

# Information Server CDC

La nuova soluzione IBM  
per l'integrazione dei dati in tempo reale

**Monica Taranto**  
SWG IM Technical Sales  
IBM Italia

IBM Information  
On Demand 2008  
>>> Comes To You

ALLA LUCE DELL'INFORMATION ON DEMAND

Milano, 15 aprile 2008



# Agenda



- Il significato del Real-Time
- Che cos'è la Change Data Capture
- Caratteristiche della soluzione:
  - Information Server CDC
  - LiveAudit
- Q&A

# The Real-Time Challenge



- Need access to information on demand
  - Streamline business processes
  - Improve customer service
- Currently, many companies use batch-based processes to gather data
  - Provides a point-in-time view of the business
    - Data only as fresh as the previous batch cycle
  - Bulk data extracts consume large amounts of system resources
    - Traditional data integration performed only during off-business hours

# Solution – Real-Time ETL Enablement



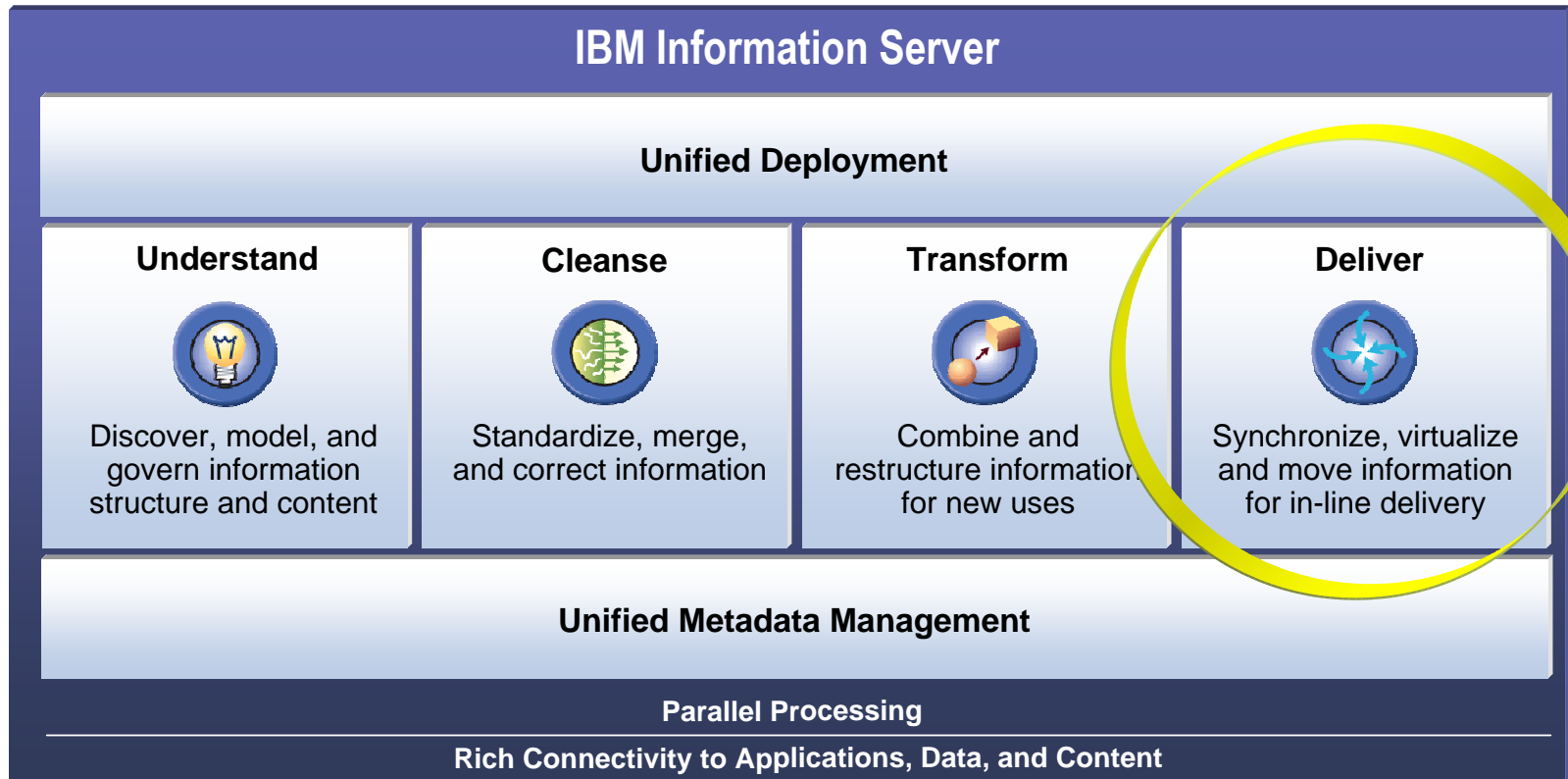
- Continuously flow information throughout the enterprise
  - Real-time data feeds into ETL tools
- Real-time event visibility and propagation
- Data traceability and trustworthiness
- Flexibility of delivery

*“Real-time data integration has a current market penetration of 5-20 percent of the target audience, it is expected to grow to more than 80% of organizations by 2010”*

*- Gartner*

# The IBM Solution: IBM Information Server

*Delivering information you can trust*





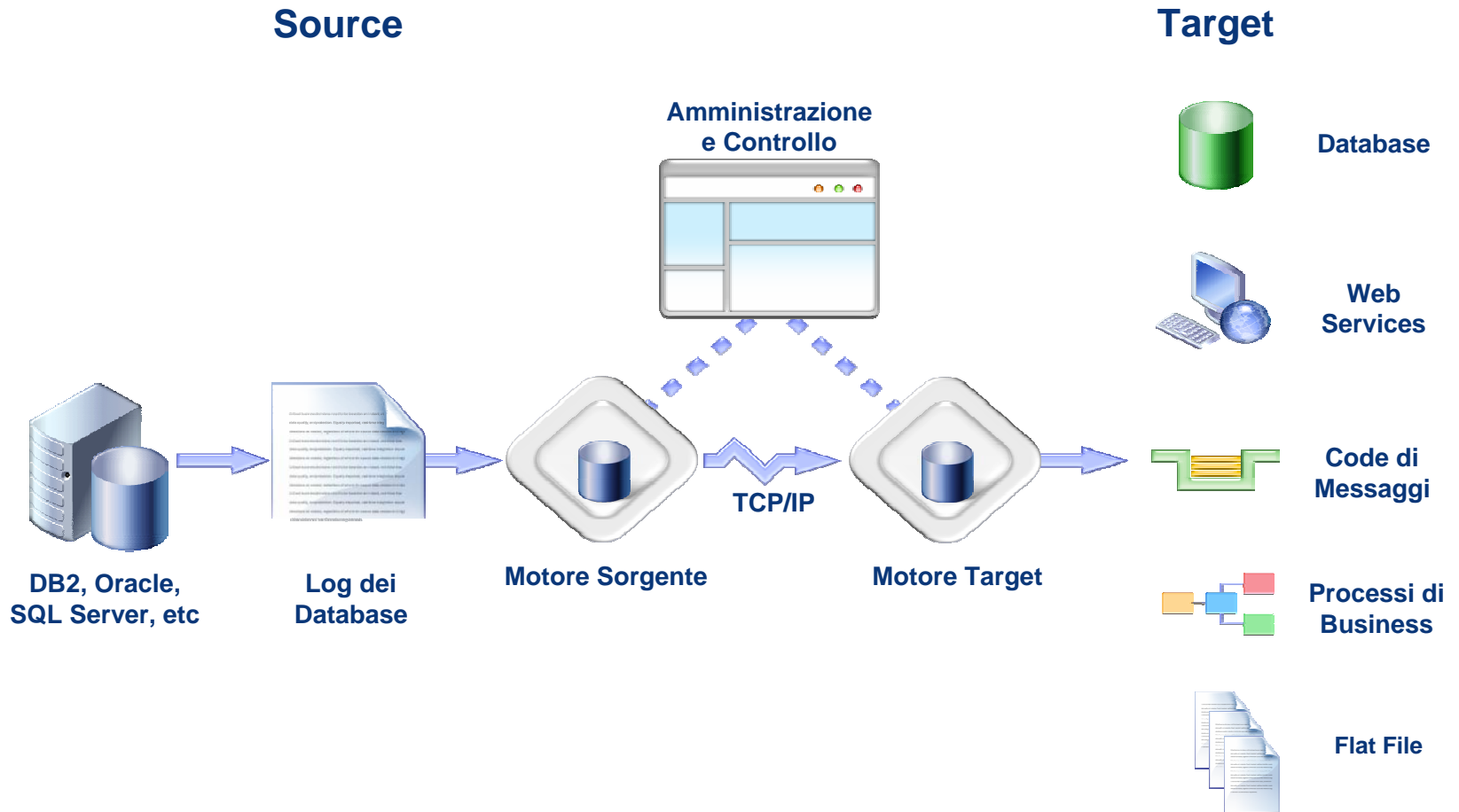
- **What is Log-Based Change Data Capture?**
- *"Use of existing native database recovery logs to capture table and row level activity occurring in a relational database".*
  - **Native Database logs**
    - Oracle = Re-do Log
    - DB2/UDB = DB2 Log
    - DB2 (i5) = OS/400 Journal
    - SQL Server = Transaction Log
    - DB2 (z/OS) = DB2 Log
    - Sybase = Transaction Log
  - **Alternatives to Log-Based CDC**
    - SQL Select - based on Date/Time Stamp
    - File Comparison – determine differences
    - Database Triggers
    - Modify Source Application

# Change Data Capture



- **What are the benefits of using Native DB Logs for CDC?**
  - Easy deployment
    - No changes to existing applications or schemas required
  - Utilizing existing component of the database
  - Resource efficient
    - Little impact to existing source application and system
    - “0.05% system resources required to process over 300 GB of data”
    - Sending only the changes – efficient use of bandwidth
  - Guaranteed Data integrity – utilizing log position (LSN/bookmark)
  - Before and After Images available
  - Additional Log Information available - documenting the change
  - Scalable – Single Scrape
  - **Enables access to changed data in a timely fashion**
    - **Differentiation from batch integrators**

# Log-based Change Data Capture





# Key Elements Of The DataMirror Value Proposition

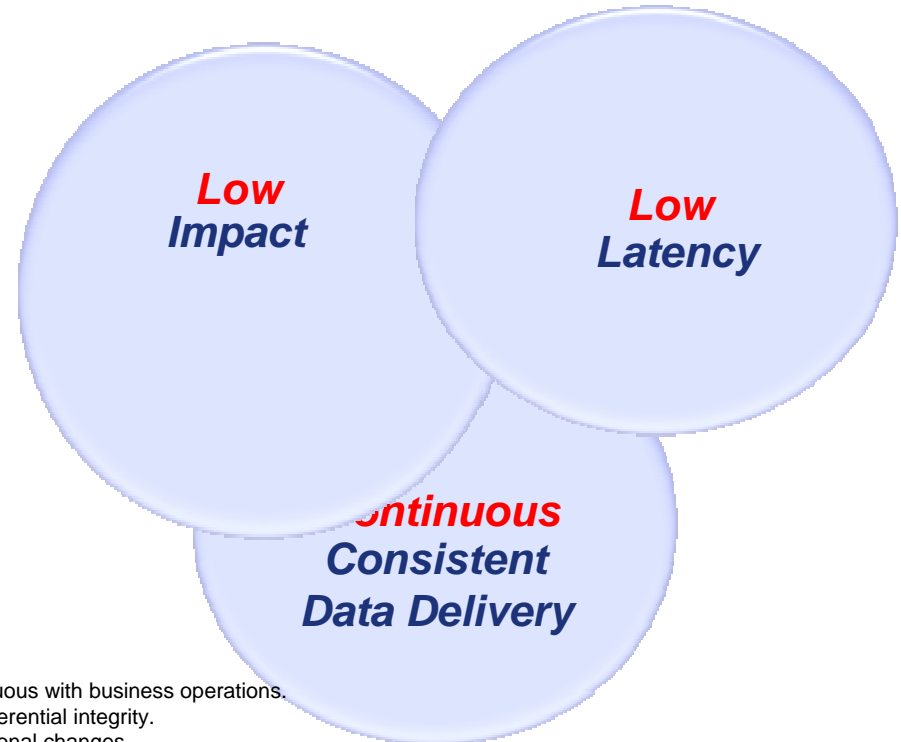


## **LATENCY**

1. Near zero latency for pervasive integration projects.
2. ETL can also deliver low latency but at significantly higher impact to product systems and mission-critical applications.

## **IMPACT**

1. Reduces risk to operational systems.
2. Non intrusive to applications and databases.
3. Use of native DB logs, documented overhead of 2-5%.
4. No use of database triggers.
5. Management easily integrated into existing IT operations.
6. Help reduce/manage operational windows.



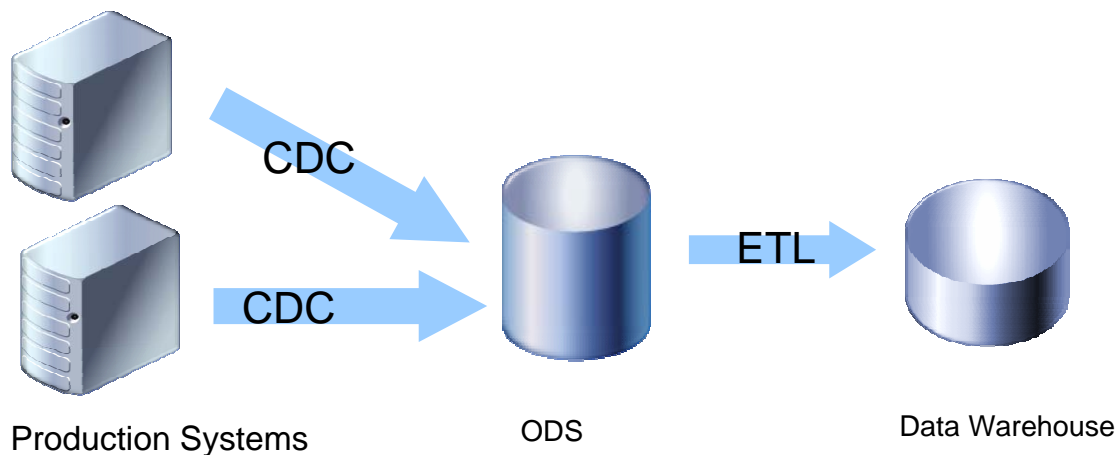
## **CONSISTENT DATA DELIVERY**

1. Data pushed from source, delivered in continuous stream, continuous with business operations.
2. Transaction consistency maintained to preserve units of work, referential integrity.
3. Full transaction granularity, before and after image of all transactional changes.
4. Data event aware, can be used to trigger specific business processes.
5. Fault tolerance, recover to last committed transaction.

# Complementary to ETL: The Shrinking Batch Window



- Problem: Batch window is not long enough to move the data.
  - Customer uses ETL tool to get data out of production systems into a warehouse but as the volume of data to be moved increases, so does the time it takes to move the it.
  - An 8 hour batch window isn't long enough.
- Solution: Move data continuously to ODS, use ETL to extract from ODS

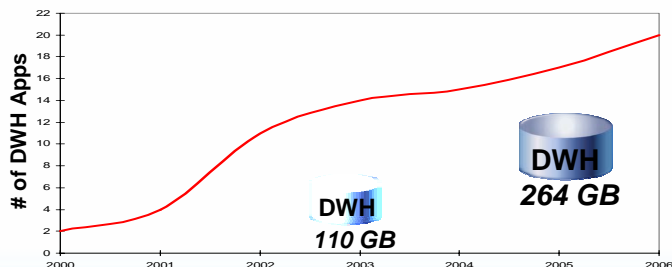


# Business Intelligence e Datawarehouse



## La necessità:

- Consolidare dati provenienti da svariate applicazioni, appartenenti a regioni e business unit differenti
- La finestra temporale non è più sufficiente per il caricamento batch del datawarehouse



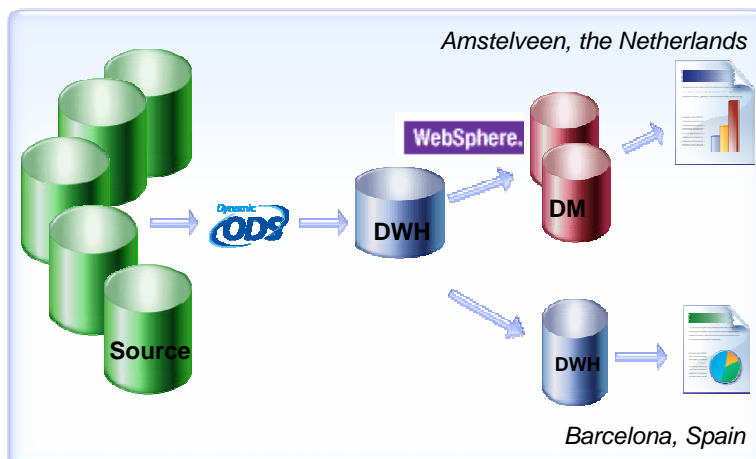
## La soluzione:

- E' stata creata una Operational Data Store che contiene la movimentazione giornaliera dei sistemi sorgenti cosicchè i volumi dei dati da trasportare verso il DWH si sono ridotti notevolmente

- Ridotti i volumi, la finestra temporale si è accorciata, rendendo possibile per gli utenti finali di consultare i propri report la mattina del giorno dopo

## I benefici:

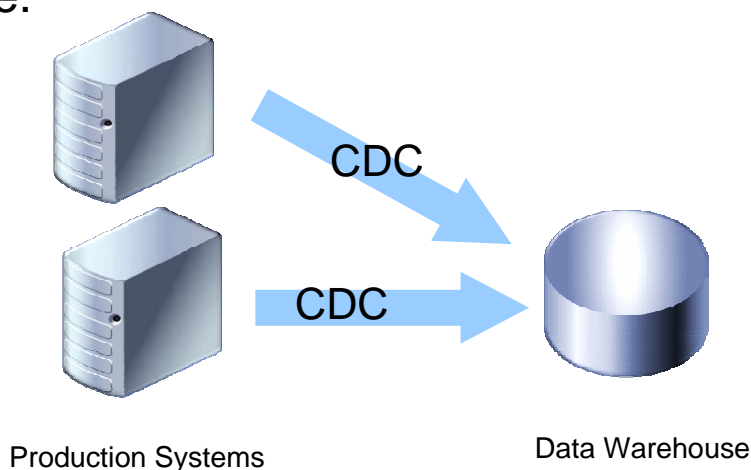
- Ricoh ha potuto ridurre il tempo totale di processing e il volume dei dati, grazie all'estrazione incrementale dei dati stessi
- La soluzione non ingenera il degrado delle prestazioni dei sistemi sorgenti, preservando l'integrità delle normali operazioni di business
- La crescita del DWH non è per il momento motivo di preoccupazione
- Per alcune metriche, è stato possibile avere report aggiornati in tempo reale (esigenza in crescita per il Cliente)



# Complementary to ETL : Once a Day is Not Enough



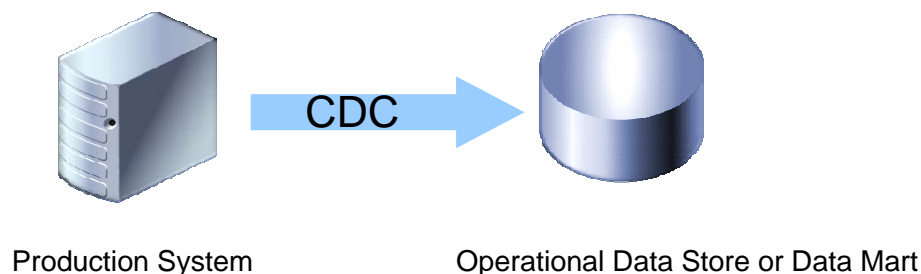
- Problem: Day old data is too stale to meet the business need.
  - Customer needs to analyze shipping, order, inventory information continuously throughout the day.
  - **BUT** – doesn't want to impact production systems by running queries or reports directly against them.
- Solution: Transformation Server to provide continuous ETL flow into Data warehouse.



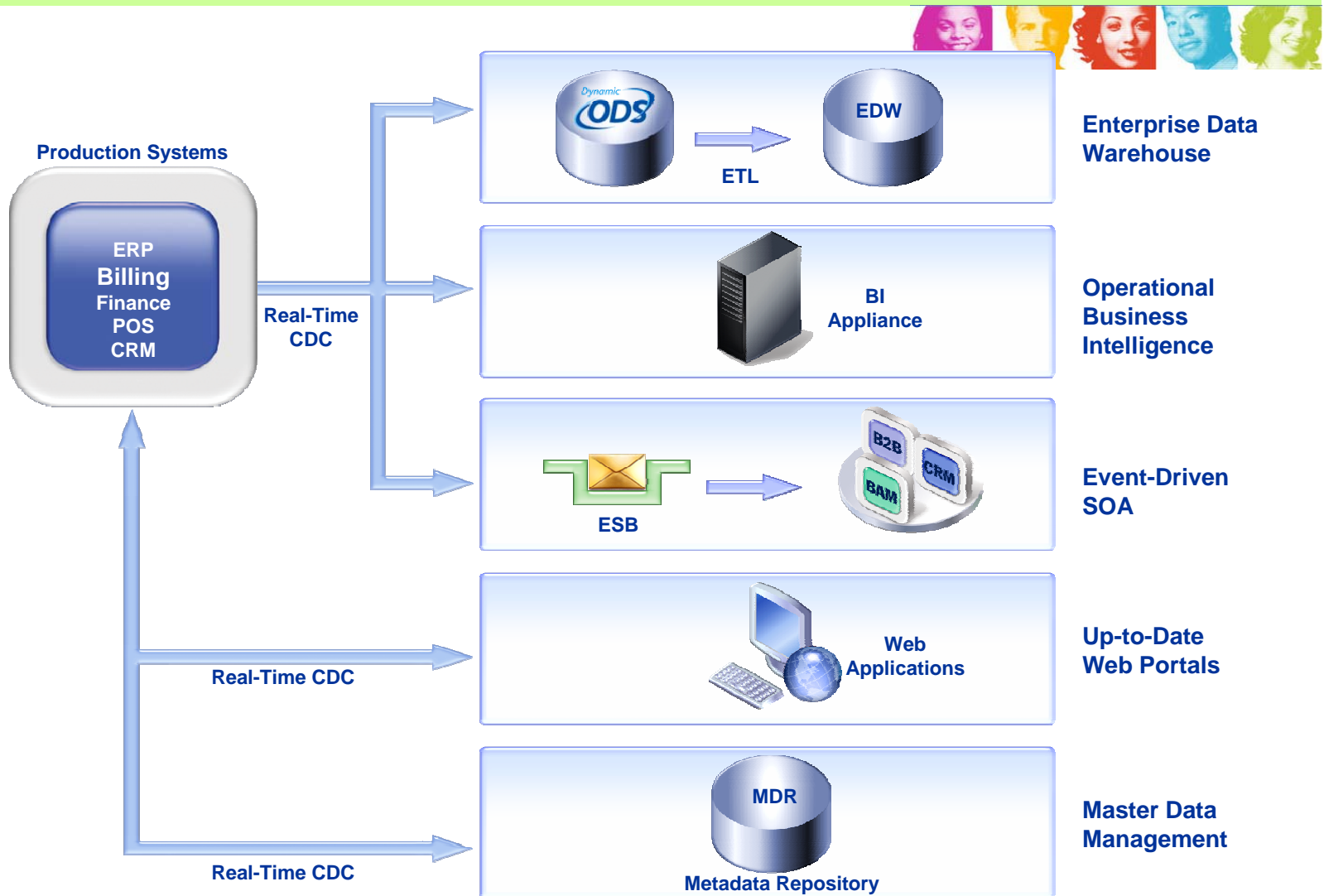
# Complementary to ETL : My Production System is Too Maxed out for Reporting



- Problem: Users running queries and reports on production systems are slowing them down.
  - Customer doesn't have a warehouse but is running queries and reports directly against the production system which is impacting system performance
  - Wants a quick, easy, cheap way to offload queries and reporting onto a secondary system or Data Mart
- Solution: Transformation Server to provide easy synchronization to secondary system.



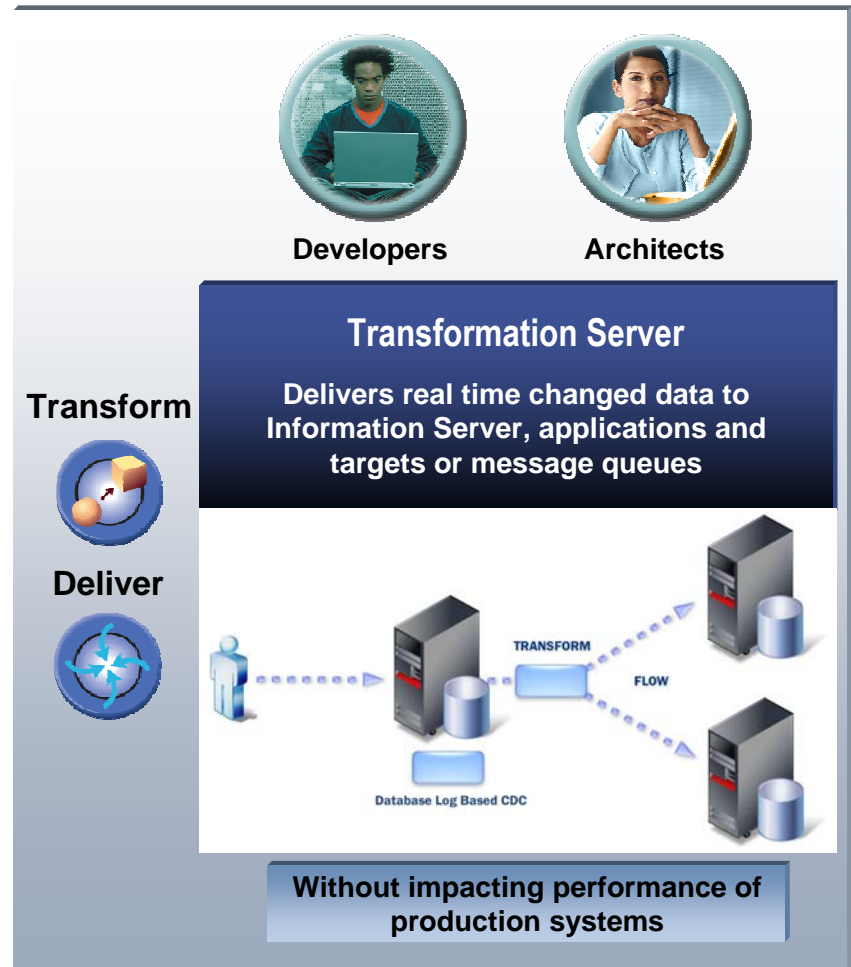
# Business Application



# What is Information Server CDC?



- Provides real time changed-data capture and delivery for
  - Dynamic changed data integration
  - Synchronization
  - Replication
- Minimal impact on production systems
- High scalability and end-to-end performance
- Wide range of RDBMS support
  - Captures for DB2 (i/z/LUW), Oracle, SQL Server, Sybase
  - Applies for these above and more



# Heterogeneous Platform Support



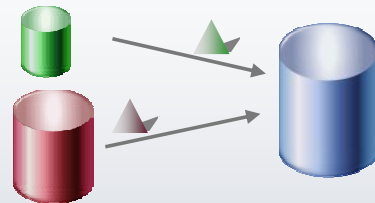
<b>DATABASES Source &amp; Target</b>	<b>DATABASES Target Only</b>	<b>MESSAGE QUEUE Support</b>	<b>OPERATING SYSTEMS</b>	<b>HARDWARE PLATFORMS</b>	<b>NETWORK PROTOCOLS</b>
DB2 / UDB	Teradata	JMS	i5/OS	IBM System i	TCP/IP
Oracle	Netezza	MQ Series	z/OS	IBM System z	
Sybase	Greenplum	TIBCO	AIX	IBM System p	
MS SQL Server	Informix	WebMethods	HP-UX	HP-9000	
PointBase	MySQL	BEA	Solaris	Intel	
	PostgreSQL		MS Windows	HP Alpha	
			Linux	Sun	
				Itanium	



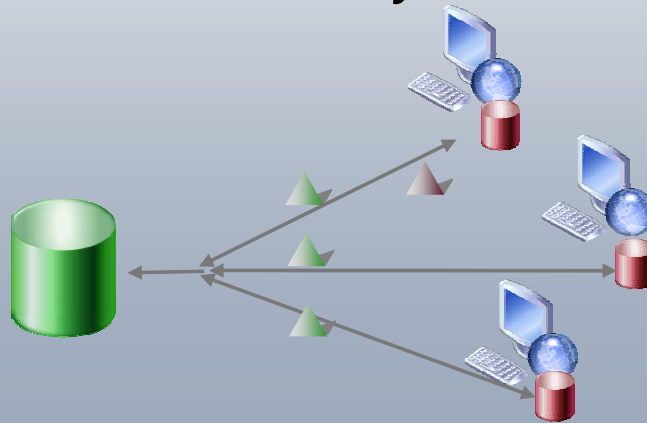


## Real-Time Integration

*Consolidation*



*Distribution Synchronization*

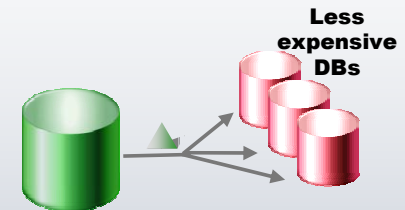
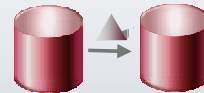
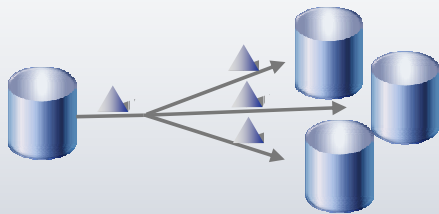


# Deployment Topologies

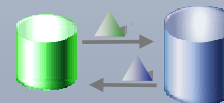
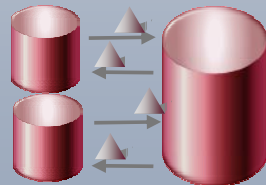


## Replication

*Need to report on operational systems but cannot impact production system – must synchronize data across 1 or more databases*



*Need to maintain business continuity during application migrations, consolidations, upgrades*



# Guaranteed Data Integrity



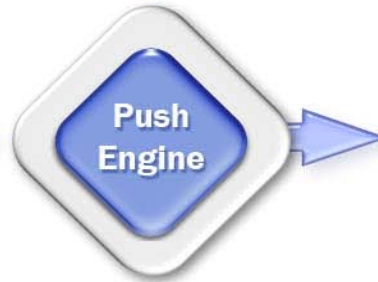
## 1. Scrape

Journal/Log



Source Tables

## 2. Push



## 3a. Apply



Target Tables

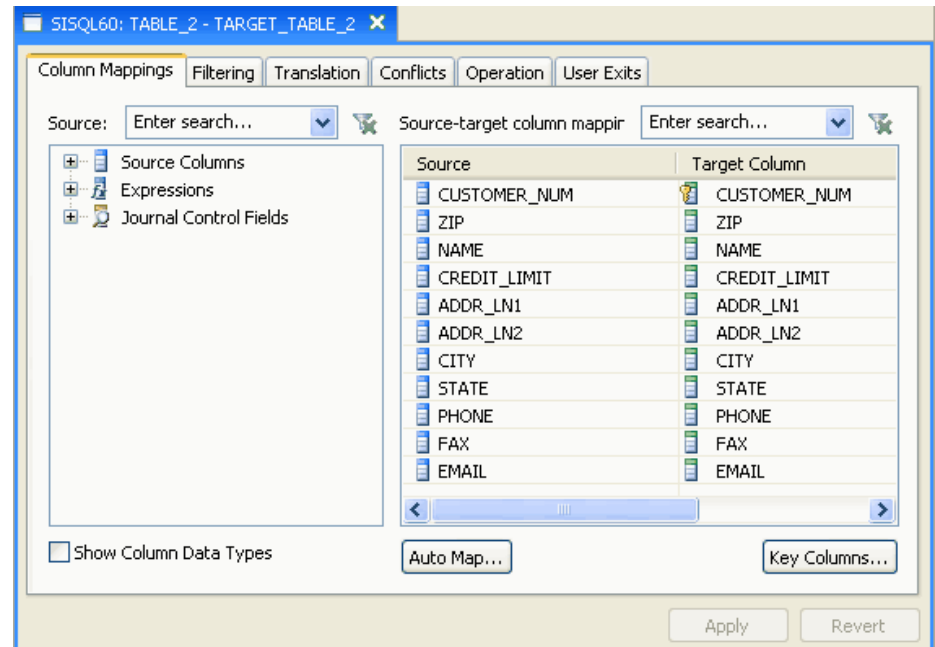
Metadata  
What was the last transaction you successfully applied?

## 3b. Confirm

# Subscription - Assignments



- Assignments
  - individual source to target table mappings
  - is associated with a subscription
  - Table and Column level configuration



# 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

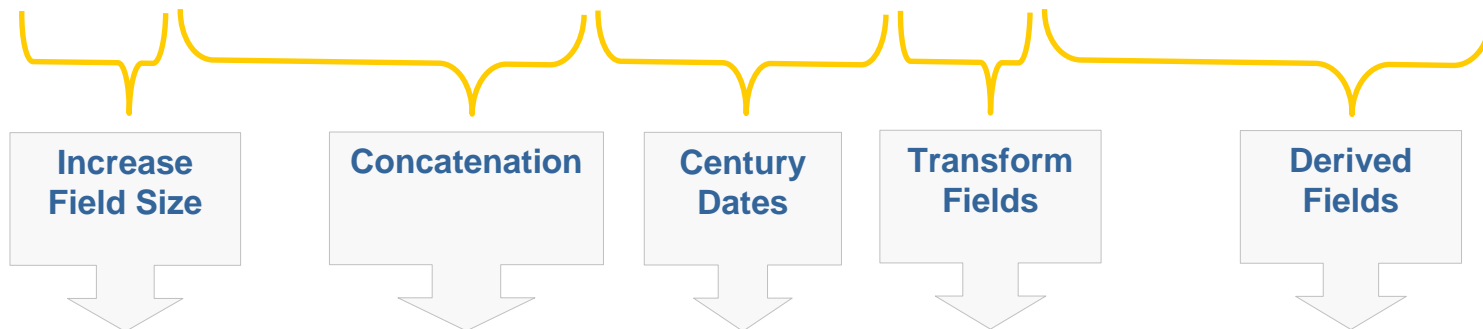


CUST_NO	L_NAME	F_NAME	REP_NO
37283	Duggan	Ira	25
89732	Muntz	Josie	25

# Transformations



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



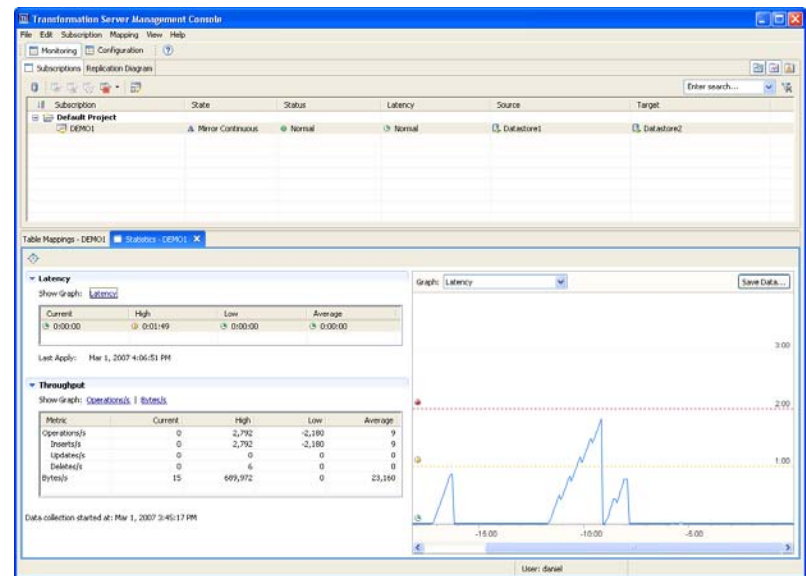
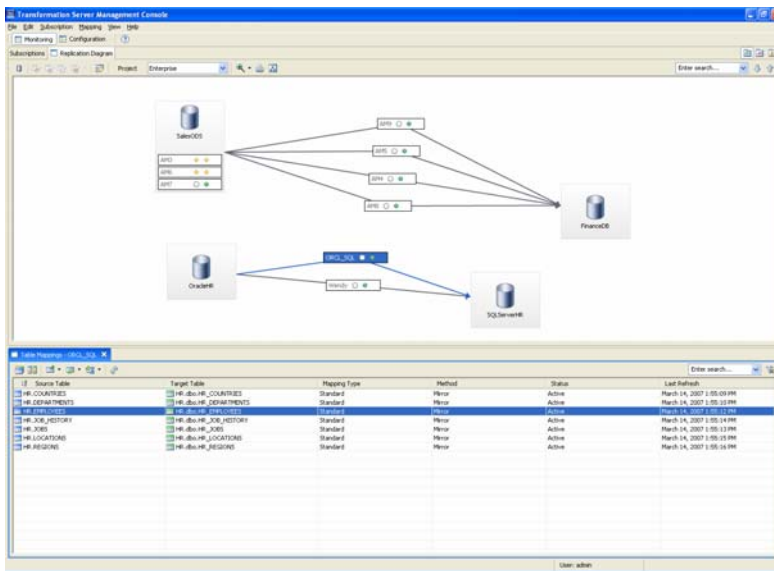
EMP_ID	FULL_NAME	HIRE_DATE	STATUS	%SALARYMAX
001234	Nicole Moreiro	01/05/1997	Active	92%
002345	Val Ellison	04/12/1997	Inactive	80%

# Monitoring & Ease of Use



Java based GUI for configuration, administration, and monitoring

- Manage data integration processes from one screen
- Graphical representation of the replication process
- Event logs and Notifications

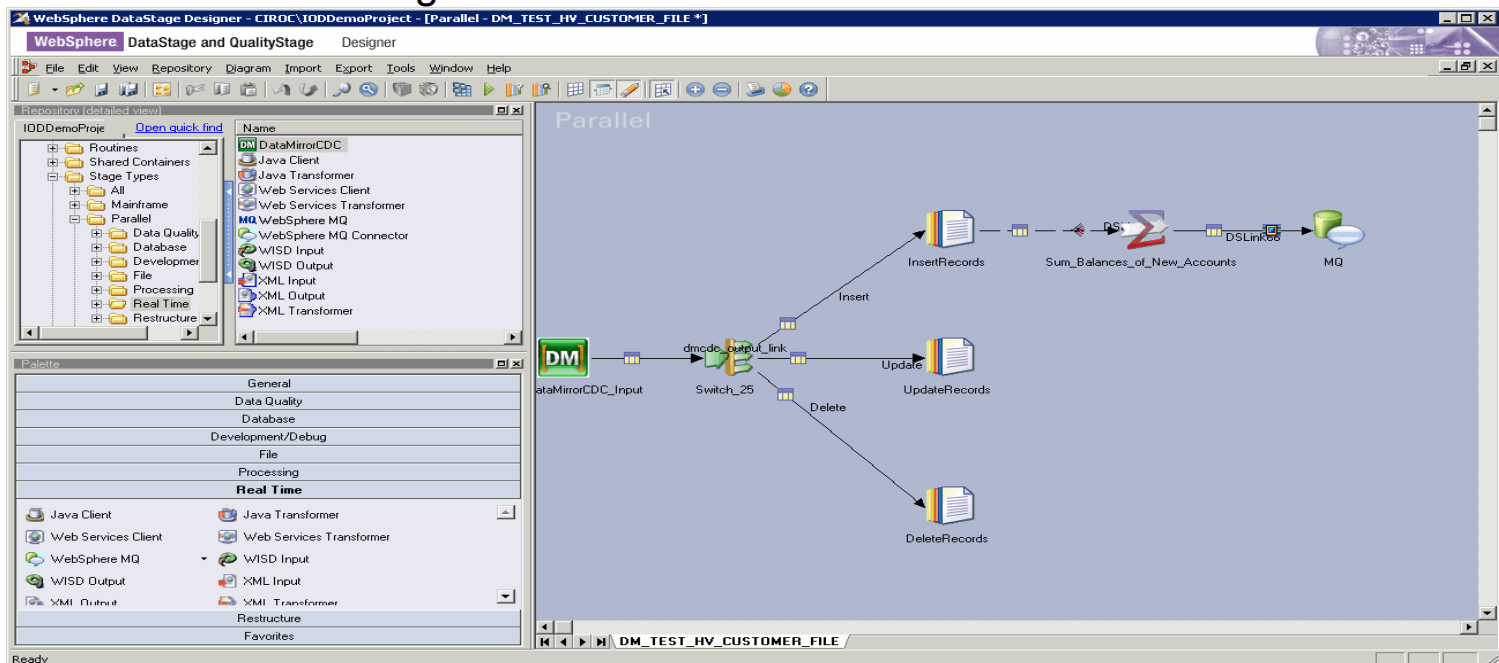


# Information Server CDC 6.2 for WebSphere DataStage



Enabling real-time response to data changes and business events

- Low impact log-based changed data capture
- New palette stages on Information Server
- Stream data changes into Information Server

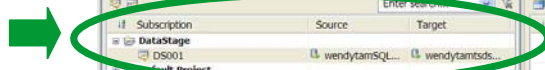




# Create Subscription & Table Mapping



DataStage  
subscription  
with  
database  
source and  
DataStage  
target



The screenshot shows the 'Transformation Server Management Console' with the 'Table Mappings - DS001' window open. The window contains a table with columns: Source Table, Target, Mapping Type, Method, and Status. A row is highlighted with a green oval. A green arrow points from the text 'DataStage subscription with database source and DataStage target' to this row. Another green arrow points from the text 'Table Mapping Wizard' to the 'Map Tables' dialog box. The dialog box has a title bar 'Map Tables' and a section 'Select Mapping Type'. It contains two options: 'Automatic' with 'WebSphere DataStage' selected, and 'Custom' with 'One table mapping' selected. The 'Mapping Type' dropdown is set to 'WebSphere DataStage'. Buttons for '< Back', 'Next >', 'Finish', and 'Cancel' are at the bottom.

Source Table	Target	Mapping Type	Method	Status
wendytamsQL...	wendytamsds...			

**Map Tables**

**Select Mapping Type**

Depending on your replication requirements, select from one of the following mapping types to apply data to the target.

**Automatic:**

- WebSphere DataStage**  
Map one or more source tables for use as inputs to WebSphere DataStage jobs.

**Custom:**

- One table mapping**  
Create a single table mapping for use as input to a WebSphere DataStage job.

Mapping Type: WebSphere DataStage

< Back Next > Finish Cancel

# Table Mapping – Connection Methods (continued)



Transformation Server Management Console

Subscription	Source	Target	Mapping Type	Method	Status
DS001	wsdytarn02.q.10101	wsdytarn02.q.10401			
DS002	wsdytarn02.q.10502	wsdytarn02.q.10401			
DS003	wsdytarn02.q.10503	wsdytarn02.q.10401			
DS004	wsdytarn02.q.10504	wsdytarn02.q.10401			

Select WebSphere DataStage Connection Method

Depending on your replication requirements, select from one of the following WebSphere DataStage connector methods.

Flat File  
Use files to deliver the source changes to WebSphere DataStage.

Direct Connect  
Use the WebSphere DataStage Transformation Server Connector stage.

Flat File

Direct Connect

Map Tables

**WebSphere DataStage Flat File**

Specify the output directory and record format for the WebSphere DataStage flat files.

Location

Directory: C:\testflatfiles

Record Format

Single Record  
Include a set of columns for the before and after images in a single record.

Multiple Records  
Include one set of columns in each record. For updates, a separate record will be generated for the before and after images.

< Back Next > Finish Cancel

Map Tables

**WebSphere DataStage Direct Connect**

Specify the WebSphere DataStage direct connect host and port and choose the output record format.

Connection

More than one source table has been selected. Each source table requires a unique port number. Specify the starting port for the 3 selected tables.

Host: localhost

Port: 7001 to 7003

Record Format

Single Record  
Include a set of columns for the before and after images in a single record.

Multiple Records  
Include one set of columns in each record. For updates, a separate record will be generated for the before and after images.

< Back Next > Finish Cancel

# Generate WebSphere DataStage Definition



The screenshot shows the IBM DataMirror Transformation Server Management Console interface. The main window is titled "IBM DataMirror Transformation Server Management Console" and has a menu bar with "File", "Edit", "Subscription", "Mapping", "View", and "Help". Below the menu bar are tabs for "Monitoring" and "Configuration". The left pane shows a tree view of "Subscriptions" and "Datastores". The right pane shows "Table Mappings - DS010" with a table of source and target tables.

Source Table	Target	Mapping Type
DBADMIN.TABLE_1	WebSphere DataStage	Flat File
DBADMIN.TABLE_2	WebSphere DataStage	Flat File
DBADMIN.TABLE_3	WebSphere DataStage	Flat File
DBADMIN.TABLE_4	WebSphere DataStage	Direct Connect
DBADMIN.TABLE_5	WebSphere DataStage	Direct Connect

A context menu is open over the "Subscriptions" tree, with the option "Generate WebSphere DataStage Job Definition..." highlighted. The menu items include:

- Map Tables...
- Notification...
- Latency Thresholds...
- Change Refresh Order...
- WebSphere DataStage Properties...
- Static SQL...
- Copy Subscription...
- Promote Subscription...
- Import Subscription...
- Export Subscription...
- Delete Subscription
- Generate WebSphere DataStage Job Definition...
- Project
- Start Refresh...
- Start Mirroring (Continuous)
- Start Mirroring (Net Change)
- End Replication
- Properties...

# POV presso una società telefonica

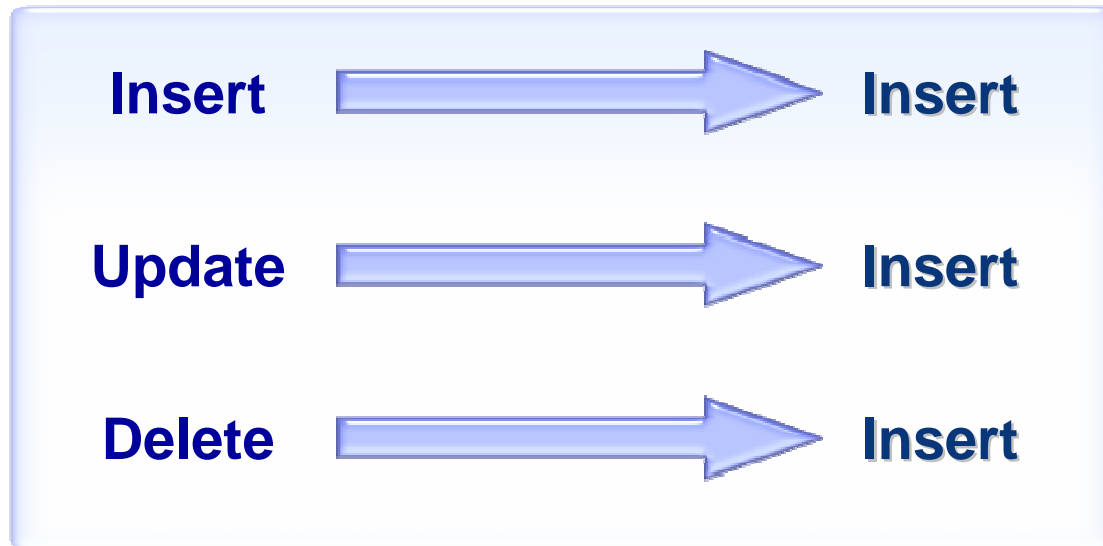
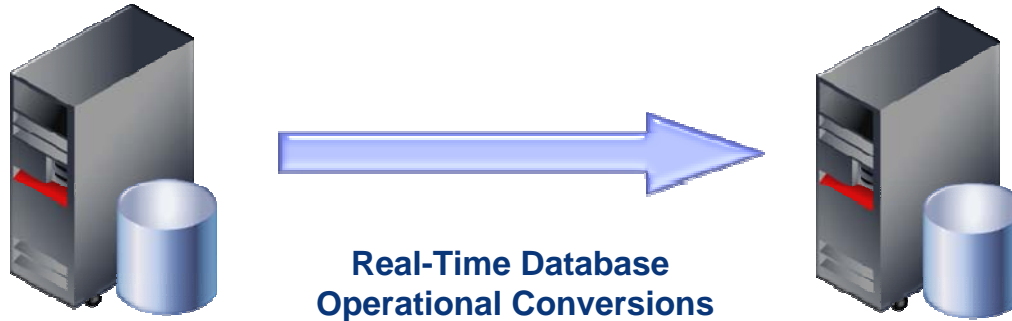


- Versatilità dell'implementazione
  - Il fine ultimo è stato quello di creare unicamente il flat file. Non era richiesta l'integrazione con DataStage.
  - Esempio di Tracciato Record:  
"2008-02-28 17:29:40","663494","I","POC",,,,"5","mariano"  
"2008-02-28  
17:31:31","269599","U","POC","5","mariano","5","monica"  
"2008-02-28 17:35:43","531801","I","POC",,,,"2","1111"



- Real-time data auditing solution
  - Captures data from production systems without impacting performance
  - Creates an audit trail and applies data to target systems in real time
  - Designed to work in compliment with Transformation Server
- Audit trails tracks information such as:
  - Who, what, when, and how data was changed
  - Preserves historical database information
  - Audits metadata as well:
    - Addition/Deletion of tables and columns
    - Changes to database properties

# How TS/LiveAudit Works



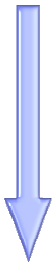
# Example



Application Database

Product ID	Action	Qty
Drug001	Make	1000
Drug001	Calibrate Test Eqmt	-
Drug001	Test Initiated	1000
Drug001	Test Result: Passed	
Drug001	Bottle	1000
Drug001	Ship	1000

Regulations require that data is not manipulated and also records that are deleted cannot be reported on to make accurate decisions



LiveAudit Database

Date / Time	Action	User	Prd ID	Event	Qty
05/31/01-0800	I	jwalker	Drug001	Make	1000
05/31/01-1300	I	jwalker	Drug001	Calibrate Test Eqmt	-
05/31/01-1500	I	jwalker	Drug001	Test Initiated	1000
06/01/01-0800	I	jwalker	Drug001	Test Result: Particles Found	1000
06/01/01-0900	D	jwalker	Drug001	Particles Found	
06/01/01-1100	U	swilson	Drug001	Test Initiated	1000
06/02/01-0800	U	swilson	Drug001	Test Result: Pass	1000
06/01/01-1600	I	jwalker	Drug001	Bottle	1000
06/05/01-0800	I	jwalker	Drug001	Ship	1000



LiveAudit replicates production data to an auditing database and can create new database columns to capture additional data

- Type of data change made, 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
&TIMESTAMP	Time of the update or refresh.
&USER	The user ID which made the update.



# Data Backup And Availability



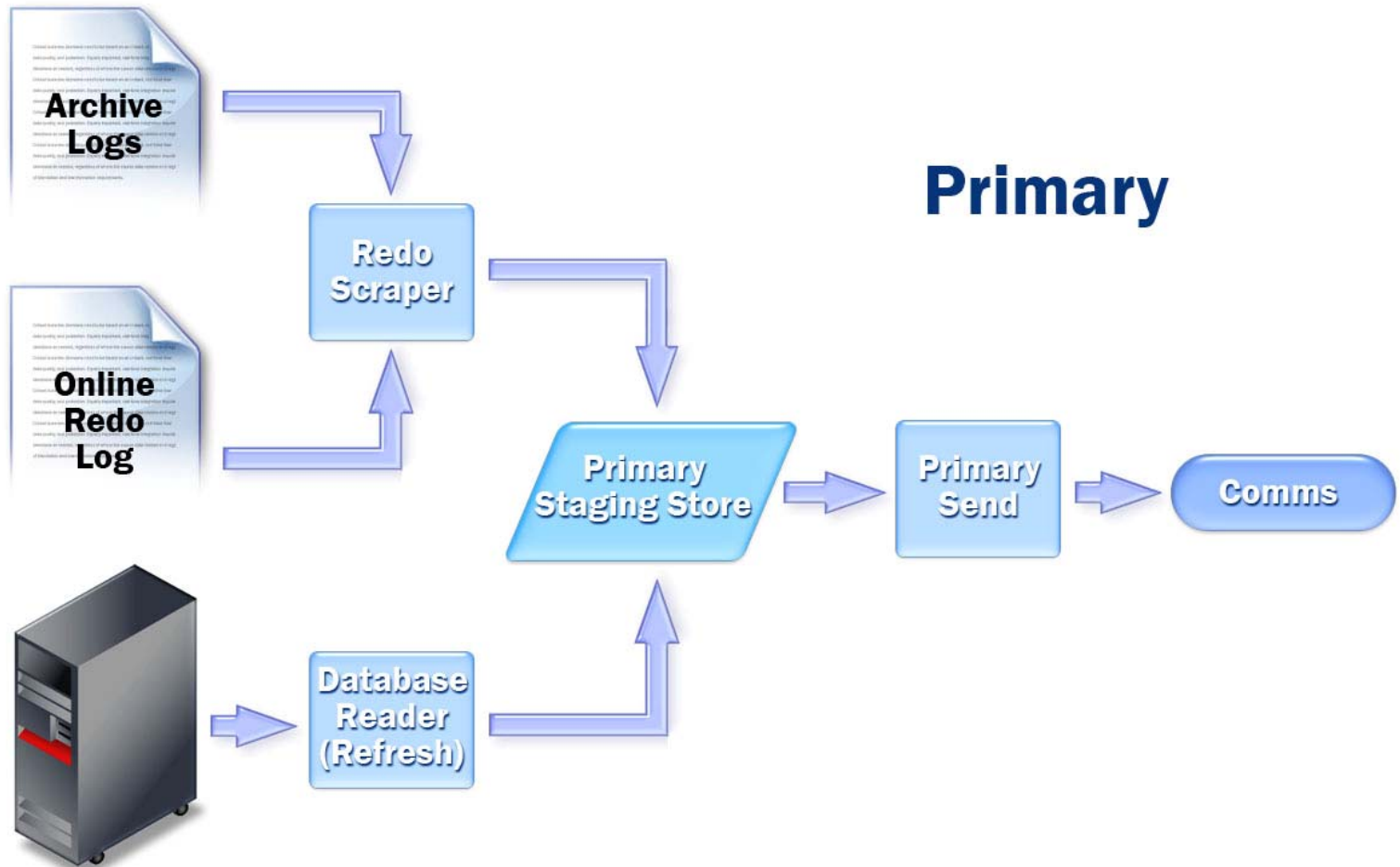
*“Solution deployed to allow organizations to backup copies of critical data for recovery where a full disaster solution is not a requirement”.*



**Feature:**  
Redo and Archive Logs

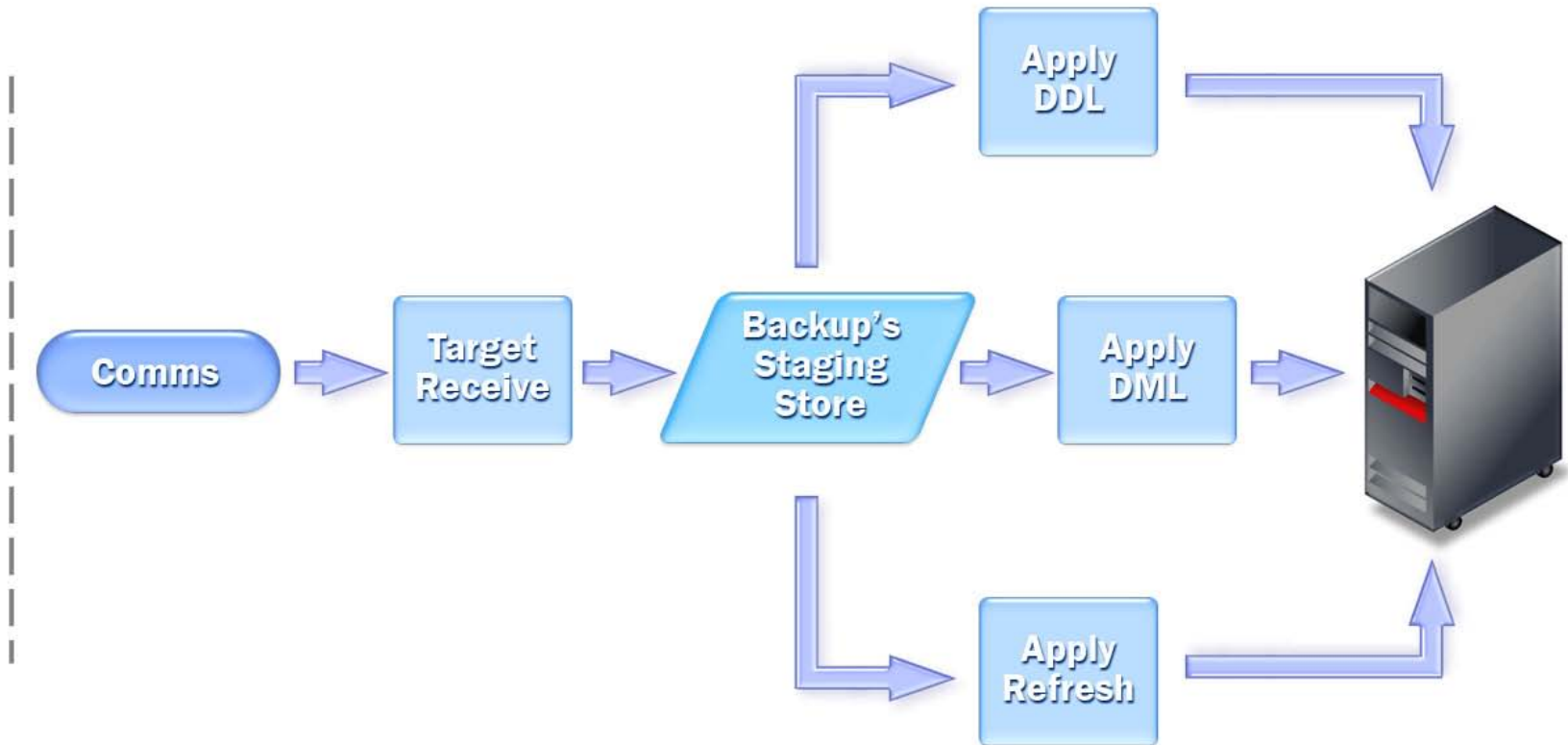
**Business Benefits:**  
Real-Time, Minimum  
Resource Utilization

# Architecture





## Backup



# DML and DDL Replication

*Oracle9i*



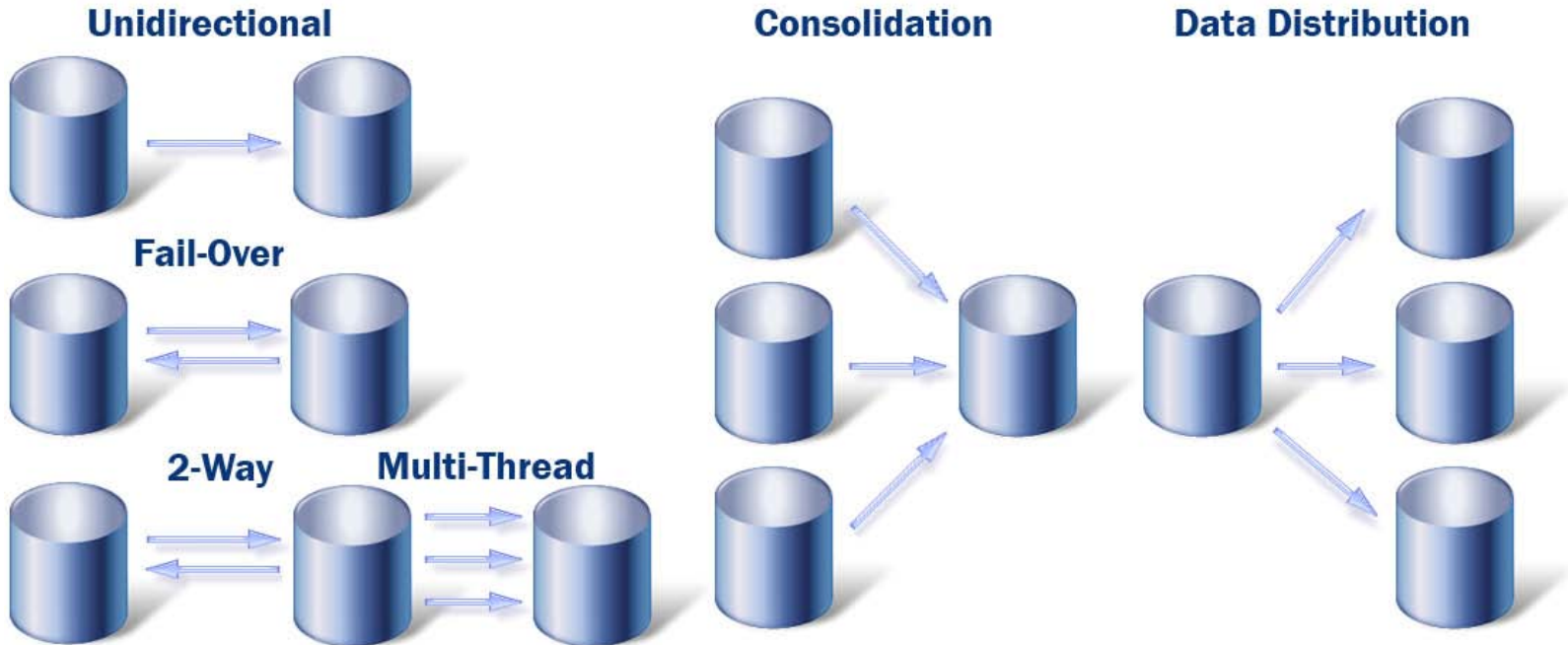
**Refresh and Mirroring**  
**Support for DDL**

**Feature:**  
Database Object/Schema  
Change Support (DDL)

- Table
- Sequence
- Context
- Directory
- Dimension
- Function
- Index
- Library
- Materialized View
- Package
- Trigger
- Procedure
- Profile
- Resource Cost
- Role
- Rollback Segment
- Synonym
- Tablespace
- User
- View

**Business Benefits:**  
Reduced User Errors and  
Admin, Better Data Integrity

# Flexible Implementation



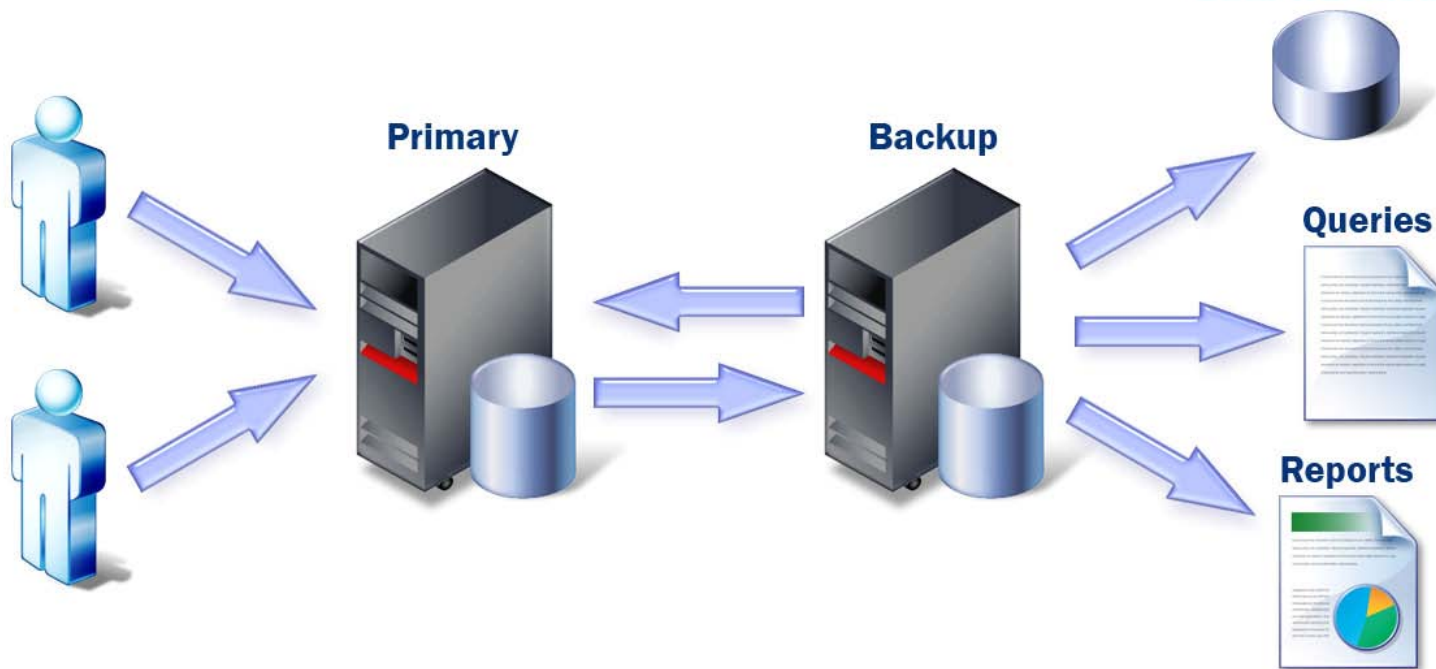
**Feature:**  
Extreme Scalability

**Business Benefits:**  
Non-intrusive / Flexible Arch.,  
Geographic Dispersion

# Workload Balancing



**Data Warehouse**



**Feature:**  
Workload Distribution

**Business Benefit:**  
Extended Resource  
Capabilities

# Master-to-Master Replication

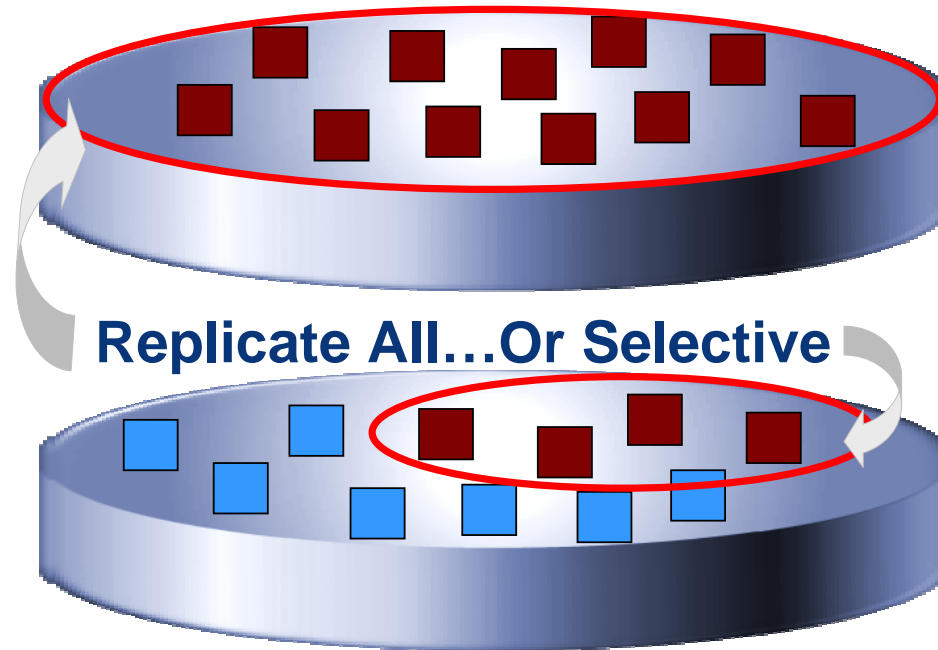


- Supports any number of masters
- Supports DDL and DML changes
- Distributed Database Applications

Property	Value
Group Name	G1
States	Active, Active, Active
Operational Status	Normal, Normal, Normal
Description	
Object Specifiers	Double click to see Object Specifiers

Request Done 9:32 AM

# Selectable Replication



**Feature:**  
Selective Data /  
Object Replication

**Business Benefit:**  
Flexible to Business Needs



# Java-Based Admin GUI



The screenshot shows the iReflect Administrator GUI. On the left is a tree view of the iReflect Hierarchy. In the center, a 'Set Role' dialog box is open, showing fields for Node, Database, and Target Name, with radio buttons for Primary and Backup. On the right, a table displays properties for the target 'SUNGSCOTT'.

Property	Value
Target Name	SUNGSCOTT
Role	Primary
Status	Active
Primary Host Name	dmsund
Primary Port	11122
Primary Database	dghda
Backup Host Name	dmsung
Backup Port	11122
	dghda
	dghda
	N
	0
	0
	0
	Scott Schema HA
	SCOTT SCHEMA

**Feature:**  
Event Driven Administrator

**Business Benefits:**  
Easier / More Efficient Admin.

# Event Viewer Java GUI



iReflect Event Viewer-dmcsund

File Edit View Command Window Help

Time	Source	Operation	Target
1/31/02 9:51:10 AM	dghda	Executive	DRE1... Infor... Proce
1/30/02 3:10:04 PM	dghda	Executive	DRE8... Oper... Target
1/30/02 3:09:43 PM	dghda	Executive	DRE8... Oper... The sc
1/30/02 3:09:42 PM	dghda	Executive	DRE8... Oper... The sc
1/30/02 3:09:42 PM	dghda	Scrape	DRE5... Infor... Redo l
1/30/02 3:09:42 PM	dghda	Executive	DRE8... Oper... The sc
1/30/02 3:09:42 PM	dghda	Executive	DRE8... Oper... The sc
1/30/02 3:09:42 PM	dghda	Executive	DRE8... Oper... The sc
1/30/02 2:55:22 PM	dghda	Executive	DRE8... Oper... Group
1/30/02 2:55:22 PM	dghda	Executive	DRE8... Oper... The sc
1/30/02 2:55:22 PM	dghda	Scrape	DRE5... Infor... Redo l
1/30/02 2:55:21 PM	dghda	Scrape	DRE5... Infor... Starte
1/30/02 2:55:21 PM	dghda	Scrape	DRE5... Infor... New s
1/30/02 2:55:21 PM	dghda	Scrape	DRE5... Infor... The ds
1/30/02 2:55:21 PM	dghda	Scrape	DRE5... Infor... The ds
1/30/02 2:55:21 PM	dghda	Scrape	DRE5... Infor... Startin
1/30/02 2:55:21 PM	dghda	Scrape	DRE5... Infor... The ds
1/30/02 2:55:20 PM	dghda	Scrape	DRE5... Infor... Startin
1/30/02 2:55:20 PM	dghda	Scrape	DRE5... Infor... The ds
1/30/02 2:55:19 PM	dghda	Scrape	DRE5... Infor... Startin
1/30/02 2:55:19 PM	dghda	Scrape	DRE5... Infor... The ds
1/30/02 2:55:19 PM	dghda	Scrape	DRE5... Infor... Startin
1/30/02 2:55:19 PM	dghda	Scrape	DRE5... Infor... The ds
1/30/02 2:55:14 PM	dghda	Scrape	DRE5... Infor... Startin
1/30/02 2:55:14 PM	dghda	Scrape	DRE5... Infor... The ds
1/30/02 2:55:08 PM	dghda	Scrape	DRE5... Infor... Startin
1/30/02 2:55:08 PM	dghda	Scrape	DRE5... Infor... The ds

Detailed Message

Database: dghda Target Name: SUNGSCOTT  
Group Name: SCOTT Zone: Scrape  
Event Type: DRE5118 Severity: Informational  
Timestamp: 1/30/02 2:55:22 PM

Message Text

Redo log scraping started at position '506430.863.204.0' timestamp 'Wed Jan 30 14:54:51 2002'.

Close Prev Next Help

10 40 0 0 Total:50

Display Zones: Executive, Scrape, Send/Receive, Apply

Database: \* Target Name: \* Group Name: \*

Display Severities: Error, Warning, Informational, Status, Operational

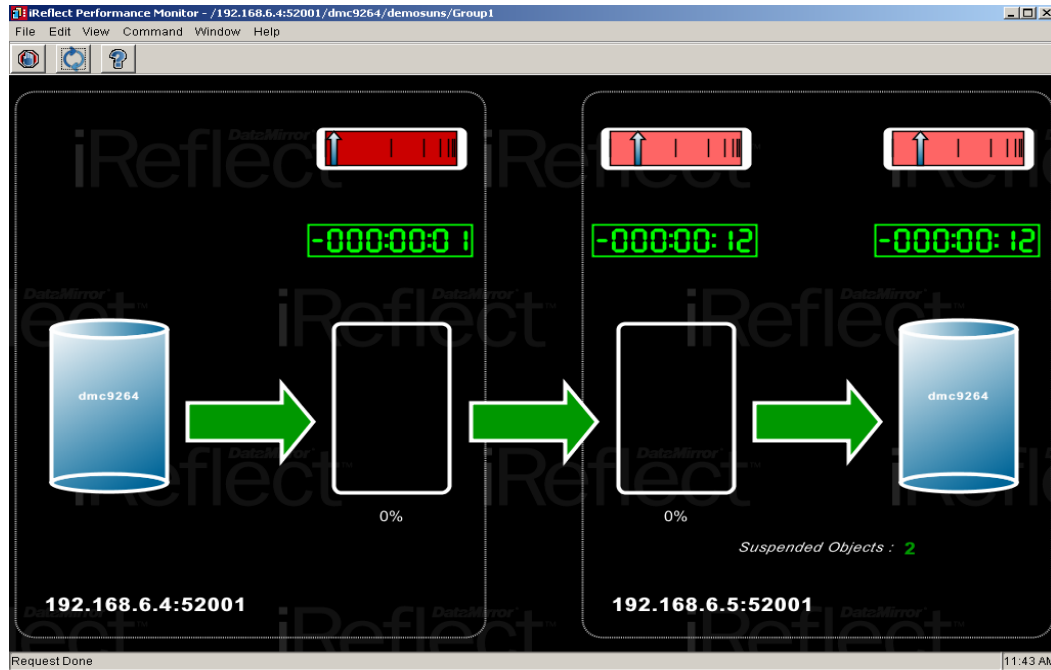
Time Range: From Time: To Time:

Request Done 9:52 AM

**Feature:**  
Event Viewer

**Business Benefits:**  
Easier / More Efficient  
Troubleshooting

# Digital Graphical Monitor



**Feature:**  
Digital Monitor

**Business Benefits:**  
Easier / More Efficient  
Monitoring



# Thank YOU

IBM Information  
On Demand 2008  
>>> Comes To You

ALLA LUCE DELL'INFORMATION ON DEMAND

Milano, 15 aprile 2008

