Best Practices Series

Maximizing the Use of Your IMS Data with Replication

Prepared for the:

Virtual IMS User Group

5 February 2013

Agenda

- Introduction
- Leveraging Replication for IMS Data Usage
- In-Depth Look at IMS Replication Components
- Sample Deployment Scenarios
- Common Challenges
- Q & A
- Conclusion

About the Speaker

- **Scott Quillicy**
  - ✔ 30+ Years Database Experience
  - ✔ Commercial Database Software Development
  - ✔ Deployment of Complex Data Integration Solutions
- **Founded SQData to Provide Customers with:**
  - ✔ An Enterprise Class Data Integration / Replication Framework
  - ✔ A Solution for Both Relational and Non-Relational Data
  - ✔ Technology Built Around Best Practices
- **Specialization**
  - ✔ Database Replication
  - ✔ IMS – the More Complex, the Better
  - ✔ Heterogeneous Database Integration
  - ✔ Continuous Availability
  - ✔ Database Performance

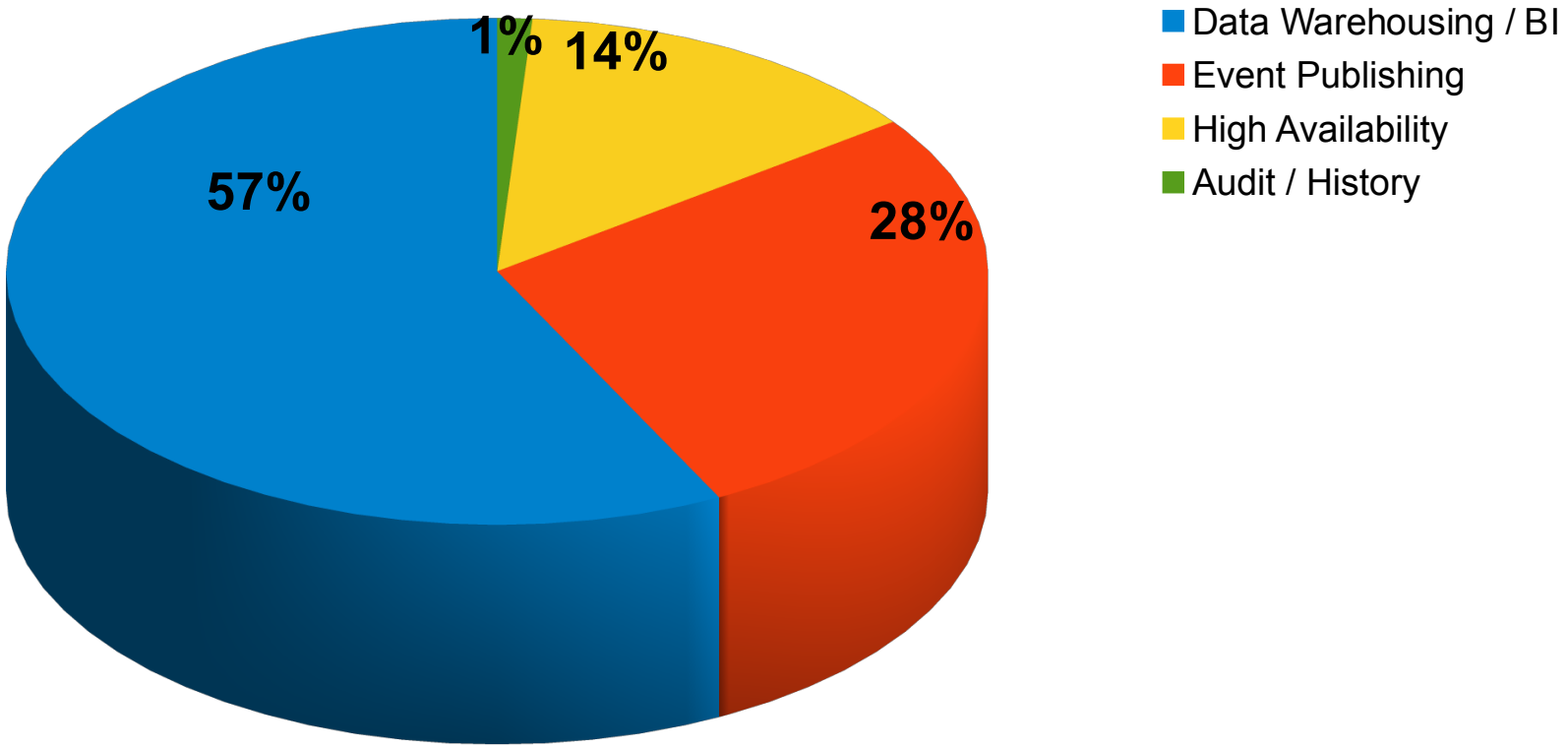
About SQData

- **"Swiss Army Knife of Data Replication Tools"**
- **Core Competencies**
  - ✔ High-Performance Changed Data Capture (CDC)
  - ✔ Non-Relational Data → IMS, VSAM, Flat Files
  - ✔ Relational Databases → DB2, Oracle, SQL Server, etc.
  - ✔ Deployment of Complex Data Integration Solutions
  - ✔ Continuous Availability of Critical Applications
  - ✔ Data Conversions / Migrations
- **Customer Usage**
  - ✔ Relational and Non-Relational Data
  - ✔ Data Replication – Relational and Non-Relational
  - ✔ ETL (Bulk Data Extracts/Loads)
  - ✔ Application Integration
  - ✔ Business Event Publishing
  - ✔ Data Conversions / Migrations

Why Replicate IMS Data?

- Provide Users with a Method of Querying Data Outside of IMS
- Continuous Availability
- Business Intelligence / Data Warehousing
- Co-Existence with Newer Applications
- Application Migration / Replacement
- Audit / Historical Data Archive
- “We're Moving Off of the Mainframe”....☺

How is IMS Replication Being Used Today?



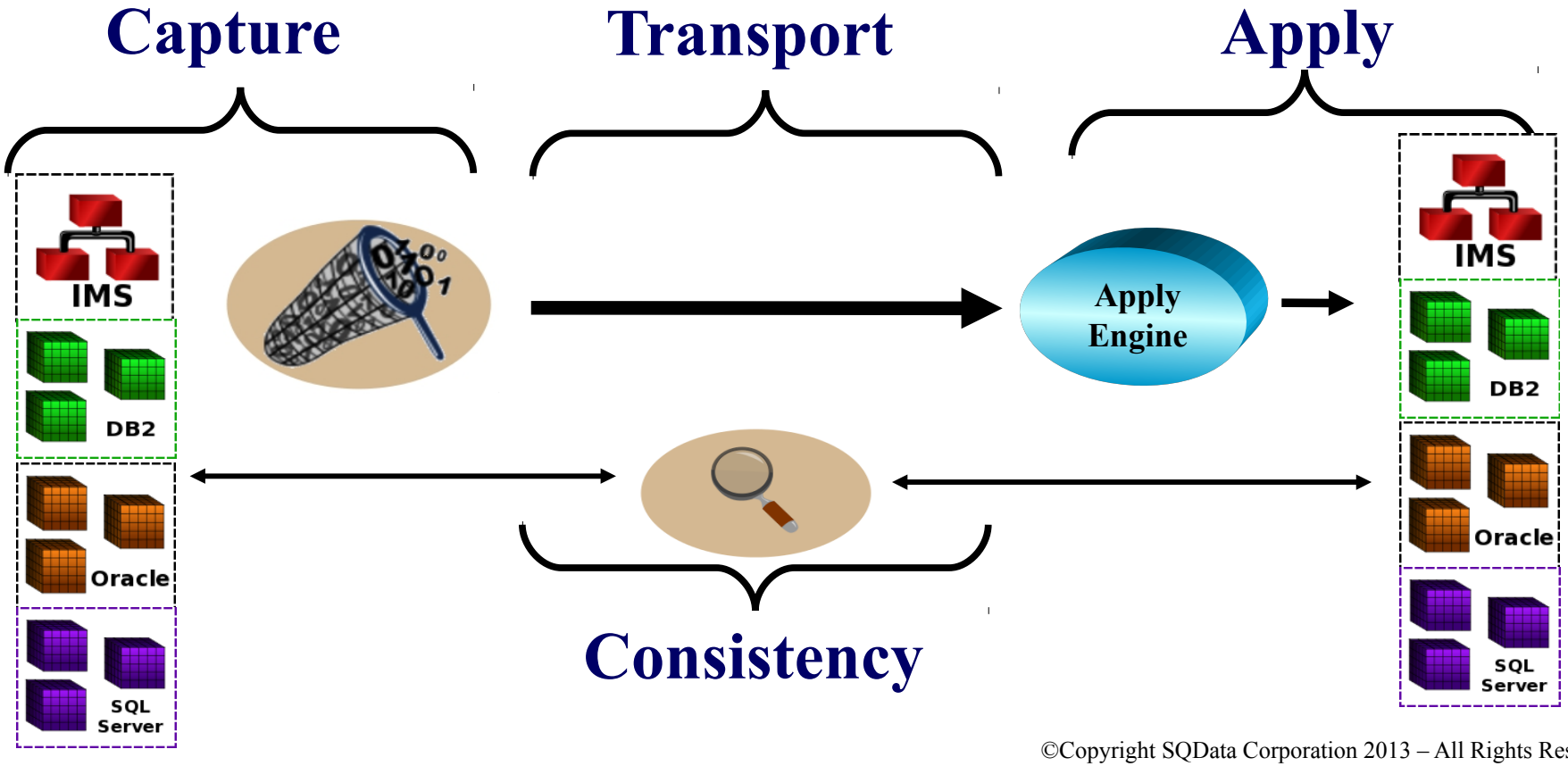
Synchronous vs Asynchronous Replication

- **Synchronous**
  - ✓ Multiple Sites Updated within the same Transaction Scope
  - ✓ Ensures Zero (0) Data Loss
  - ✓ Downside
    - Transaction Latency MUCH Longer
    - Transactions Fail if All Sites Not Connected (defeats purpose of continuous)
- **Asynchronous**
  - ✓ Single Site Updates
  - ✓ Data Captured on Updating Site and Propagated to other Sites
  - ✓ Latency Typically Sub-Second
  - ✓ Downside
    - Minimal Data Loss May Occur

Replication Basics

Four (4) Primary Components

- ✓ Source Capture
- ✓ Data Transport
- ✓ Target Apply
- ✓ Consistency Monitoring



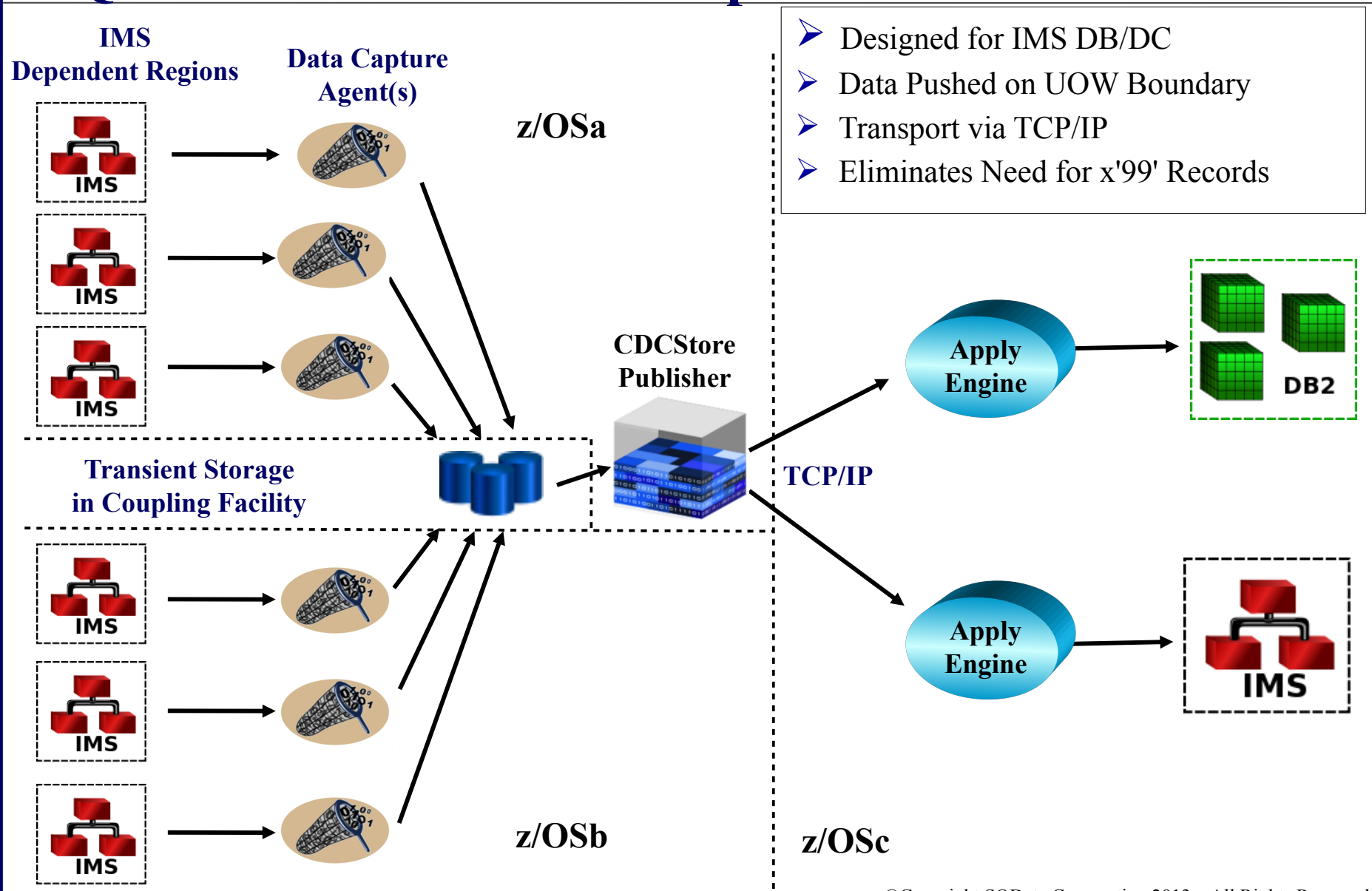
IMS Data Capture Methods

- **Primary Methods of Capture**
  - ✓ Data Capture Exit Routines
  - ✓ Log Based
  - ✓ Application Based
  - ✓ Hardware
- **Database Exit Routines**
  - ✓ Near-Real-Time
  - ✓ Scalability → Capture / Apply by FP Area, HALDB Partition, PSB, Database
  - ✓ Can Use MQ for Persistent Storage and Transport
  - ✓ Do Not Require x'99' Log Records
  - ✓ Executes in Dependent Region as Part of Transaction
- **Log Based**
  - ✓ Near-Real-Time or Asynchronous
  - ✓ Requires x'99' Log Records
  - ✓ Allows for Recapture
  - ✓ Scalability → Single Capture Point...Apply by PSB
  - ✓ Executes in Control Region or in Separate Address Space

Data Transport Methods

- **Two (2) Primary Methods of Transport**
  - ✓ Queue Based
  - ✓ Native TCP/IP
- **Queue Based**
  - ✓ Handles Persistent Storage in Addition to Transport
  - ✓ Resilient
  - ✓ Can Handle Moderate Data Volume on a Continuous Basis
  - ✓ Operates Independently of Capture and Apply
  - ✓ Adds Overhead to Replication Process
- **Native TCP/IP**
  - ✓ Transport Typically Faster than Queue Based
  - ✓ Can Handle High Data Volume on a Continuous Basis
  - ✓ Requires Separate Transient Storage for CDC Data
  - ✓ Resiliency Must be Built In to CDC Storage
  - ✓ Operation not Always Independent

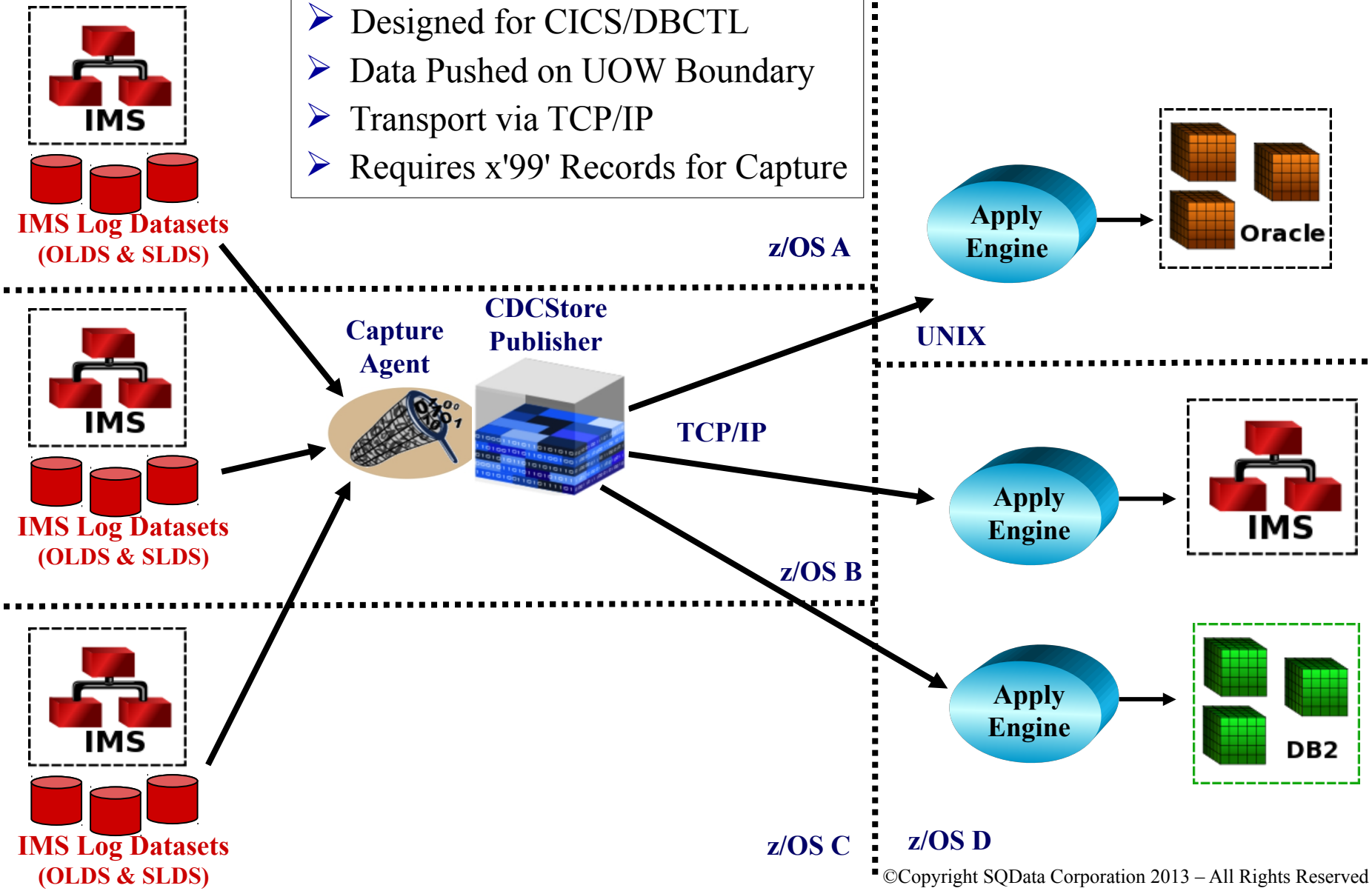
SQData IMS CDCStore Replication



- Designed for IMS DB/DC
- Data Pushed on UOW Boundary
- Transport via TCP/IP
- Eliminates Need for x'99' Records

SQData IMS CDCLog Replication

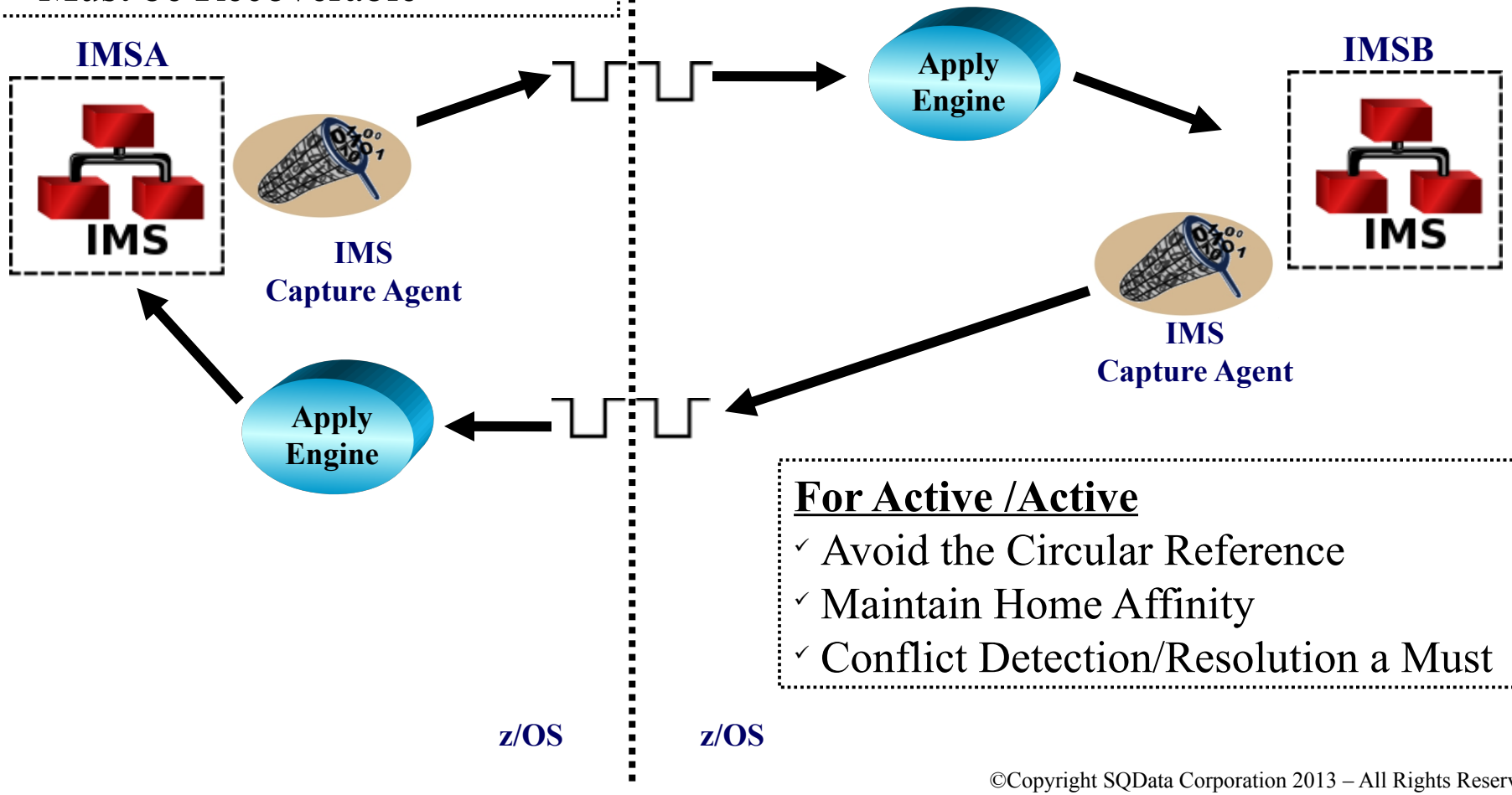
- Designed for CICS/DBCTL
- Data Pushed on UOW Boundary
- Transport via TCP/IP
- Requires x'99' Records for Capture



Target Apply Process

Basic Rules

- ✓ Data Must be Applied in Order
- ✓ Must be Able to Keep Up
- ✓ Must be Recoverable



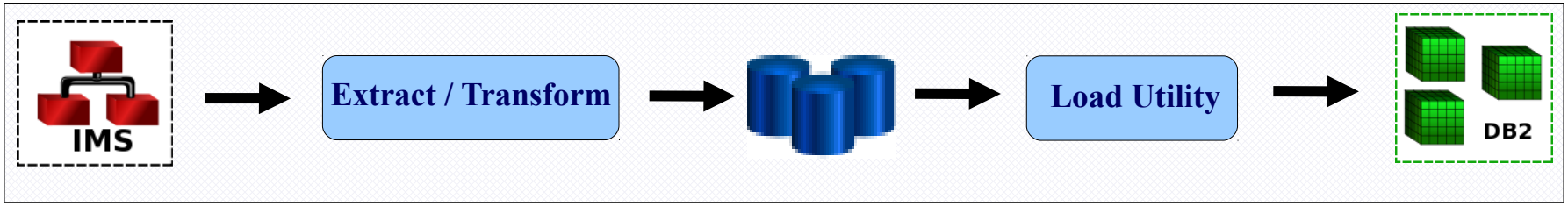
For Active /Active

- ✓ Avoid the Circular Reference
- ✓ Maintain Home Affinity
- ✓ Conflict Detection/Resolution a Must

The Role of ETL and CDC

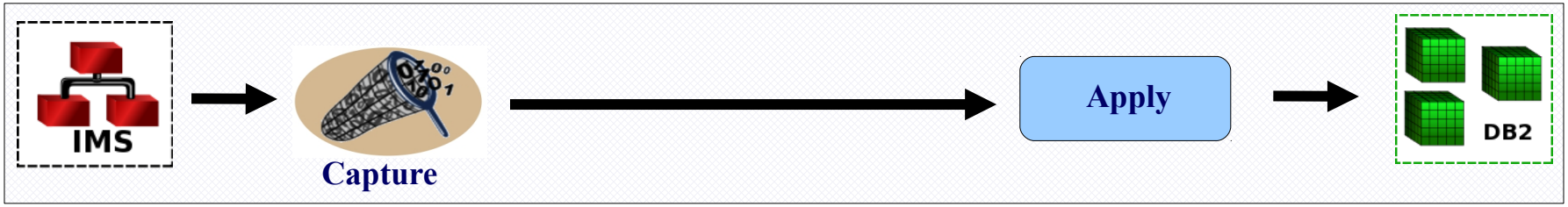
ETL (Extract, Transform, Load):

- ✓ Full Data Extract / Load
- ✓ Data Transformation Logic Defined in this Step
- ✓ Iterative Process – Must be Fast and Efficient
- ✓ Should Minimize Data Landing



CDC (Changed Data Capture):

- ✓ Keeps Data In-Sync After Initial Load – Allows for a Phased Implementation
- ✓ Should be Able to Re-Use Data Transformation Logic from ETL
- ✓ Should be Able to Replicate Both Ways (Active/Active)

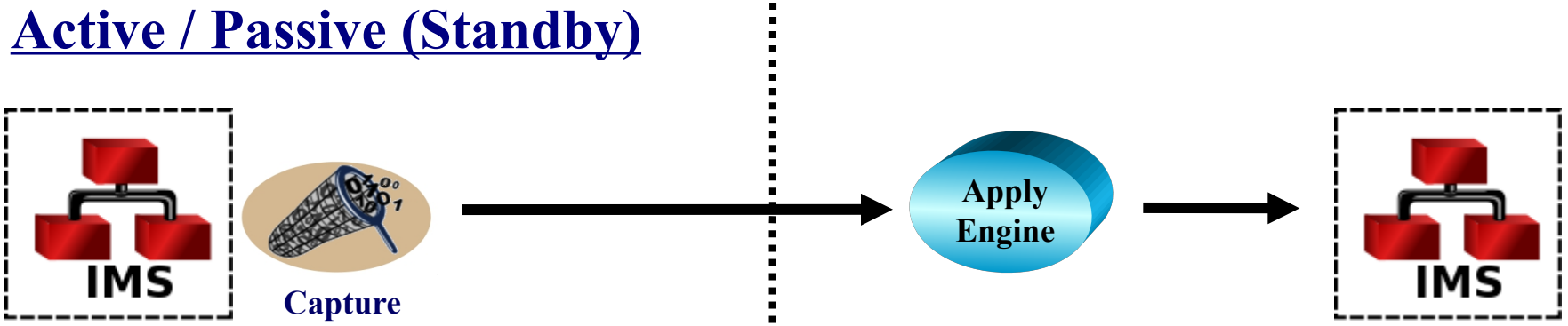


Agenda

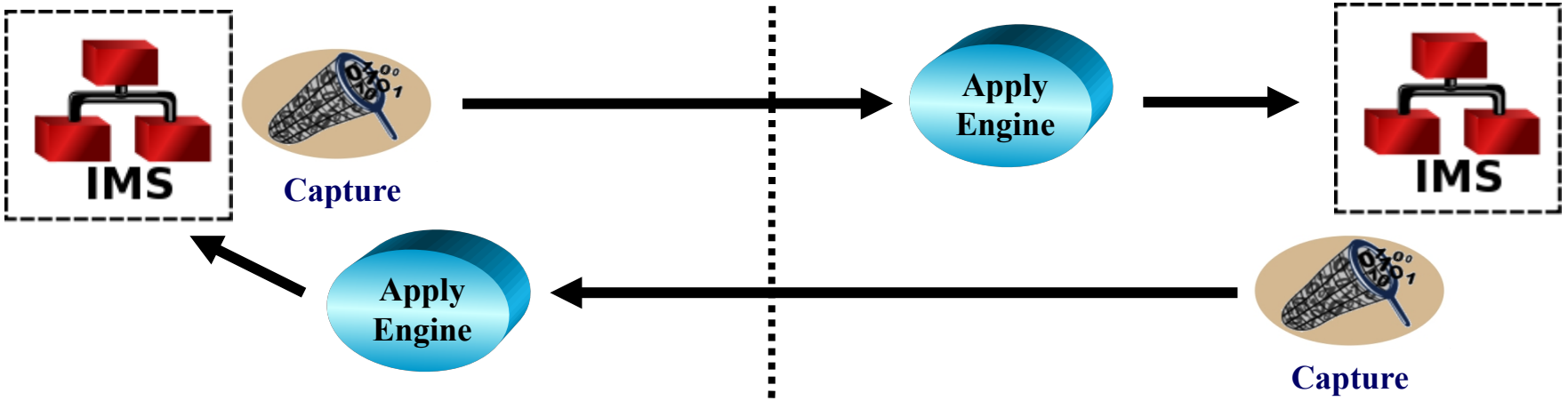
- Introduction
- Leveraging Replication for IMS Data Usage
- In-Depth Look at IMS Replication Components
- **Sample Deployment Scenarios**
- Common Challenges
- Q & A
- Conclusion

High / Continuous Availability

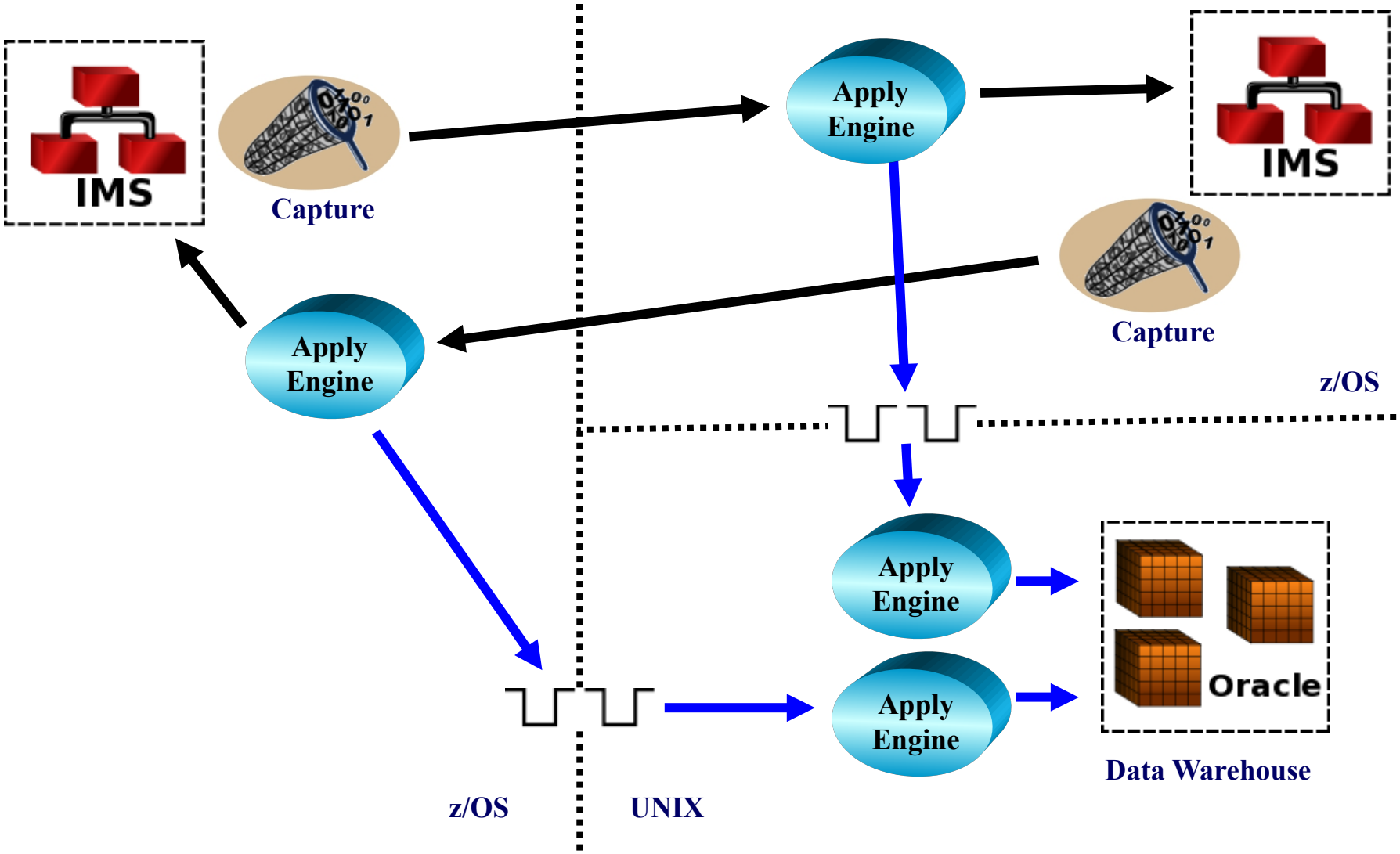
Active / Passive (Standby)



Active / Active



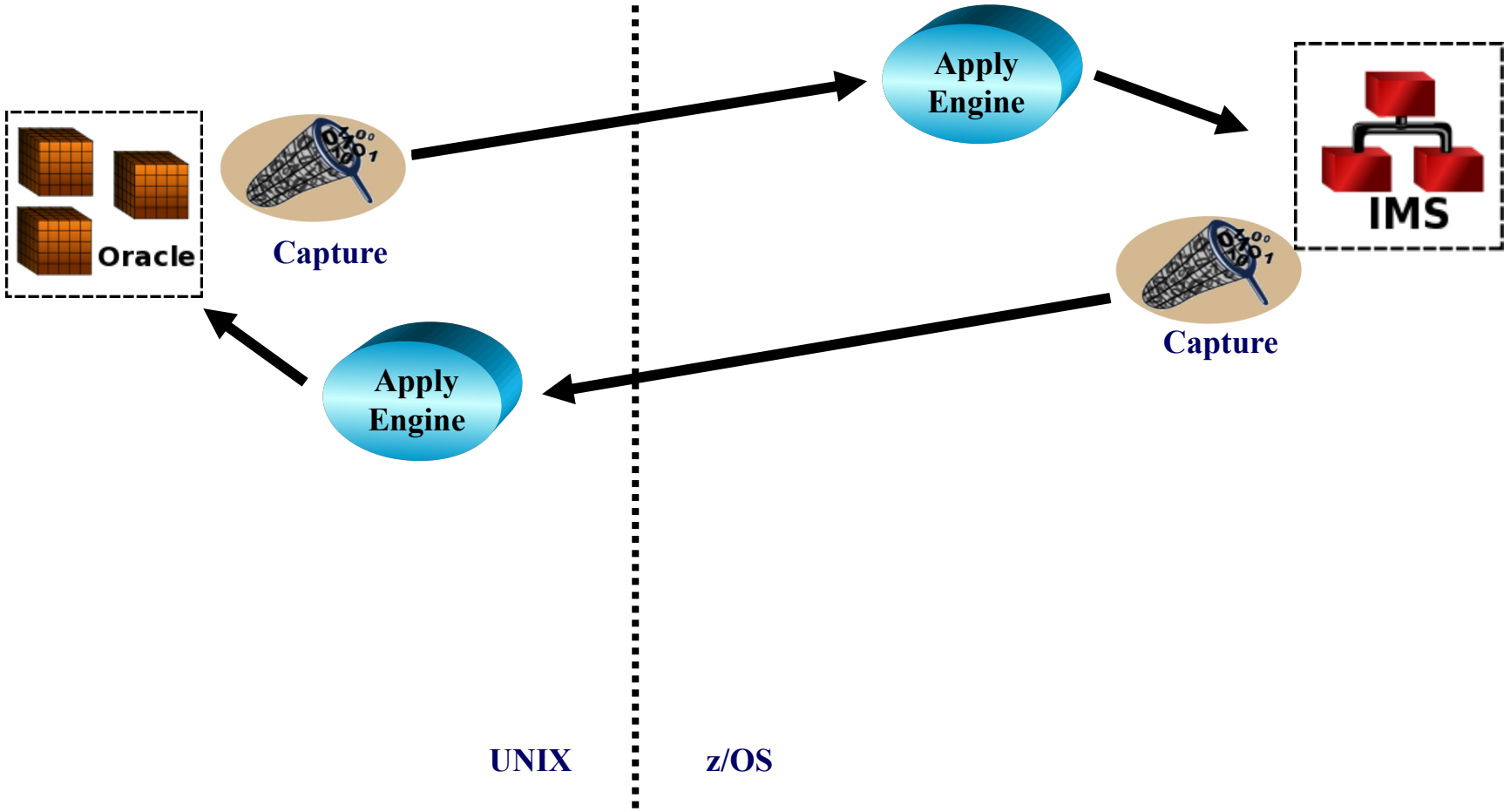
Active/Active with Selective Routing



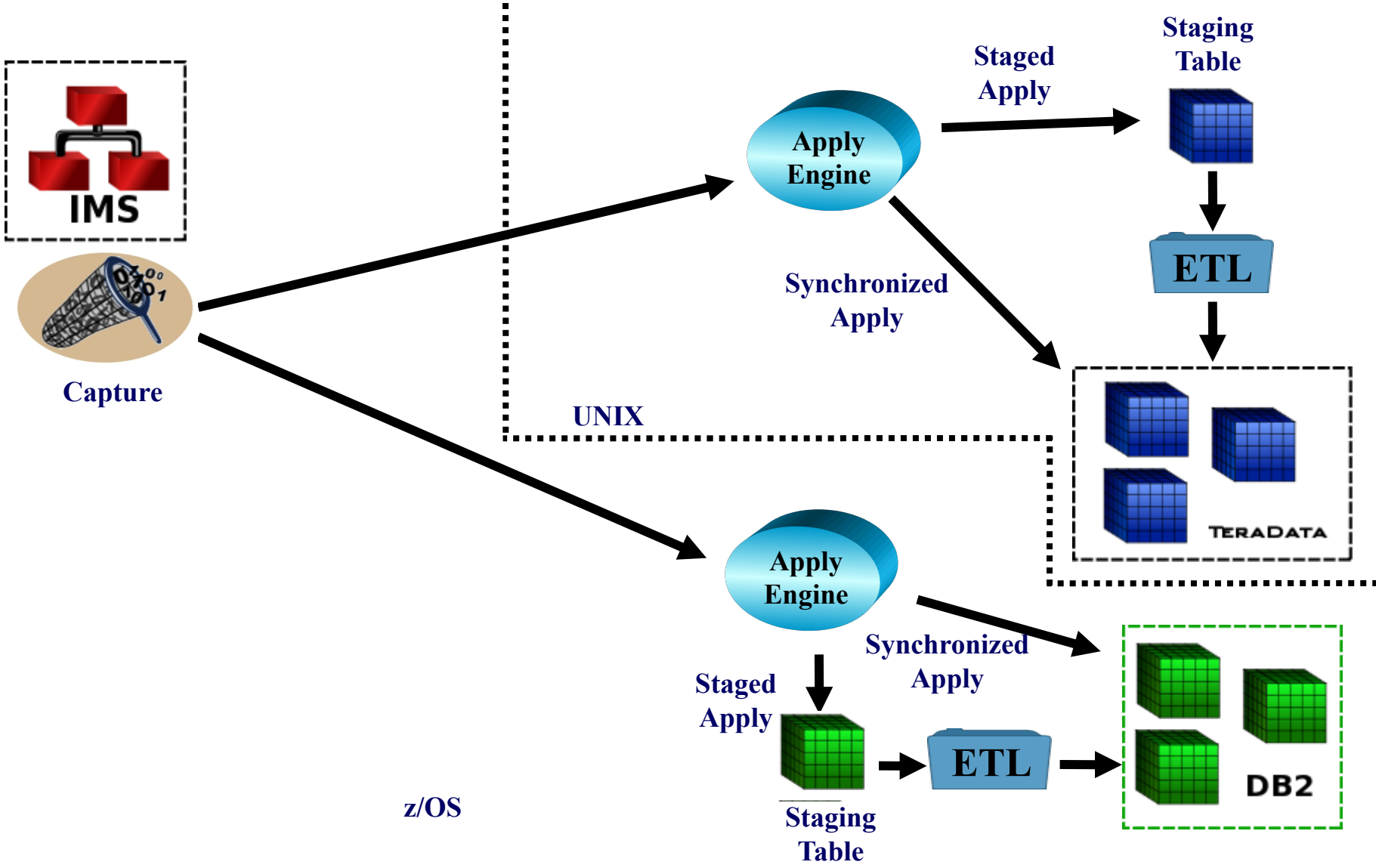
Active/Active: Oracle & IMS

Intended Use for this Setup

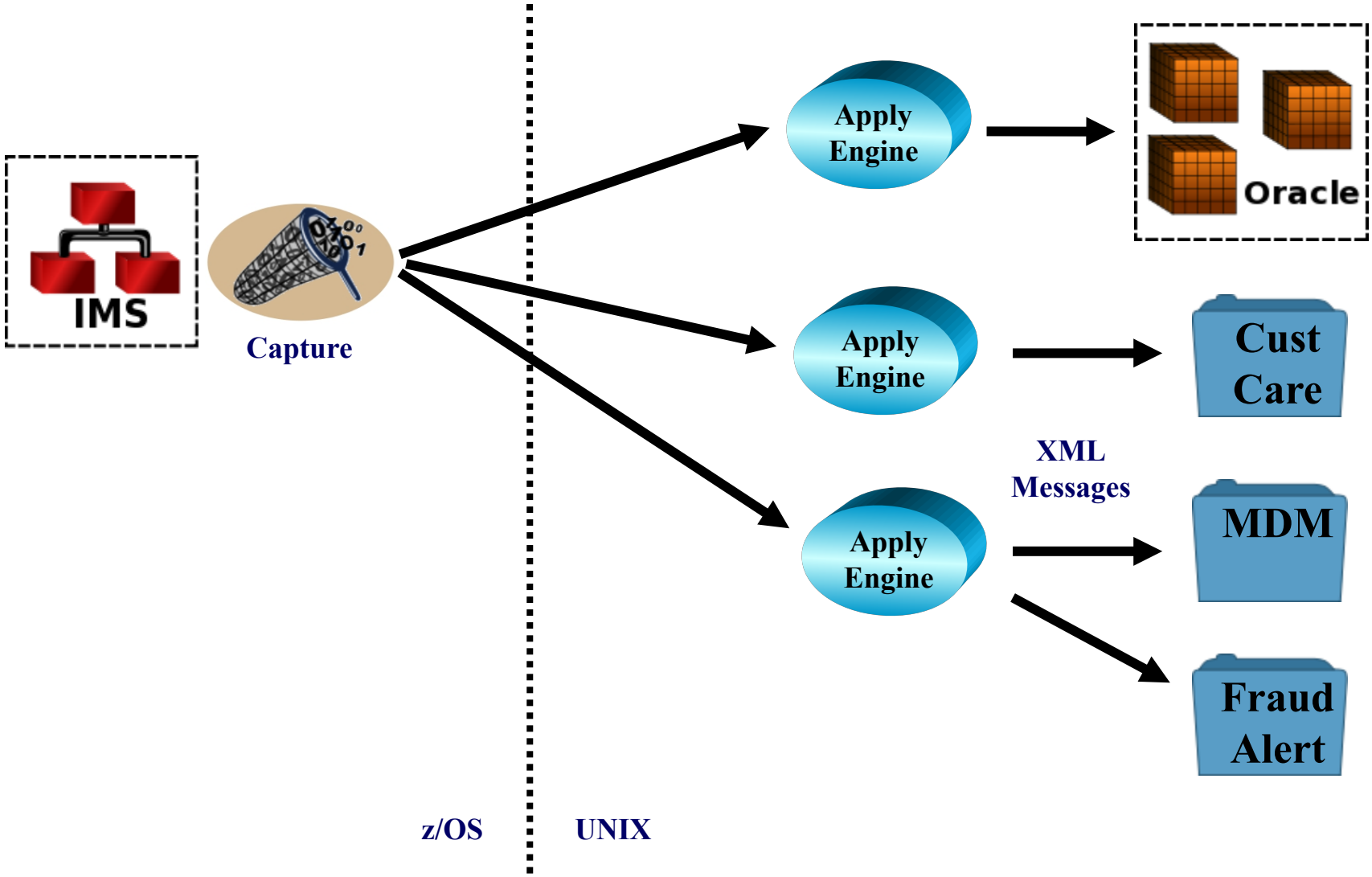
- Application Migration



CDC Enhanced ETL



Event Publishing



Agenda

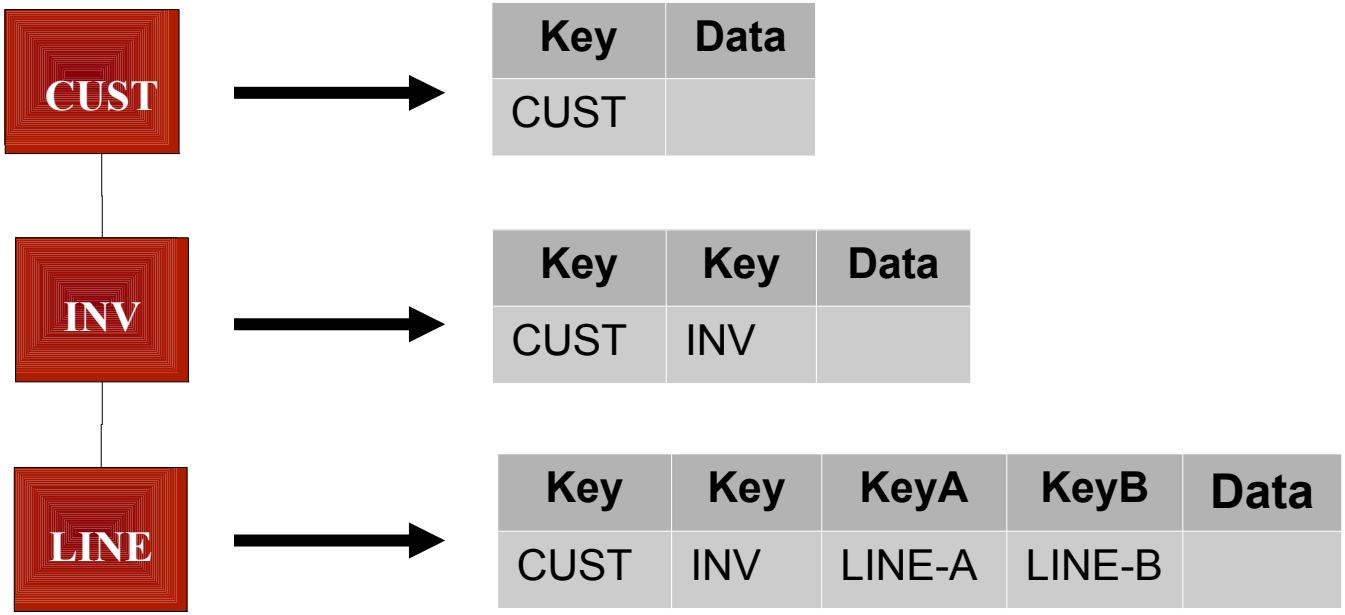
- Introduction
- Leveraging Replication for IMS Data Usage
- In-Depth Look at IMS Replication Components
- Sample Deployment Scenarios
- **Common Challenges**
- Q & A
- Conclusion

Common Challenges

- **Invalid Data**
  - ✔ Non-Numeric Data in Numeric Fields
  - ✔ Binary Zeros in Packed Fields (or Any Field)
  - ✔ Invalid Data in Character Fields
  - ✔ Business Rule Violation – Requires Assistance from SME
- **Dates**
  - ✔ Must be Decoded / Validated if Target Column is DATE or TIMESTAMP
  - ✔ May Require Knowledge of Y2K Implementation
  - ✔ Allow Extra Time for Date Intensive Applications
- **Non-Keyed Segments**
  - ✔ Typical Deployment is to Treat Entire Segment as the Key
  - ✔ Must be able to Tolerate Out-of-Logical Order Condition in Target
- **Logical Relationships → Delete Rule Restrictions for IMS Capture**
  - ✔ Logical Parents Must Have a Delete Rule of Physical or Logical
  - ✔ Logical Children Must Have a Delete Rule of Virtual

Notes on Approach

- ✓ Each Segment Maps to One (1) or More Tables
- ✓ Helpful → Keep Source Fields and Target Column Names Similar
- ✓ Design Considerations
  - Duration → Lower for Rehost...Higher for BI/DW
  - Strong Target Data Types will Require Additional Transformation
  - Be Careful to Avoid the 'Over Design'
- ✓ **Best Practice**: Keep Things as Simple as Possible



Redefined Fields

- ✓ Extends Analysis Timeline More Often than Not
- ✓ Requires Consult with SME and/or Research to Determine Which Field to Use
- ✓ Options for Simple Redefines:
  - Map Least Restrictive Field (PIC X)
  - Map Both Fields

```
05   ACCOUNT-ID                                PIC 9(7).
05   ACCOUNT-ID REDEFINES ACCOUNT-NO           PIC X(7).
```

- ✓ Options for Complex Redefines:
  - Map More Granular Field(s) → Will Require More Data Cleansing / Transformation
  - Map All Fields

```
05   ACCOUNT-ID                                PIC X(5).
05   ACCOUNT-ID REDEFINES ACCOUNT-NO.
       10  ACCOUNT-PREFIX              PIC X(1).
       10 ACCOUNT-NUMBER              PIC S9(7) COMP-3.
```

Redefined Segments: Full

- ✓ Redefine Generally Identified by One (1) or More Code Fields
- ✓ Each Redefine Mapped to a Separate Target Table

Code Field = Event Type

Key	Fairways	Greens	Hazards
Participant #	10	12	3

Key	At Bats	Hits	Runs
Participant #	10	8	2

Key	Blocks	Digs	Kills
Participant #	13	7	6

Redefined Segments: Partial

- ✓ Redefine Generally Identified by One (1) or More Code Fields
- ✓ Redefines can be Mapped to the Same Target Table if Enough Fields in Common

  or
- ✓ Each Redefine Mapped to a Separate Target Table

Premise

Code Field = Premise Type

Residential →

Key1	Key2	Addr	Pool Size	Tenants	Crop
PR#	PR_Type	123 Elm	25,000	null	null

Commecial →

Key1	Key2	Addr	Pool Size	Tenants	Crop
PR#	PR_Type	456 Ash	null	38	null

Farm/Ranch →

Key1	Key2	Addr	Pool Size	Tenants	Crop
PR#	PR_Type	456 Ash	null	null	Corn

Repeating Groups / Occurs

- ✓ Typical Candidates for Normalization Based on # Occurs
- ✓ Options:
  - Low # Occurs → Keep in Same Table as Rest of Segment
  - Map to Separate Table – Requires a Sequence Number
- ✓ Be Prepared to Handle Sparse Arrays

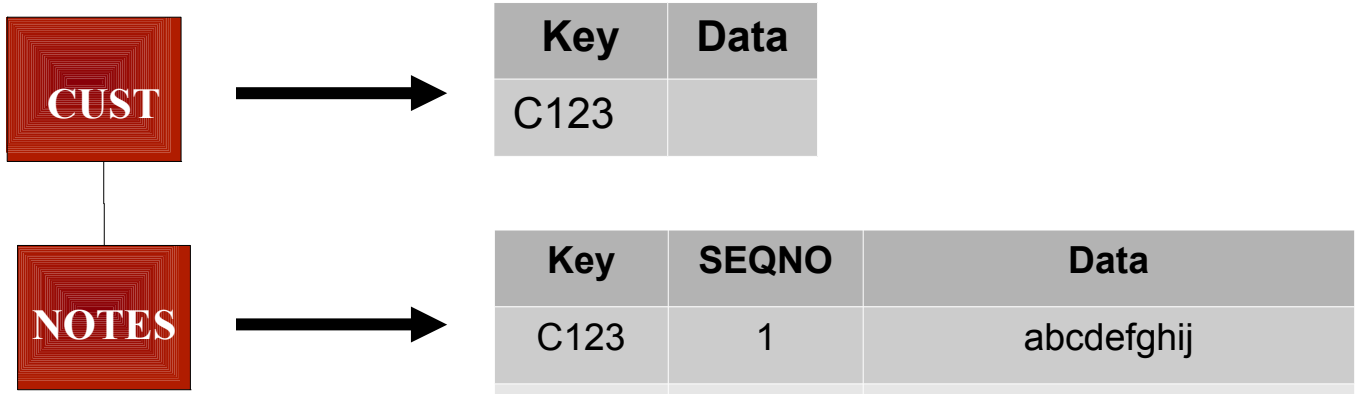
```
05  ACCT-ID                              PIC 9(7).
05  ACCT-CRDATE                          PIC X(8).
05  ACCT-BALANCE                         PIC S9(13)V99 COMP-3.
05  ACCT-ACTIVITY OCCURS 100 TIMES.
      10 ACT-DATE                        PIC 9(8).
      10 ACT-TYPE                        PIC X.
      10 ACT-AMOUNT                      PIC S9(11)V99  COMP-3.
```

ACCT_ID	ACCT_CRDATE	ACCT_BALANCE
12345	20120617	9000.00

ACCT_ID	SEQNO	ACT_DATE	ACT_TYPE	ACT_AMOUNT
12345	1	20120618	D	8000.00
12345	2	20120622	D	1000.00

Non-Keyed Segments

- ✓ Commonly Used for Text / Comments
- ✓ Straightforward for ETL
  - Unload in Order of Occurrence
  - Optional: Use a Sequence Number to Keep Things in Order on Target Side
- ✓ Tricky for CDC
  - Only Have Access to Parent Key(s)
  - Option 1: Set Apply Key to Include All Non-Keyed Data (exclude sequence #)
  - Option 2: Fully Materialize All Non-Keyed Segments when 1 Changes
  - Make Sure Your ETL/CDC Tool Can Handle Non-Keyed Segments



Key	Data
C123	

Key	SEQNO	Data
C123	1	abcdefghij
C123	2	klmnopqrs

Summary

- **Replication is a Simple Concept, but Filled with Nuances**
- **Best Practices are the Key to Success**
- **For IMS Replication, there are Few Viable Options**
- **Make Sure You Evaluate All Options**
- **Make Sure You Include SQData in Your Proof-of-Concepts**

Questions?

Where to Find Additional Information

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

Best Practices Series

Maximizing the Use of Your IMS Data with Replication

Prepared for the:

Virtual IMS User Group

5 February 2013