

InfoSphereTM software

Trusted Information



Information Management software

Best Practices in Data Warehouse loading and synchronization with System z data - A US State Govt. Case Study

March 10, 2009

Agenda

- **Customer Story**
 - Background
 - Problem
 - Solution
 - Benefits

- **Technical Solution**
 - Solution overview
 - Product overviews

Customer Story

Background

- The Judiciary Department in an US State Government wanted to provide up-to-date court case information to their end users including clerks, commissioners, clearinghouses, and general public
- Operational data is stored in IMS on z/OS
- Data warehouse (DWH) in Oracle on AIX serving web access applications
- IMS data is constantly being updated throughout the day
- Batch updates are also done during evenings and weekends



The Problem

- Daily batch loads to Oracle led to 24-hour data latency
- Customer's end users (commissioners, legal data clearinghouses, general public) demanded more current case information
- Customer needed a solution to transport operational data to data warehouse in near real-time
- Solution needed to facilitate a seamless integration with customer's existing systems

The Solution

- Provide a near real-time solution using IBM's InfoSphere suite of products:
 - Classic Data Event Publisher for z/OS
 - Classic Federation Server for z/OS
 - DataStage
- Classic Data Event Publisher/Classic Federation used to capture data changes and access IMS
- WebSphere MQ is being used to transport captured IMS changes from z/OS to the DataStage server on AIX
- InfoSphere DataStage used to transform data and populate the Oracle DWH

The Benefits

- Tight integration with existing architecture
- GUI-based technology with no or minimal programming requirements for ongoing maintenance
- Transmission of both CICS online and batch updates to Oracle
- Near real-time data currency provided to end users via customer's web portal application, Case Search



Technical Solution

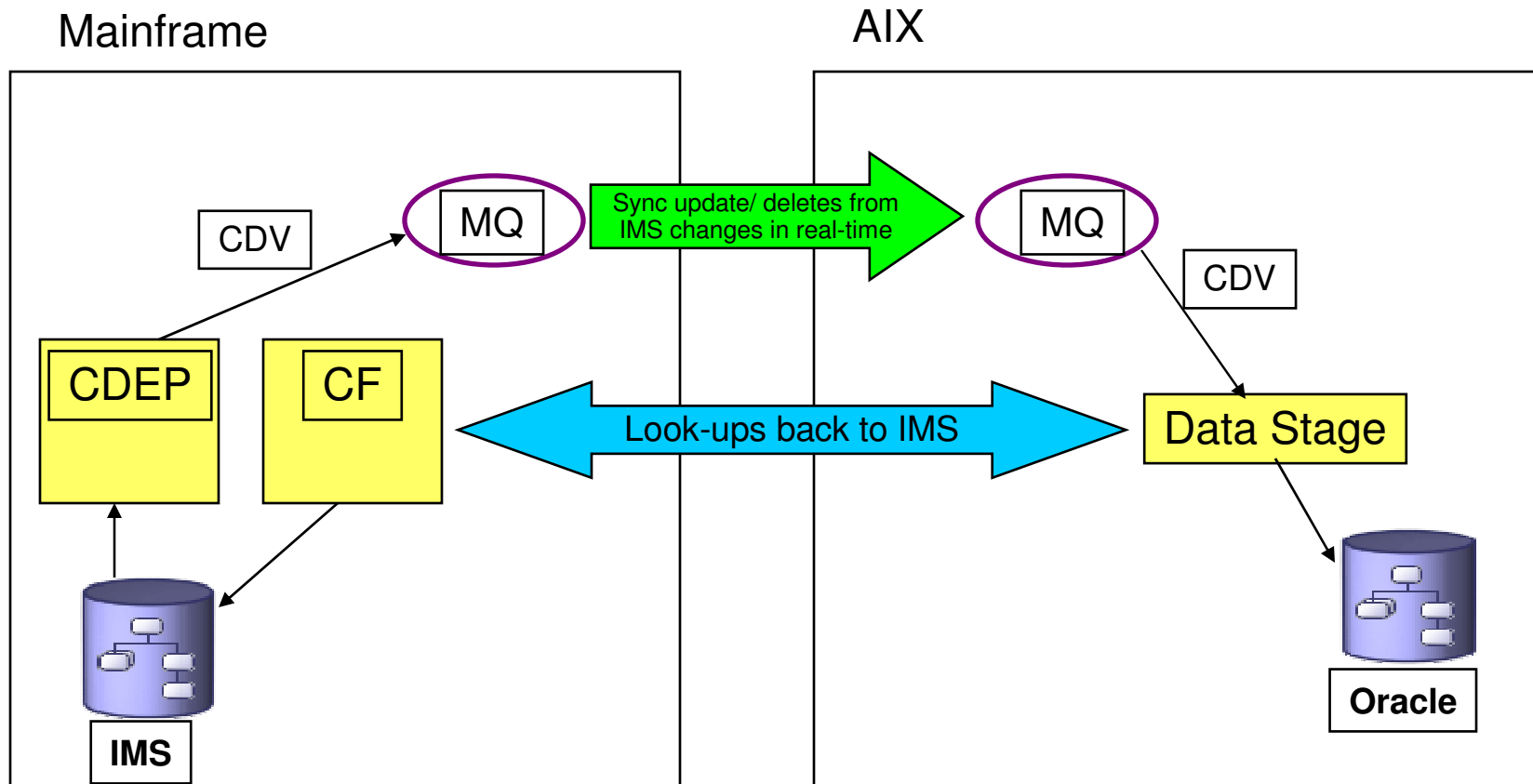
- **Solution Details**
- **Product Overviews**
 - Classic Data Event Publisher for z/OS
 - Classic Federation Server for z/OS
 - DataStage

Solution Details

Overview

- Initial loads executed to synchronize Oracle DWH with IMS
- IMS changes processed in near real-time to keep Oracle in sync
- Audit processing captures transaction types and message timestamps
- "Bad" messages are rejected and reprocessed via error handling jobs
- Initial loads are also run for disaster recovery to re-synchronize Oracle with IMS

DWH load and sync for US State Gov. Judicial Information Systems



Classic Data Event Publisher

- Created logical relational table mappings, using Classic Data Architect, of IMS segments (case information) to be monitored for changes
- IMS change capture using the IMS logger exit
- Changes captured and placed on a mainframe MQ queue in comma delimited format and transmitted to a remote queue on AIX
- Changes processed via a DataStage transaction job



Classic Federation Server

- IMS segments mapped for performing look-ups from the DataStage transaction job
- Look-ups also used in error handling for refreshing Oracle tables based on key fields



InfoSphere DataStage

- Initial loads to Oracle are automated using a master DataStage sequencer job
- Mainframe jobs extract IMS data to files and send a 'start' message to DataStage
- Jobs FTP IMS extract files to the DataStage server
- Load jobs process data to the Oracle staging database
- Approx 22 million rows per database are loaded into Oracle
- Data is swapped from the staging database to the transaction database
- Sequencer job starts transaction job



InfoSphere DataStage (cont.)

- Two Oracle database schemas used:
 - Staging schema
 - Transaction schema
- Tables in transaction schema used by Case Search web portal for detailed data
- Design allows for uninterrupted operation of customer's Case Search web portal

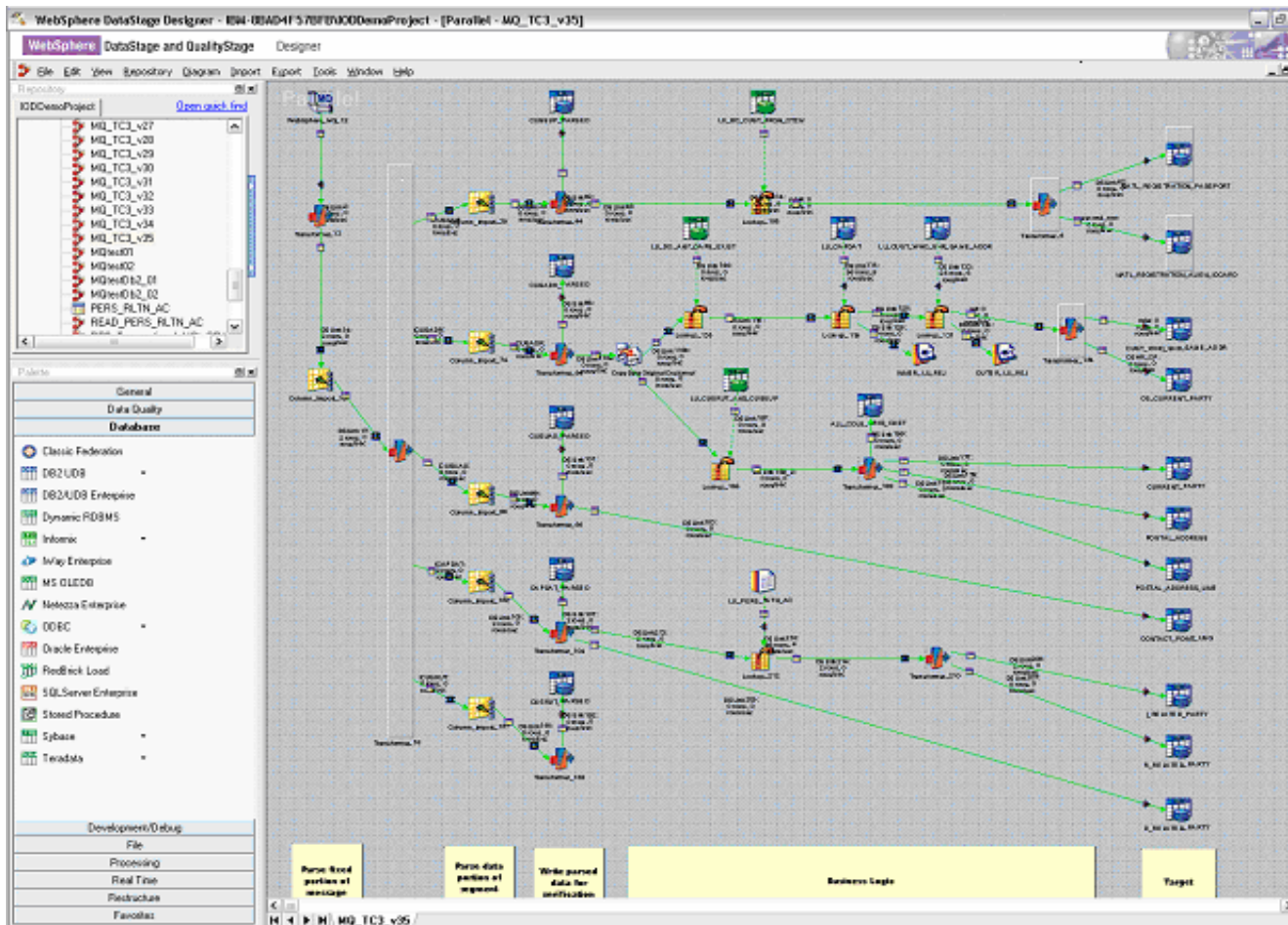


DataStage Transaction Processing Job

- Reads messages from AIX queue and temporarily moves them to a work queue before processing them
- Based on business rules:
 - Transforms data
 - Validates key fields
 - Cleanses data
- Messages with invalid data sent to reject queue for reprocessing
- Oracle DWH refreshed for non-keyed data
- Guaranteed transaction integrity using XA 2-phase commit
- Transaction job always running for maintaining near real-time data currency

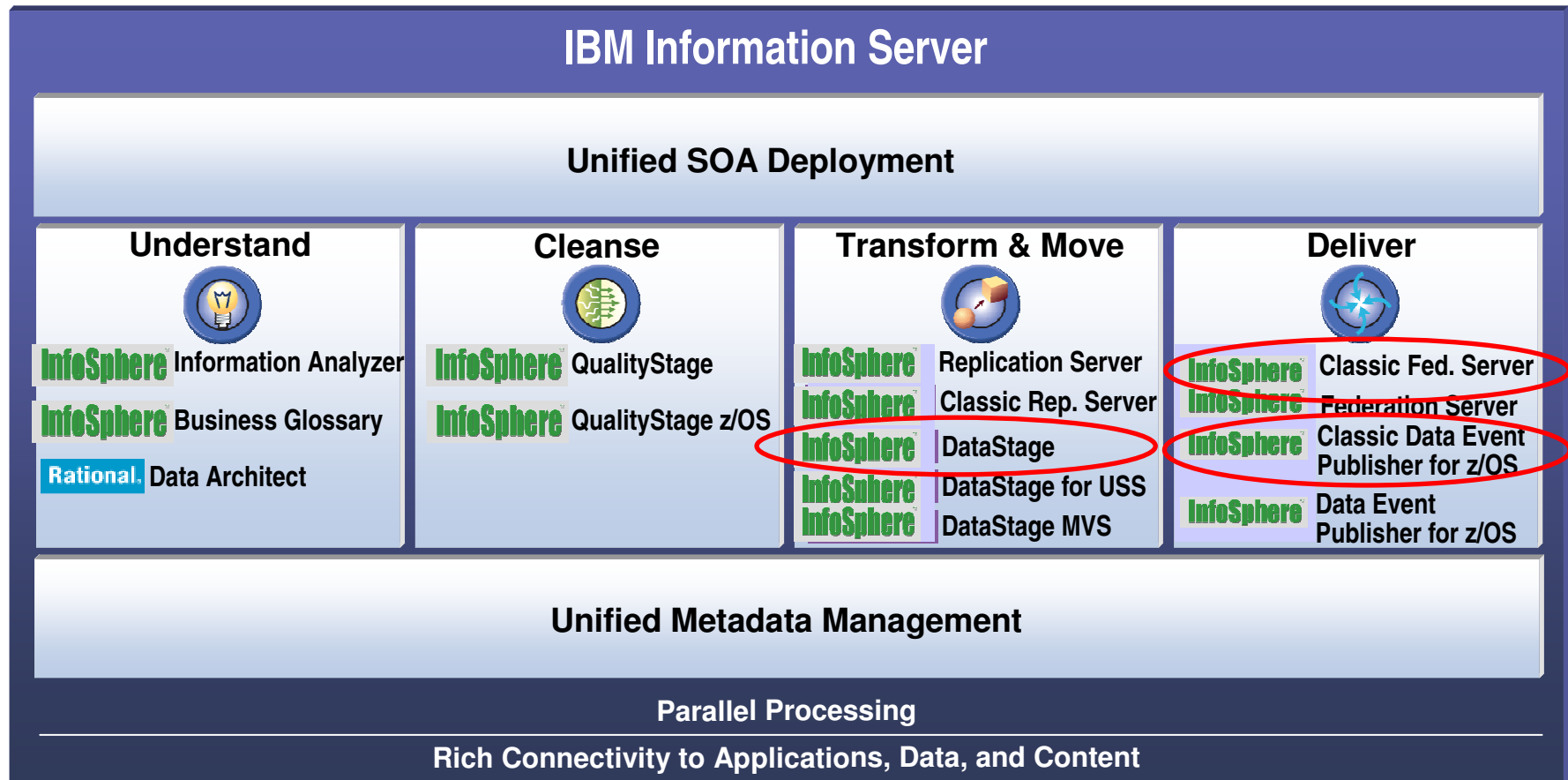


Sample DataStage Transaction Processing Job



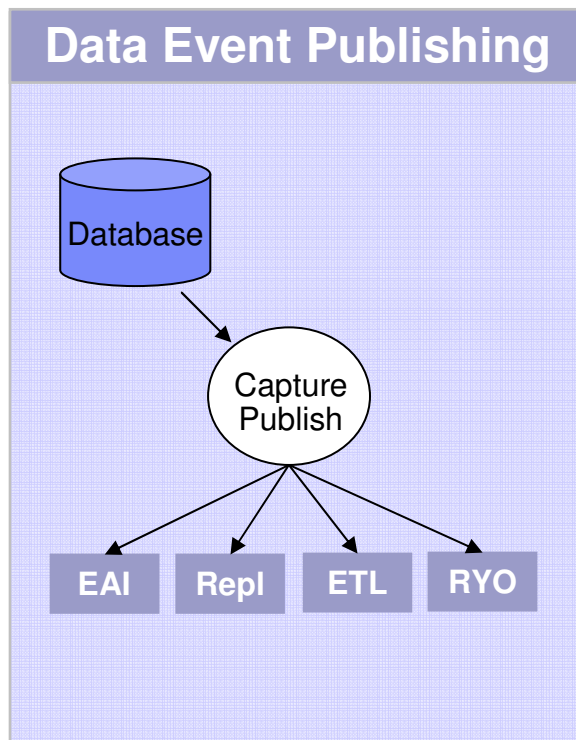
Product Overviews

IBM Information Integration



Data Event Publishing

Message-based publishing driven
by changed-data-event capture



- **Warehouse / Business Intelligence**
 - Integrate captured changed data with an ETL tool
 - Perform very complex transformations
 - Use a specific transaction format to update target
- **Application to Application Messaging**
 - Drive downstream applications or APIs based on transactional data events
 - Reduce application development and maintenance
 - Reduce performance impact to source applications
 - Reduce availability impact to source applications
- **Event Notification**
 - Stream changed data information to Web interfaces
 - Stream only particular events of interest (filter data)

InfoSphere Classic Data Event Publisher for z/OS

Changed-Data-Capture

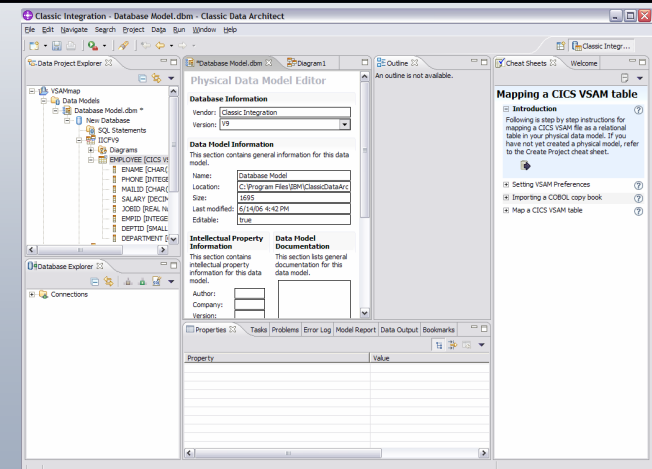
- **Integration using data events not application "hooks"**
- **Capture data changes in real time, publish as "data events" to drive integration & incremental updating**
- **Flexible and efficient:**
 - Low-latency or scheduled data capture
 - Multiple publication formats:
 - Consistent relational format for ease of use
 - XML for ease of consumption
 - Delimited values for reduced message size
 - Recoverable
 - WebSphere MQ - Assured delivery
 - Eliminates dependence on batch window!
 - Loosely coupled approach

Deliver



InfoSphere Classic Data Event Publisher

Detect and respond to data changes in source systems, and publish changes to subscribed systems, to ETL or to other modules for event-based processing

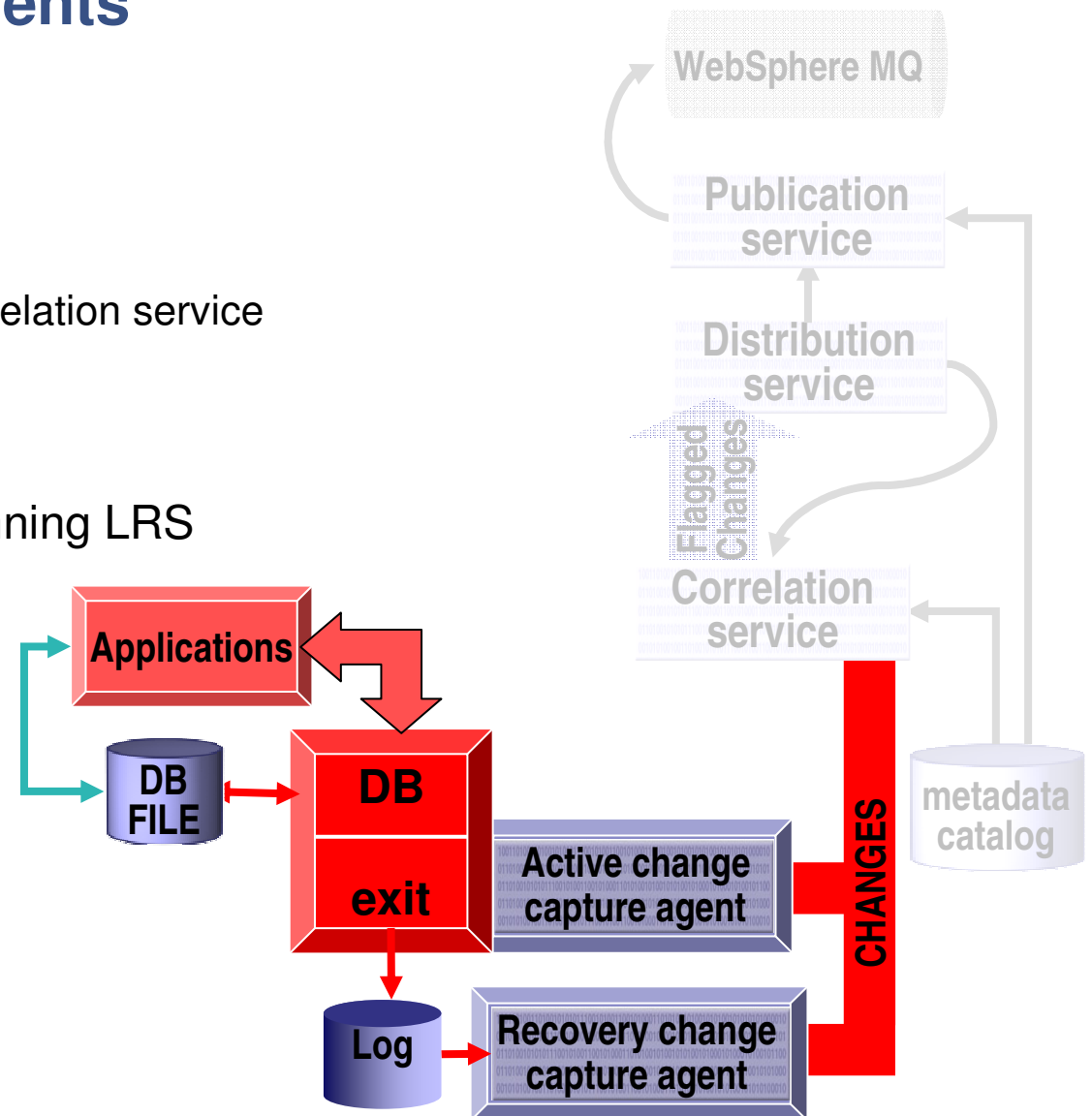


Easy Changed Data Capture

IMS Change Capture Agents

▪ IMS agent/service

- Exit runs within IMS region
 - All changes forwards to Correlation service
- Log reader service (LRS)
 - Automatically read logs
- Recovery – batch job or running LRS



Data Event Correlation – Distribution - Publication

Correlation Service

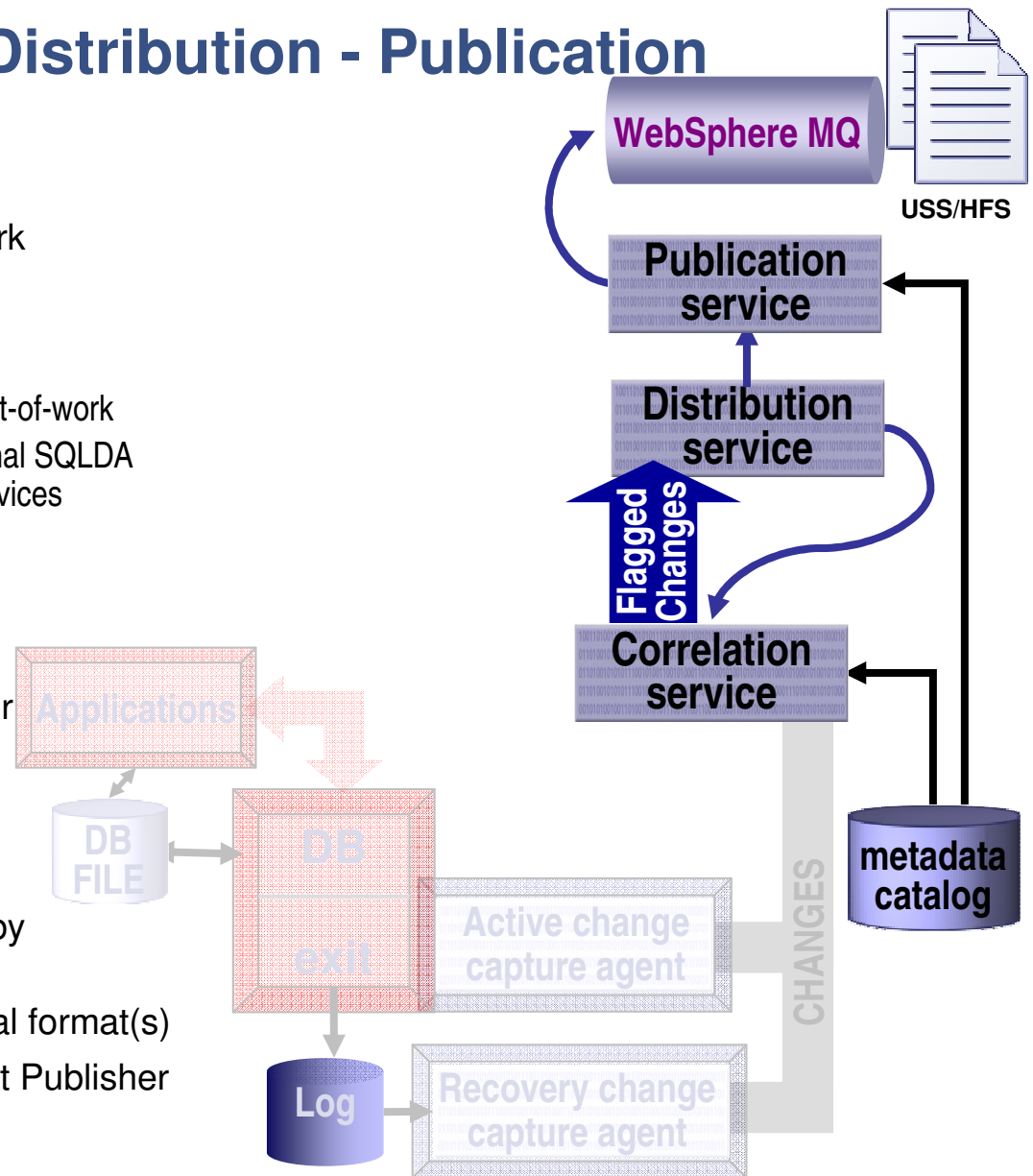
- Manage data publishing by unit-of-work
- Sort data by unit-of-work identifiers
- At end of unit-of-work, CCA returns:
 - Rollback – flush all data for this unit-of-work
 - Commit – reformat data into internal SQLDA
 - send to Distribution Services

Distribution Service

- Orders changes by transaction Unit-of-Recovery
- Metadata is received from Data Server
- Distributes changes to Publication Service threads

Publication Service

- Manage distribution of data changes by unit-of-work
- Handle SQLDA transformations to final format(s)
- XML & CDV delivered as part of Event Publisher product



InfoSphere Classic Federation Server for z/OS

Virtualize

- **Standardized ODBC/JDBC/CLI SQL access to:**
 - VSAM, IAM, sequential files
 - IMS, CA-IDMS, CA-Datcom, Adabas and DB2/z databases
- **Metadata-driven, so there's no mainframe programming needed**
- **Works with mainframe infrastructure and "modern" applications and tools**
- **Deliver mainframe data to**
 - IBM's own data profiling, cleansing and transformation solutions
 - Self-service portals
 - e-commerce solutions
 - Reporting and analytical tools

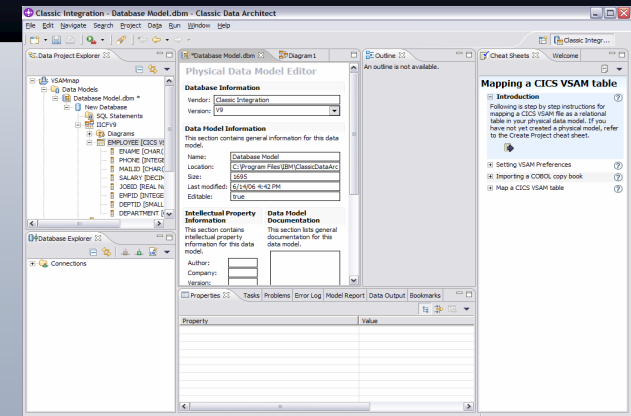
Deliver



InfoSphere Classic Federation Server for z/OS

Read-from and write-to mainframe data sources using SQL from Unix, Windows, Linux and JVM platforms

Empowers mainframe data integration with Information Server components, your applications as well as IBM and 3rd party tools and applications

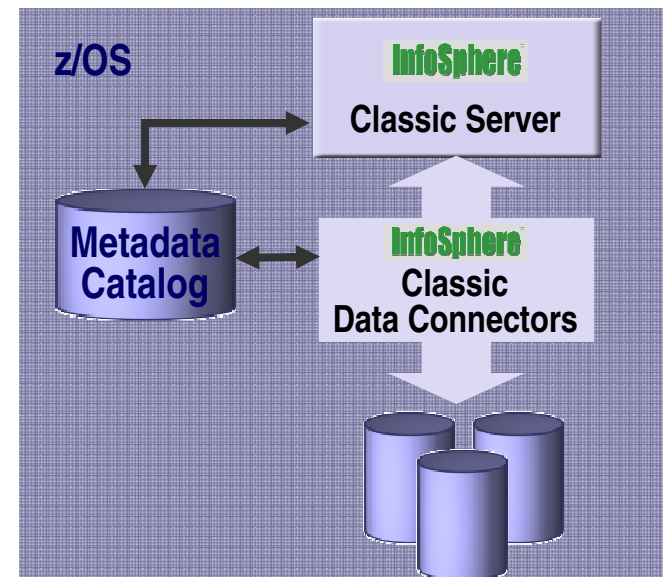


Dynamic Visual Metadata Management

Solution Profile:

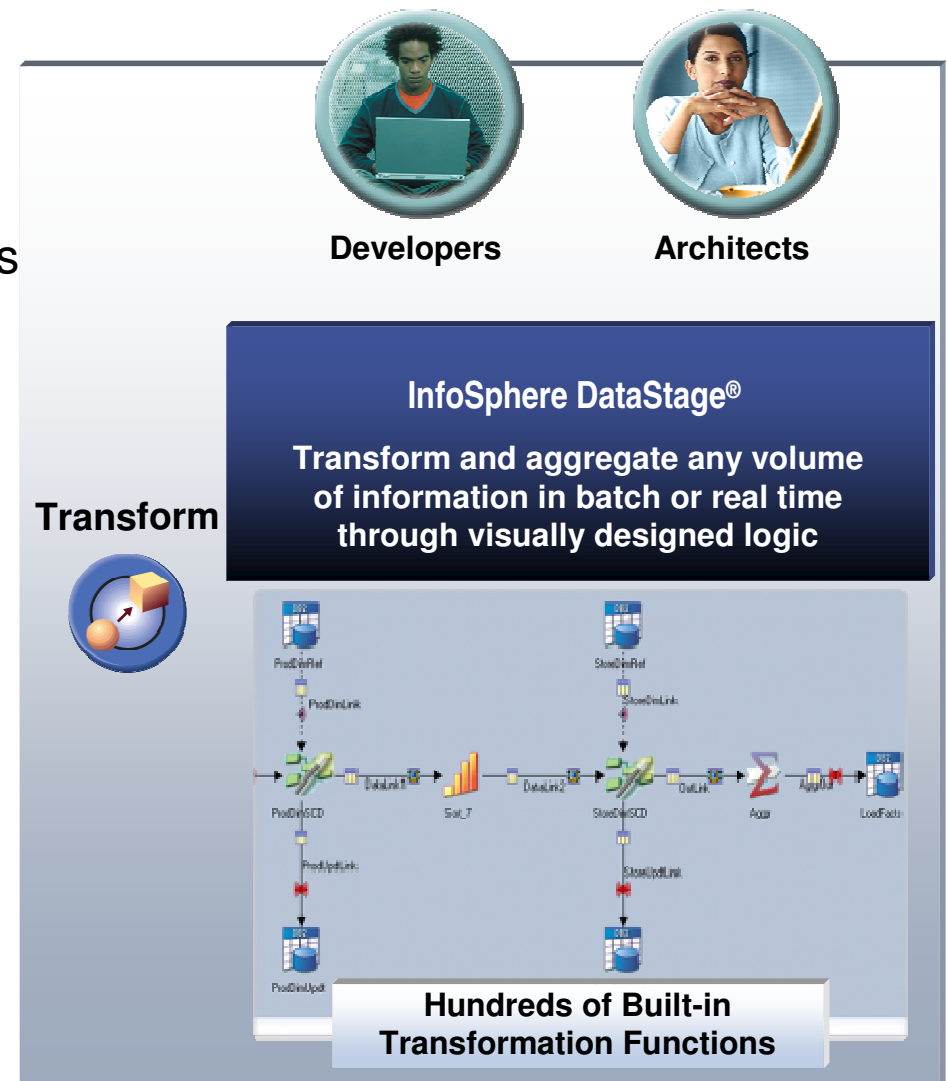
Fits into Mainframe Control Infrastructure

- **Governors – prevent the runaway user**
 - Limit resource utilization
 - Set at installation, server and/or user levels
- **Security – ensure only authorized access**
 - SAF interface for RACF, ACF and Top Secret
 - Optional embedded DB2-like security
 - Grant/Revoke SQL privileges
 - Supported at table and view level
- **Monitor and Control Usage**
 - Use existing tools (e.g. OMEGAMON)
- **Embedded SMF exit support**
 - Charge-back and other accounting purposes
 - Delivers user defined record types
- **Tracing and Logging by level**
 - Remote and local monitoring
 - Triggered tracing



IBM InfoSphere DataStage

- Graphical, codeless data transformation and movement
- Processes and transforms large amounts of data in real-time or bulk mode as well as in a services model
- Handles all transformations from simple to complex
- Manages multiple integration processes
- Integrates data from the widest range of enterprise and external data sources as sources and/or targets
- Leverages meta data for cross tool impact analysis and easy maintenance



Summary

- Customer wanted to improve their process of keeping their Oracle data warehouse in sync with their operational data in IMS
- End users demanded more current case information due to data latency
- Products (Classic Data Event Publisher, Classic Federation, DataStage) from the InfoSphere suite are able to provide near-real time synchronization between the mainframe data and the data warehouse
- This provides end users with the most current, complete, and trusted case information

Thank You