IBM

InfoSphere™

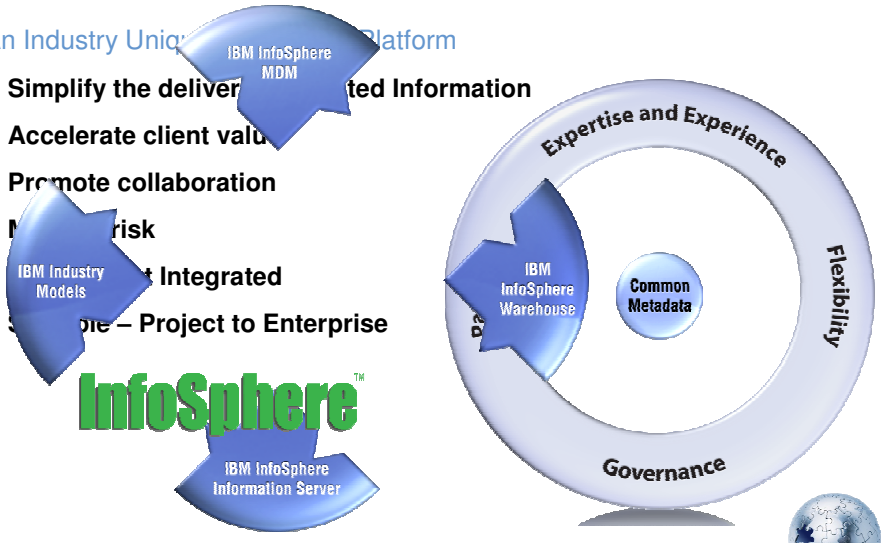
Easy and hassle-free real-time data delivery and integration on System Z

InfoSphere™ Information Management Software IBM

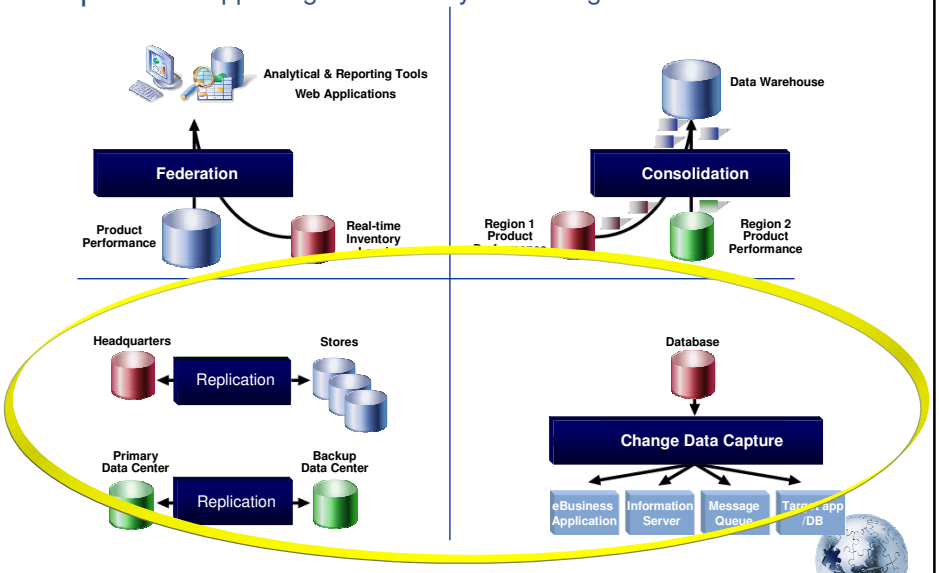
The IBM InfoSphere Vision

An Industry Uniq[illegible]Platform

- Simplify the deliver[illegible]ted Information
- Accelerate client valu[illegible]
- Pr[illegible]mote collaboration
- M[illegible]risk
- [illegible]t Integrated
- S[illegible]ple – Project to Enterprise



InfoSphere: Supporting Different Styles of Integration



Business Challenges

Dynamic Warehousing & Business Intelligence and Reporting

- ***Yesterday's*** **data inadequate for inventory and purchasing decisions**

Data Synchronization and Replication

- **We need up to date information flowing between applications and to ensure an up-to-date version is always available**

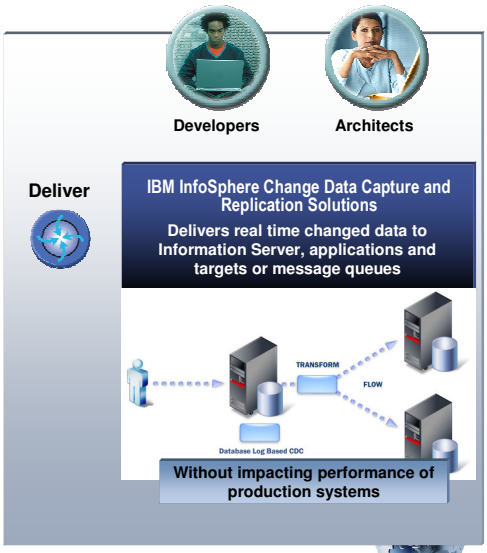
Real-time Event Detection

- **Need to pro-actively monitor and respond to business changes**

……Without Impacting the Performance of Production Systems

What is IBM's CDC and Replication Solutions?

- **Family of products to support any environment**
  - Each product built to support specific client environment – customers can choose product based on their source and target environment
- **Provides real-time change data capture and delivery for**
  - Active Data Warehousing
  - Master Data Management
  - Application Migration/Consolidation
  - Operational Business Intelligence
  - Services Oriented Architecture (SOA)
- **Minimal impact on production systems**
- **High scalability and performance**
- **Guaranteed data integrity**

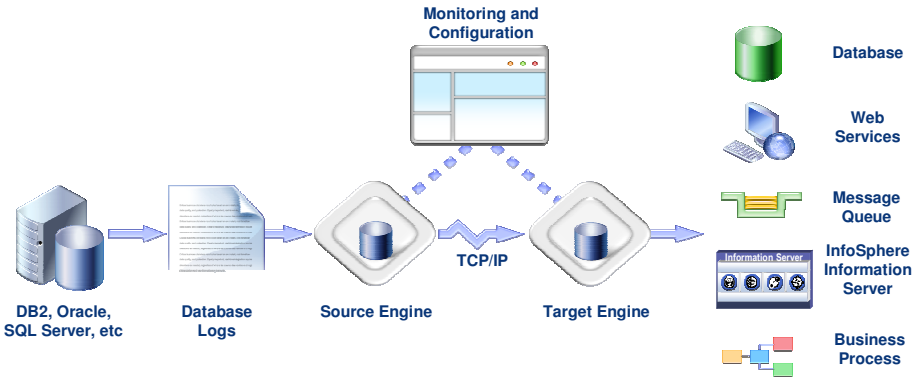


InfoSphere Change Data Capture - Platform Support

DATABASES Source & Target	TARGETS	MESSAGE QUEUE	OPERATING SYSTEMS	HARDWARE PLATFORMS
DB2 z/OS	Teradata	JMS	IBM i OS	IBM i OS
Oracle	Information Server	MQ Series	z/OS	IBM System z
Sybase	Cognos Now!	TIBCO	AIX	IBM System p
MS SQL Server	Netezza*	WebMethods	HP-UX	HP PA-RISC
DB2 LUW		BEA	Solaris	HP Itanium
DB2 i			MS Windows	Intel / AMD
Informix			Red Hat, SUSE Linux	Sun SPARC

** Customized solution, limited requirements*

Log-Based Change Data Capture



- **Key Benefits:**
  - Low impact
  - Flexible implementation
  - Heterogeneous platform support
  - Easy to use

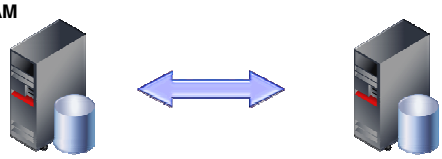
Low Impact

- Log-based CDC captures data without interacting with database
- No changes or upgrades to applications and schemas required
- Peer-to-peer architecture does not require additional hardware
- Sending only changed data requires minimal network bandwidth

Migrations, HA, Data Distribution

- DB2 z, LUW, System i
- Oracle
- SQL Server
- Sybase
- VSAM, IMS, ADABAS, CA-IDMS, CA-Datacom

- DB2 z, LUW, System i
- Oracle
- SQL Server
- Sybase
- VSAM



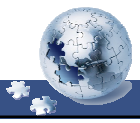
DB2 LUW, z
Oracle

DB2 LUW, z
Oracle

Migrations, HA, Data Distribution

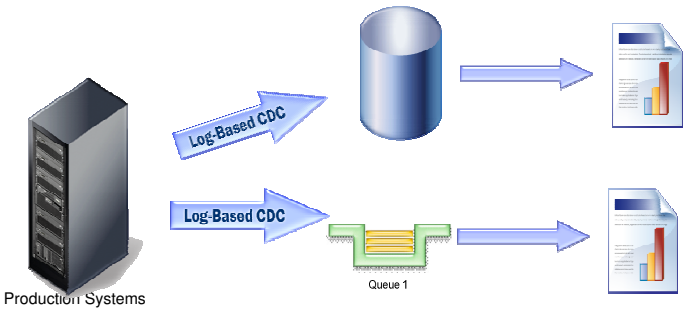
IBM Solution Advantages

- Broad support for heterogeneous environments
- Cross-platform support
- Low impact on source systems
- No batch window required
- Target system available for use

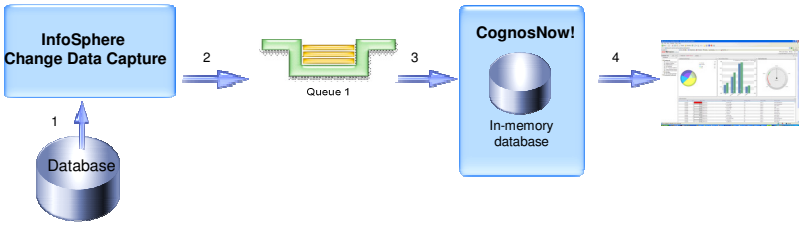


Live, Operational Reporting

- Distribute reporting workload over existing environments
- Real-time data for Business Intelligence



Feeding Real-Time Changes to IBM Cognos Now!



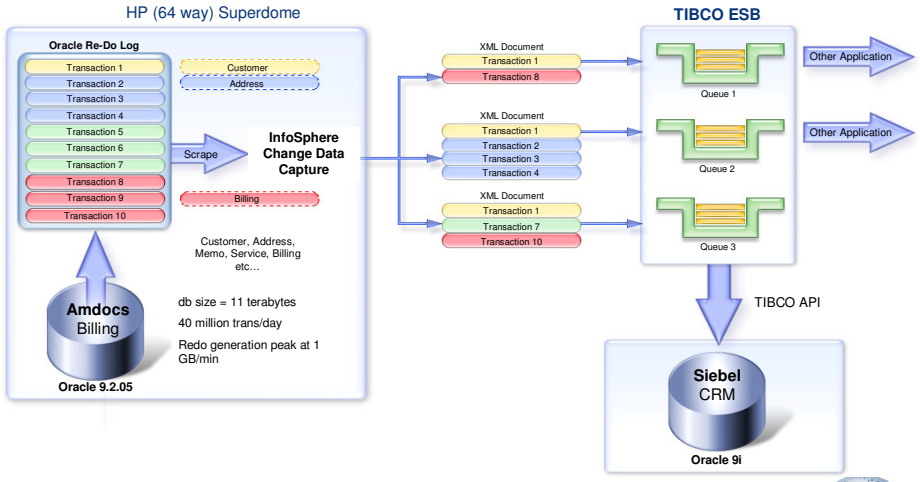
1. Capture source database changes
2. Send changes to JMS message queue
3. CognosNow! receives changes from message queue; stores them in in-memory database
4. CognosNow! dashboard updates and displays data as it changes

Live, Operational Reporting

IBM Solution Advantages

- Low impact on source systems
- Safer, faster, more accurate BI
- Supports multiple topologies
- Real-time delivery of changes to IBM Cognos Now! for immediate analysis and display

Event Driven Architectures

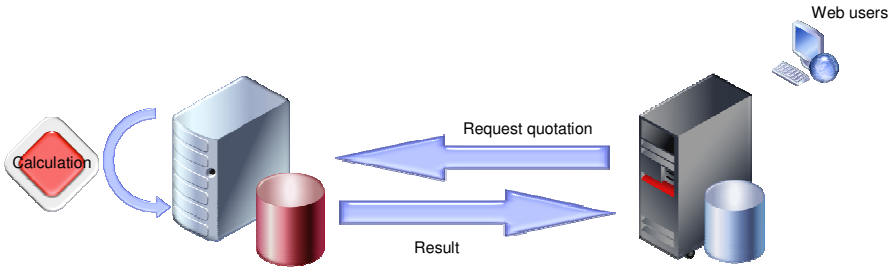


Event Driven Architectures

IBM Solution Advantages

- Minimal impact on operational system
- Minimum Latency (1100 transactions per second with no latency)
- Scalability
- Supports any JMS-compliant message queue

Application Integration



Back-Office application **(AIX-Oracle)**

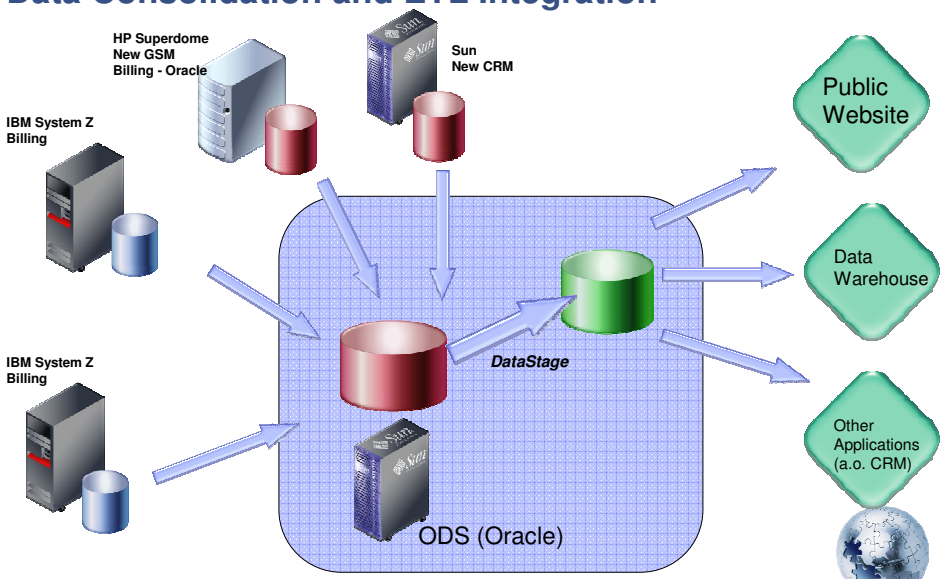
Front-Office application **(DB2 Z)**

Application Integration

IBM Solution Advantages

- Heterogeneous support
- No application/database changes required
- Leverage existing investments

Data Consolidation and ETL Integration

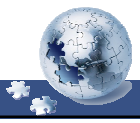


InfoSphere Change Data Capture

Product Features

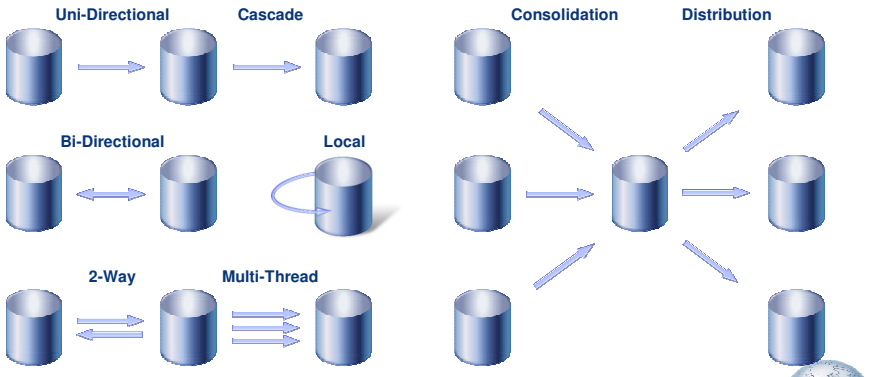
InfoSphere Change Data Capture

- **Real-time changed data capture across database systems**
  - Captures data from production systems without impacting performance
  - Applies data to target systems in real time
- **Transforms database operations into XML documents**
  - Supports simple or composite XML transactions
- **Creates audit trails for full data traceability**



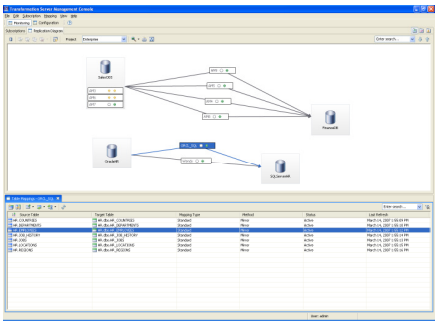
Flexible Implementation

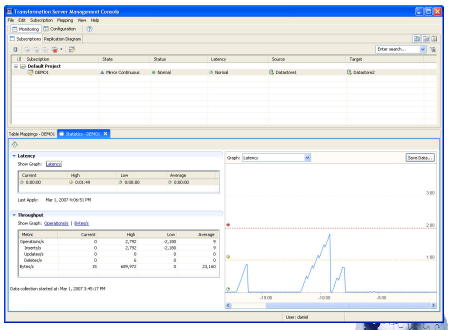
- Supports all topologies and environments
- Conflict detection & resolution maintains data integrity



Easy to Use

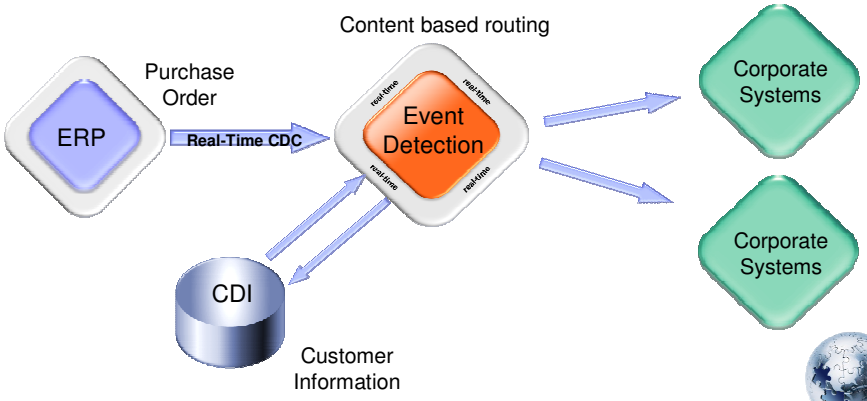
- Java-based GUI for configuration, administration, and monitoring
- Manage data integration processes from one screen
- Wizards and task automation
- No programming required





Context-Rich Composite Events

- Combines content associated with the event from other systems
- Routes the data to different applications to initiate business processes based on content of a message



Replication

- **3 Modes of Replication**
  - Continuous mirroring
    - Apply data changes at the target as it is generated at the source
  - Periodic mirroring
    - Apply net changes on a scheduled basis
  - Refresh
    - Apply a snapshot version of source system

Filtering

CUST_NO	L_NAME	F_NAME	PHONE	REP_NO
58699	Smith	John	404-555-3874	45
37283	Duggan	Ira	613-555-8367	25
89863	Quinn	Fran	905-555-1296	11
89732	Muntz	Muntz	704-555-2738	25

- Integrate entire systems or only a subset of data
- Table/row/column-level filtering options available

ROW SELECT
REP_NO = 25

CUST_NO	L_NAME	F_NAME	REP_NO
37283	Duggan	Ira	25
89732	Muntz	Josie	25

Table Mapping Methods

One-to-one
- Source and target tables have similar table structures

LiveAudit
- Generates audit trail of data transactions from source

Adaptive Apply
- Automatically synchronizes data for dissimilar sources and targets

Summarization
- Keeps a running total of numerical values at the target

Consolidation: One-to-One
- Merges data from several tables into a single row

Consolidation: One-to-Many
- Used to apply a source lookup table change to all affected target rows

Data Translations

- Convert data representations on the fly to integrate disparate systems and database table structures

EMP	LAST	FIRST	HIRE_DATE	STAT	SALARY	MAX
1234	Moreiro	Nicole	01/05/97	A	$55,000	$60,000
2345	Ellison	Val	04/12/97	I	$40,000	$50,000

Increase Field Size | Concatenation | Century Dates | Transform Fields | Derived Fields

EMP_ID	FULL_NAME	HIRE_DATE	STATUS	%SALARYMAX
001234	Nicole Moreiro	01/05/1997	Active	92%
002345	Val Ellison	04/12/1997	Inactive	80%

Auditing

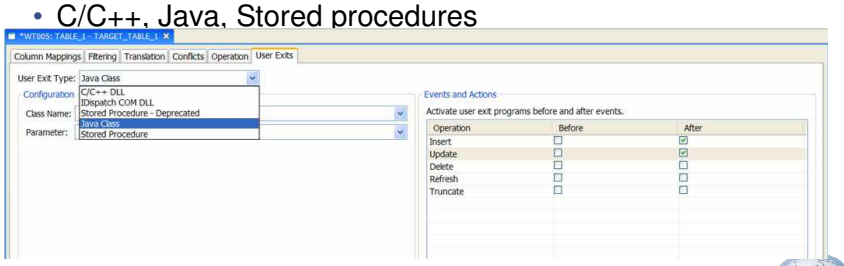
- Switch all operations into INSERT to keep transactional history
- Capture additional data for full data traceability
- Type of data change, origin of data change, etc

JOURNAL CONTROL COLUMNS

&CCID	An identifier for the transaction with the update.
&CNTRRN	Source table relative record number
&CODE	Always "U" for refresh. Always "R" for mirror.
&ENTTYP	Indicates the type of update.
&JOB	The name of the source job that made the update.
&JOBNO	The operating system user Id of the update process.
&JOBUSER	The operating system user at the time of the update.
&JOURNAL	The name of the journal, as described in Properties.
&JRNFLG	Indicates if before image is present
&JRNLIB	The name of the journal schema.
&LIBRARY	The source table schema or its alias.
&MEMBER	The source table name or its alias.
&PROGRAM	The name of source program that made the update.
&OBJECT	The source table name or its alias.
&SEQNO	The sequence number of this update in the journal.
&SYSTEM	The hostname of the source system
&TIMSTAMP	Time of the update or refresh.
&USER	The user ID which made the update.

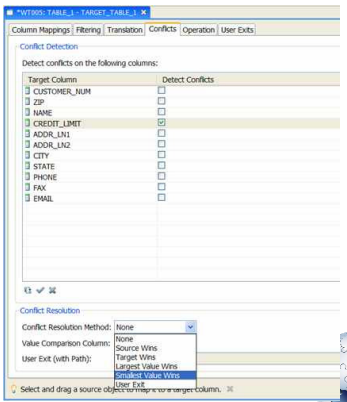
User Exits

- **Execute custom business logic**
  - React to business events in real time
  - Automate business processes
- **Multiple implementation methods available:**
  - C/C++, Java, Stored procedures



Conflict Detection & Resolution

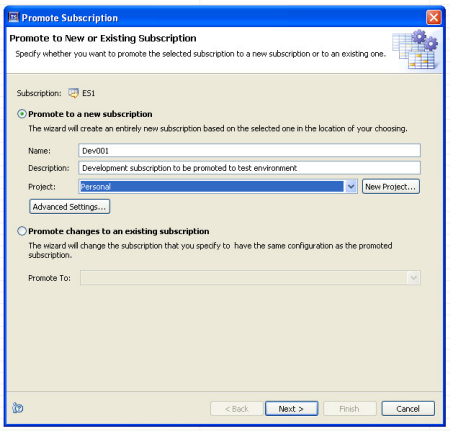
- Ensures data integrity when multiple systems change the same data simultaneously
- Conflicts can be resolved in various ways:
  - Source wins, target wins
  - By data value
  - Execute user exit



Change Management

- **Promote test and development integration processes into production without risk**
  - Eliminates potential user error
  - Enables faster rollout of new business

  processes
  - Rollback capabilities available
  - Changes are tracked for compliance

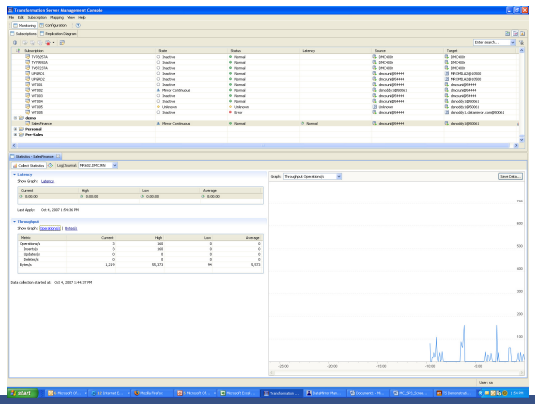


Guaranteed Data Integrity

- Data transactions are applied at the target in the same order as it was generated at the source
- Target acknowledges each apply operation to ensure delivery
- No data is lost even if communications link becomes unavailable

Monitoring

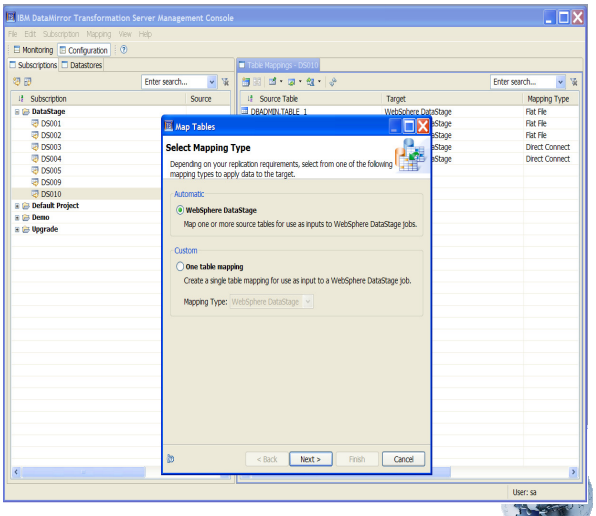
- Graphical visualization of replication processes
- Event logs, alerts & alarms
- Exportable throughput & latency statistics



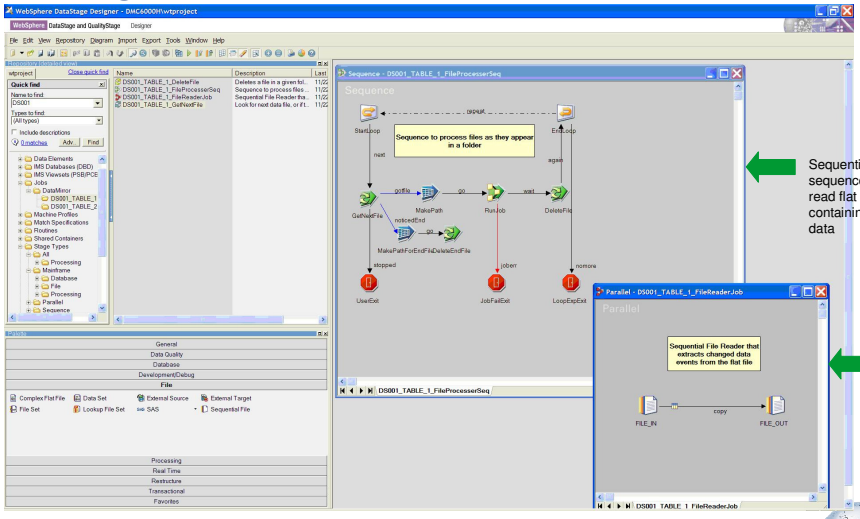
InfoSphere DataStage Integration

To Enable Real-time Response to Data Changes & Business Events

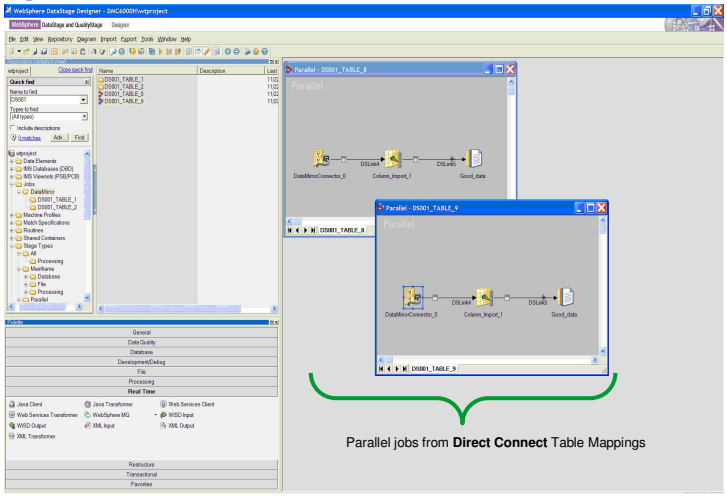
- Low impact log-based changed data capture
- New palette stages on Information Server
- Extremely low impact on sourcing for ETL processing into data warehouse
- Leverage existing Data ETL and Data Cleansing investments



DataStage Jobs - Flat File Connection Method



DataStage Jobs – Direct Connect Method



CDC z/OS Performance

Work Load Test

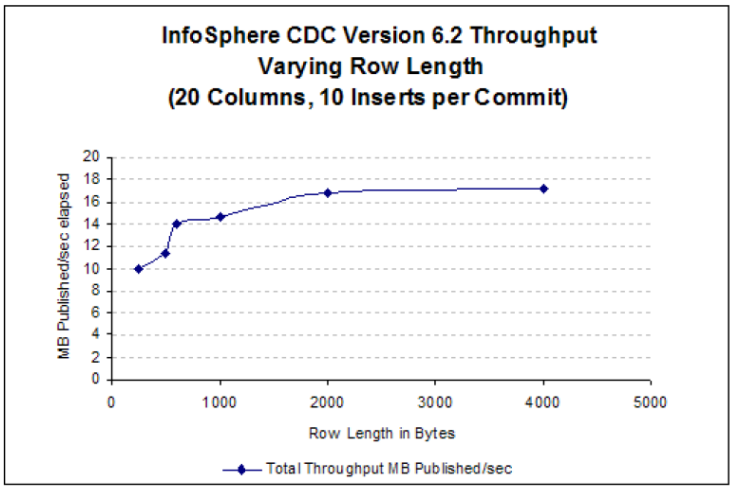
- The intent of this workload test was to see how quickly data could be published and applied, and to determine the CPU footprint of the source and target engines.
- A simple insert-only1 test was run, keeping numbers of columns and commit size consistent while varying the row size.
  - **10 Inserts per commit, 20 columns**
  - **240, 480, 600, 1000, 2000, 4000 byte rows**
- For all tests, one million rows were replicated with the rows equally distributed into ten tables.
- CPU measurements include the InfoSphere CDC and communication processes for both the source and target systems. The CPU for the target database itself is not included.

Testing Environment

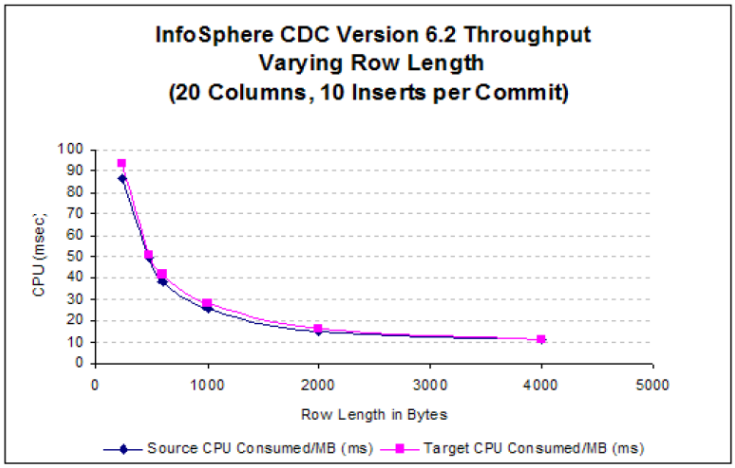
- **DB2 for z/OS (Source and Target)**
  - Utilized a z10 2097 E12. System capacity number is 708 - approximate SI MIPS = 6439.
  - Used 24 GB memory with another 10 reserved, 2 TB DASD in ESS 6800 with RAID 10 configuration.
  - Source LPAR had 3 dedicated CPUs
  - Target LPAR had 3 dedicated CPUs.
- **DB2 Version 8 subsystem utilized on both source and target. No data sharing defined. DB2 has 890 cylinders.**

Workload Test Results By Row Length

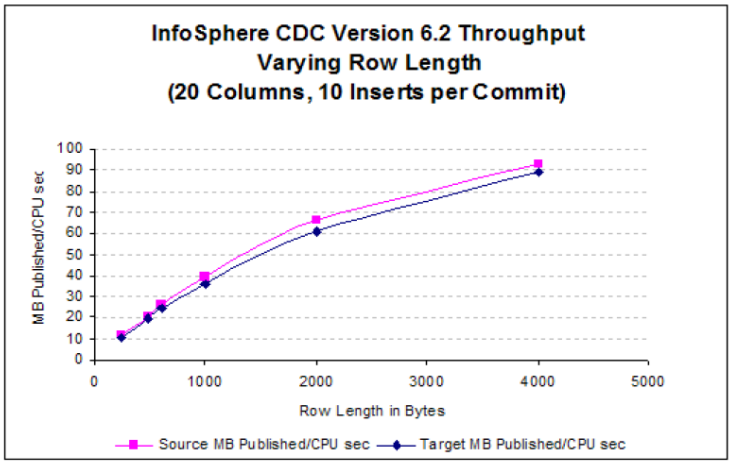
Row Length in Bytes	240	480	600	1000	2000	4000
Elapsed Time (seconds)	24	42	43	68	120	233
Source CPU Consumed (seconds)	20.7	23.7	23.0	25.3	30.4	43.0
Target CPU Consumed (seconds)	22.5	24.2	24.9	27.7	32.8	45.0
Rows Published per second elapsed	41667	23810	23256	14706	8333	4292
Source Rows Published per CPU seconds	48309	42194	43478	39526	32895	23256
Target Rows Published per CPU seconds	44444	41322	40161	36101	30488	22222
Total Throughput MB Published per second elapsed	10.00	11.43	13.95	14.71	16.67	17.17
Source MB Published per CPU second	11.59	20.25	26.09	39.53	65.79	93.02
Target MB Published per CPU second	10.67	19.83	24.10	36.10	60.98	88.89
Source CPU Consumed per MB (ms)	86.25	49.38	38.33	25.30	15.20	10.75
Target CPU Consumed per MB (ms)	93.75	50.42	41.50	27.70	16.40	11.25



The chart illustrates how the InfoSphere CDC product scales with respect to the amount of data that can be published per second of elapsed time.



The chart illustrates the efficiency of the InfoSphere CDC product in terms of CPU usage for this scenario. In particular, you can clearly see the decrease in CPU usage per MB as row size increases.



The chart shows a trend with respect to the scalability of the InfoSphere CDC product. In particular you can see that the volume of data that can be replicated per CPU second is very significant and increases as the row size increases.

Summary

- **InfoSphere Information Server can address any type of data integration style**
- **Change Data Capture and Replication products provides real-time changed data capture across the enterprise**
- **Key Benefits:**
  - Low impact
    - **Does not impact performance and requires no changes to applications**
  - Heterogeneous
    - **Integrates data from all platforms and databases**
  - Flexible
    - **Supports any topology**
  - Easy to use
    - **Fast deployment with low risk**
  - Integrated with Information Server
    - **Single solution for all data integration requirements**


Thank
You