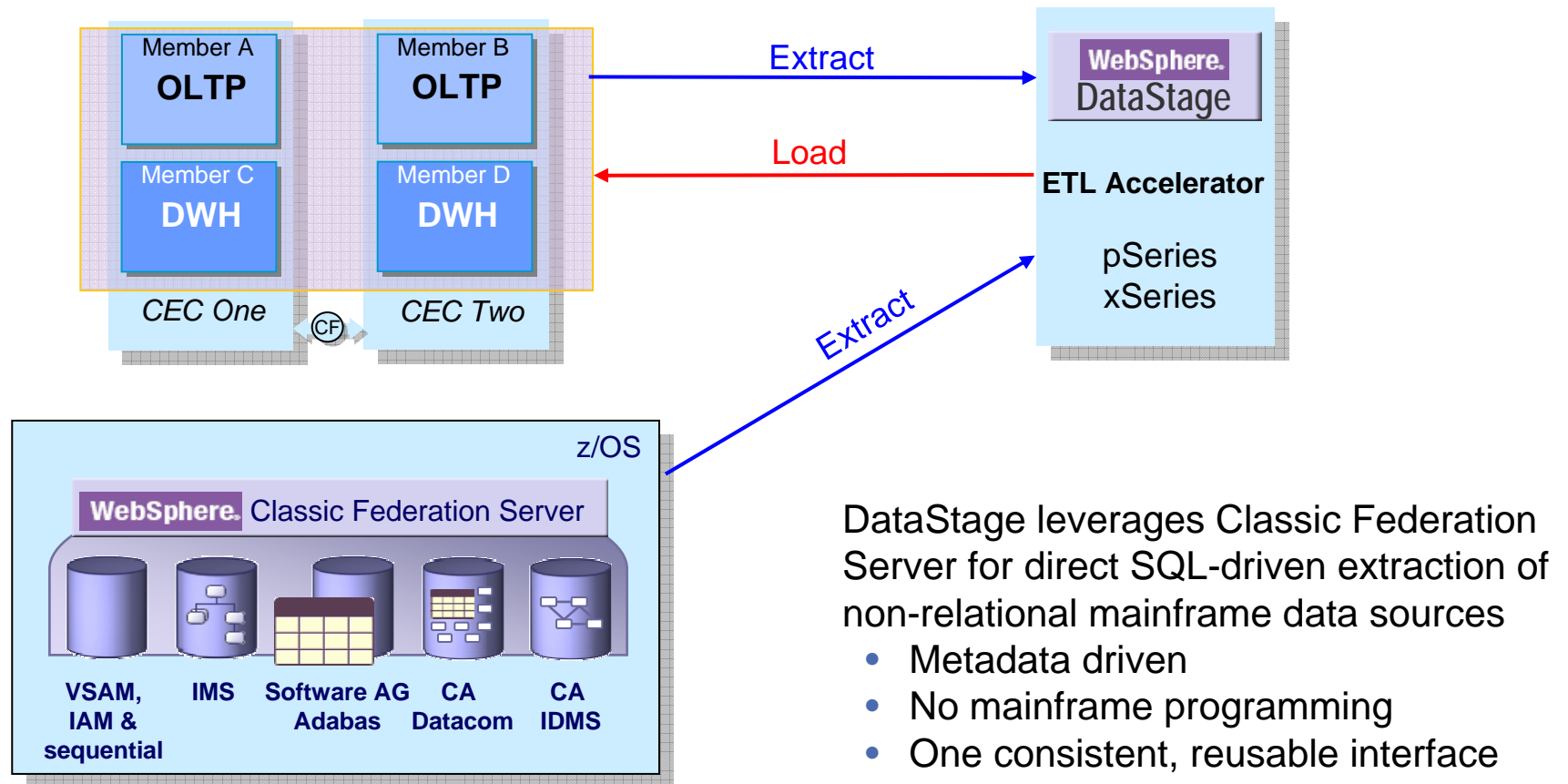
DataStage leverages Classic Federation Server for direct SQL-driven extraction of non-relational mainframe data sources
- Metadata driven
- No mainframe programming
- One consistent, reusable interface

"cut development time in half"

Managing the Dynamic Data Warehouse Content
Distributed Data Integrated via Native Connectors
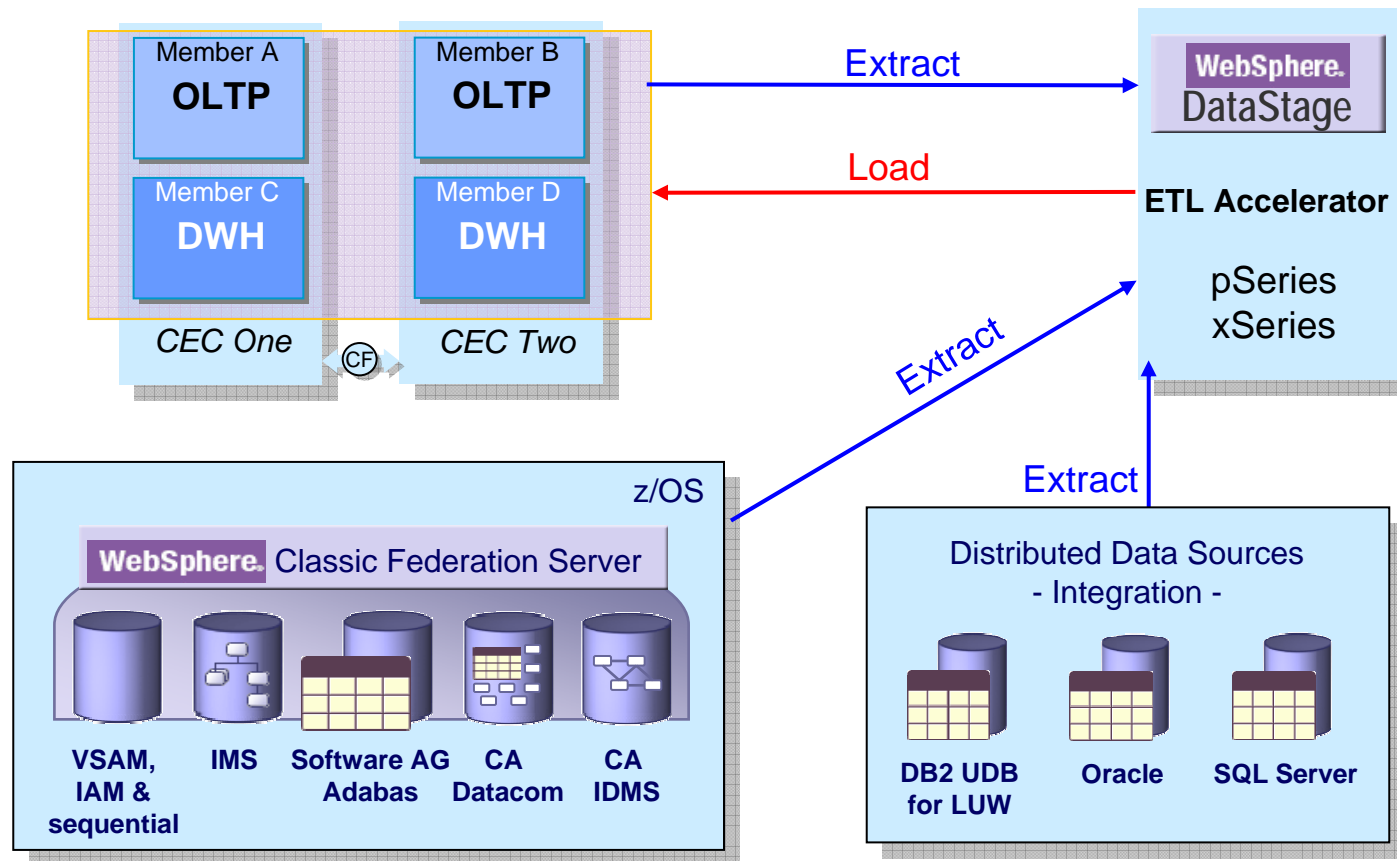
Distributed data sources are integrated directly through the IBM Information Server.

Maintenance of the Dynamic Data Warehouse
Incremental Updating is Essential

Member A
OLTP

Member B
OLTP

Member C
DWH

Member D
DWH

CEC One   CF   *CEC Two*

Change Capture

Update & Insert

WebSphere DataStage

ETL Accelerator

pSeries
xSeries

MQ

MQ

z/OS

WebSphere Data Event Publisher   Capture

VSAM   IMS

WebSphere Classic Data Event Publisher   Capture

CA IDMS   Software AG Adabas

MQ

MQ
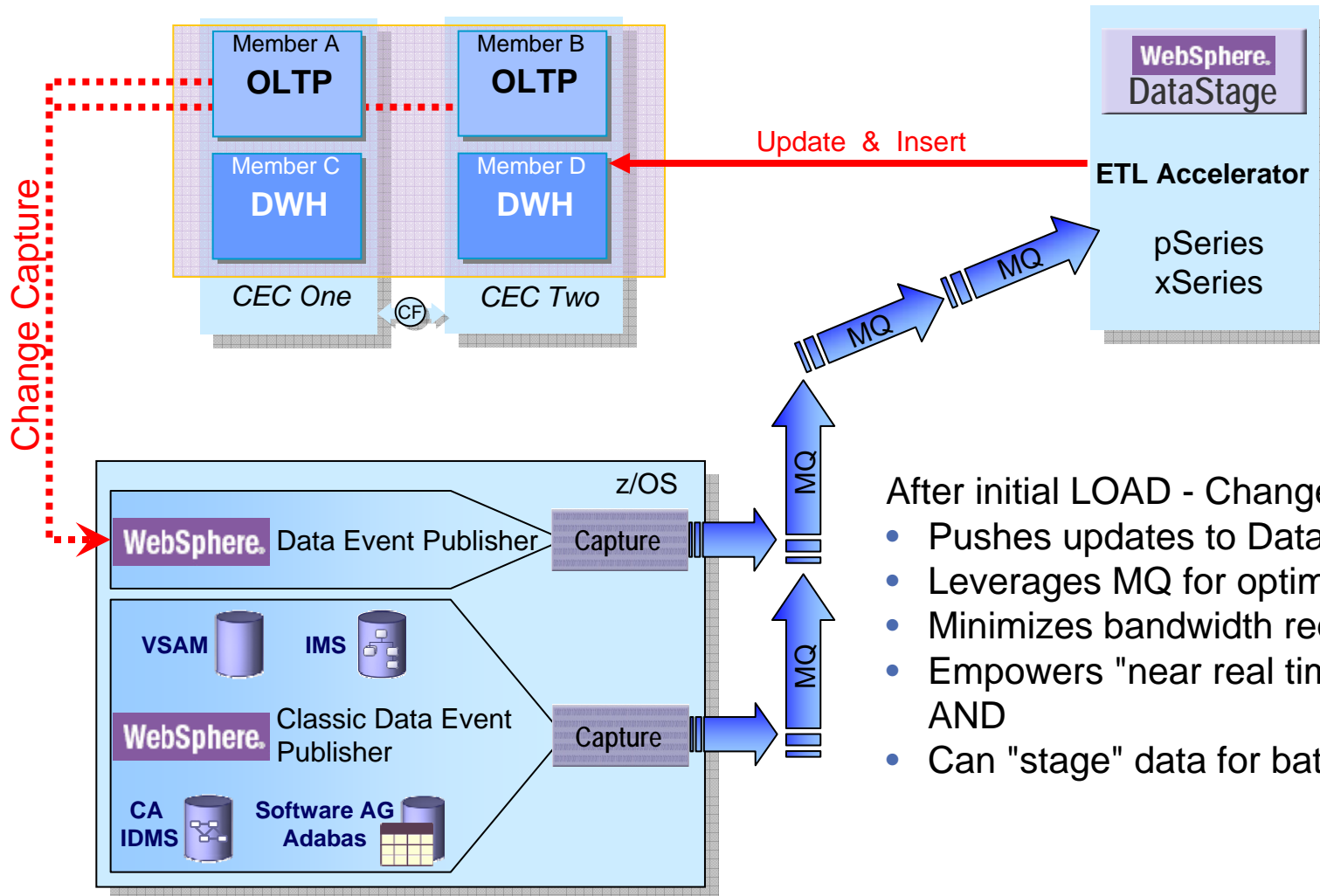
After initial LOAD - Changed-Data-Capture:
- Pushes updates to DataStage
- Leverages MQ for optimal delivery
- Minimizes bandwidth requirements
- Empowers "near real time" updating AND
- Can "stage" data for batch aggregation

TAKE BACK CONTROL

Maintenance of the Dynamic Data Warehouse
Incremental Updating is Essential

Member A **OLTP**	Member B **OLTP**
Member C **DWH**	Member D **DWH**
CEC One	*CEC Two*

CF

Update & Insert →

WebSphere. DataStage

ETL Accelerator

pSeries
xSeries
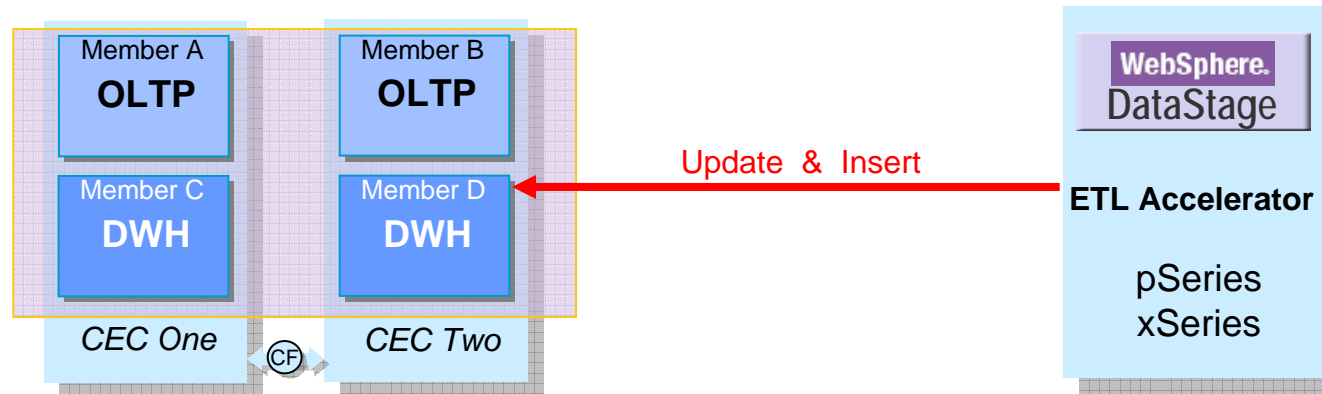
- Batch windows are shrinking
- Data volumes are rapidly expanding
- Requirements demand current information
- "Pull" integration is no longer enough

Maintenance of the Dynamic Data Warehouse
Incremental Updating is Essential
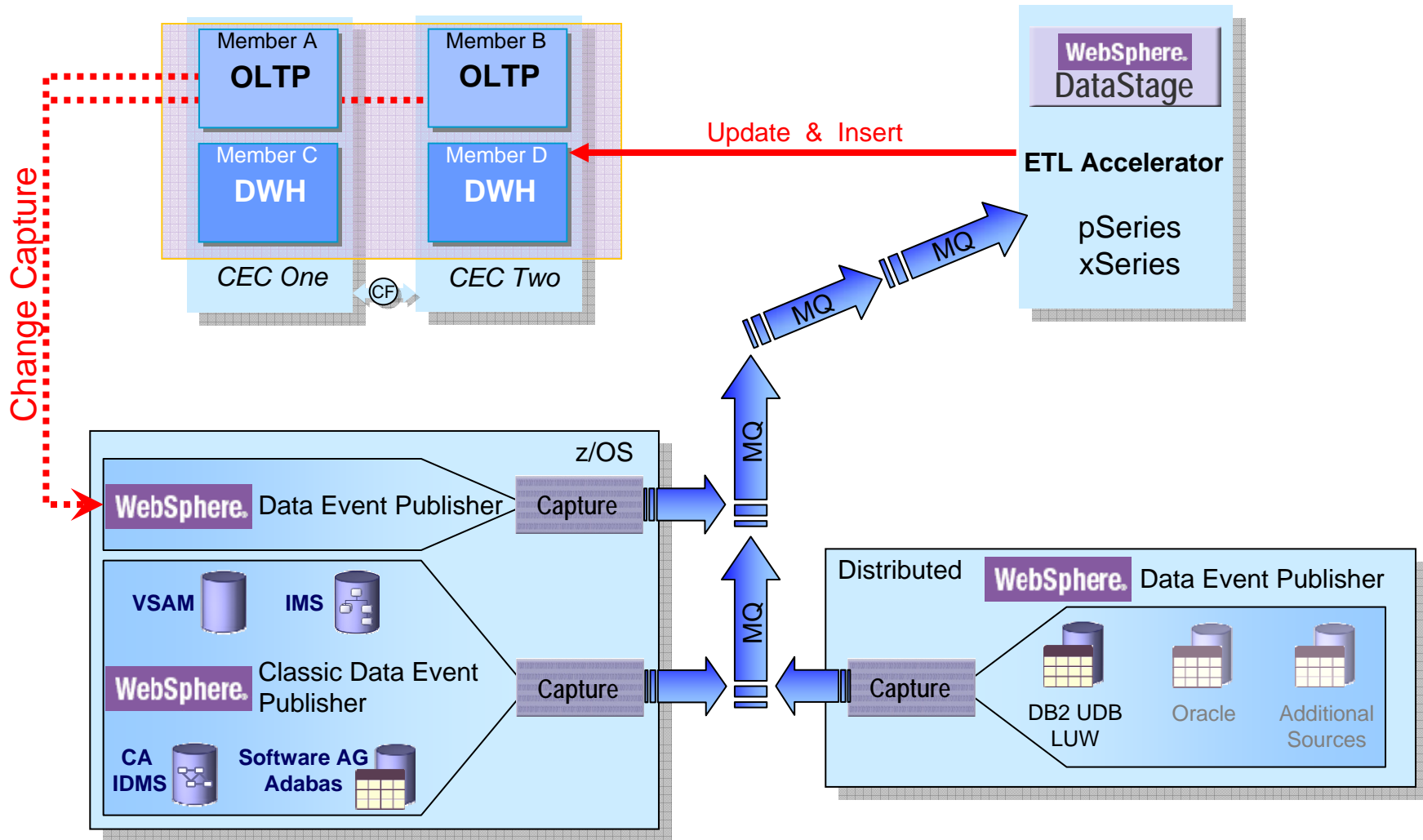
Replication Complements an ETL Architecture
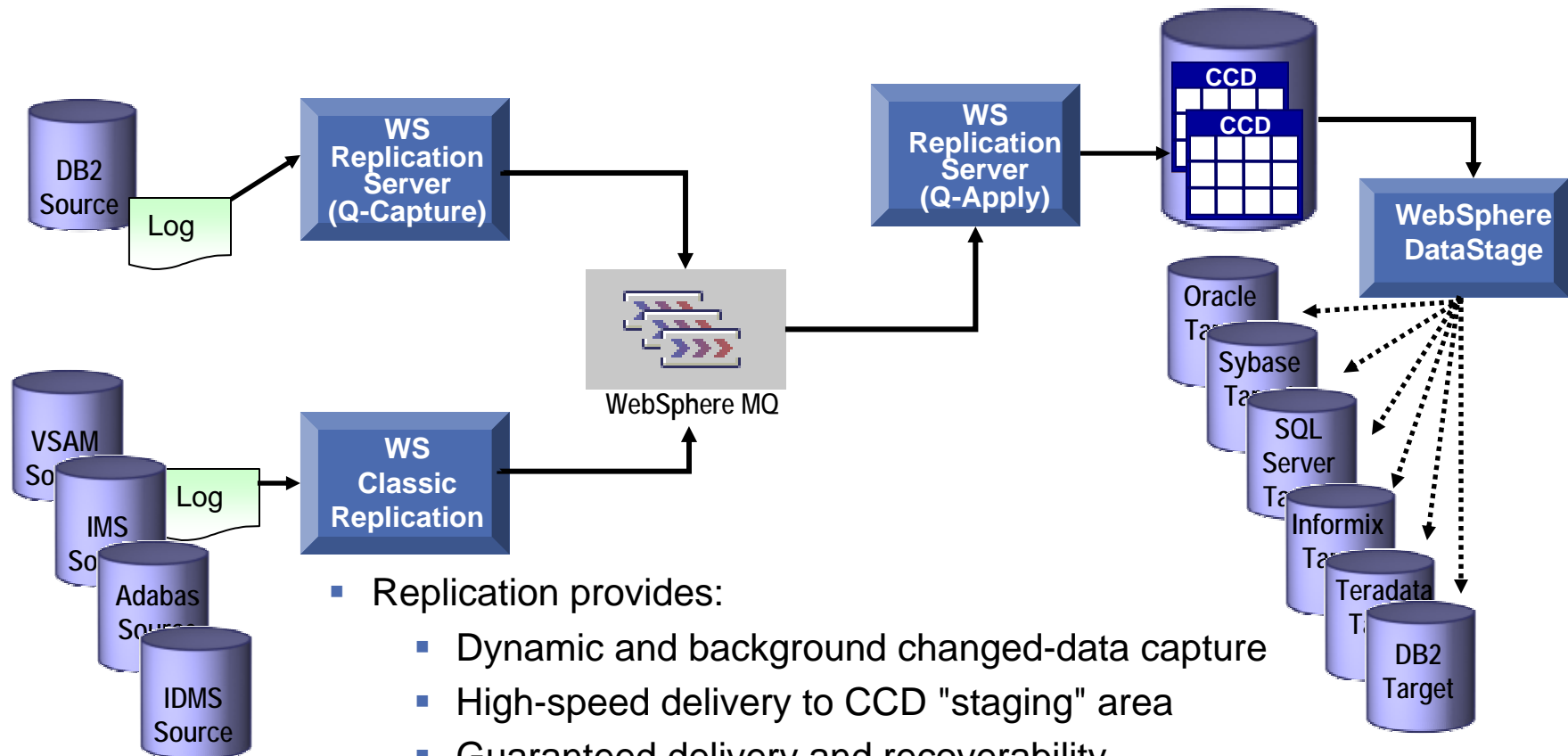
- Replication provides:
  - Dynamic and background changed-data capture
  - High-speed delivery to CCD "staging" area
  - Guaranteed delivery and recoverability
- DataStage provides:
  - Complex transformations
  - "Set based" aggregations
  - Highly heterogeneous targeting

TAKE BACK CONTROL

The Role of IBM Information Server
… Why and What

TAKE BACK CONTROL

Recognize that …
Information On Demand Requires an Eco-system of Services

…Creating business value by integrating, analyzing and optimizing heterogeneous types & sources of information throughout its lifecycle to manage risk and create new business insight.

Strategy

Business Driven Development

- Business Processes
- Business Applications
- Information Accelerators
- Information Services
- Information Assets & Systems

IT Service Management

Governance

Model / Assemble Deploy Manage
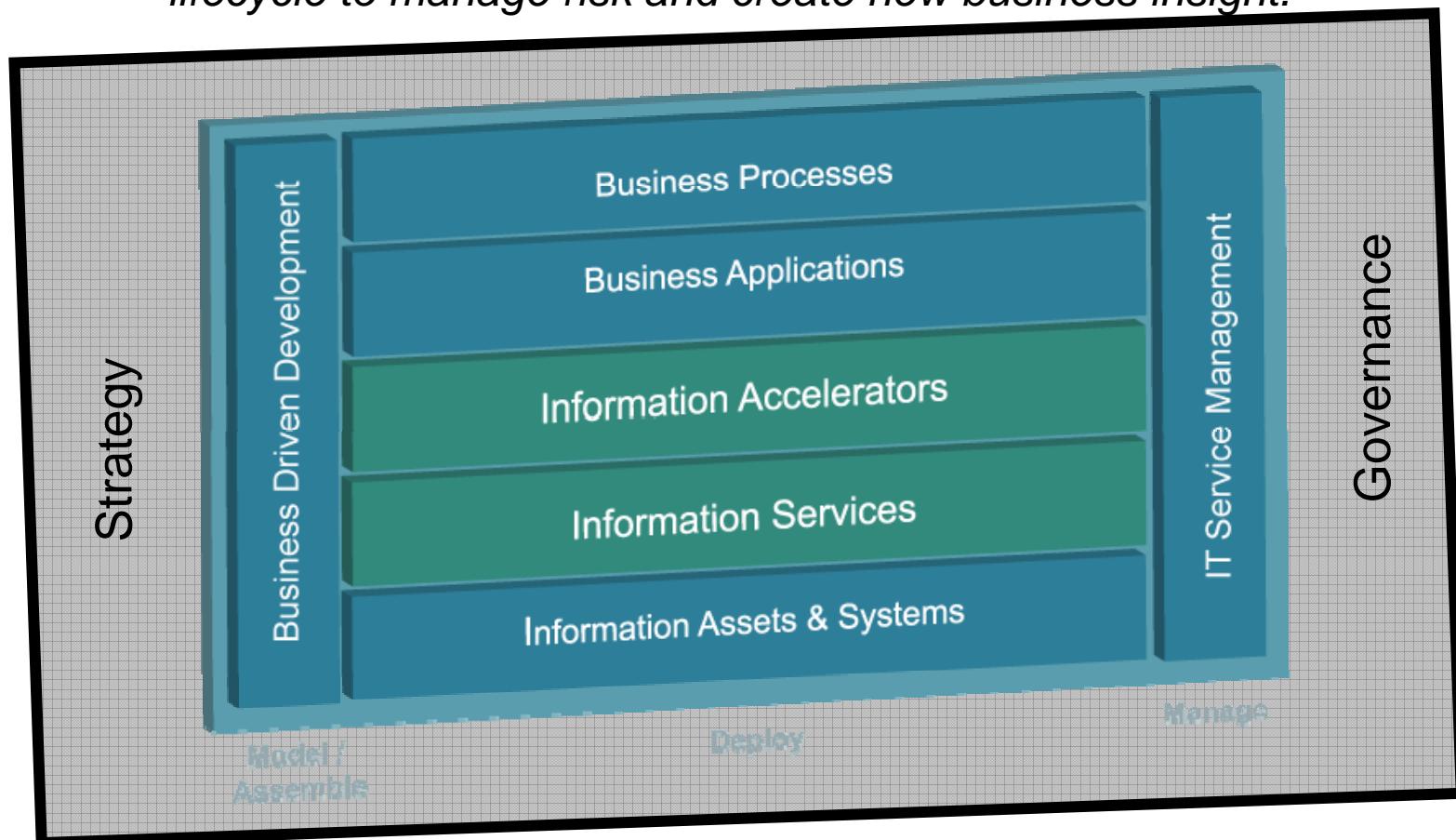
TAKE BACK **CONTROL**

And, like an Historic Process Inflection Point
That drove the creation of the Application Server …

APPLICATION SERVERS

1996

Logging
Load Balancing
Java
Execution Handling
Transaction Monitoring
Security
JDBC
Messaging
Session Management

Security
Scalability
Reliability
Reuse
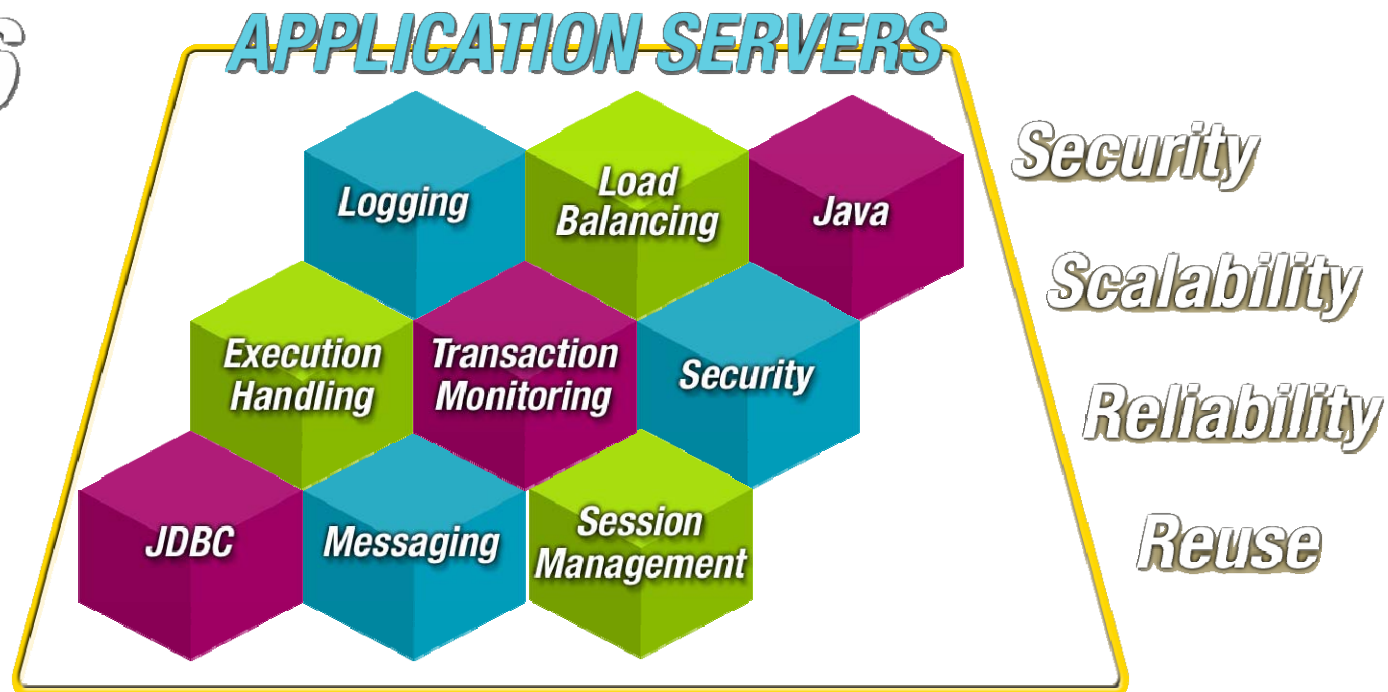
Today's Information Inflection Point
Requires a Platform Approach
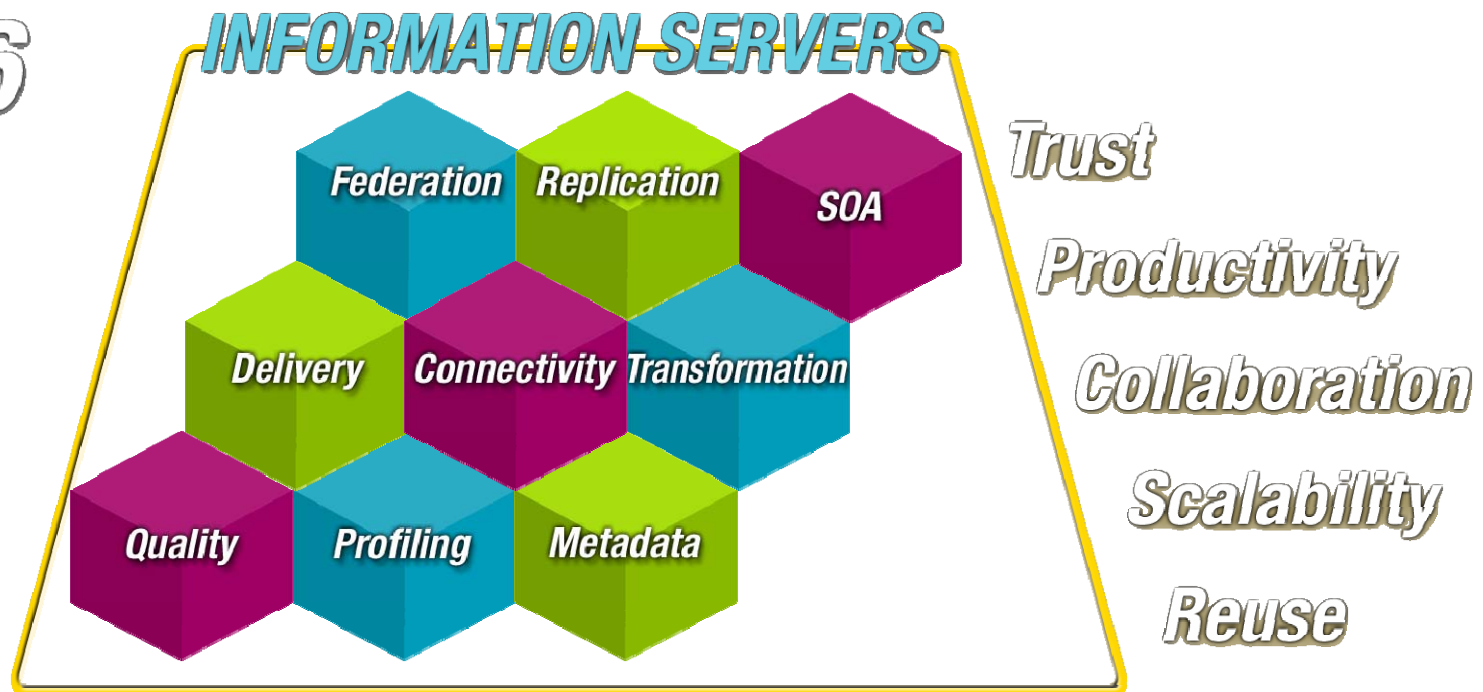
And... IBM Information Server is ready Today!
Delivering information you can trust

IBM Information Server

Unified SOA Deployment

Understand	**Cleanse**	**Transform**	**Deliver**
Discover, model, and govern information structure and content	Standardize, merge, and correct information	Combine and restructure information for new uses	Synchronize, virtualize and move information for in-line delivery

Unified Metadata Management

Parallel Processing

Rich Connectivity to Applications, Data, and Content

IBM Information Server
A "z" work in progress

IBM Information Server

Unified SOA Deployment

Understand	**Cleanse**	**Transform**	**Deliver**
WebSphere Information Analyzer	WebSphere QualityStage	WebSphere DataStage	WebSphere Classic Fed. Server
WebSphere Business Glossary	WebSphere QualityStage z/OS	WebSphere DataStage for z/OS	WebSphere Data Even Publisher
Rational Data Architect		WebSphere DataStage MVS	WebSphere Classic Replication
			WebSphere Replication Server

Unified Metadata Management

Parallel Processing

Rich Connectivity to Applications, Data, and Content

IBM Information Server Architecture

UNIFIED USER INTERFACE

Analysis Interface
Development Interface
Web Admin Interface

COMMON SERVICES

| Metadata Services | Unified Service Deployment | Security Services | Logging & Reporting Services |

UNIFIED PARALLEL PROCESSING

Understand
Cleanse
Transform
Deliver

UNIFIED METADATA

Design   Operational

COMMON CONNECTIVITY

Structured, Unstructured, Applications, Mainframe

Three-tier implementation:
- Design on Wintel
- Mid-tier Server for metadata & scheduling
- Deployment platforms are USS and/or distributed Linux, Unix and Windows

Information Server and System z –
What do we have today?

System z-centric ETL

- WebSphere DataStage for z/OS – 3-tier Enterprise ETL
  - Leverages USS as an ETL operational environment
- WebSphere DataStage MVS
  - z-centric ETL using a COBOL Code generator

Robust System z data connectivity and integration

- WebSphere Classic Federation
  - Read and write SQL access to z/OS data from Linux, Unix and Windows
  - Direct connectivity with VSAM, IMS, CA-IDMS, Adabas, Datacom, & sequential files
- WebSphere Data Event Publishers
  - Changed-data capture and publishing for DB2, VSAM, IMS, CA-IDMS and Adabas
  - Drive change-only updating for data warehousing
  - Drive business integration using data events

TAKE BACK CONTROL

Today's DataStage z/OS v7.5.2 & DataStage MVS v8
Multi-platform Solutions

Windows

DS

DataStage Client

DS MVS

COBOL, JCL

upload

DataStage MVS v8.0.1
- Design: ▸ Windows
- Deploy: ▸ COBOL apps

DataStage z/OS v7.5.2
- Design: ▸ Windows
- Server: ▸ Mid-tier
- Deploy: ▸ Mid-tier
  - ▸ z/OS (USS)

Unix, Linux, Wintel

Metadata

DataStage Server

DSEngine

PX Engine

Non-z/OS Connectors

z/OS Connectors

DSMVS Jobs

Run Time Library DSMVS

z/OS

SFTP or Batch Pipes

DB2

PX Engine USS

z/OS Operators

VSAM

SEQ.

IMS

SQL

Classic Federation Server

IMS   SEQ.   VSAM   IDMS   Adabas   Datacom

WebSphere MQ

Data Event Publisher

TAKE BACK CONTROL

What are the challenges?

Challenges

- Known issues with DataStage for z/OS v7.5.2
  - Can be MIPs intensive……this translates to costly
  - Resource restrictions such as number and size of jobs on USS
  - z/OS job management and balancing can be complex and challenging
  - Requires a non-z mid-tier
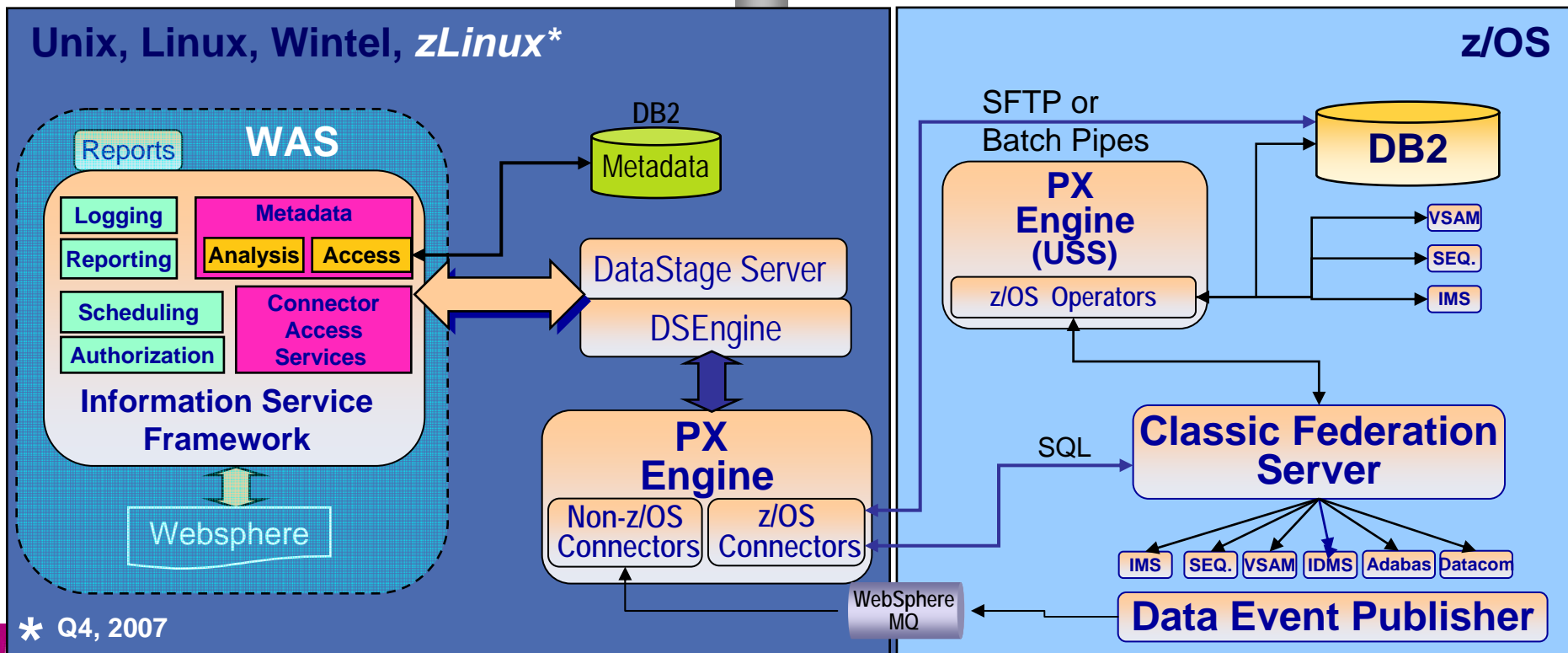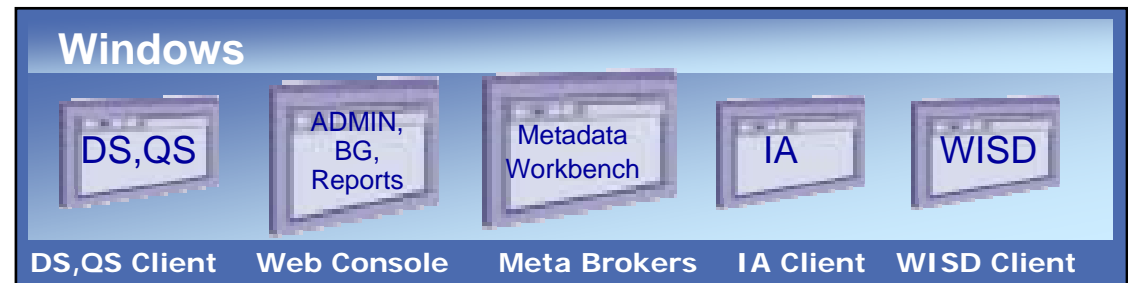
How do we address them?

Solutions

- Expand & improve our System z capabilities by delivering a "z" mid-tier to:
  - Keep System z data processing on System z
  - Provide lower-cost z-centric deployment environments
  - Improve scalability and resource utilization
  - Automate and improve resource management
  - Provide consistency with the Information Server across platforms

The IBM Information Platform Technology Model
Leveraging the power of a multi-platform Environment

- Tight Integration w/ z/OS
- zLinux mid-tier: robust, complete, z-centric
- Enhanced USS Engine

Windows

DS,QS	ADMIN, BG, Reports	Metadata Workbench	IA	WISD
DS,QS Client	Web Console	Meta Brokers	IA Client	WISD Client

Unix, Linux, Wintel, *zLinux*

WAS

Reports

Logging	Metadata
Reporting	Analysis / Access
Scheduling	Connector Access Services
Authorization	

Information Service Framework

Websphere

DB2
Metadata

DataStage Server
DSEngine

PX Engine

Non-z/OS Connectors | z/OS Connectors

WebSphere MQ

z/OS

SFTP or Batch Pipes

DB2

PX Engine (USS)

z/OS Operators

VSAM
SEQ.
IMS

SQL

Classic Federation Server

IMS | SEQ. | VSAM | IDMS | Adabas | Datacom

Data Event Publisher

∗ Q4, 2007

TAKE BACK CONTROL

Information Server on System z v8.0.1

Benefits of this hybrid architecture that leverages zLinux
- Significant cost savings:
  - z/OS MIPs consumption dramatically reduced!
    All Job Processing is on zLinux, except the z/OS data access
  - DB2 access qualifies for ZIIP specialty engine
  - MIPs charged at IFL rate … NOT z/OS rate
  - Reduced z/OS CPU minimizes impact on other z/OS software costs

- High performance z data connectivity:
  - Batch Pipes for DB2 load, DRDA to DB2 over hipersockets
  - SQL to Classic over hipersockets, Integration with Data Event Publishers

- Added Functionality:
  - Full function Information Server scheduler
  - Expanded Job Footprint: more operators, more jobs, less resource
  - QualityStage, Information Analyzer, WISD, Business Glossary, …

- Seamless integration with other Information Server platforms
  - Same operational architecture and meta data Repository
  - Eliminates deployment issues
  - Maintains value of DataStage for z/OS investments

TAKE BACK CONTROL

Summary

IBM can help with an
Information Integration Assessment

On-site workshop for your team providing:
- Information Integration needs assessment
- Report and recommendations

What is it?

An interactive session led by an IBM Information Integration Solutions specialist, to help you take advantage of emerging technologies, standards, and current trends in information integration

What will you get as a result?

- Outline of your current information integration environment
- Identification of key projects where information integration could improve IT productivity and speed development
- Exploration of alternative approaches
- Recommendations for architecture enhancements

TAKE BACK **CONTROL**

Workshop Basics

Typical Agenda

- Part One:
  Information gathering and sharing
  - Information Integration Overview
  - Discovery –
    Extensive interviews and data collection

- Part Two:
  Analysis, summary, recommendations
  - Summary of information collected
  - Review of:
    challenges, issues, problems, needs
  - Approaches considered
  - Tradeoffs
  - Recommendations

Typical Workshop Participants

- Up to 15 team members:
  for example:
  - Architects
    - Data architects, information architects
    - Enterprise architects
    - Data architects or information architects
  - LOB managers in areas like
    - Compliance (financial services and other industries)
    - Supply chain management (CPG)
    - Clinical intelligence (healthcare)
  - Application owners
  - Middleware managers
  - Data warehouse managers and developers