



Putting Data Where You Need It: The Options Session Number 2708

Kim May, The Fillmore Group

Information On Demand 2010

The Premier Forum for Information & Analytics

Gain Insight. Optimize Results.



Session 2708: Putting Data Where You Need It: The Options

- The Fillmore Group
 - IBM Premier Business Partner
 - Working with data since 1987
 - Consulting, IBM Authorized education and software
 - Data replication and interoperability specialty practices

THE FILLMORE GROUP
Relational Database Solutions



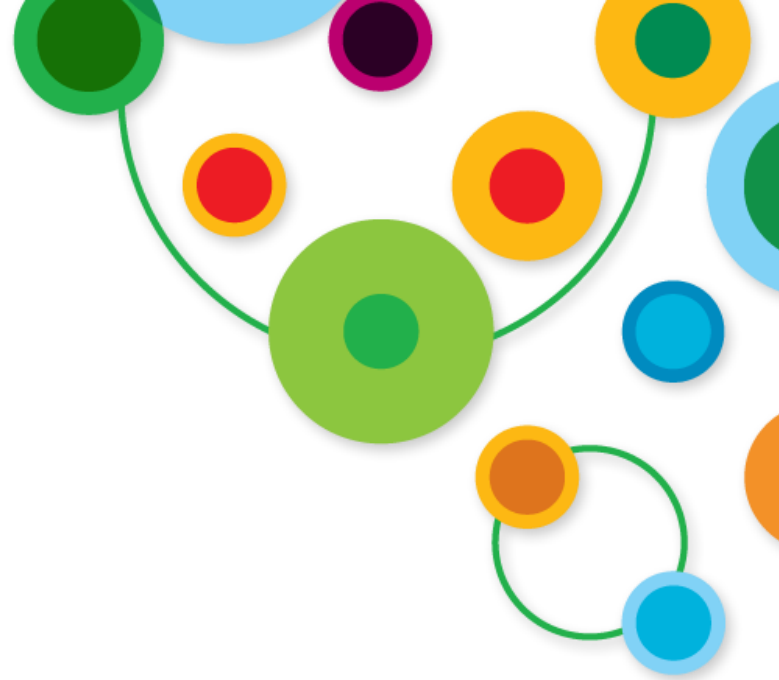
What is Replication?

- **Business Continuity**
- **Disaster Recovery**

- **Data Replication**

From the Information Management Glossary:

The process of copying a portion of a database from one environment to another and keeping the subsequent copies of the data in sync with the original source. Changes made to the original source are propagated to the copies of the data in other environments.

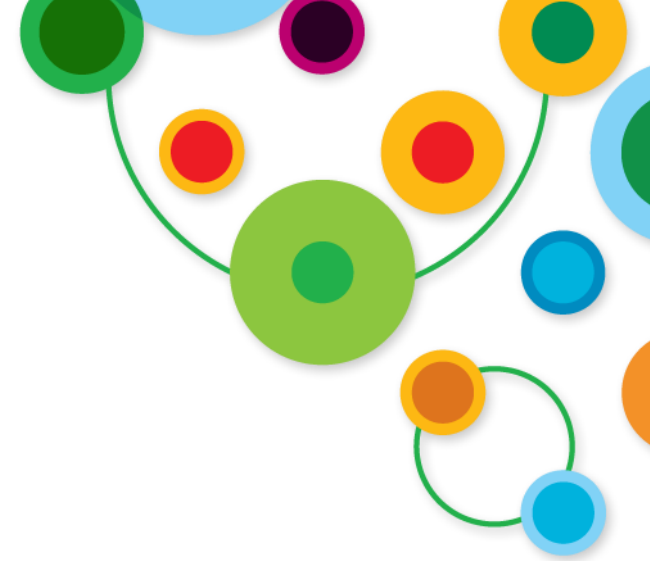


Why Replication?

- ✓ **Disaster Recovery**
- ✓ **Load Balancing**
- ✓ **Data Mart**
- ✓ **Data Warehouse**



Replication



History



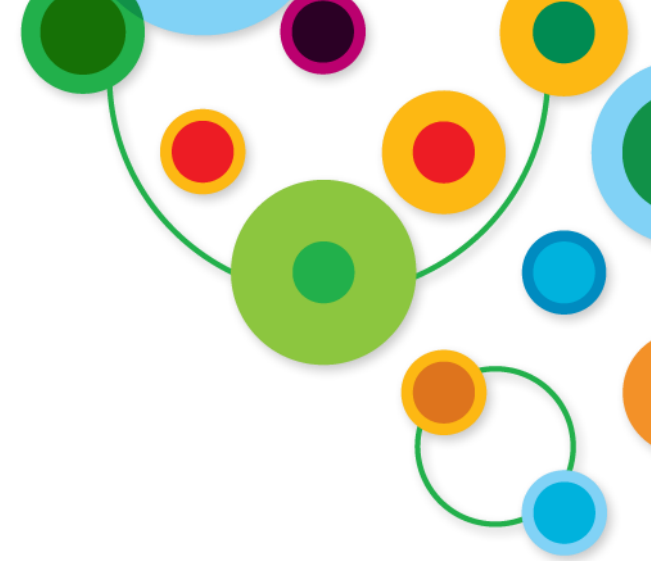
- **SQL Replication** – released in the mid-1990's as DataPropagator Relational
- **Queue Replication** - released along with WebSphere Information Integrator V8.2 in 2005, incorporating MQ and SQL Replication to deliver high-speed replication
- **DataMirror/ICDC** – IBM acquired DataMirror in 2007 and rebranded DataMirror Transformation Server as InfoSphere Change Data Capture (ICDC) in 2008.

SQL Replication

Q Replication

ICDC

Options We Aren't Considering Today



pureScale



SAN and hardware-based storage (Tivoli, StorServer)



High Availability Disaster Recovery (HADR)





Terminology

- **Latency**
 - The time it takes for data to get from one point to another
 - Synonymous with delay; measured in microseconds, seconds, minutes
- **Source and Target databases**
 - The “source” database on which data is initially stored and the “target” is the database to which data is replicated
- **Capture and Apply**
 - Terms used to describe the process of collecting and delivering changed data
- **Transport**
 - The methodology used to move the data from source to target
- **Uni- directional, bi-directional and peer-to-peer**
 - Description of data movement from either source to target (uni) or from both source to target and target to source (bi), or to n-tier





Evaluating the Options

- Document and prioritize your needs - factors to consider:
 - Source and target databases
 - Are the source and target homogeneous or heterogeneous?
 - Uni-directional, bi-direction or peer-to-peer replication?
 - Speed and latency
 - Is there an SLA for replication speed?
 - What is the data volume on the source and target databases?
 - GB/TB
 - What is your anticipated transaction volume?
 - INSERTS/UPDATES/DELETES





Evaluating the Options (con't)

— Resiliency

- Are outages acceptable? How often? For how long?
- Do you need automated failover in the event of an unplanned outage?

— Ease of use

- Monitoring, installation, administration
- How frequently do you change your data model?
- How complex is conflict resolution? Is it based on:
 - Value/Source/Timestamp/Application Logic

— Cost

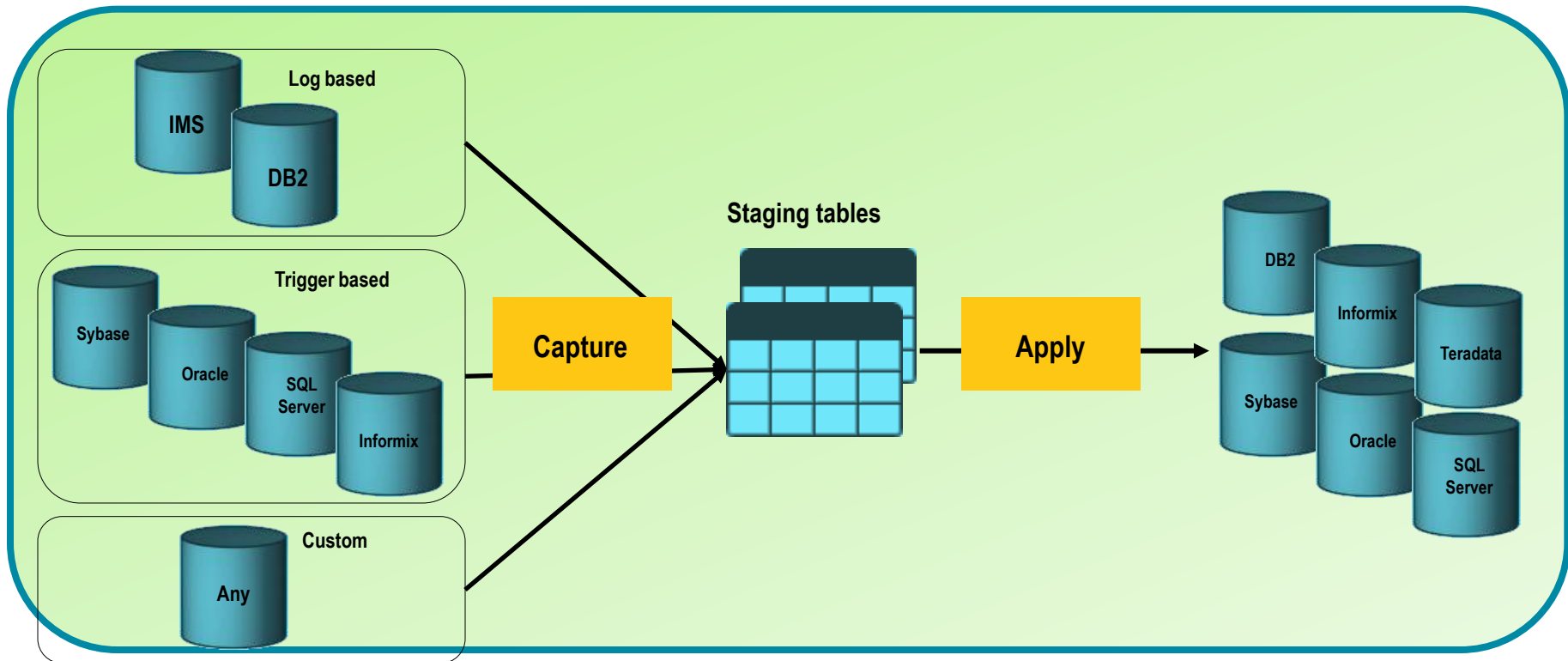
- What is your budget for software, implementation and training?
- What is your budget for annual maintenance?



SQL Replication

- Sources: DB2 (all platforms), Oracle, SQL Server, Sybase, Informix
- Targets: DB2 (all platforms), Oracle, SQL Server, Sybase, Informix
- Uni-direction, bi-directional, peer-to-peer replication supported
- Latency – 3X slower than Q Replication; not recommended for very large transaction volumes
- Resiliency – works with HADR
- Ease of use – Monitoring console
- **Cost - *FREE* with DB2 for LUW; Capture component included with DB2 for z/OS but requires purchase of Apply; may need InfoSphere Federation Server for heterogeneous targets.**

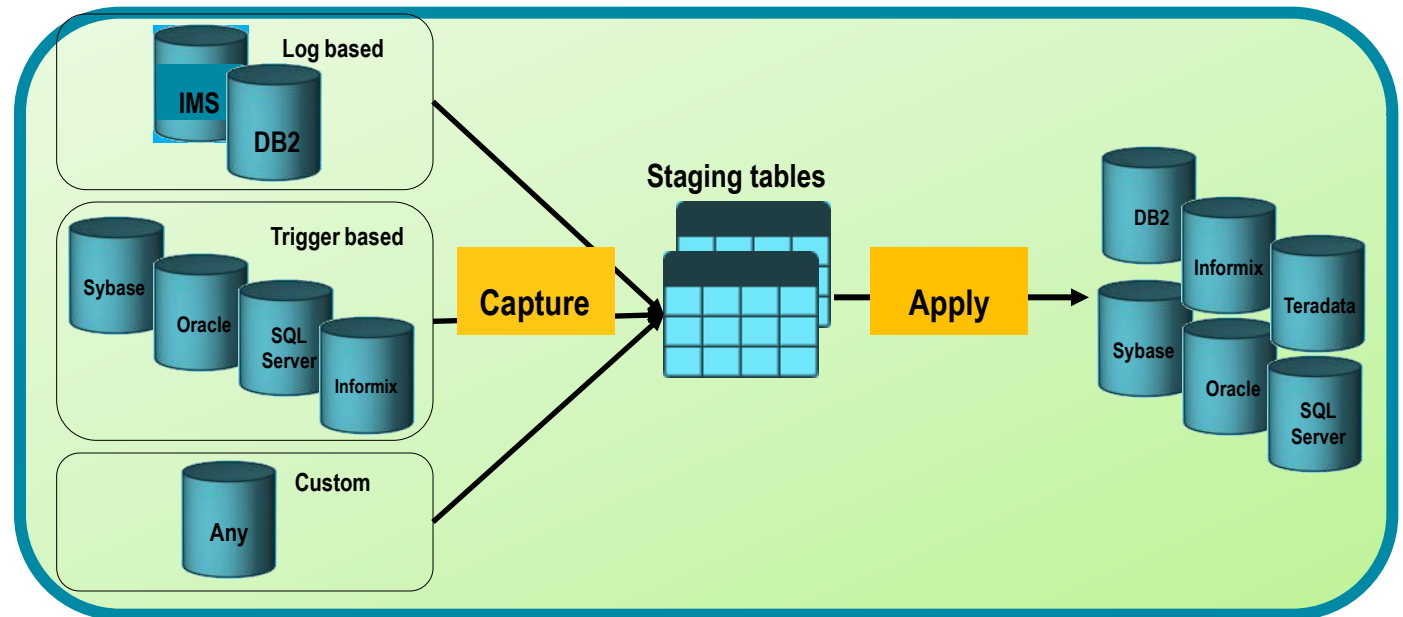
SQL Replication



SQL Replication

Important

- Triggers are needed when the source is not DB2.
- InfoSphere Federation Server is needed when the target is not DB2.



Q Replication

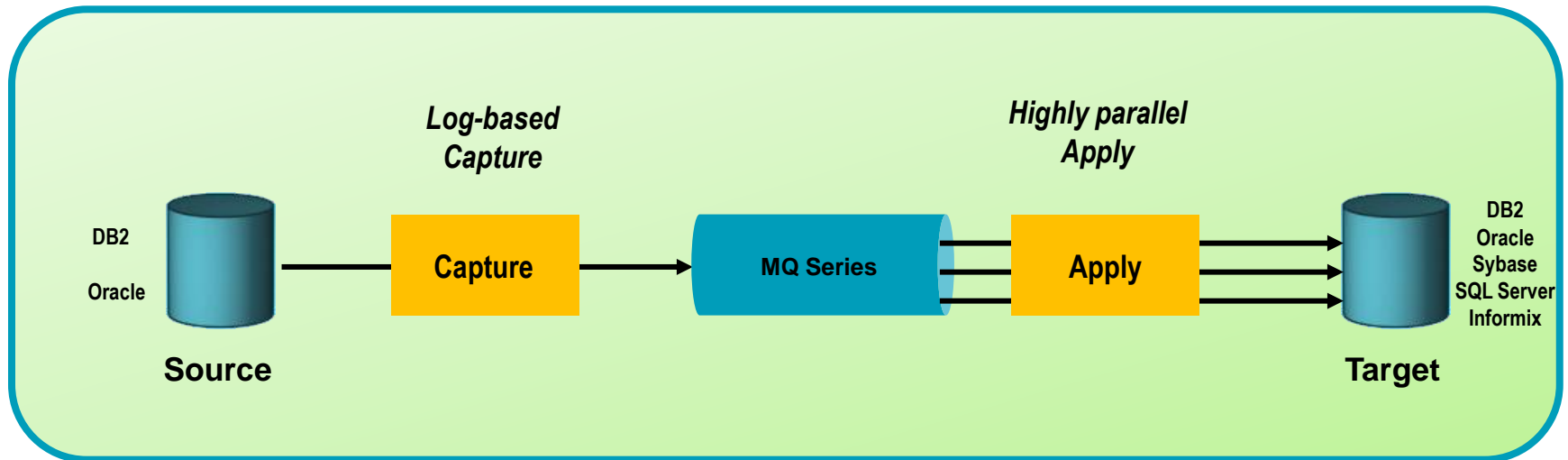
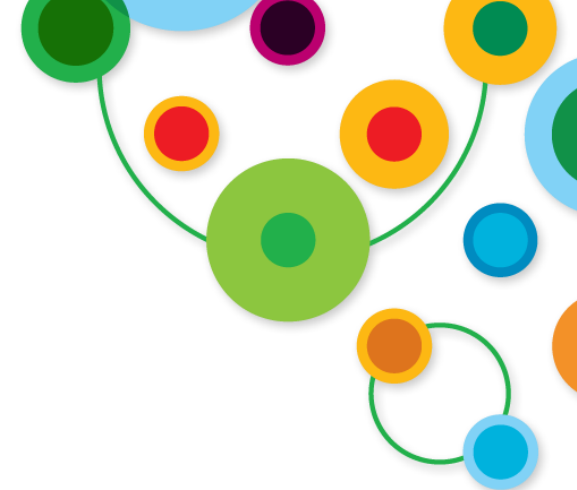
- Sources: DB2 for z/OS and LUW, and Oracle on all platforms
- Targets: DB2 for z and LUW, Oracle, Sybase, others with Federation



Can invoke Stored Procedures or publish XML

- Uni-direction, bi-directional and peer-to-peer replication supported
- Latency – 3X faster than SQL Replication; recommended for very large transaction volumes; uses MQ Series for speedy delivery
- Resiliency – works with HADR, Q Replication Dashboard monitoring
- Ease of use – ASNCLP scripting language
- **Cost – P/N D59ILLL is \$158.00 per Processor Value Unit; No requirement for InfoSphere Federation Server for targets listed above.**

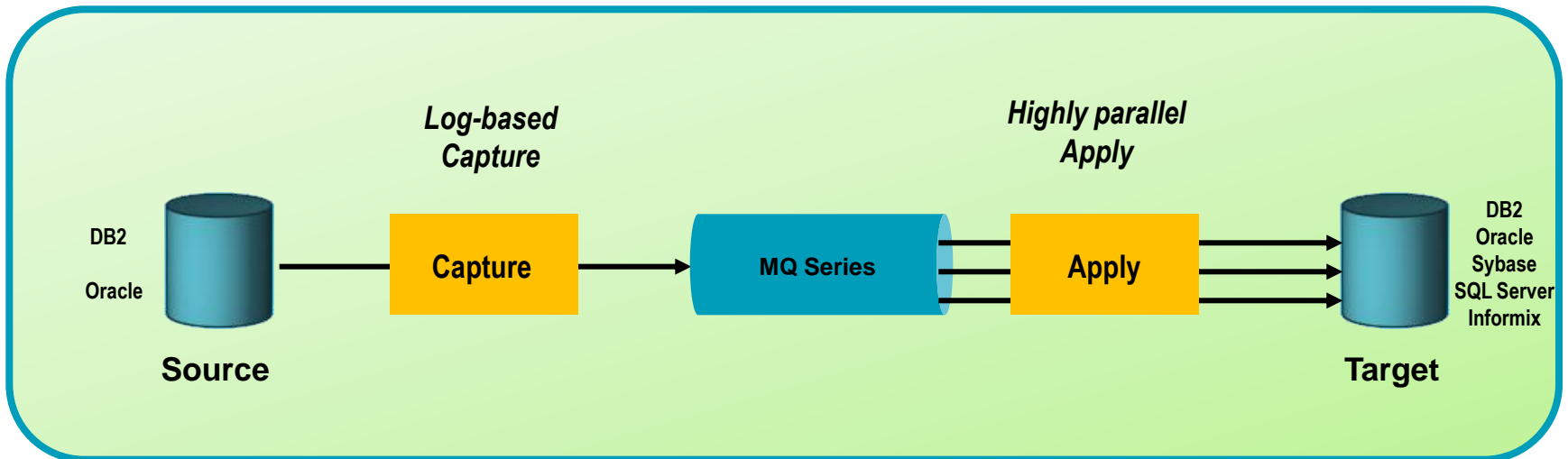
Q Replication



Q Replication

Important

- Additional complexity due to MQ Series
- Fastest; lowest latency
- Replicates compressed DB2 v9.7 data
- No support (source or target) for DB2 on Series i



ICDC

- Sources: DB2 (all platforms), Oracle, SQL Server, Sybase, Informix
- Targets: DB2, Oracle, SQL Server, Sybase, Teradata, Netezza

➔ And non-relational data targets such as DataStage and MQ

- Uni-direction and bi-directional replication supported
- Latency – faster than SQL Rep; vs. Q Replication scalability issues may impact performance at very high volumes
- Resiliency – works with HADR for DB2
- Ease of use – great GUI interface; no equivalent scripting to ASNCLP
- Cost – **P/N D0406LL is \$158.00 per Processor Value Unit, bundled with InfoSphere Information Server**



InfoSphere Change Data Capture Supported Sources and Targets

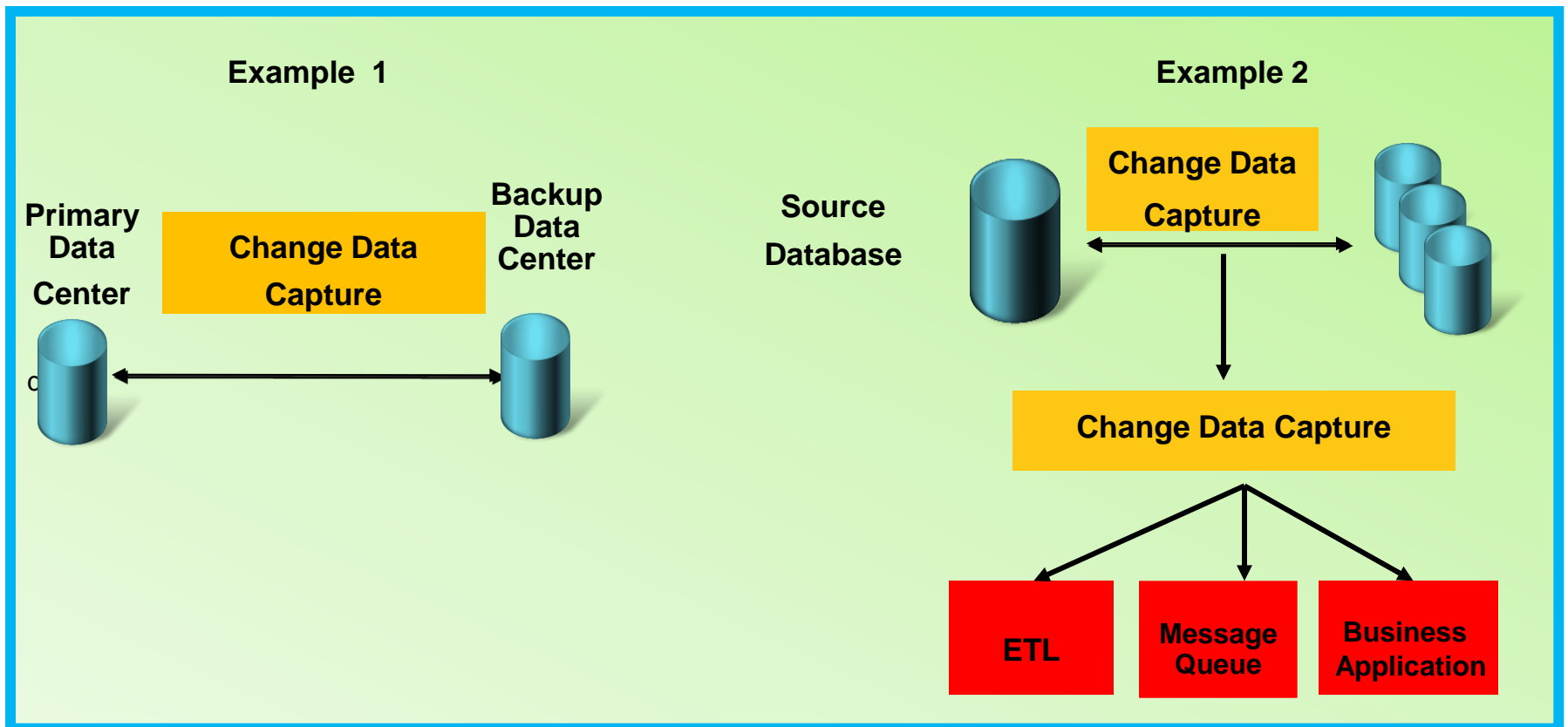
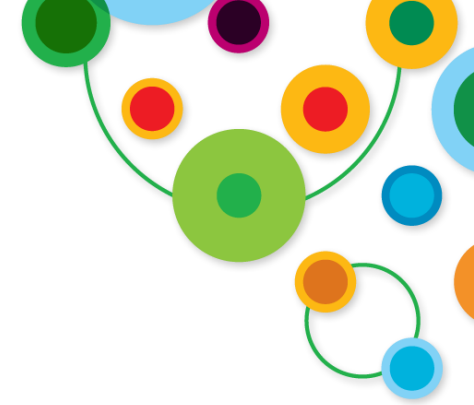
SOURCE Databases	TARGET Databases	Message Queues	Operating Systems	Hardware
DB2 z/OS	DB2 z/OS	MQ Series	z/OS	IBM System z
DB2 LUW	DB2 LUW	JMS	AIX	IBM System p
DB2 i	DB2 i	TIBCO	IBM i OS	IBM i Series
IMS	VSAM*	WebMethods	Red Hat, SUSE Linux for System Z	Intel / AMD
VSAM	Informix	BEA	Red Hat, SUSE Linux	HP PA-RISC
Informix	Information Server		HP-UX	HP Itanium
SolidDB	Cognos Now!		Solaris	Sun SPARC
Oracle	SolidDB		MS Windows	
MS SQL Server	Oracle			
Sybase	Teradata			
ADABAS	MS SQL Server			
IDMS	Sybase			
	Netezza, MySQL, Greenplum**			

* VSAM target only valid with VSAM source

** Customized solution, limited requirements



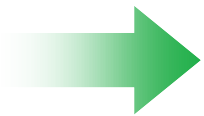
InfoSphere Change Data Capture



ICDC

Important

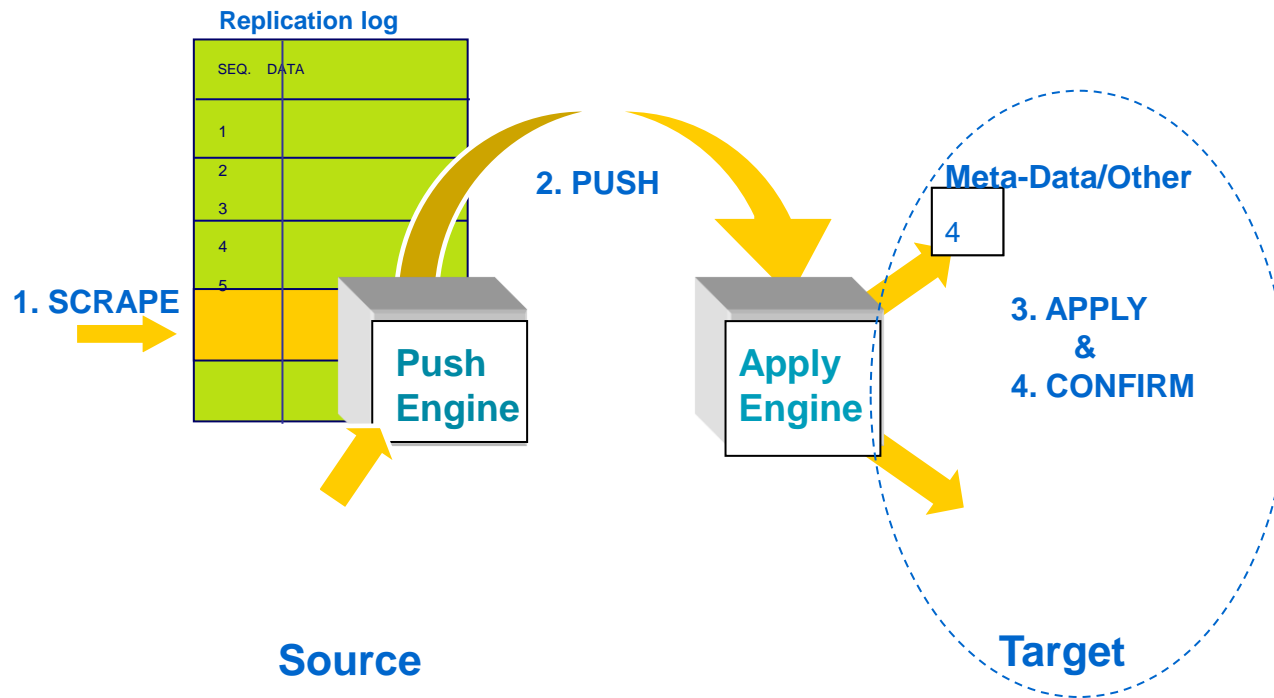
- Updates a bookmark with any change being delivered to the target.
- ICDC uses a small, proprietary database to manage the log position of each change.
- ICDC's performance is dependent on this database and it may constrain the volume and speed of transactions.



And non-relational data sources such as DataStage and MQ

- Does not require triggers on a heterogeneous source.
- Does not require InfoSphere Federation Server for heterogeneous targets.

ICDC Architecture





Customer Example – SQL Replication

- State government web portal
 - Source and target = DB2 Enterprise Server Edition for LUW – both source and target on z/Linux
 - Uni- directional replication from a database behind a firewall in the state capital to the public portal databases located in each county
 - Speed and latency – Not a critical issue; data volume low
 - Resiliency – Not a critical issue; replacing an old system with a planned 2 hour outage each night
 - Ease of use – relatively important; heritage mainframe DBA staff already being tasked with developing new DB2 for LUW skills and frequent changes need to be incorporated
 - Cost – Agency had a minimal IT budget



Customer Example – Q Replication

- National government customs and immigration
 - Source and target = DB2 for z/OS
 - Bi-directional replication for warm failover
 - Speed and Latency - built to ensure continuous availability during anticipated peak traffic times
 - Resiliency – critical particularly during high traffic periods
 - Ease of use – not as important as performance
 - Cost – well funded customer



Customer Example – ICDC

- Appliance Parts Retailer extending web presence
 - Source = DB2 on System i, target = DB2 Express C
 - Bi-directional, using replication to populate data warehouse
 - Speed and Latency – up-to-the-second replication not critical
 - Resiliency – important during peak web use hours but not a priority
 - Ease of use – critical as data inventory items change constantly
 - Cost – budget < \$25K



Customer Example – SQL Replication

- International Bank with branches in several countries
 - Multiple sources and targets – all DB2 for LUW
 - Peer-to-peer replication in an update anywhere scenario - a transaction can occur at any location and be replicated to all others
 - Speed and latency – location (Latin America) suffers with low Internet speed so sluggish transmission was a factor regardless of the replication solution
 - Resiliency – the customer wanted to be able to continue operating in all other locations if one server went down, and then have the server catch up when it was restored
 - Ease of use – DBA's needed to be able to support at each location
 - Cost – not a factor



Customer Example – Q Replication

- Regional bank web transaction portal database
 - Source and target = DB2 for z/OS
 - Uni-directional replication
 - Speed and Latency – critical – supporting popular web application
 - Resiliency – critical – uptime SLA = 24X7
 - Ease of use – not as important as performance
 - Cost – well funded customer



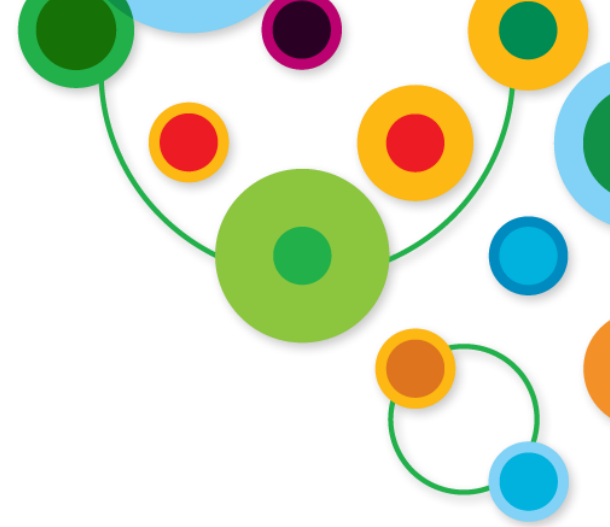
Customer Example – ICDC Replication

- Services vendor upgrading hardware enterprise-wide
 - Sources - various Oracle versions, targets - various Oracle
 - Uni-directional replication populating databases on new hardware
 - Speed and Latency – not critical – used during non peak hours
 - Resiliency – not a priority as migrations are one shot and done
 - Ease of use – critical - data sources and targets change each time
 - Cost – affordable as not initially budgeted for hardware migration

You Make The Call

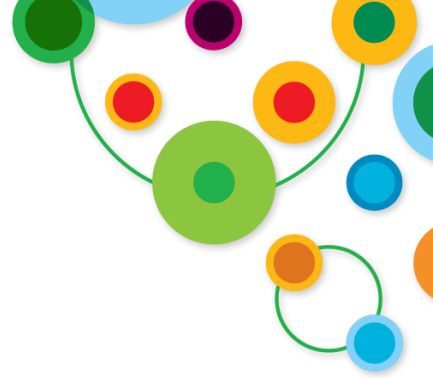
- Source and Target
- Uni-direction, bi-directional or peer-to-peer
- Latency
- Resiliency
- Ease of use
- Cost

The best solution is:



Resources

- IBM Redbooks:
www.redbooks.ibm.com
- IBM Authorized Training:
www.ibm.com/training



SQL Replication Resources

- ✓ **“My Mother Thinks I'm a DBA! Cross-Platform, Multi-Vendor, Distributed Relational Data Replication with IBM DB2 DataPropagator and IBM DataJoiner Made Easy!”**
(SG24-5463)
- ✓ **“A Practical Guide to DB2 UDB Data Replication V8”**
(SG24-6828)
- ✓ Course CE221 **“Using SQL Replication”** May 9, 2011
- ✓ Course CE231 **“SQL Replication: Advanced Topics”** May 12, 2011

Q Replication Resources

- ✓ **“WebSphere Replication Server Using Q Replication High Availability Scenarios for the AIX Platform”** (SG24-7216)
- ✓ **“WebSphere Information Integrator Q Replication: Fast Track Implementation Scenarios”** (SG24-6487)
- ✓ **Developerworks**
<http://www.ibm.com/developerworks/data/roadmaps/qrepl-roadmap.html>
- ✓ **Q Replication Tools:**
<http://www.01.ibm.com/support/docview.wss?uid=swg27007070>
- ✓ **Course CE243 “Using Queue Replication”**, February 14, 2011

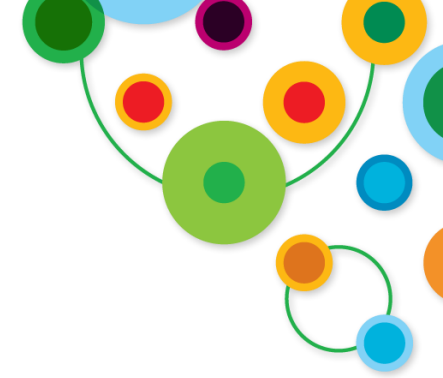
ICDC Resources

- ✓ **Documentation:**
<https://www-304.ibm.com/support/entdocview.wss?uid=swg27018235>
- ✓ **Developerworks**
<https://www.ibm.com/developerworks/mydeveloperworks/groups/service/html/communityview?communityUuid=a9b542e4-7c66-4cf3-8f7b-8a37a4fdef0c>
- ✓ Course DX550 “**Change Data Capture Essentials**”, January 31, 2011

Follow-up Questions

Kim May, The Fillmore Group

kim.may@thefillmoregroup.com
(410) 465-6335 office
www.thefillmoregroup.com/blog





Thank You!

Your Feedback is Important to Us

- Access your personal session survey list and complete via SmartSite
 - Your smart phone or web browser at: iodsmartsite.com
 - Any SmartSite kiosk onsite
 - Each completed session survey increases your chance to win an Apple iPod Touch with daily drawing sponsored by Alliance Tech

