



DB2 z/OS Replication & Data Capture

Are You Making the Most of This?

September 17th, 2010



Agenda

- Background
- DB2 CDC / Replication Overview
- Data Replication Success Factors
- Common DB2 Replication Scenarios
- Data Replication Issues / Pitfalls
- DB2 Replication: How to be Production Ready in 30 Days
- Q&A
- Wrap Up

SQData the Company



➤ **Founded in 2000**

- ✓ Team of Database Software Developers and Consultants
- ✓ Goal → Offer Enterprise Class Data Replication Solutions

➤ **Core Competencies**

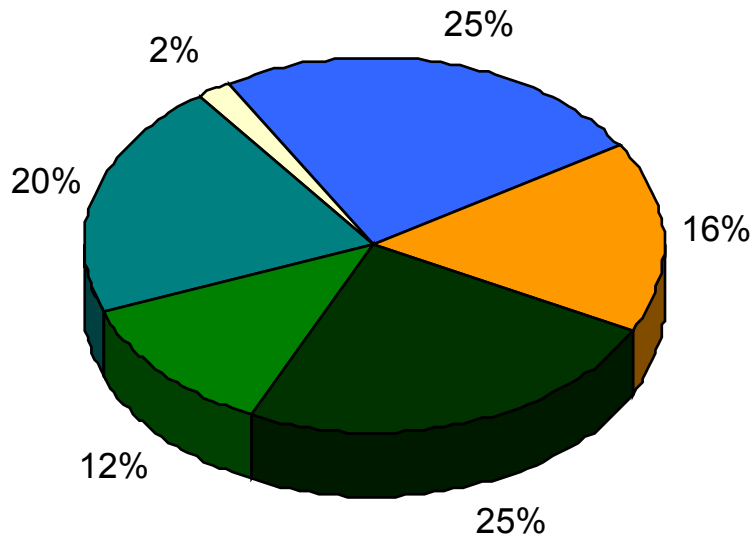
- ✓ High-Performance Changed Data Capture (CDC)
- ✓ Non-Relational Data → IMS, VSAM, Flat Files
- ✓ Relational Databases → DB2, Oracle, etc.
- ✓ Planning & Deployment of Complex Data Integration Solutions
- ✓ Replacing Existing Replication Products

➤ **Customer Value**

- ✓ Comprehensive Replication Solution
- ✓ Significant Cost Savings
- ✓ Easy to Deploy & Maintain
- ✓ Superior Customer Service



Why DB2 CDC / Replication?



- High Availability
- Operational Data Analytics (ODS)
- Data Warehouse / Business Intelligence
- Event Publishing
- Data Integration
- Audit

Replication Terminology

- **CDC**: changed data capture
- **Source**: tables whose changes are being captured
- **Target**: tables that are updated with changed data
- **Capture**: process that captures changed data
- **Apply**: process that updates target datastores
- **Transport**: method of moving data from source to target
- **Synchronized Target**: operation on target reflects source operation
 - ✓ Insert = Insert, Update = Update, Delete = Delete
- **Staged Target**: all source changes inserted into target (i.e. audit trail)
 - ✓ Requires 'Staging' Attributes: Time of Change, Change Operation, etc.

Replication Terminology...

- **Interface**: a unique source to target mapping / processing procedure
- **Refresh**: bulk data unload / load (ETL)...full or partial
- **Latency**: time between source update and target apply
 - ✓ Near-Real-Time
 - ✓ Trickle Feed
 - ✓ Asynchronous
- **Active/Passive**: one-way replication → 1 source to 1 or more targets
- **Active/Active**: multi-way replication between 2 or more sources / target pairs
- **Conflict**: target operation that is inconsistent w/ source change
- **Homogeneous**: source/target same technology (i.e. DB2 to DB2)
- **Heterogeneous**: source & target are unlike (i.e. IMS to DB2)

DB2 Data Capture Methods

➤ Three (3) Primary Methods of Capture

- ✓ Trigger Based
- ✓ Application Based
- ✓ Log Based

➤ Trigger Based

- ✓ Insert, Update, Delete Triggers on Source Tables
- ✓ Typically Requires Interim Staging Tables
- ✓ Some Customers Using MQ w/ User Defined Functions (UDF)
- ✓ Impacts Application Performance
- ✓ Apply Can be a Challenge

➤ Application Based

- ✓ Data Capture Built-In to Application Code
- ✓ Can be Customized to Meet Specific Customer Requirements
- ✓ More Costly to Implement than Other Capture Methods
- ✓ Must be Careful to Include ALL Programs that Update DB2 Tables
- ✓ Apply is Typically Application Specific

DB2 Data Capture Methods...

➤ Log Based

- ✓ Preferred Method of Data Capture
- ✓ IFCID 306
- ✓ Typically Used by Advanced Commercial Products

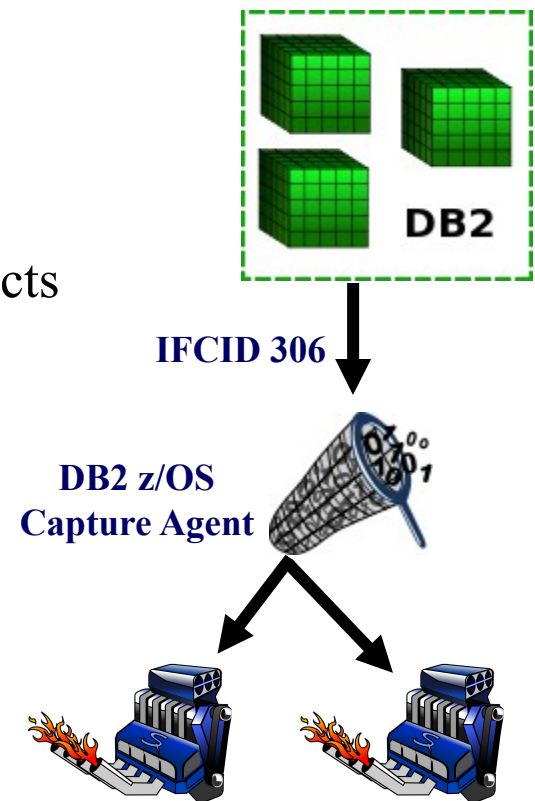
➤ Non-Intrusive to Applications

➤ Simple to Manage

➤ Allows for Recapture

➤ Data Sharing

- ✓ Capture on One (1) Member
- ✓ Option to Capture Changes from Multiple Members Individually
 - Applicable to Staging Targets – Data can Arrive Out of Order
 - Not Recommended for Synchronized Targets



Categories of Log Based Replication Tools

➤ **Log Analyzers**

- ✓ Full Function DB2 Capture
- ✓ Apply Not So Full Function (SQL Based)

➤ **Single Purpose Data Replication Products**

- ✓ Replication Only
- ✓ Typically Homogeneous (i.e. DB2 to DB2)
- ✓ Require Additional Products for More Advanced Functionality

➤ **Multi-Dimensional Data Integration Products**

- ✓ Full Function ETL, CDC, XML, Event Publishing and Data Integration
- ✓ May Require Purchase of Multiple Components
- ✓ SQData → Comprehensive Solution (DB2, IMS, VSAM)

Top 10 Attributes of a Good Replication Tool

- 1) Operates Properly**
- 2) Operates Fast and Efficiently**
- 3) Cost Effective**
- 4) Easy to Use & Understand**
- 5) Allows Multiple Options for Replication**
- 6) Rapid Deployment**
- 7) Multi-Dimensional: Replication, ETL, Integration, Event Pub**
- 8) Requires Minimal Personnel to Operate / Maintain**
- 9) Works Well with Other Integration Products**
- 10) Superior Vendor Support**

Data Replication Success Factors

- **Planning**
 - ✓ Understanding Data Movement Requirements
 - ✓ Understanding Latency Requirements
 - ✓ Changed Data Volumes
 - ✓ Infrastructure Requirements

- **Minimize the Number of ‘Moving Parts’**

- **Define / Develop Recovery Strategy**

- **Define Exception Processing Criteria**

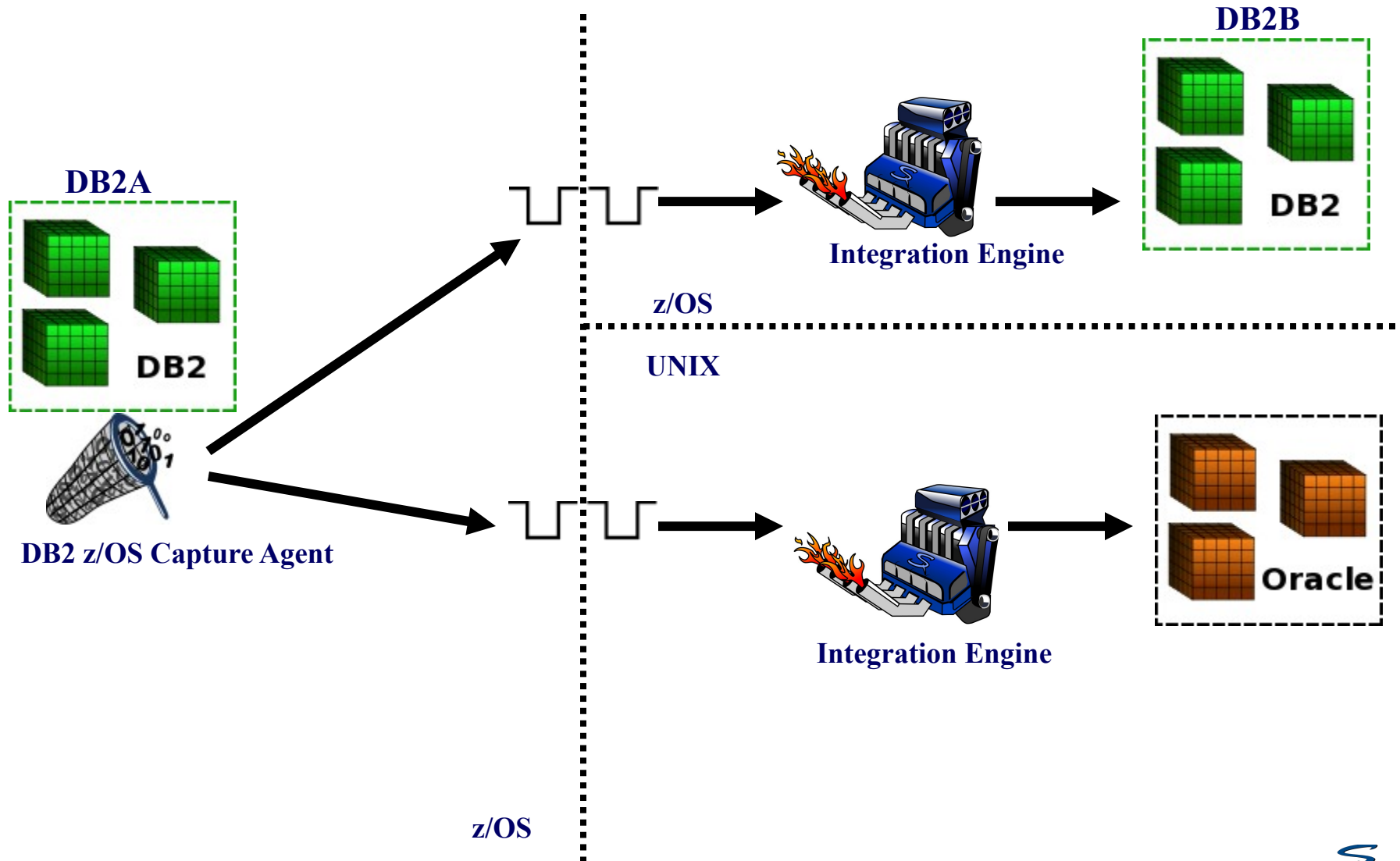
- **Identify Method of Verifying Source/Target Consistency**

- **Selecting the Right Tool for the Job**

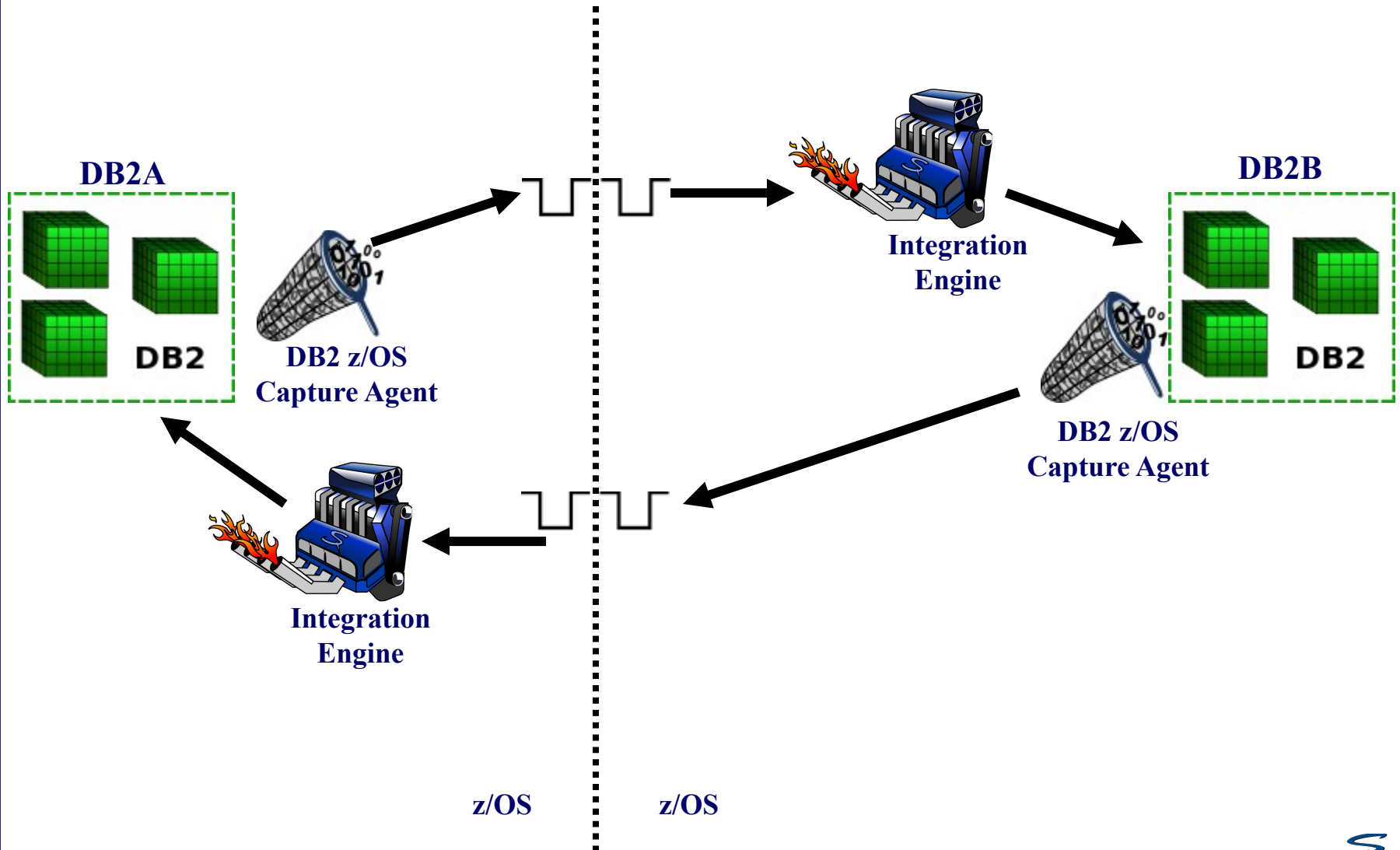
Agenda

- Background
- DB2 CDC / Replication Overview
- Data Replication Success Factors
- **Common DB2 Replication Scenarios**
- Data Replication Issues / Pitfalls
- DB2 Replication: How to be Production Ready in 30 Days
- Q&A
- Wrap Up

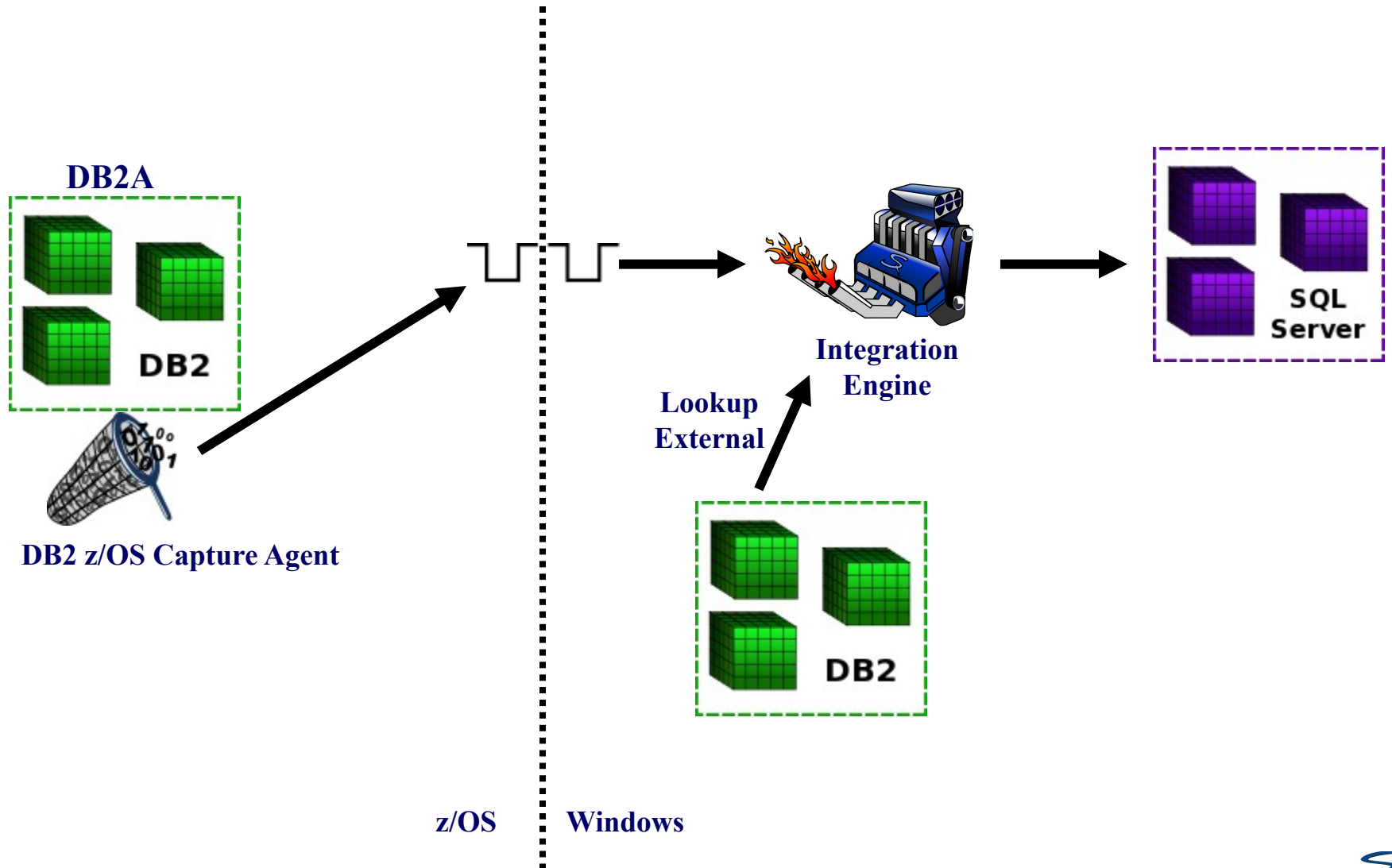
Active/Passive: Straight Replication



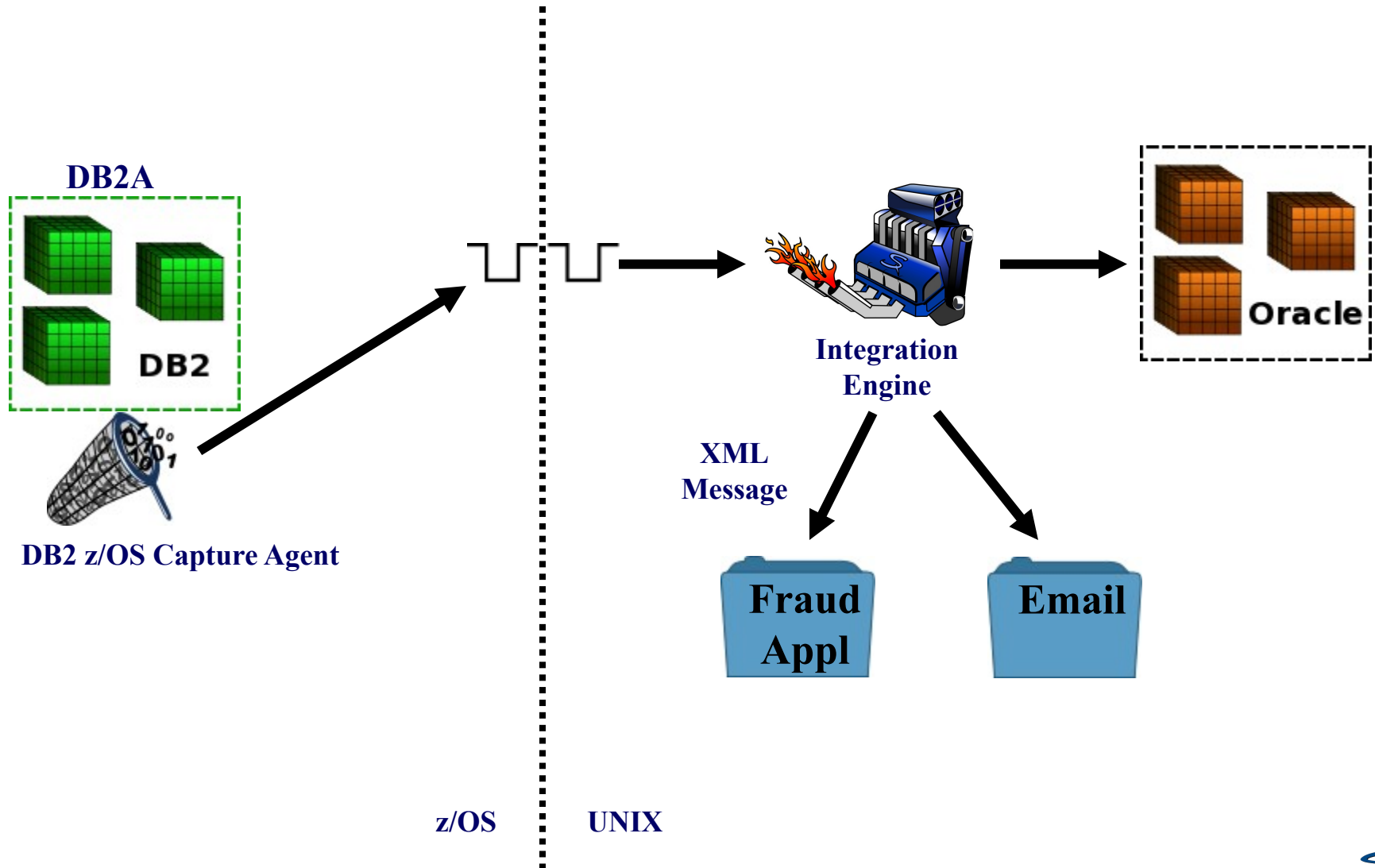
Active/Active: Straight Replication



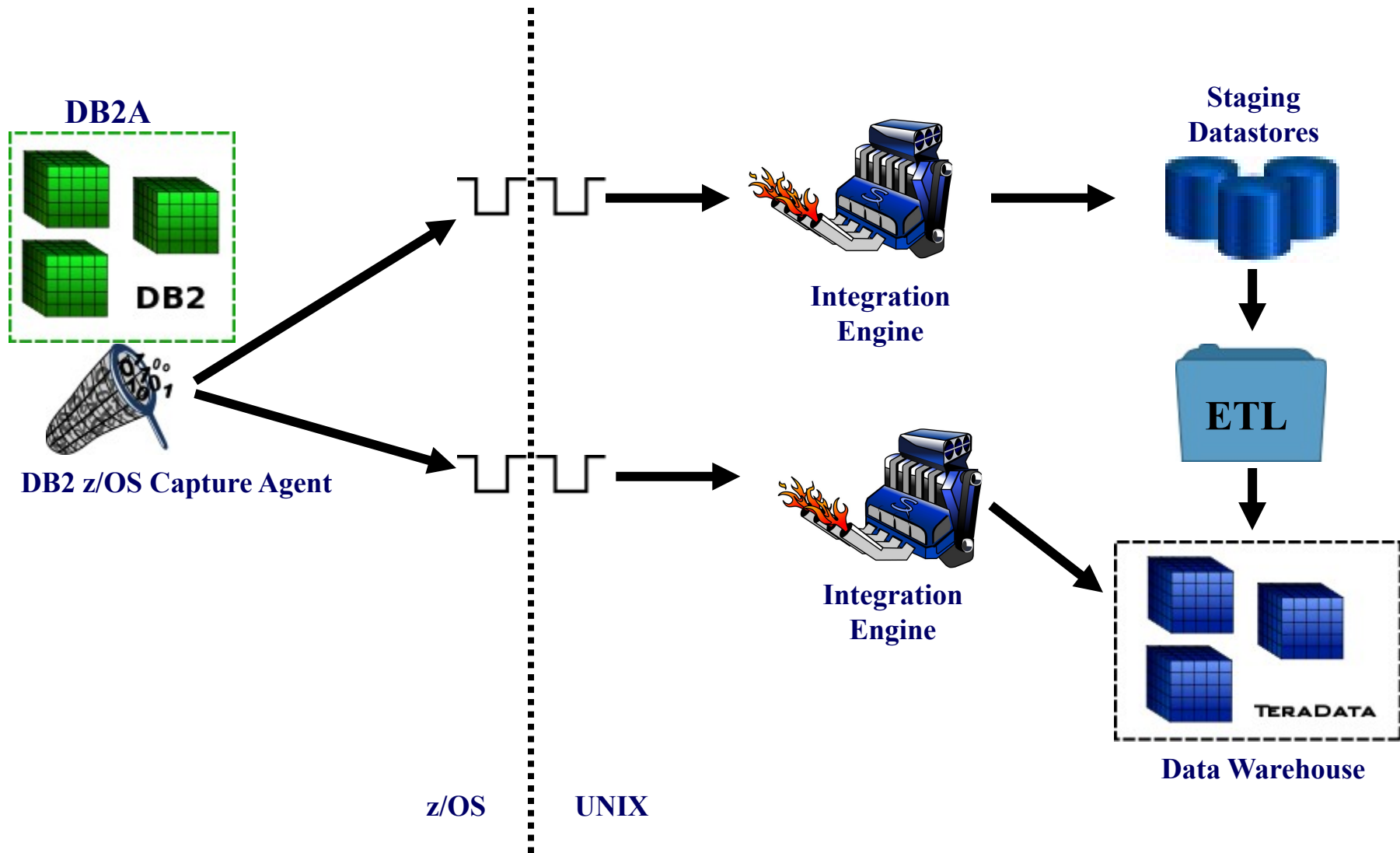
Replication with Augmentation



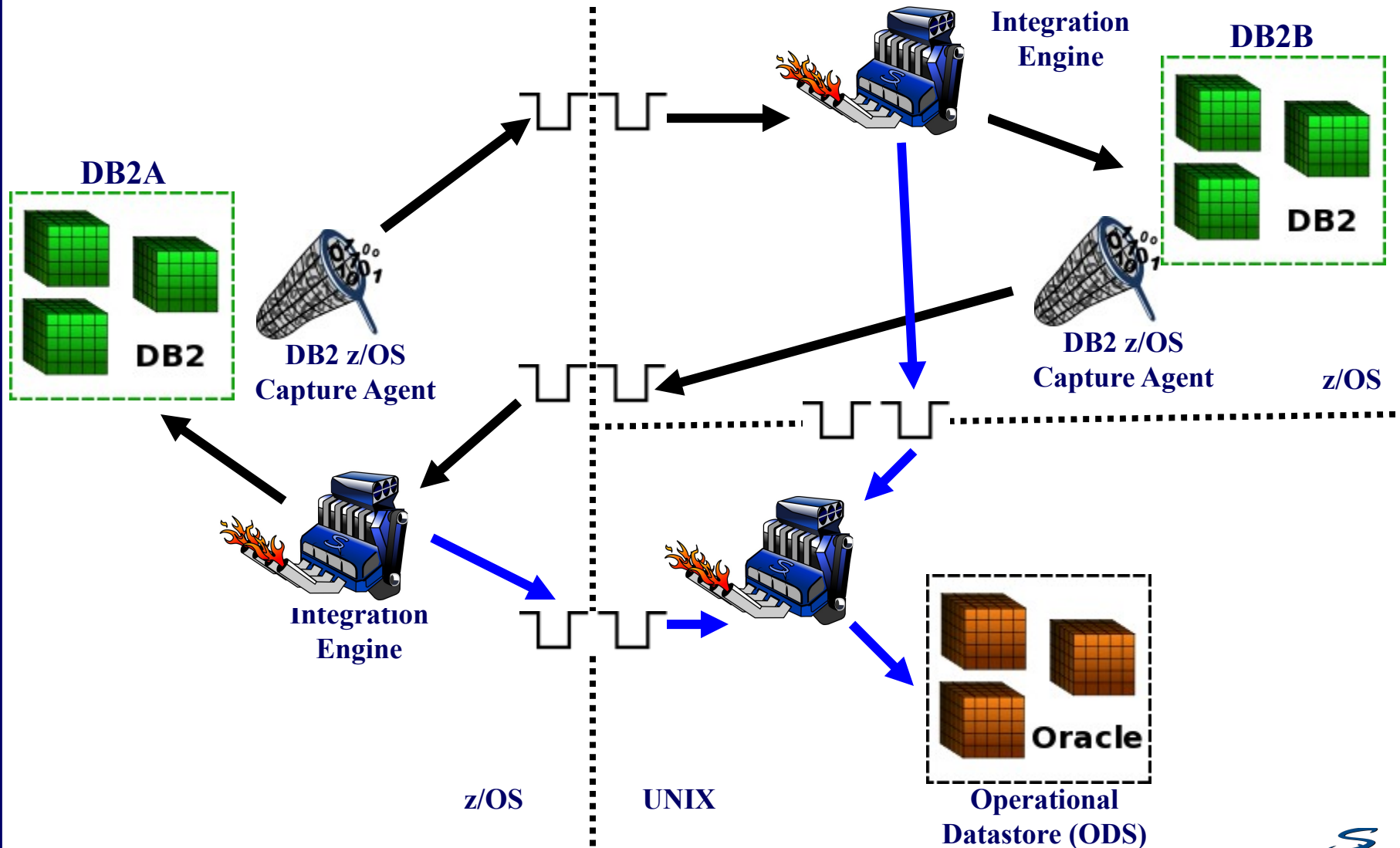
Replication with Event Publishing



CDC Enhanced ETL w/ Near-Real-Time



Advanced Integration



Agenda

- Background
- DB2 CDC / Replication Overview
- Data Replication Success Factors
- Common DB2 Replication Scenarios
- **Data Replication Issues / Pitfalls**
- DB2 Replication: How to be Production Ready in 30 Days
- Q&A
- Wrap Up

Conflict Detection and Resolution

- **Conflicts Indicate an Out-of-Sync Condition**

- **Relational Replication Conflicts**
 - ✓ Insert for Row that Already Exists in Target
 - ✓ Update for Row that Does Not Exist in Target
 - ✓ Update for Row where Before Image does not Match Current Row in Target
 - ✓ Delete for Row that Does Not Exist in Target

- **Conflicts that can be Compensated**
 - ✓ Insert for Row that Already Exists → Turn into Update
 - ✓ Update for Row that Does Not Exist → Turn into Insert
 - ✓ Deletes for Row that Does Not Exist → Ignore

- ****Not Applicable to Staging Targets**

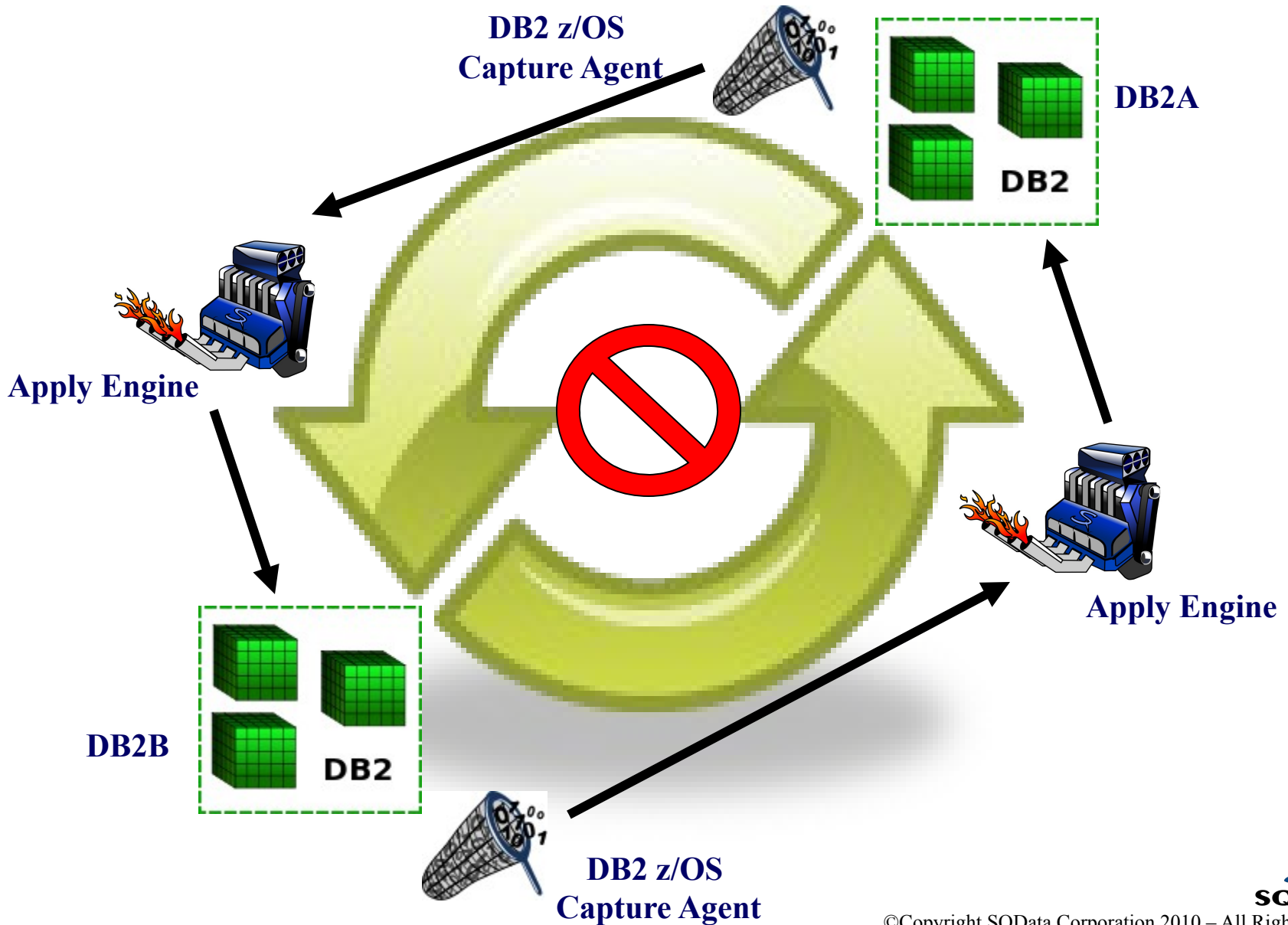
Conflict Reporting

➤ SQData Engine Runtime Report if Compensation Occurs

- ✓ Compensated Insert: Inserts that were Turned into Updates
- ✓ Compensated Update: Updates that were Turned into Inserts
- ✓ Compensated Delete: Deletes that were Ignored

```
Source Data Store.....: SQDATA.DB2CDC.QUEUE@MQS
  Alias.....: CDCIN
  Start At.....: 0
  Records Selected.....: 0
  Records Read.....: 30166
Target Data Store.....: SQDATA.DB2CDC.EXCEPTION.QUEUE@MQS
  Alias.....: EXPDS
  Records Inserted.....: 0
  Records Updated.....: 0
  Records Deleted.....: 0
Target Data Store.....: PROD.EMPLOYEE
  Alias.....: EMPLOYEE
  Records Inserted.....: 5582
  Records Updated.....: 11190
  Records Deleted.....: 389
  Compensated Insert.....: 9915
  Compensated Update.....: 11
  Compensated Delete.....: 13005
17.55.17 SQD0041I 00      0 21 Processing completed with return code 0
```

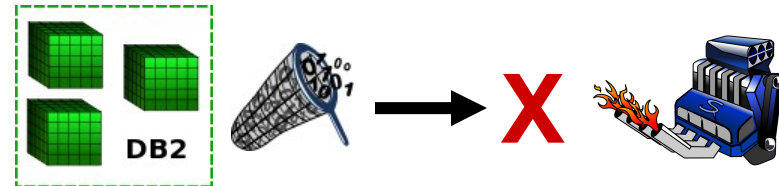
Circular Reference



Exception Processing

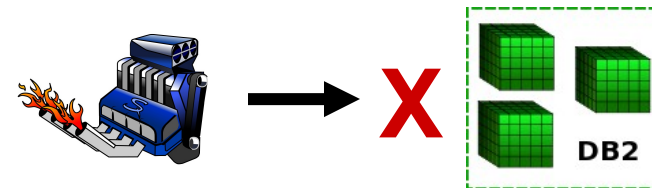
➤ Source Exceptions

- ✓ Invalid Data in Source Change Record
- ✓ Not Common for Relational CDC



➤ Target Exceptions

- ✓ Issue with Target Datastore (i.e. Unavailable)
- ✓ Conflicts
- ✓ Application Specific Logical Errors



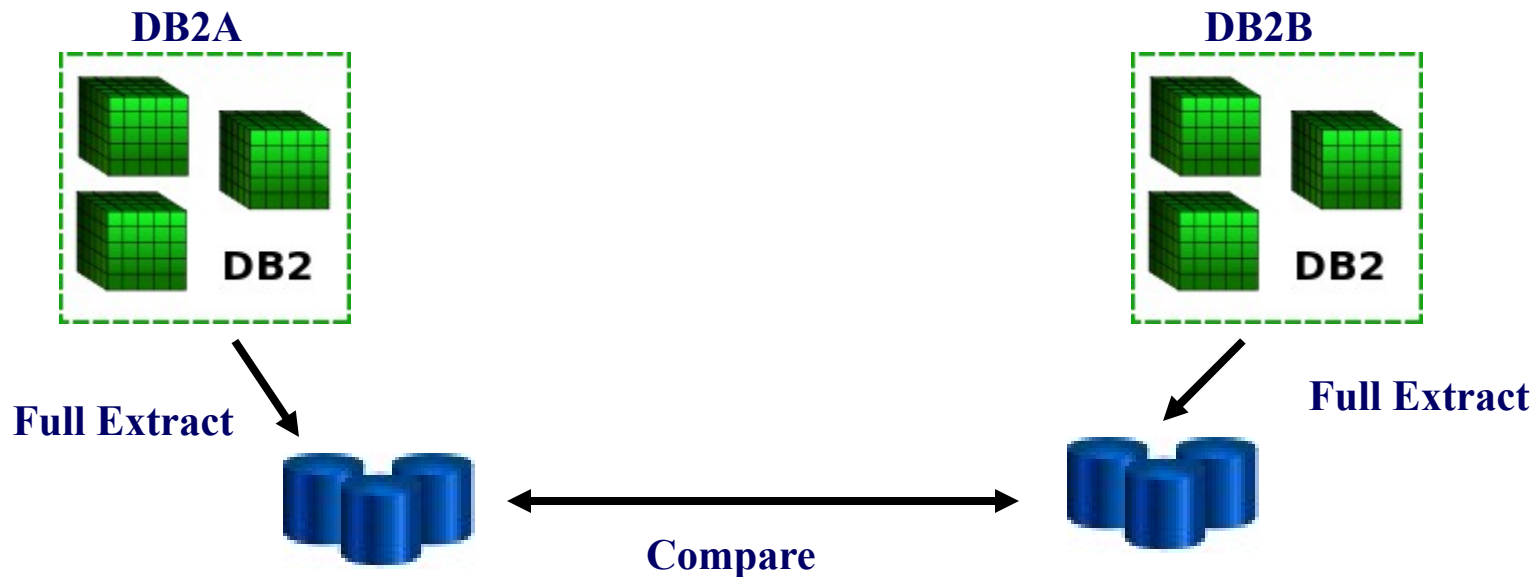
➤ Replication Tool Should Allow for Continued Operation

- ✓ Include the Ability to Provide Notification of Exceptions
 - ✓ Save Exception Records to a Separate Datastore
- ** Some Exceptions Should Force Replication Shutdown ****

Verifying Source / Target Consistency

➤ Classic Method #1: Requires Source/Target to be Read-Only

- ✓ Unload Source
- ✓ Unload Target
- ✓ Transmit Source Unload to Target (or Vice-Versa), if Required
- ✓ Compare Source Unload to Target Unload

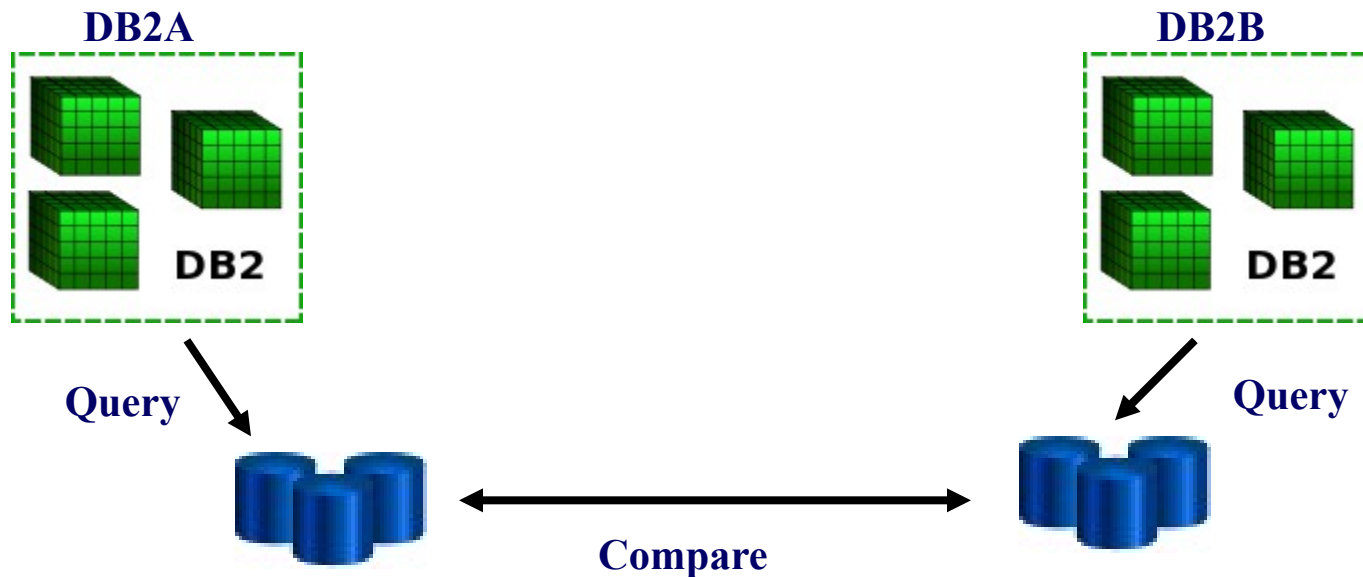


Verifying Source / Target Consistency...

➤ Classic Method #2: Smaller Tables: Data Can be 'Moving'

- ✓ Create 'Consistency Check' Query for each Source/Target Table Pair
- ✓ Best Done Using Date/Time Ranges or Primary Key Grouping
- ✓ Run CC Query on Source and Target Tables
- ✓ Compare Results

**May Require Multiple Iterations if CDC Volume High



Verifying Source / Target Consistency...

➤ Conflict Monitoring

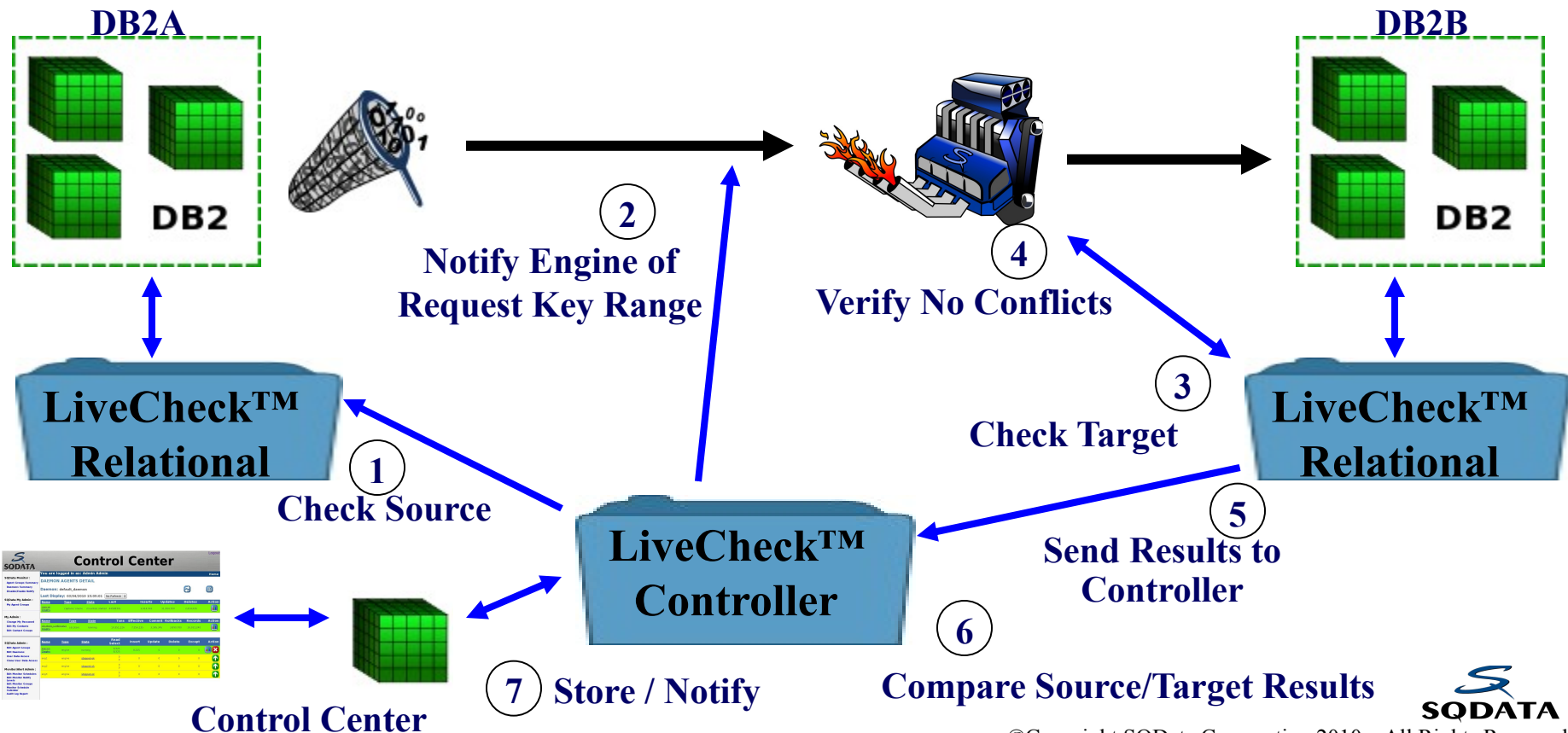
- ✓ Done during Active Replication
- ✓ Assumes Replication Tool Detects and Reports Conflicts

```
Source Data Store.....: SQDATA.DB2CDC.QUEUE@MQS
  Alias.....: CDCIN
  Start At.....: 0
  Records Selected.....: 0
  Records Read.....: 30166
Target Data Store.....: SQDATA.DB2CDC.EXCEPTION.QUEUE@MQS
  Alias.....: EXPDS
  Records Inserted.....: 0
  Records Updated.....: 0
  Records Deleted.....: 0
Target Data Store.....: PROD.EMPLOYEE
  Alias.....: EMPLOYEE
  Records Inserted.....: 5582
  Records Updated.....: 11190
  Records Deleted.....: 389
  Compensated Insert.....: 9915
  Compensated Update.....: 11
  Compensated Delete.....: 13005
17.55.17 SQD0041I 00      0 21 Processing completed with return code 0
```

Verifying Source / Target Consistency...

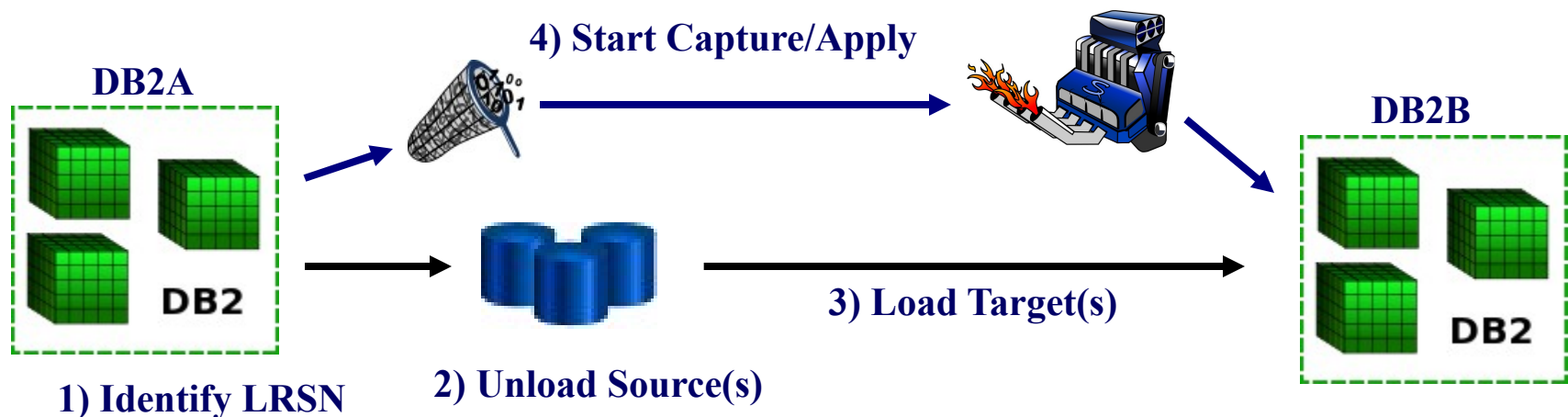
➤ SQData LiveCheck™

- ✓ Intended for High-Volume, 24x7 Setups
- ✓ Checks Source Target Integrity during Active Replication
- ✓ 1st Iteration: IMS to IMS - Relational Planned: 2Q 2011



Data Synchronization / Materialization

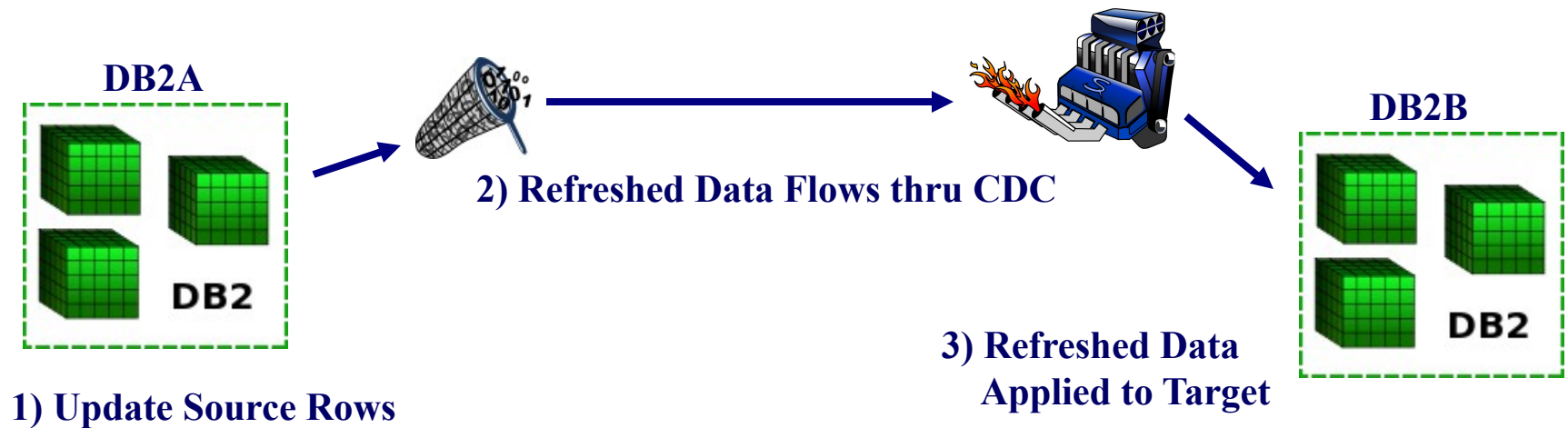
- **Having a Refresh Process is Highly Recommended**
- **Needs to be Done Against Active Source Databases**
- **Classic Method #1: Full Loads or Partial Loads**
 - 1) Identify Start LRSN (RBA) – BEFORE Unloads Start – Quiesce Preferred
 - 2) Unload Source
 - 3) Load Target
 - 4) Start Capture / Apply Processes at LRSN Identified in Step #1



Data Synchronization / Materialization...

➤ Classic Method #2: CDC by Forcing Row Updates

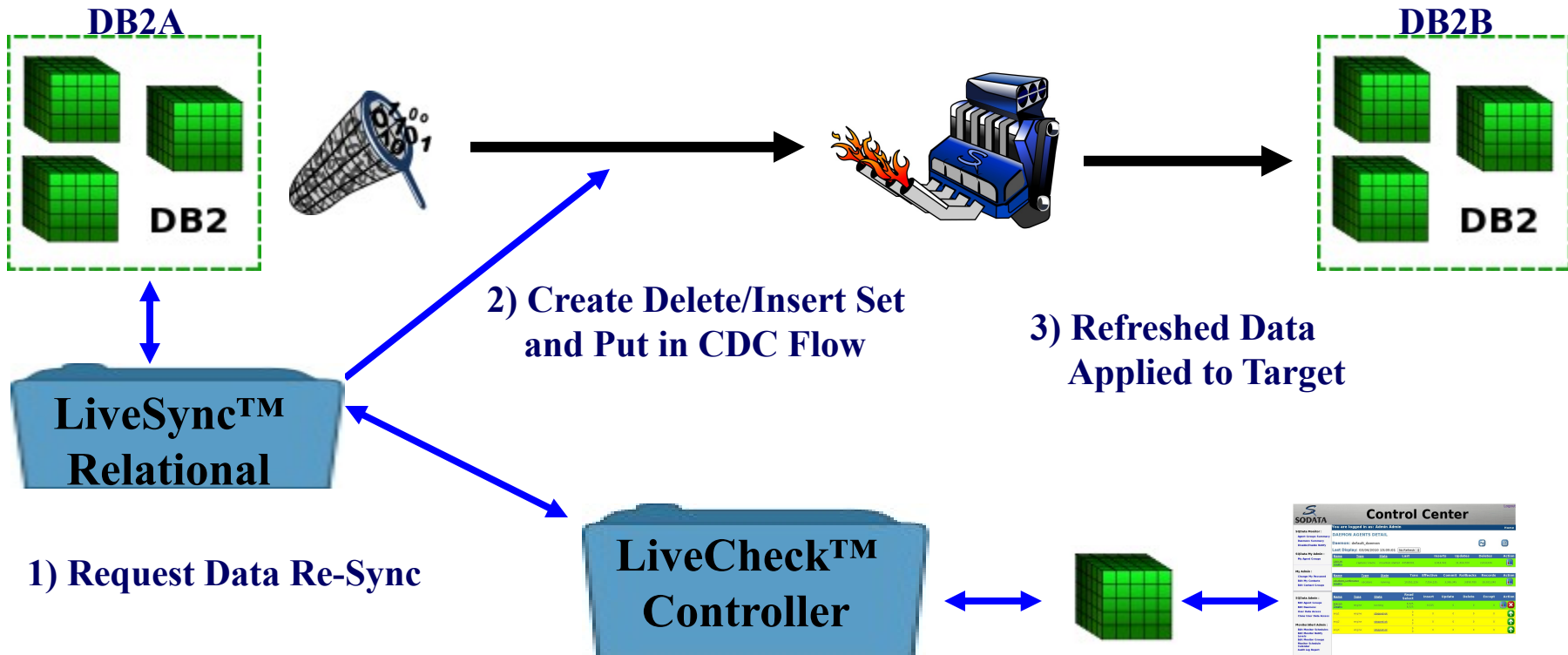
- ✓ Refresh through Capture / Apply Process
- ✓ Works Best for Smaller Tables
- ✓ Requires Update of Non-Business Critical Column
- ✓ May Interfere with Operational Transactions if Done with Large UOWs



Data Synchronization / Materialization...

➤ SQData LiveSync™

- ✓ Compliments SQData LiveCheck
- ✓ Intended for High-Volume, 24x7 Setups
- ✓ Checks Source Target Integrity during Active Replication



Production Ready in 30 Days or Less

- **SQData Implementations Only**
- **Synchronized / Staged Replication with Straight Mapping**
- **Task Summary**
 - ✓ Identify Source / Target Tables of Interest
 - ✓ Determine Infrastructure Requirements /Activities
 - ✓ Install / Configure SQData Product
 - ✓ Create Source to Target Mappings
 - ✓ Generate Interfaces*
 - Refresh
 - CDC
 - ✓ Define Transport Components (queues, ports, etc.)
 - ✓ Create Source / Target Validation Procedures
 - ✓ Verify Proper Operation of the Interfaces

*Performed by Jointly by Customer and SQData Professional Services

Questions?



Next Steps

- **Send You the Presentation Materials for this Briefing**

- **Contact You to Discuss:**
 - ✓ Your Current Replication Requirements
 - ✓ Replication Requirements for the Next 18 - 24 Months

- **Gather Feedback for Future Briefings**

Where to Find Additional Information

- **SQData Website: www.sqdata.com**
- **Phone Requests: 866-252-3575**
- **Email Requests: info@sqdata.com**

DB2 z/OS Replication & Data Capture

Are You Making the Most of This?

September 17th, 2010