Best Practices Series
IMS to Relational
Data Movement

Prepared for the:
Virtual IMS User Group

7 August 2012

Agenda

➢ Introduction

➢ IMS to Relational: Success / Risk Factors

➢ Data Migration: Common Analysis / Design Challenges
  ✔ Keys
  ✔ Data Field Challenges
  ✔ Redefined Segments / Fields
  ✔ Repeating Groups
  ✔ Non-Keyed Segments

➢ 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 that Handles 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 Integration 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 IMS to Relational?

➢ Provide Users with a Method of Querying Data Outside of IMS

➢ Business Intelligence / Data Warehousing

➢ Co-Existence with Newer Applications

➢ Application Migration / Replacement

➢ "We're Moving Off of the Mainframe"....☺

SQDATA

Success Factors

➢ **Access to Subject Matter Expert(s)**
- ✔ Significantly Decreases Risk
- ✔ Leverage Knowledge of Data / Business Rules
- ✔ Becoming More Difficult to Obtain with Outsourcing, Retirement, etc.

➢ **Planning**
- ✔ Required to Keep Risk at a Minimum
- ✔ Secure the Proper Personnel
- ✔ 40%: Analysis and Design
- ✔ 20%: Conversion → Assuming a Tool is Used
- ✔ 40%: Testing / Validation

➢ **Analysis / Design**
- ✔ IMS to Relational Data Modeling
- ✔ Source to Target Mapping Specifications
- ✔ Validation Criteria / Test Plan

➢ **Validation**

High-Risk Elements

➢ **No Access to Subject Matter Expert(s)**
  - ✔ Significantly Increases Risk
  - ✔ Extends the Project Timeline
  - ✔ Results in Guesswork for Design and Mapping

➢ **Underestimating the Complexity of IMS to Relational**

➢ **Big Bang Approach - Attempting to Migrate Everything at Once**
  - ✔ Recommend Phased Implementations
  - ✔ Subsequent Migrations become Shorter: Experience & Lessons Learned

➢ **Fast-Tracking Planning and Analysis**
  - ✔ Causes Unnecessary Rework and Waste
  - ✔ More Time Spent on the Front End Saves on the Back End

➢ **High-Transaction Workload on the IMS Side**
  - ✔ Applies Primarily to Application Conversion
  - ✔ Performance will NOT be the Same as with IMS

Common Implementations

➢ **Simple Conversion**
  - ✔ Relational Model Closely Resembles IMS Structures
  - ✔ Shortest Migration Timeline
  - ✔ Highest Chance for Success if SMEs are Not Available

➢ **Business Intelligence / Data Warehousing**
  - ✔ Relational Models can Diverge from Existing IMS Structures
  - ✔ Master Data Management (MDM) Comes into Play
  - ✔ More 'Moving Parts' / Dependencies than Simple Conversions

➢ **Application Integration**
  - ✔ Relational Models are Dictated by New Application
  - ✔ Usually Requires More Data Transformation: SMEs Critical

➢ **Application Conversion**
  - ✔ Most Complicated Implementation
  - ✔ Relational Model Depends on Extent of Application Design
  - ✔ Significant Time Must be Allocated for Testing / Validation

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

IMS → **Extract / Transform** → → **Load Utility** → DB2
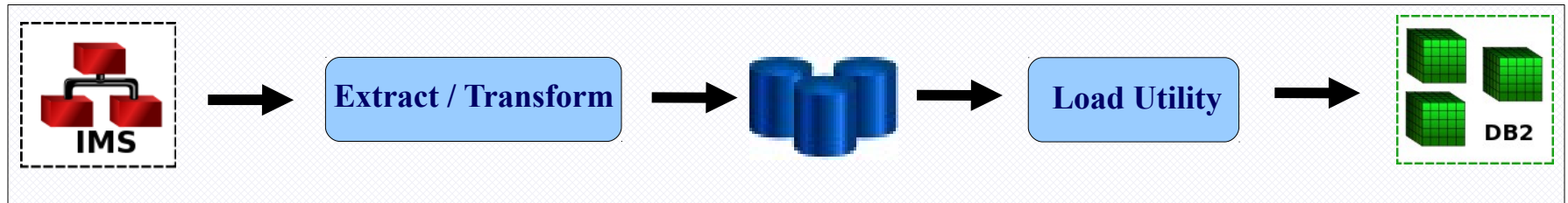
CDC (Changed Data Capture):

✓ Keeps Data In-Sync After Initial Load – Allows for a Phased Implementation

✓ Should be Able to Re-Use Data Transformation Logic from ETL

✓ Helpful to be Able to Replicate Both Ways
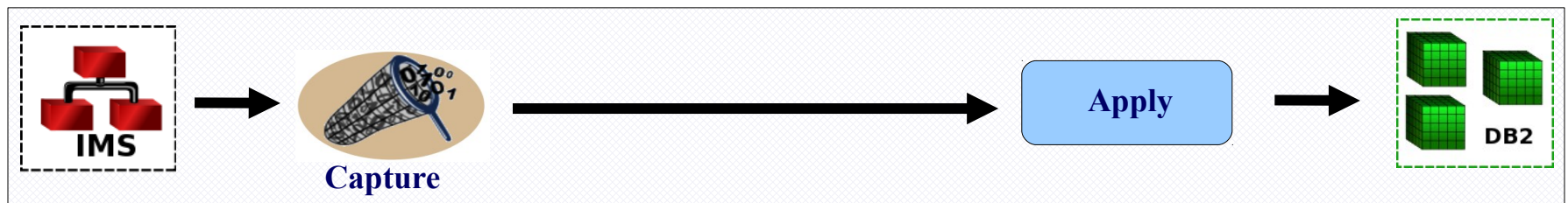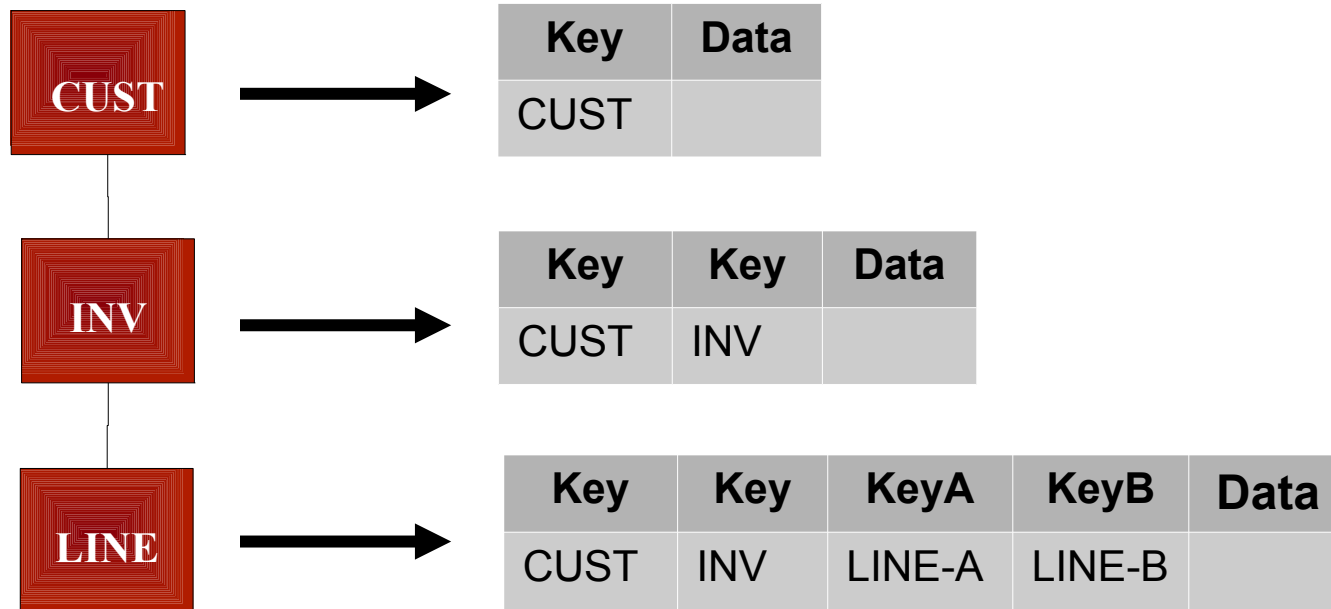
IMS → **Capture** → **Apply** → DB2

Agenda

➢ Introduction

➢ IMS to Relational: Success / Risk Factors

➢ Data Migration: Common Analysis / Design Challenges
   ✔ Keys
   ✔ Data Field Challenges
   ✔ Redefined Segments / Fields
   ✔ Repeating Groups
   ✔ Non-Keyed Segments

➢ Q & A

➢ Conclusion

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'
- ✓ **<u>Best Practice</u>**: Keep Things as Simple as Possible

CUST →

Key	Data
CUST	

INV →

Key	Key	Data
CUST	INV	

LINE →

Key	Key	KeyA	KeyB	Data
CUST	INV	LINE-A	LINE-B	

Keys

- ✓ Fairly Straightforward → IMS Key Structure Simplifies Things
- ✓ Carry Parent Keys in Dependent Tables
- ✓ Plan on Keys being Comprised of Multiple Fields with Different Data Types
  - – Character, Packed, Binary

Key	Data
CUST	

Key	Key	Data
CUST	INV	

Key	Key	KeyA	KeyB	Data
CUST	INV	LINE-A	LINE-B	
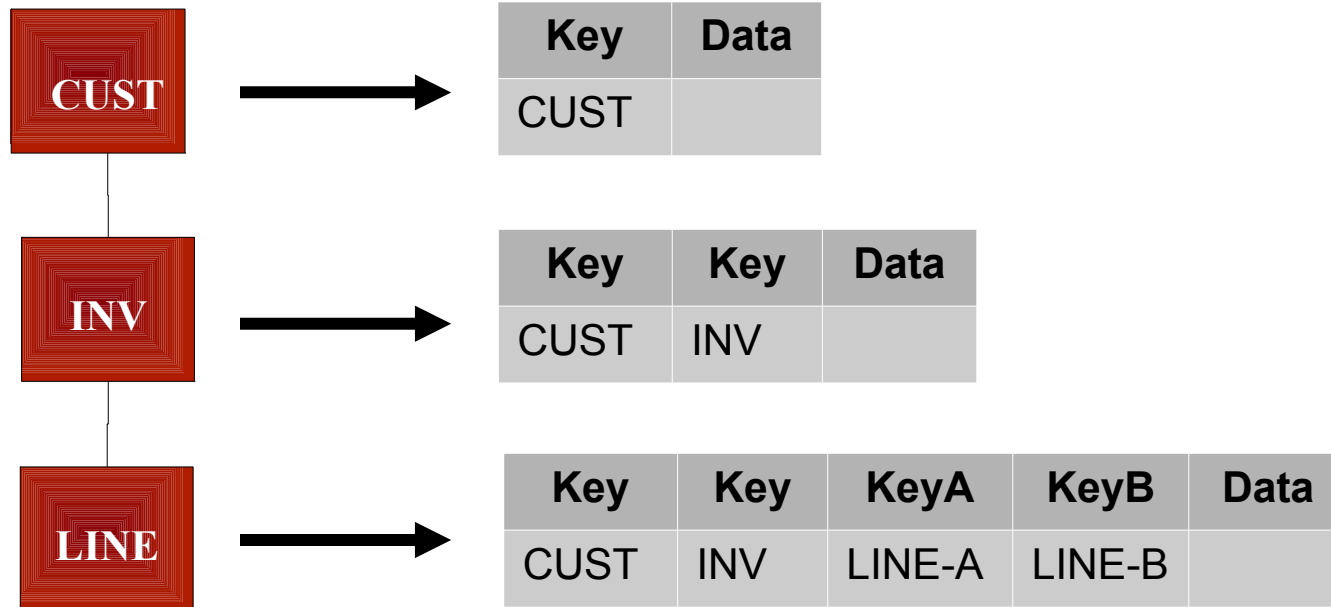
Common Data Challenges

➢ **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

➢ **Text / Comment Fields**
  ✔ Usually Mapped to VARCHAR
  ✔ Stop Mapping at First Non-Printable, Non-Control Character

➢ **Binary / 'Special' Fields**
  ✔ Very Common in Older Applications Developed in 1970s / 80s
  ✔ Generally Requires Application Specific Translation

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

Event Stats

Code Field = Event Type

Golf →

Key	Fairways	Greens	Hazards
Participant #	10	12	3

Baseball →

Key	At Bats	Hits	Runs
Participant #	10	8	2

Volleyball →

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

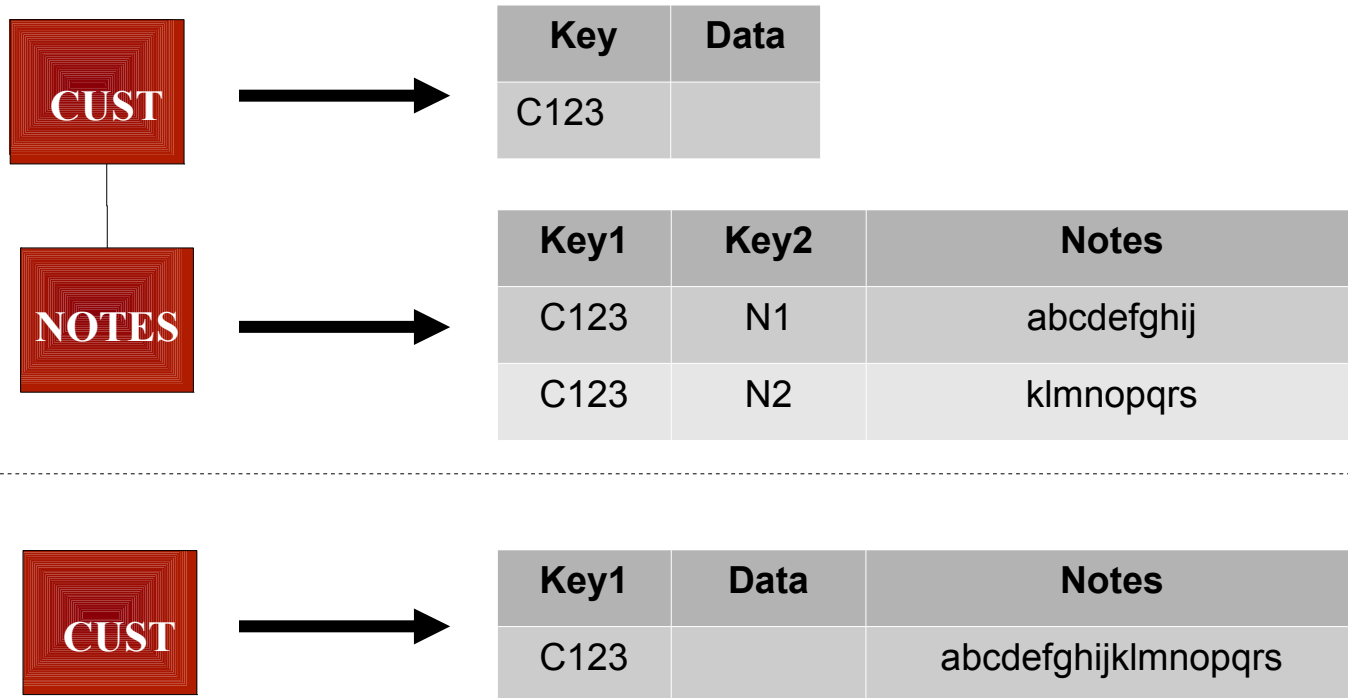| ACCT_ID	SEQNO	ACT_DATE	ACT_TYPE	ACT_AMOUNT
12345	1	20120618	D	8000.00
12345	2	20120622	D	1000.00

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

CUST →

Key	Data
C123	

NOTES →

Key	SEQNO	Data
C123	1	abcdefghij
C123	2	klmnopqrs
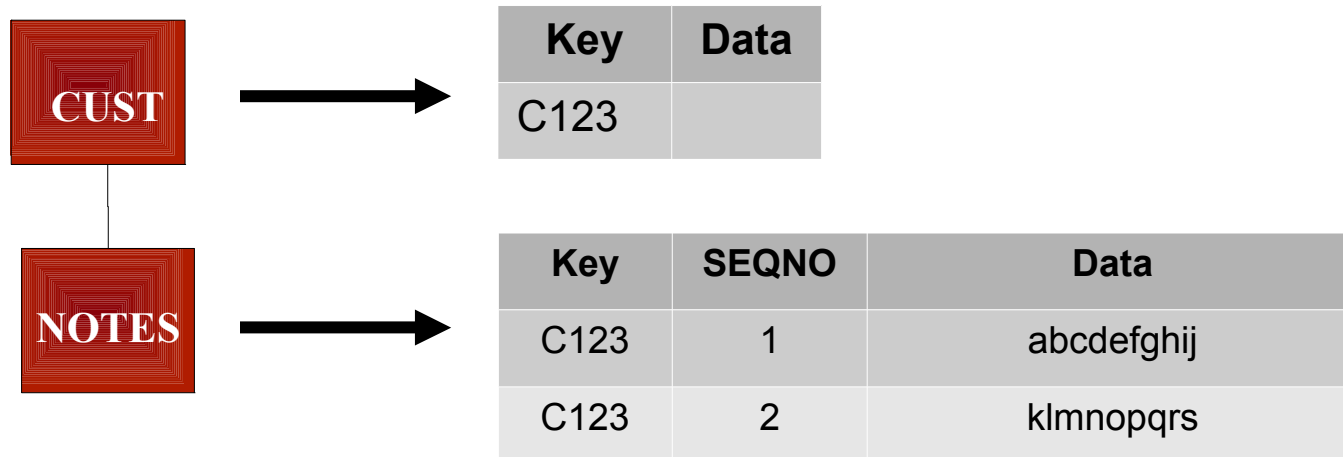
Continuation Segments

✓ Common in Older Applications

✓ Text / Comment Field Split Across Multiple Segments

✓ Options:

  – Map Each Segment Instance to a Separate Table

  – Combine and Map to Same Table (ETL Trickier than CDC for this Option)

CUST →

Key	Data
C123	

NOTES →

Key1	Key2	Notes
C123	N1	abcdefghij
C123	N2	klmnopqrs

CUST →

Key1	Data	Notes
C123		abcdefghijklmnopqrs

Data Validation

- **Does Not Have to be as Challenging as You May Think**
  - ✓ Human Verification → Required During Initial Conversion
  - ✓ Automated Verification → May Require Utilizing ETL / CDC Scripts
    - Counts
    - Check Sums
    - Compare Source / Target Fields with Same Attributes

- **Helpful → Make Sure Your Vendor Can Assist You with This**

IMS

DB2

Extract / Summarize

Extract / Summarize

Compare
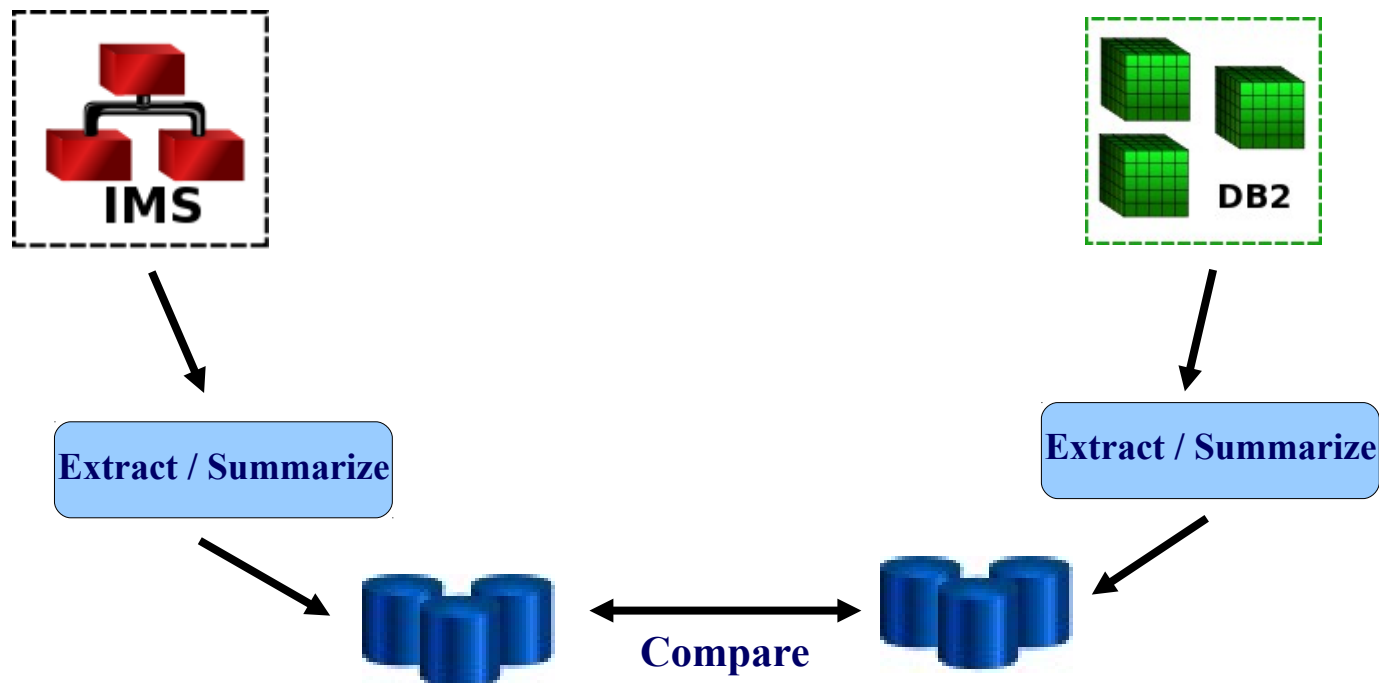
Summary

➤ **Secure Access to Subject Matter Expert(s) if Possible**
- ✔ Significantly Decreases Risk
- ✔ Leverage Knowledge of Data / Business Rules
- ✔ Becoming More Difficult to Obtain with Outsourcing, Retirement, etc.

➤ **Don't Shortcut**
- ✔ Planning
- ✔ Analysis / Design

➤ **Don't Overdo Database Design**
- ✔ Never Ending Project
- ✔ End Result Too Complicated for Users

➤ Make Sure Your Conversion Tool Does Most of the Work

➤ Have a Reliable Method of Data Validation

➤ Make Sure that Your Tool Vendor has the Capability to Assist You

Where to Find Additional Information

➢ Email Requests
  – info@sqdata.com

➢ Phone Requests
  – 866-252-3575

➢ Website
  – www.sqdata.com

Best Practices Series

IMS to Relational
Data Movement

7 August 2012