

Key Considerations for Mainframe Data Integration

-- One size does not fit all

Karen Durward
Information Integration Product Management

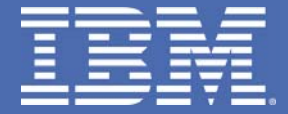


TAKE BACK CONTROL

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
- Summary
 - Sneak Peek: Information Server for System z
 - Where do you go from here?





Information on Demand ... and Data Delivery

TAKE BACK CONTROL



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



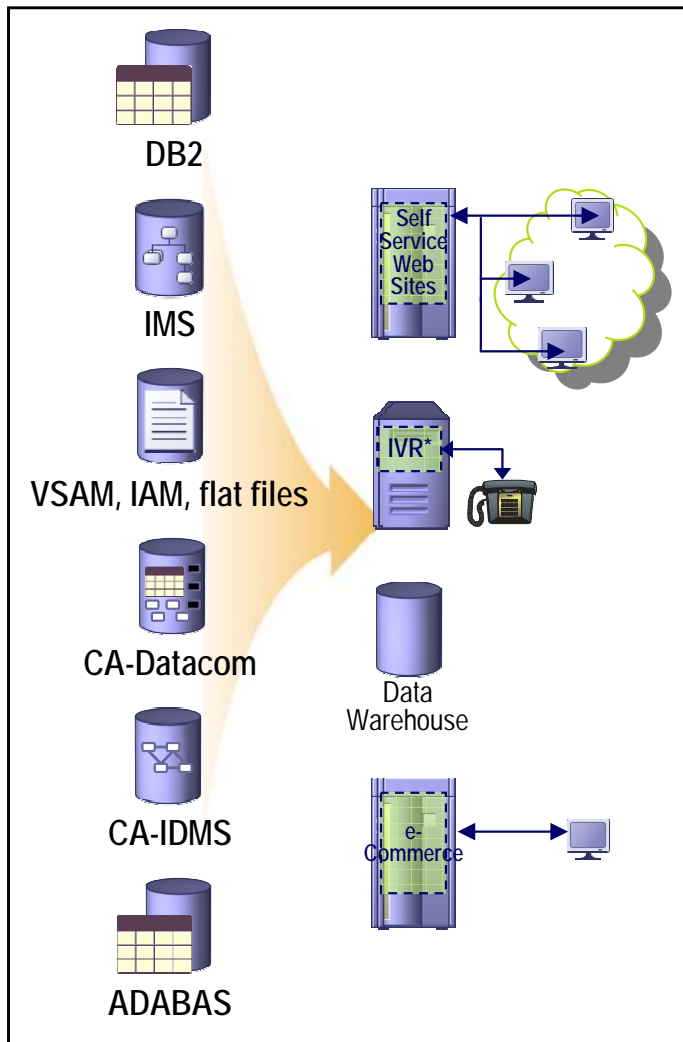
Data Integration Challenges You Face Every Day

- 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?



Mission Critical "z" Data Can Be Difficult

Over 60% of operational data is sourced from System z



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

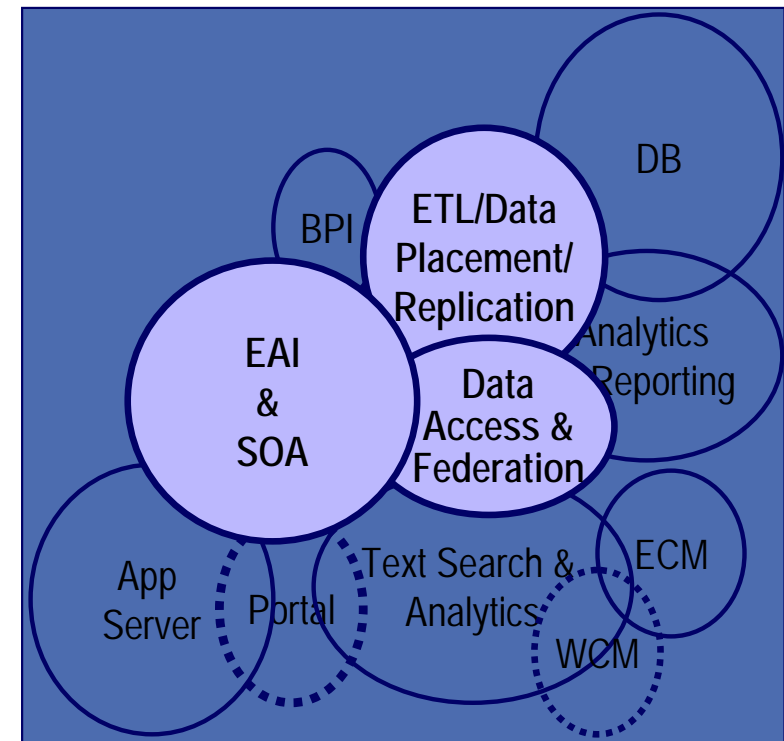
What are you doing to meet these challenges?

- *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



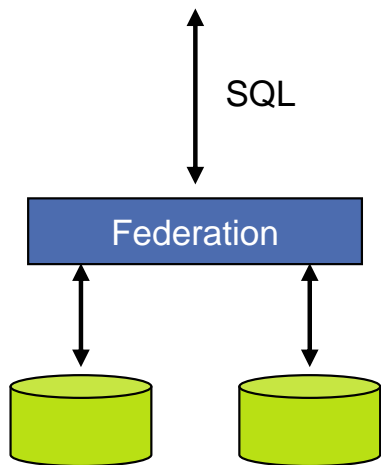
Business Integration Market Landscape

- 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

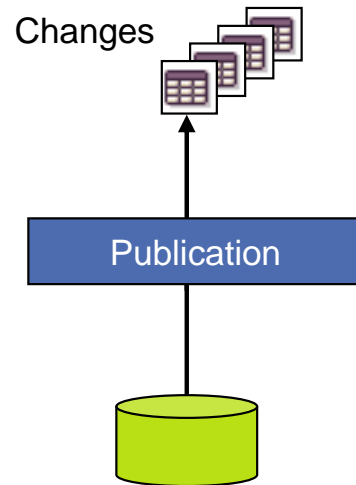


Seamlessly integrate with:

- e-commerce
- Self service

All data is an equal participant

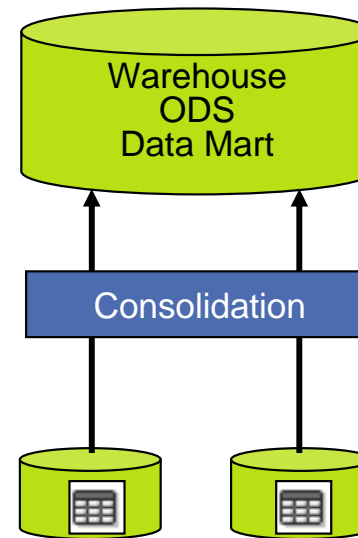
Move



Proactively deliver data to:

- Drive integration
- Enable change only updating to:
 - Reduce batch window dependencies
 - Reduce data latency

Data changes trigger action

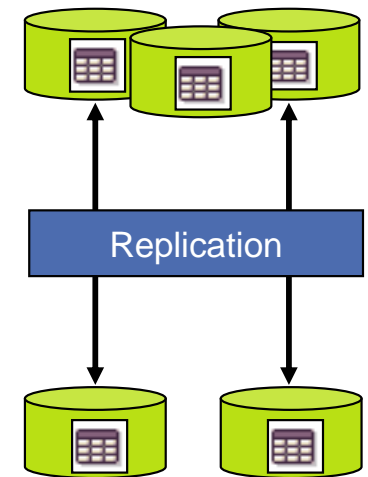


Repackage & repurpose

- Data Warehousing
- Operational Data Stores
- MDM
- CDI

Transformation rich copies

Synchronize

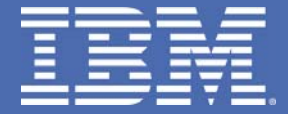


Create & synchronize copies

- Distributed operations
- High availability data
- Local copies for performance
- DBMS migrations

Like to Like copies





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

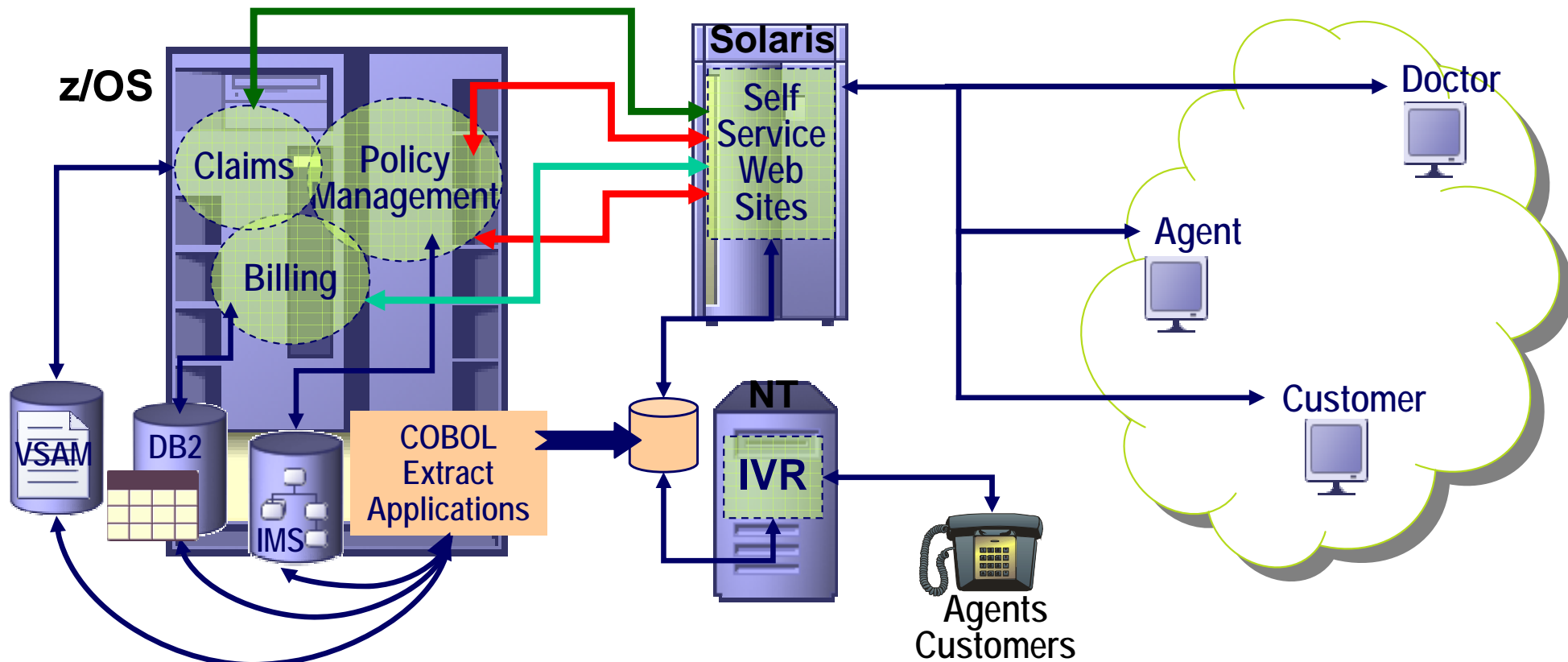


A Multi-Dimensional Self-Service Environment

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

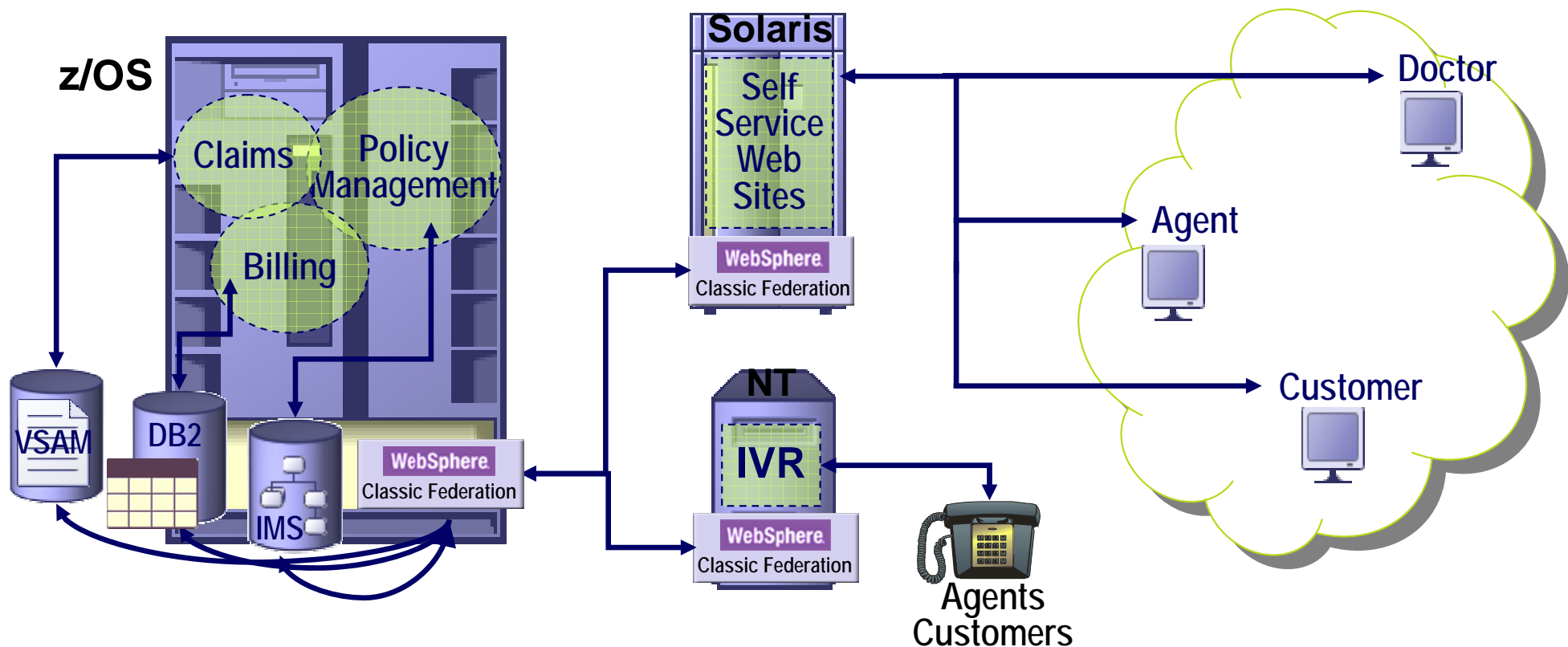


A Multi-Dimensional Self-Service Environment

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

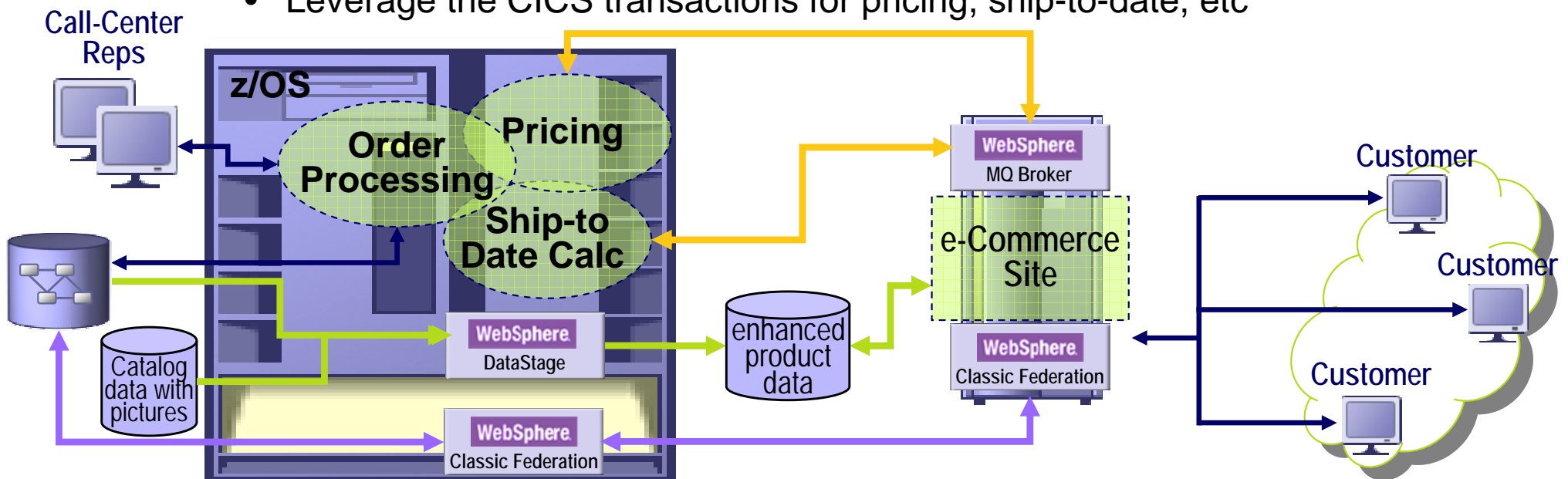
- Requirements
 - Minimal disruption on existing call-center order processing
 - Must support “shopping”, not just order taking
 - E-commerce site must be optimized for unskilled customer usage
 - Performance and accuracy are essential to business continuity and reputation



"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

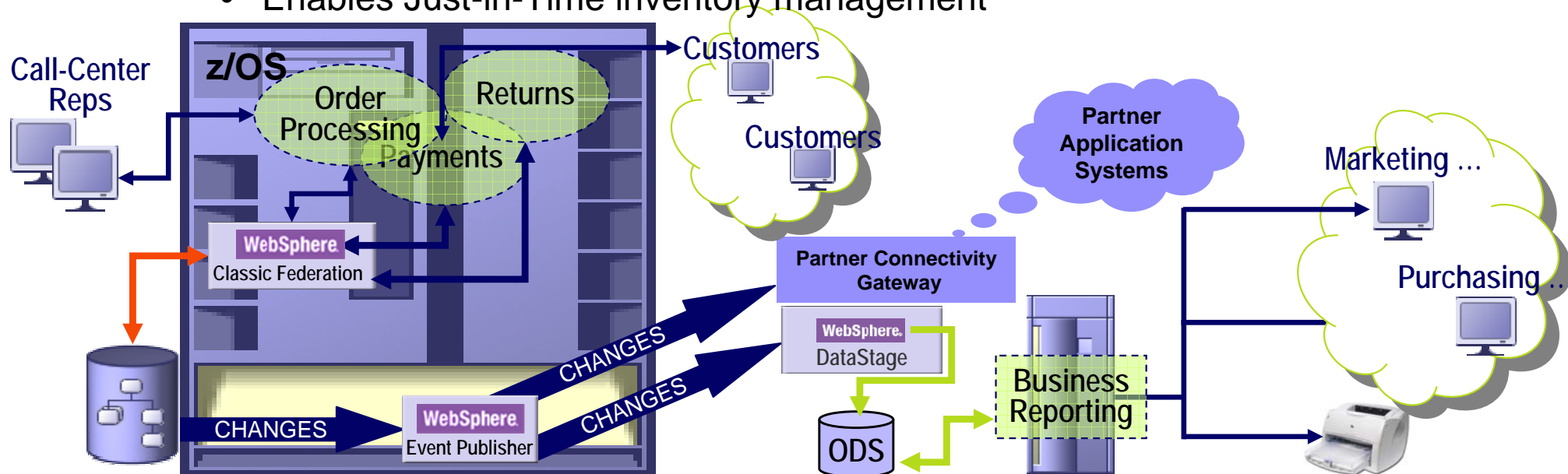
- Implement their Next-Gen transaction platform on WebSphere
 - Staying with z/OS & IMS data for performance, reliability and security
- Improve interfaces with their suppliers
 - Implement "real time" interfaces for a just-in-time inventory environment
- Leverage reduced cost of an "off shore" data 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

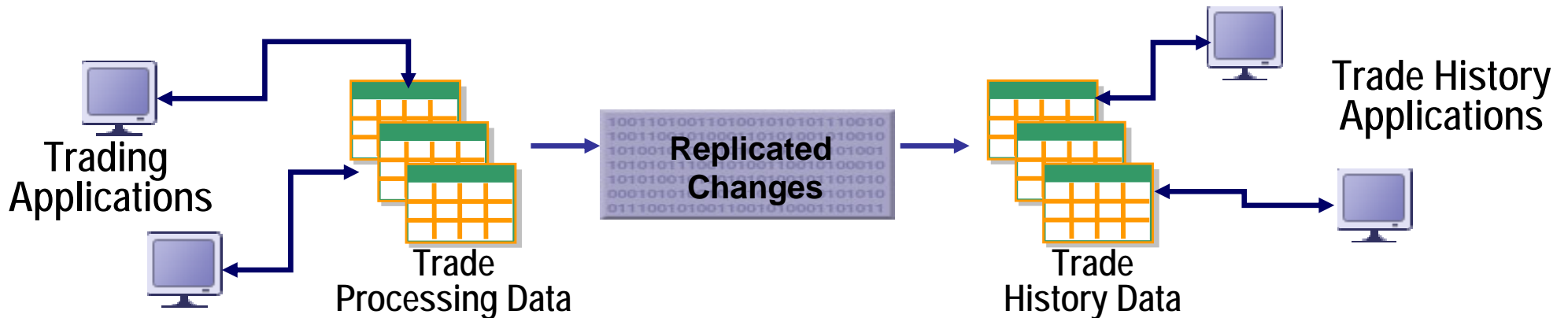
Replication is "best" because:

- *Low latency is critical*
- *No transformations are required*
- *Easy to switch source & target during a failure*



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



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

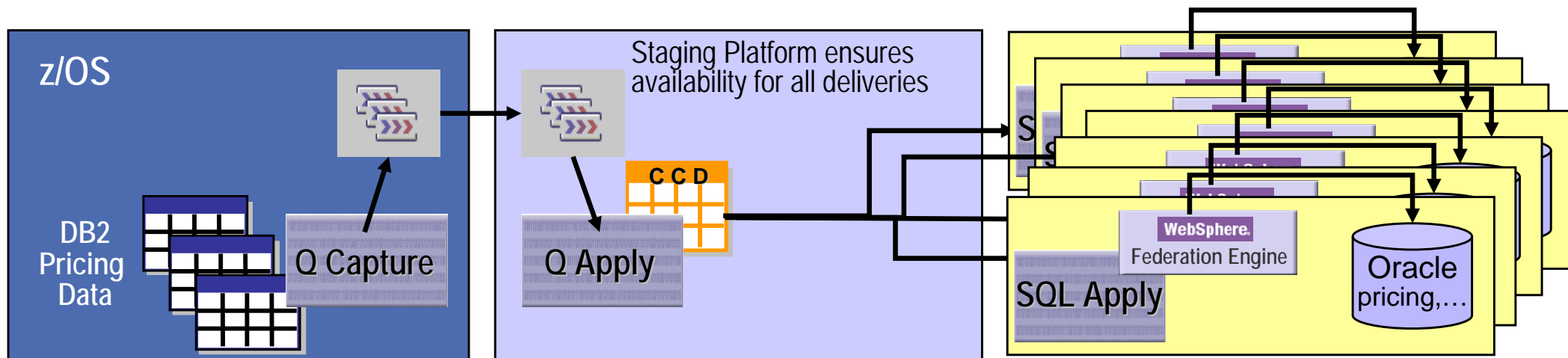
Replication is "best" because:

- *Automatic recovery with guaranteed delivery*
- *Minimal transformations are required*
- *Bidirectional capabilities inherent in Replication*



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



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*



Service-Enabling Heterogeneous Data

Financial Industry Regulatory Compliance

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

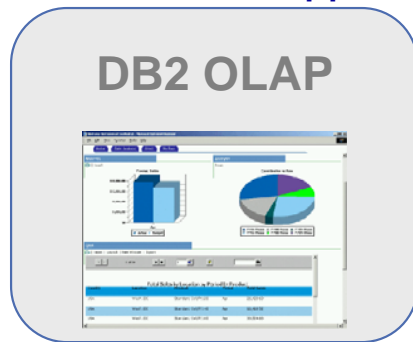
- Starting Environment
 - Information in 23 different retail systems from over 2500 branch offices
 - Multiple portals and user applications
- Requirements
 - Enhanced risk management and increased efficiency of data collection for Basel II required data
 - Ability to view data in operational systems spread across the enterprise including third party information without disrupting retail systems



Service-Enabling Heterogeneous Data

Finance: Regulatory Compliance

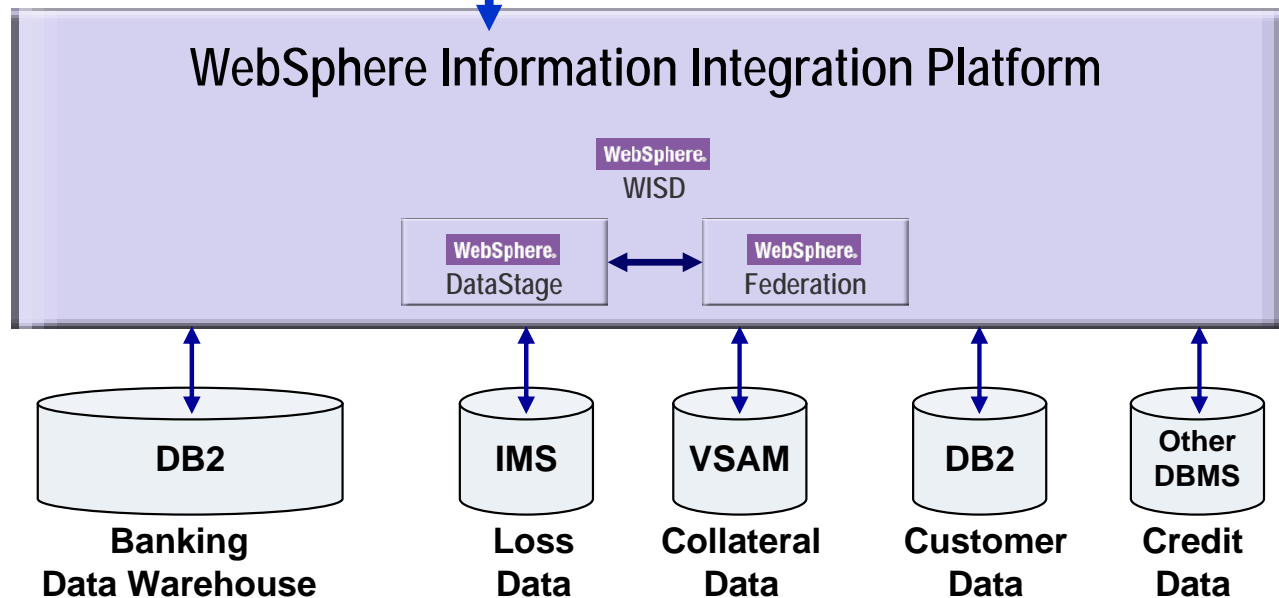
Risk measurement application

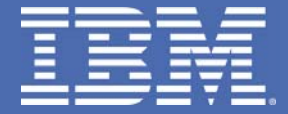


Solution – Basel II Exception Monitoring

- 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





The Role of IBM Information Server ... Why and What

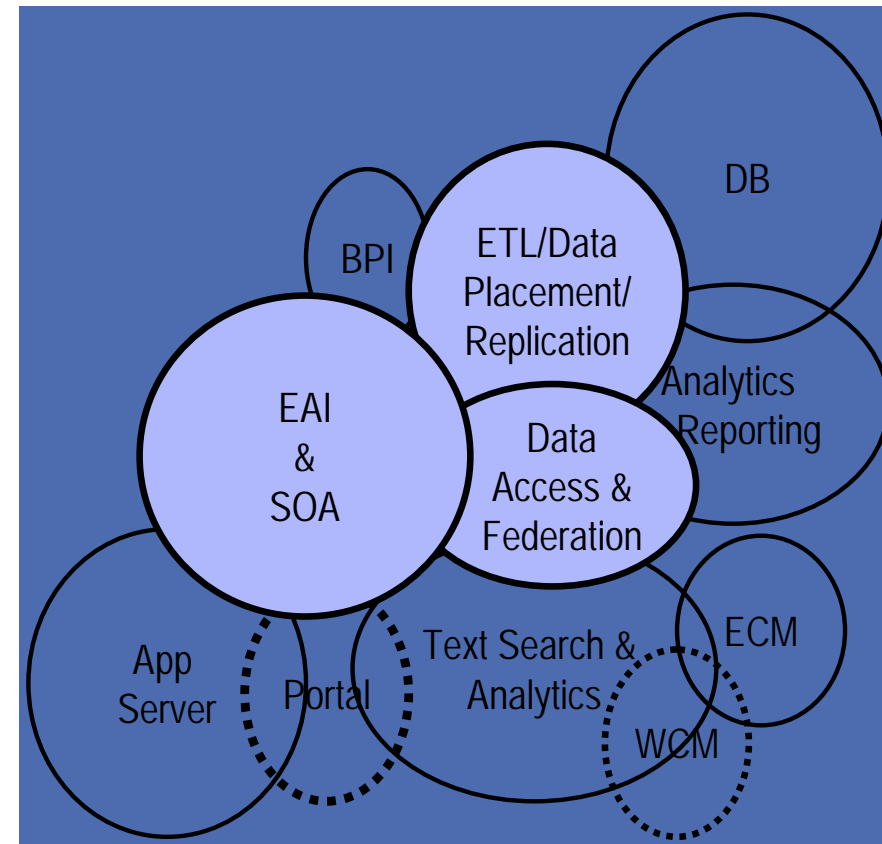
TAKE BACK CONTROL



Market Dynamics are Constantly Changing

Combinations are the norm

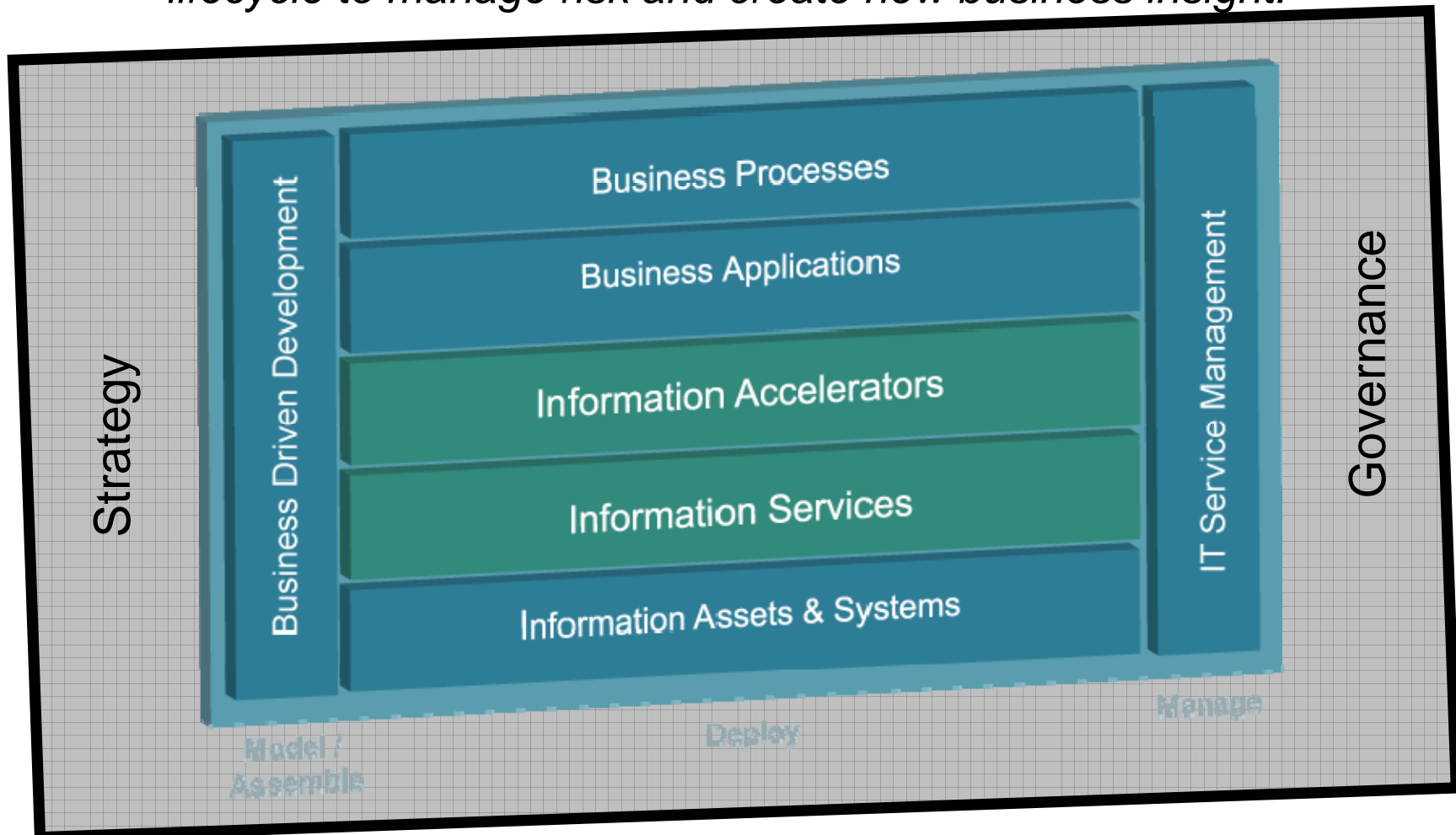
- Alphabet Soup
 - Numerous “distinct” technologies
 - Significant overlap
- Convergence is the norm
 - Vendors adding broader capabilities to their infrastructure
 - App Server growth in EAI
 - Information Integration is a marriage of many:
 - ETL
 - Federation
 - Replication
 - Publishing
- SOA is impacting the landscape
 - Use an ETL process as a service
 - Leverage Federation within a service
 - Use data events to drive execution of a service
 - Incorporate a “report” as a service



Recognize that ...

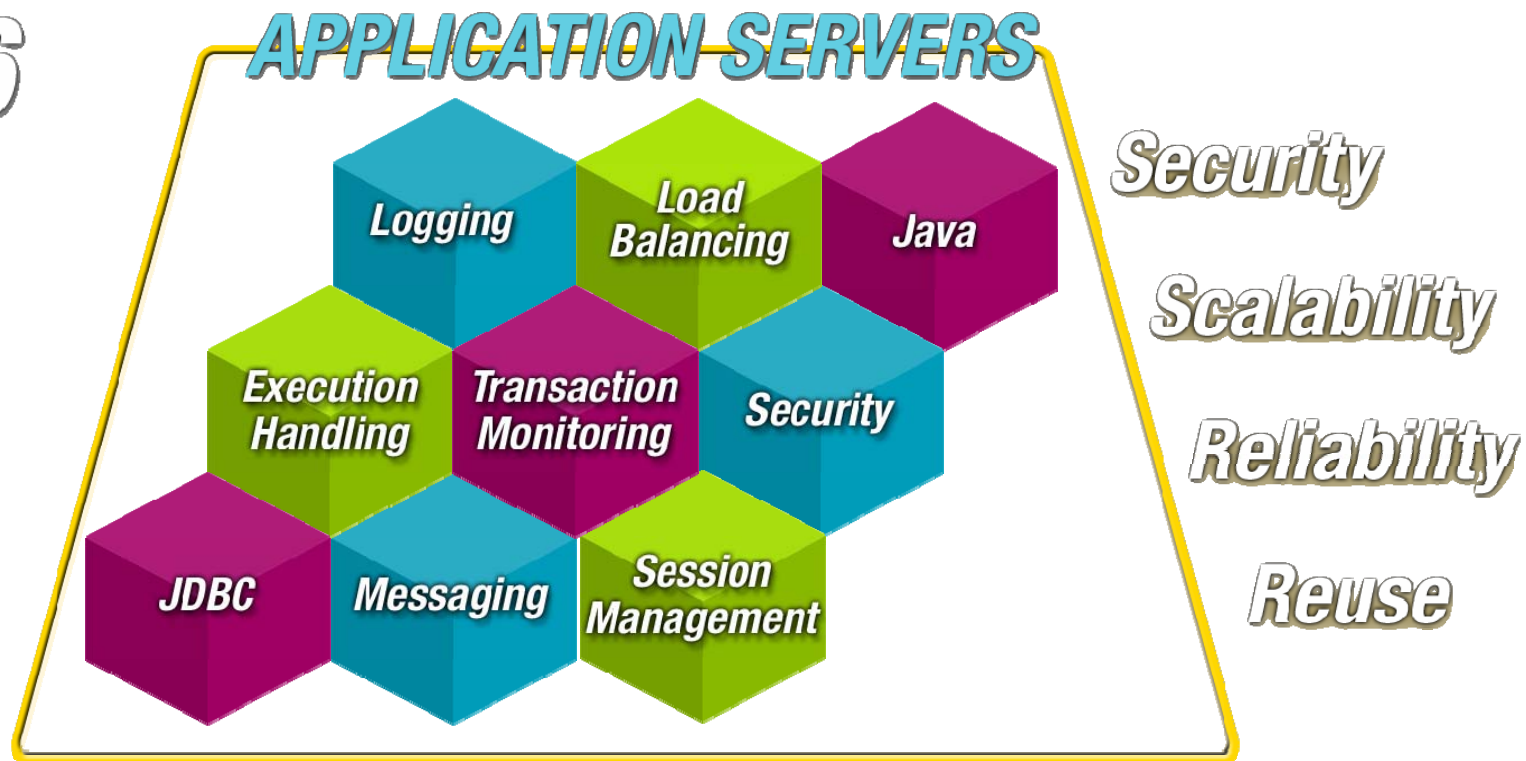
Information On Demand Requires an Eco-system of Services

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



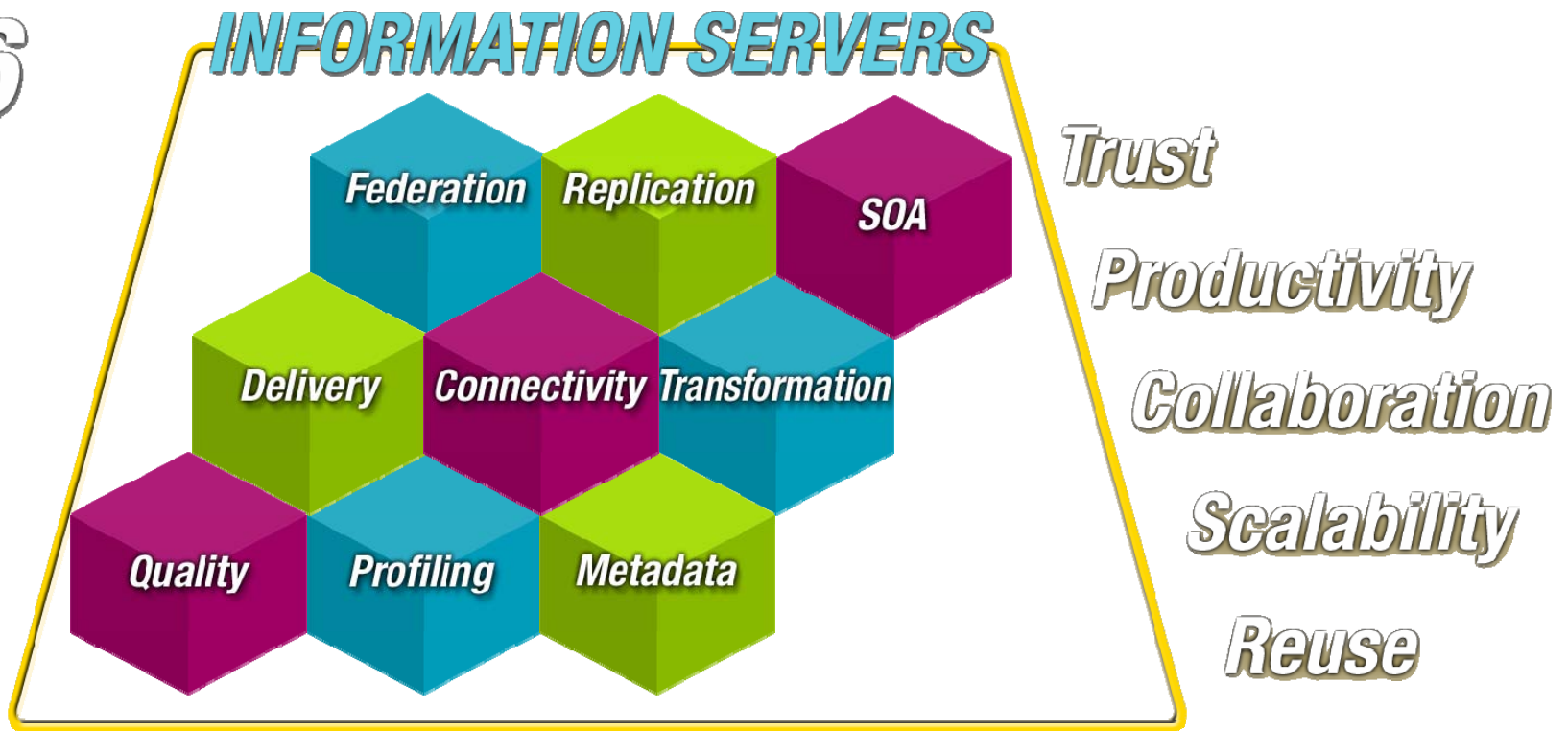
And, like an Historic Process Inflection Point *That drove the creation of the Application Server ...*

1996



Today's Information Inflection Point *Requires a Platform Approach*

2006



And... IBM Information Server is ready Today!

Delivering information you can trust

IBM Information Server

Unified SOA Deployment

Understand



Discover, model, and govern information structure and content

Cleanse



Standardize, merge, and correct information

Transform



Combine and restructure information for new uses

Deliver



Synchronize, virtualize and move information for in-line delivery

Unified Metadata Management

Parallel Processing

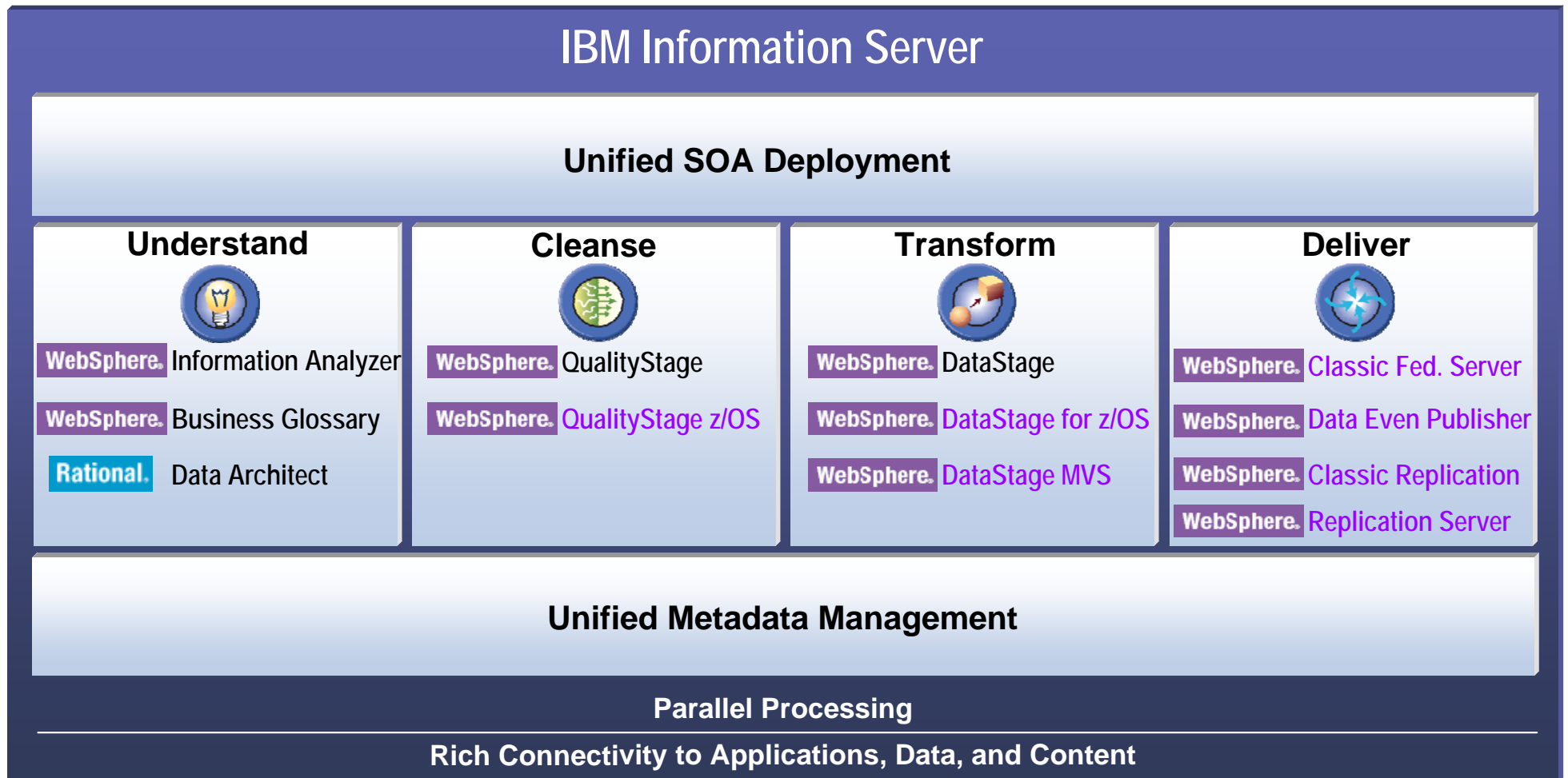
Rich Connectivity to Applications, Data, and Content



TAKE BACK CONTROL

IBM Information Server

A "z" work in progress





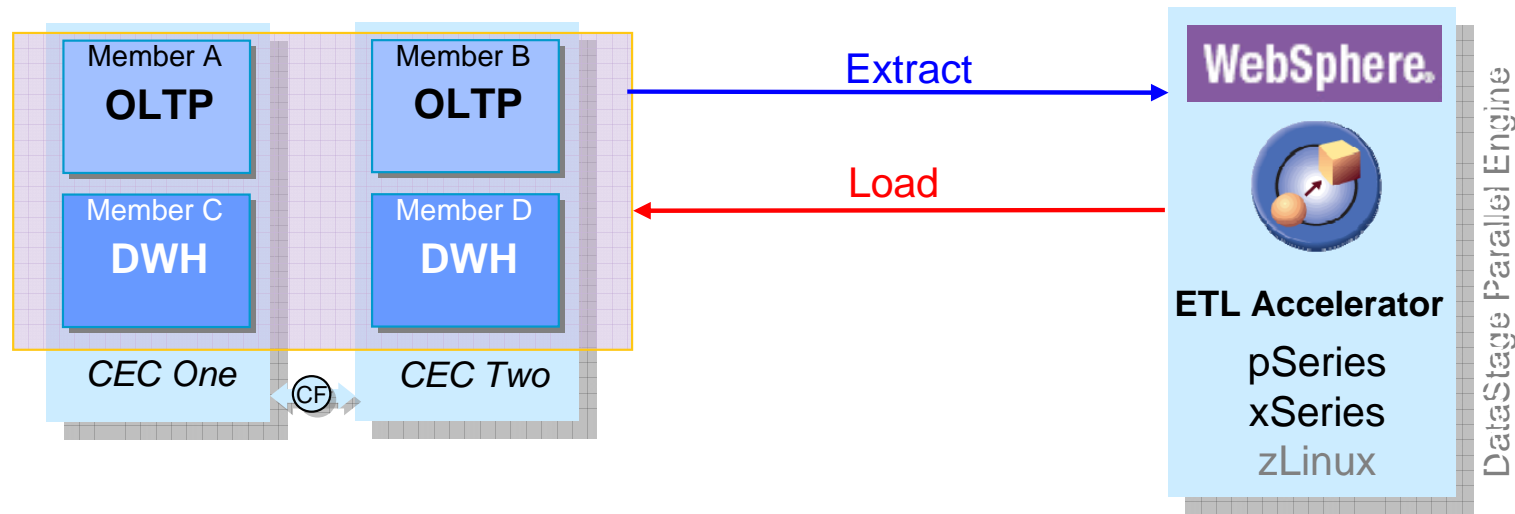
IBM Information Server at Work ... Managing a Dynamic Data Warehouse

TAKE BACK CONTROL



Managing the Dynamic Data Warehouse Content

ETL Solution Accelerates with Automation & Reuse



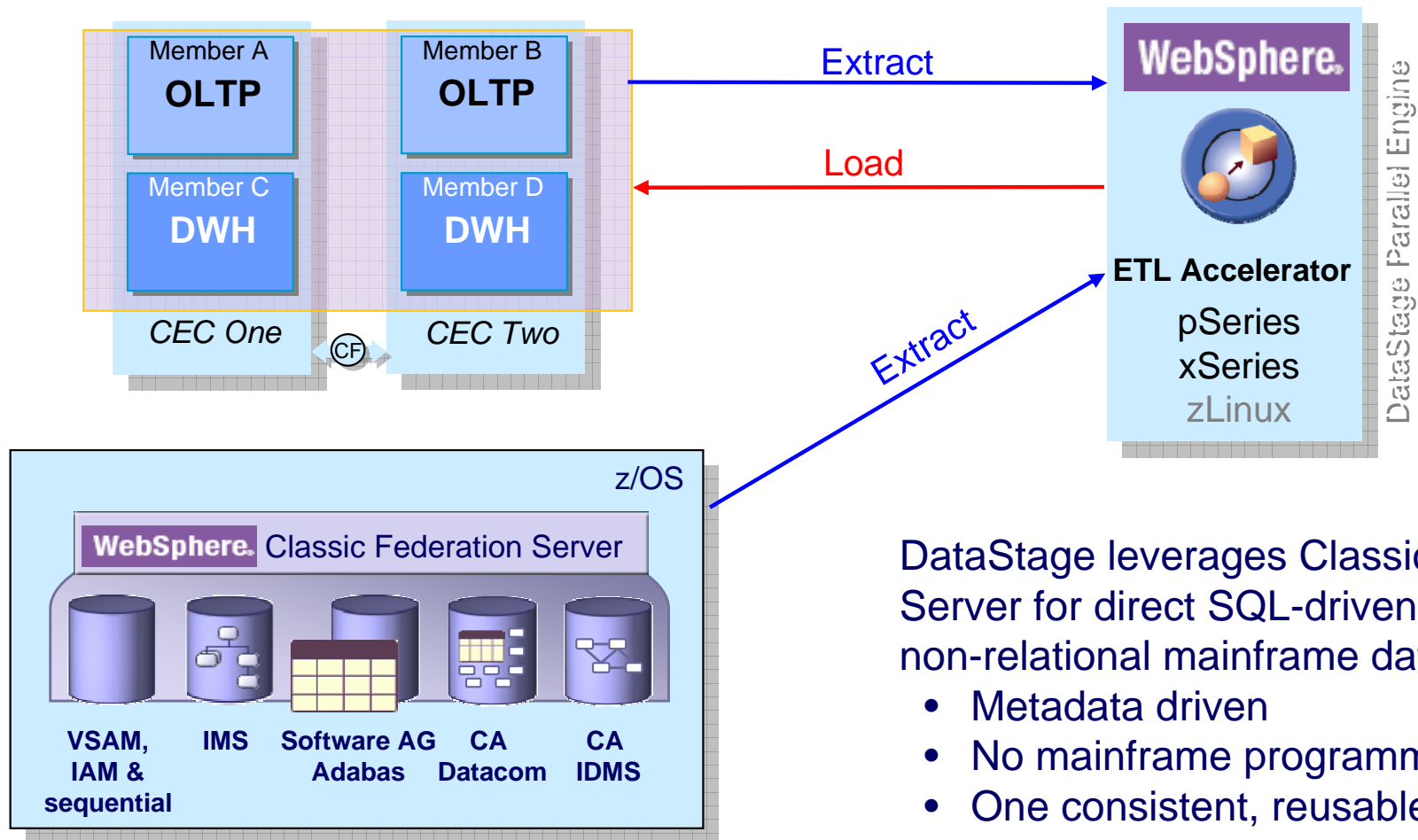
ETL is done using an ETL Accelerator, **WebSphere DataStage**.

- **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 back into the warehouse tables



Managing the Dynamic Data Warehouse Content

Non-Relational Mainframe Data is an Equal Participant



DataStage leverages Classic Federation Server for direct SQL-driven extraction of non-relational mainframe data sources

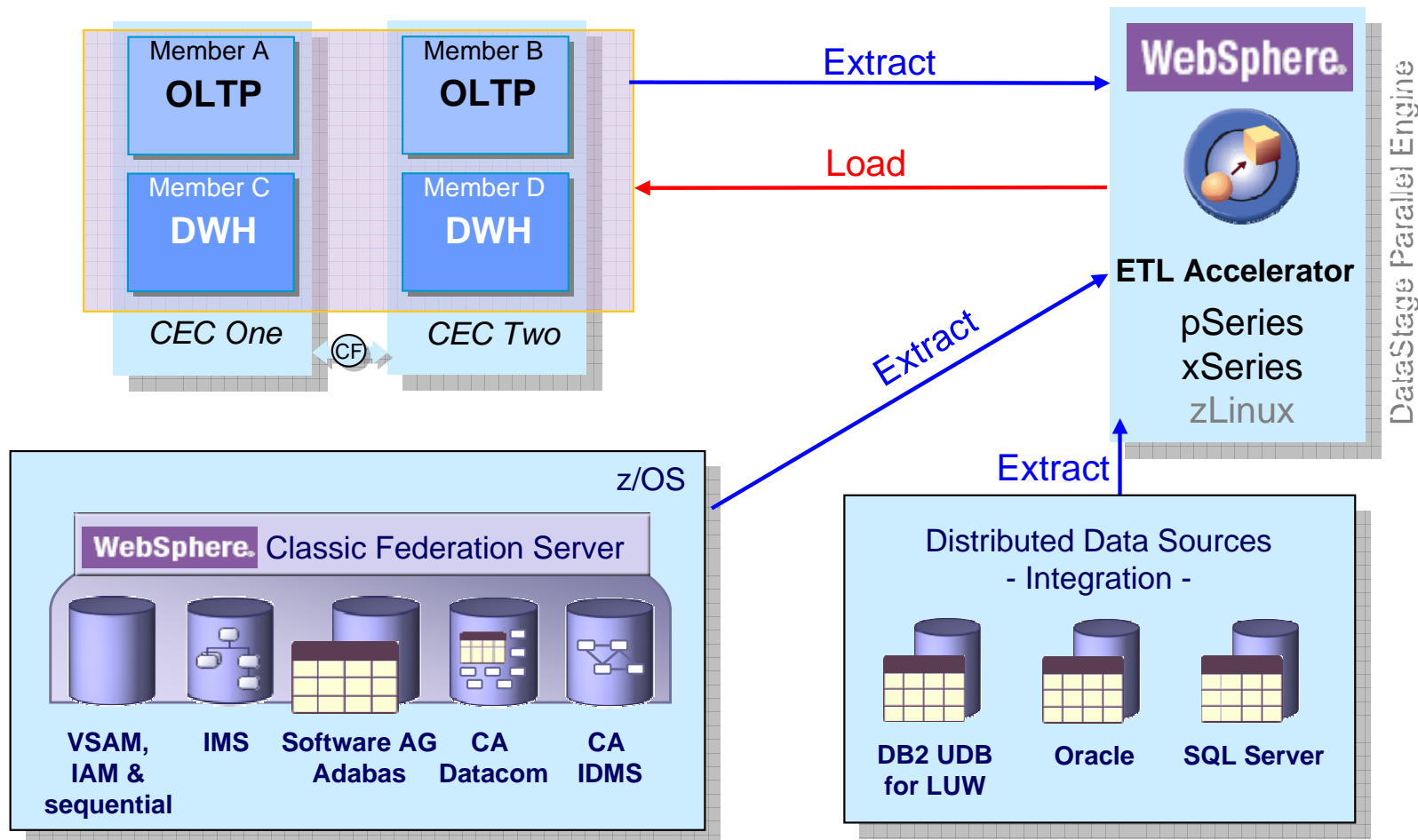
- Metadata driven
- No mainframe programming
- One consistent, reusable interface

"cut development time in half"



Managing the Dynamic Data Warehouse Content

Distributed Data Integrated via Native Connectors

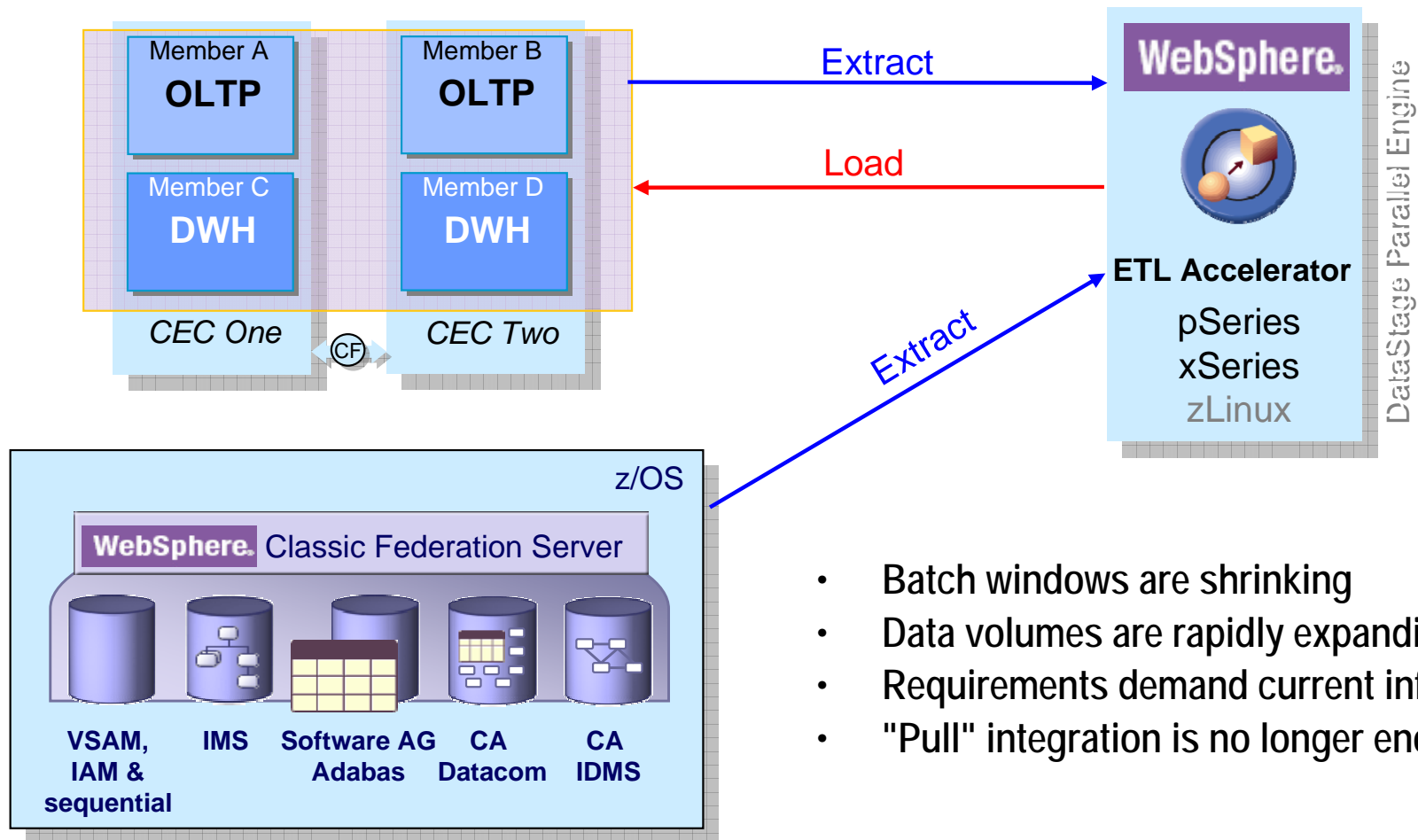


Distributed data sources are integrated directly through the IBM Information Server.



Managing the Dynamic Data Warehouse Content

Distributed Data Integrated via Native Connectors

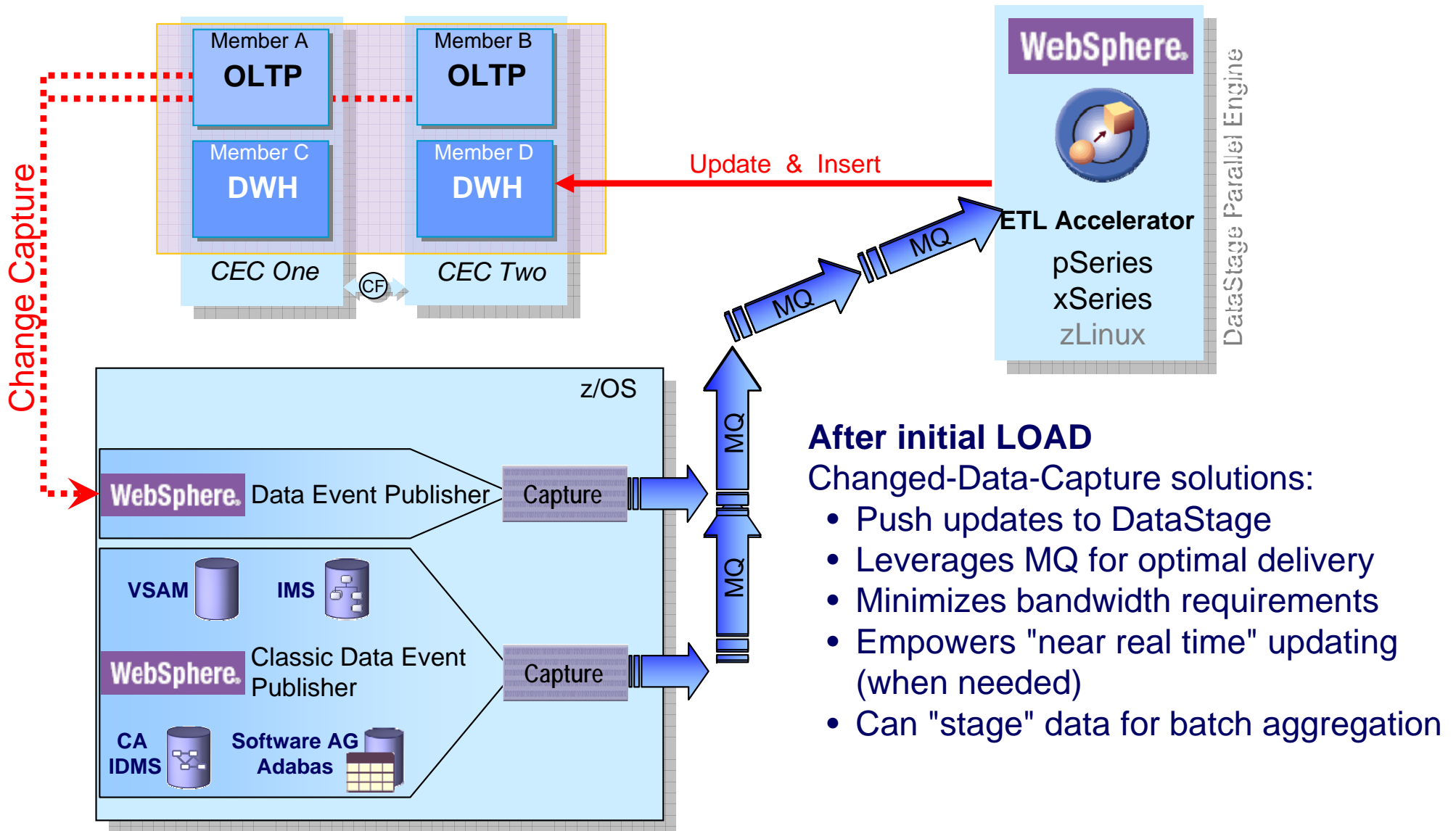


- Batch windows are shrinking
- Data volumes are rapidly expanding
- Requirements demand current information
- "Pull" integration is no longer enough



Maintenance of the Dynamic Data Warehouse

Incremental Updating is Essential



After initial LOAD

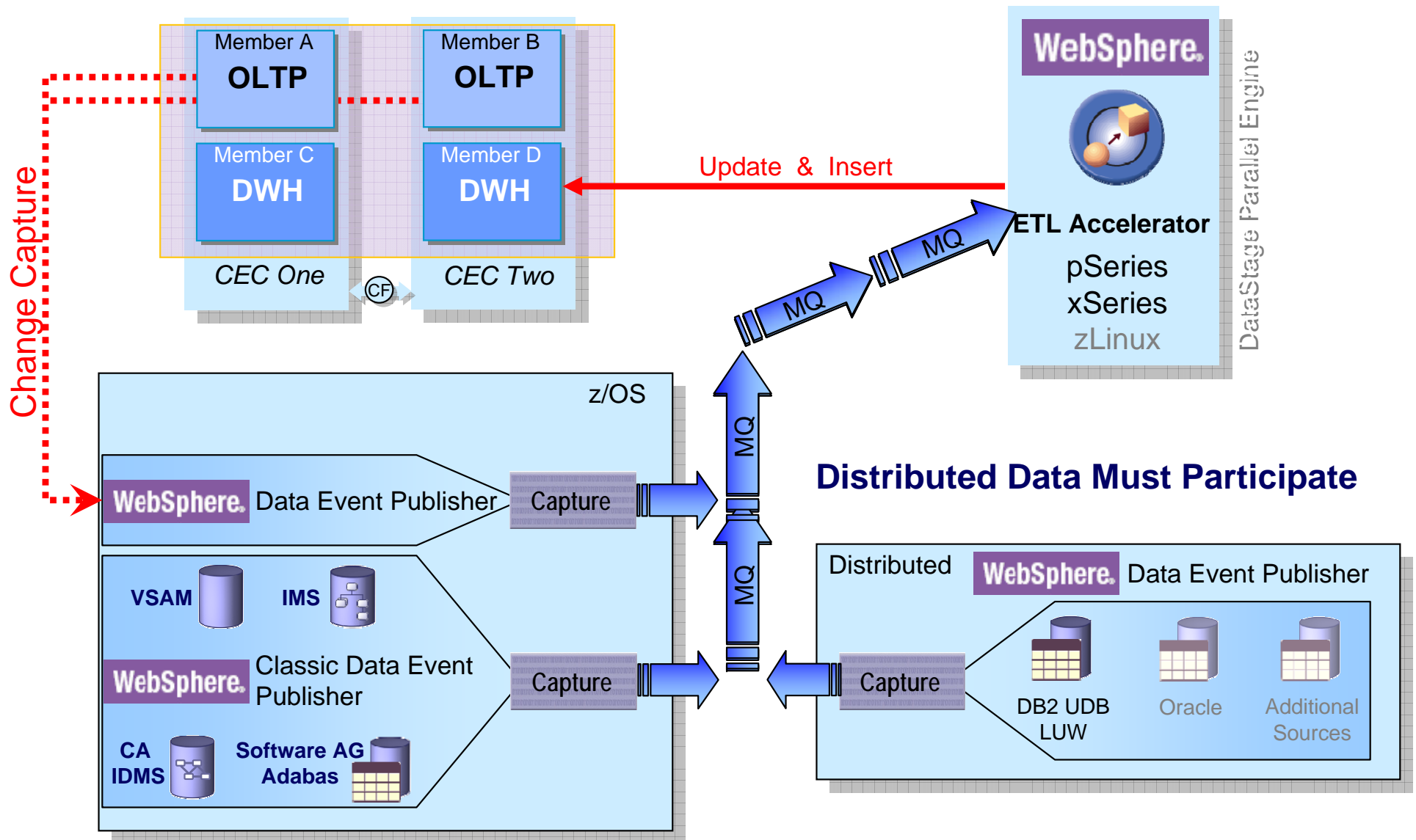
Changed-Data-Capture solutions:

- Push updates to DataStage
- Leverages MQ for optimal delivery
- Minimizes bandwidth requirements
- Empowers "near real time" updating (when needed)
- Can "stage" data for batch aggregation



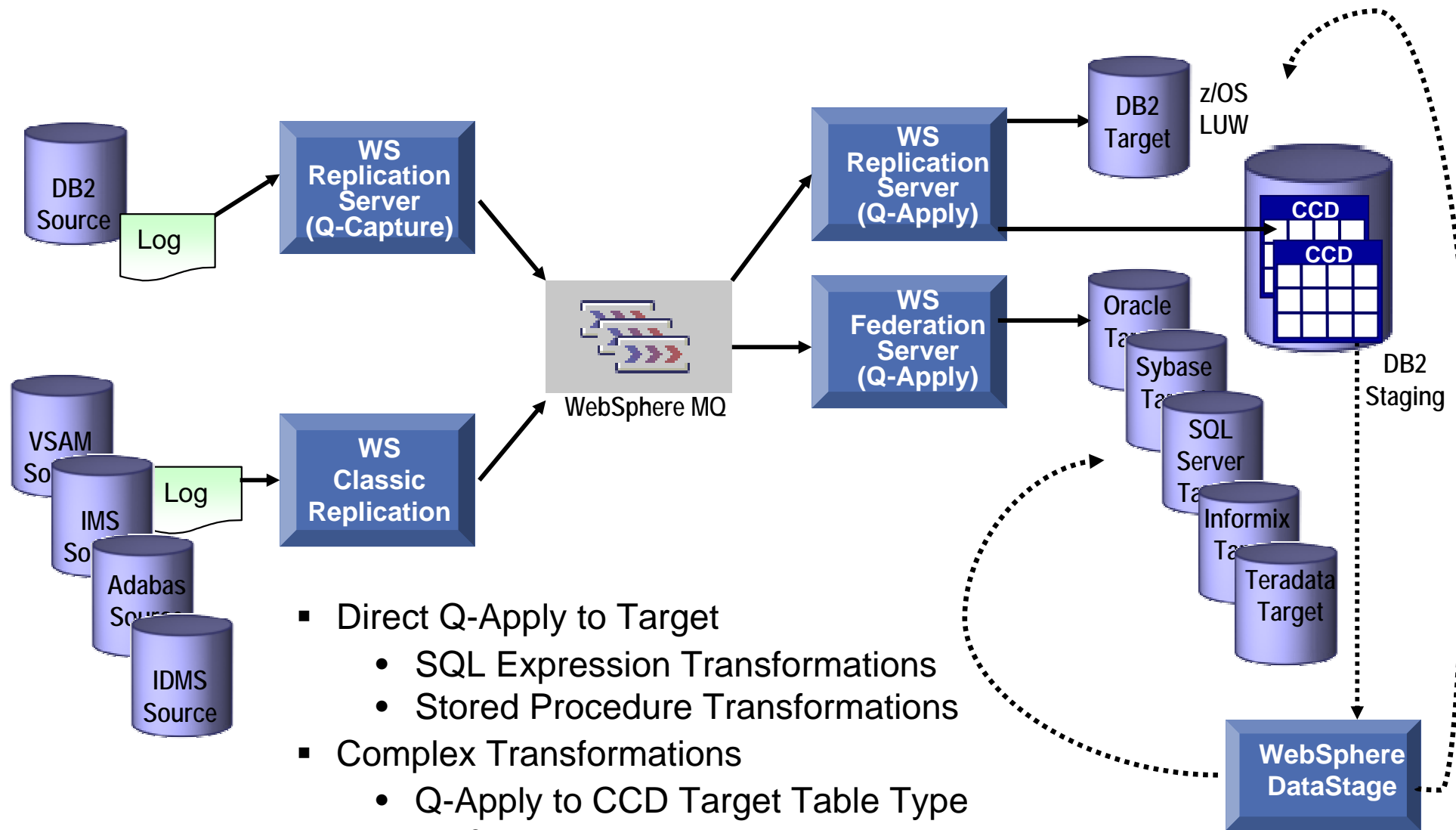
Maintenance of the Dynamic Data Warehouse

Incremental Updating is Essential



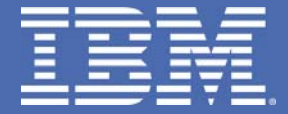
Distributed Data Must Participate

Replication Complements an ETL Architecture



- Direct Q-Apply to Target
 - SQL Expression Transformations
 - Stored Procedure Transformations
- Complex Transformations
 - Q-Apply to CCD Target Table Type &
 - WebSphere DataStage with transformations to target database(s)



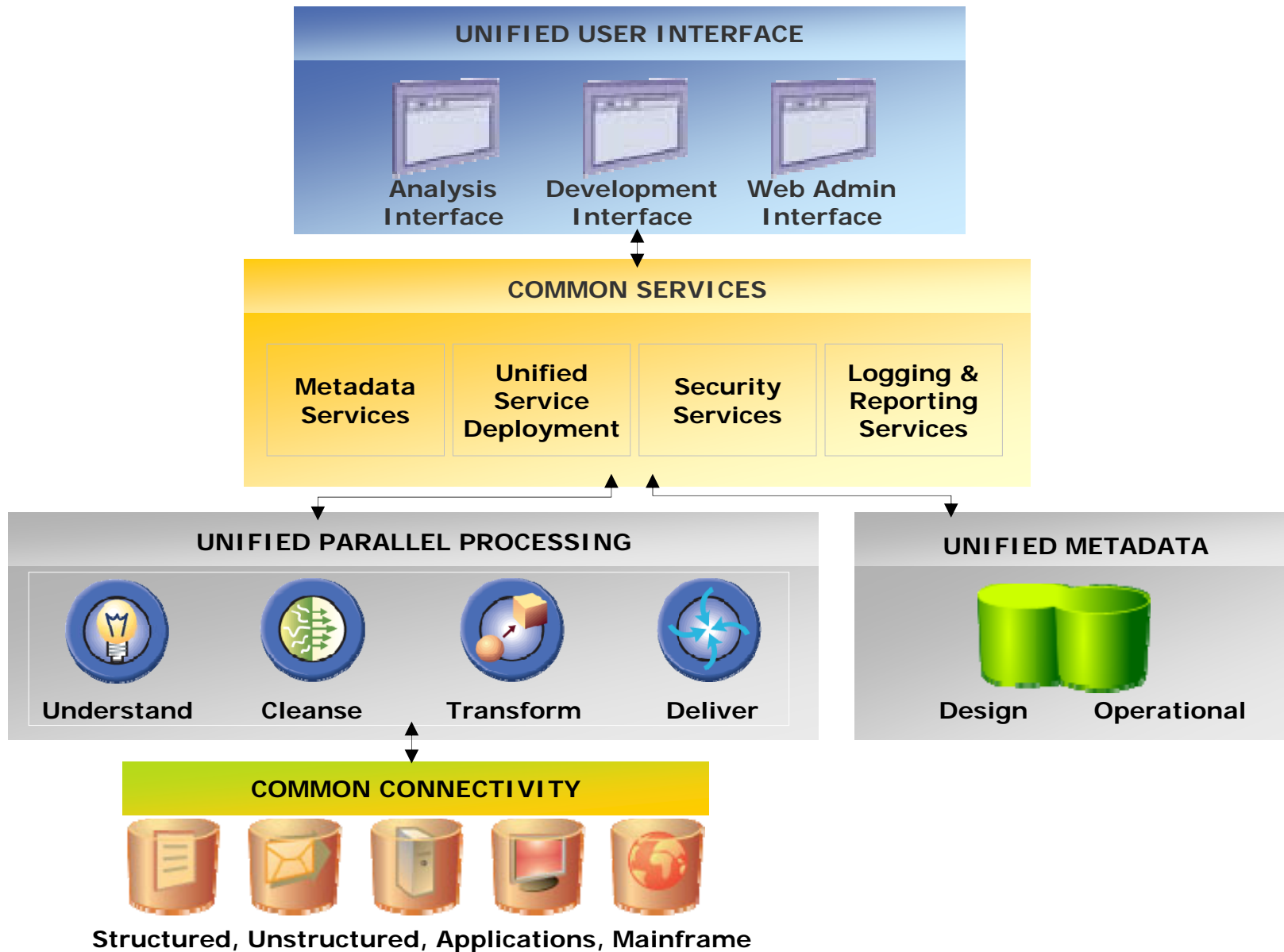


IBM Information Server ... How

TAKE BACK CONTROL



IBM Information Server Architecture



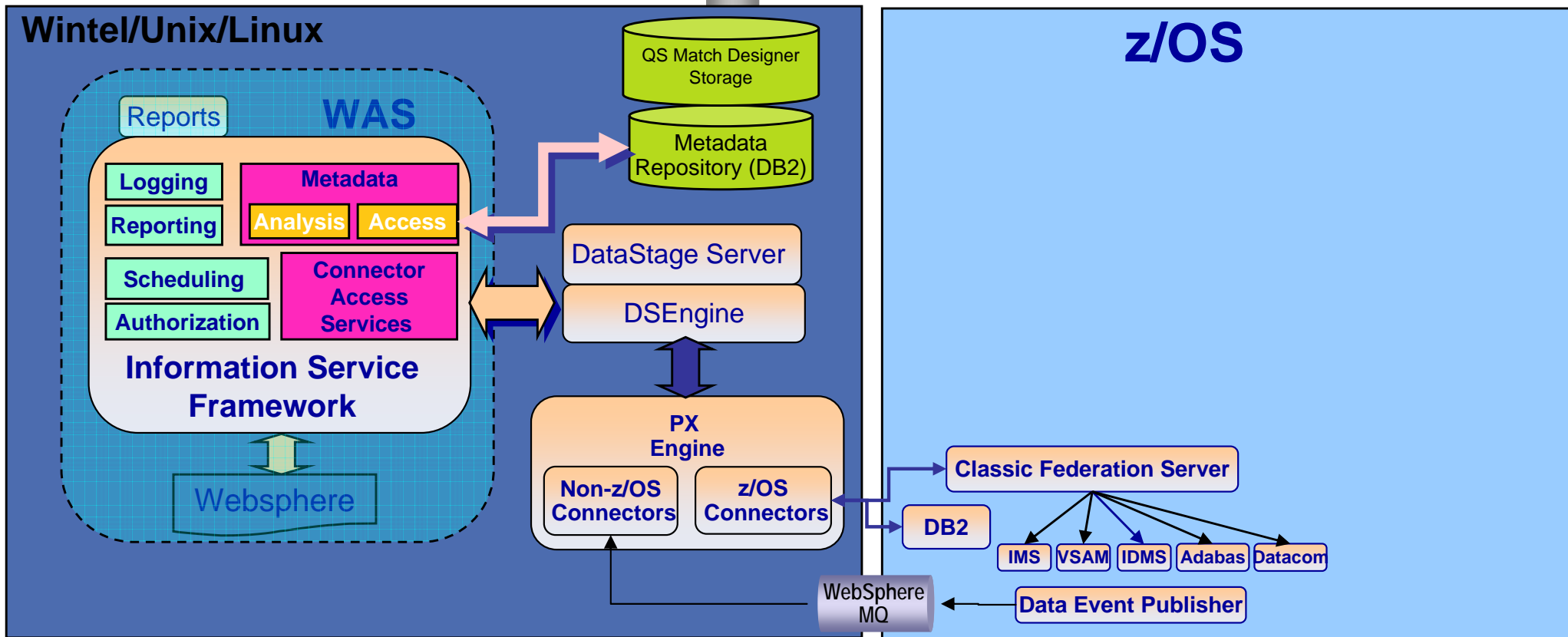
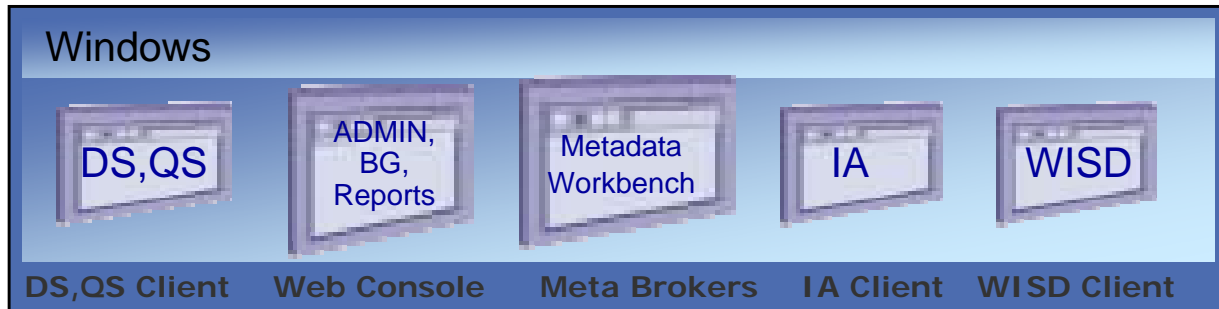
Architectural Considerations

- Design Platform
 - Leverages distributed GUI Environments
 - Interacts with WAS-based Platform Infrastructure
 - Metadata repository uses RDBMS (default is DB2)
- Operational Platforms
 - Parallel processing engine for MIPS intensive activity
 - Wintel/Unix/distributed Linux today
 - DataStage "scheduling" engine
 - Connectivity with System z Data Delivery components as well as a broad spectrum of other data sources



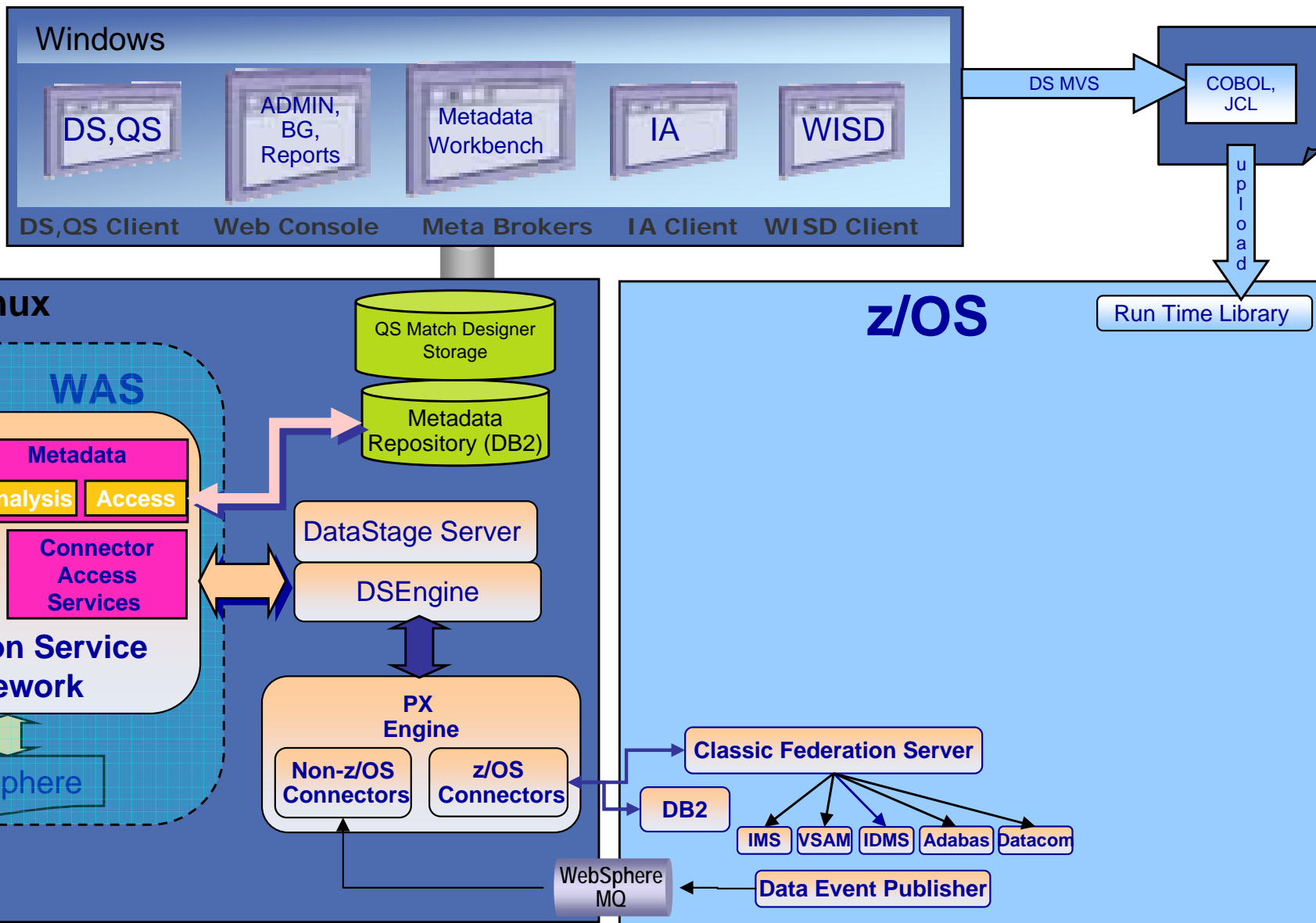
Information Server Implementation

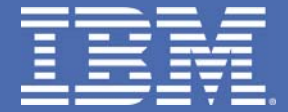
Today



Information Server Implementation

Today





Information Server and System z

Technology Preview



TAKE BACK CONTROL

What do System z Users Need?

- z-Centric Operational Platform

- Keep z-data on z

Do not move System z data off platform for Information Server functions – leverages reliability, scalability and security of z

- Leverage DB2 on z/OS for:

- Information Server Repository
- QualityStage Metadata
- Primary target for Transformation

- Optimize price performance by:

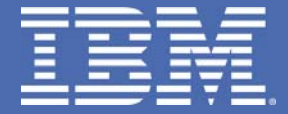
- Leveraging specialty engines:
 - zLinux IFLs, zIIP, zAAP



Information Server Implementation

Proposed System z





Summary

TAKE BACK CONTROL



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



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



THANK YOU



TAKE BACK CONTROL