Data Migration & Conversion Methodology Toronto User Group – September 20, 2017

Thibault Dambrine

Agenda - Data Migration & Conversion Methodology

Introduction: *Choices*

- Preparation: *Foundation for success*
- Methods: *Ensure no surprises*
- Execution: *A non-event!*

Conclusion: *Points to remember*

Introduction
Data Conversion Methods: What are the Choices?

1) **Big Bang: Do it all, Do it once!**

    + Done in shortest period of time

    - Biggest Risk of "Load and Explode"

2) Trickle-In: Bring in portions of the data over time

    + No Down Time

    -    Having to track what has been migrated and what has not

    -    Having to maintain two system for extended periods

    -    Does not protect from "Load and Explode", as larger volume processing is only coming in late

3) Three Practice-Cycles, go/no-go decision, Go live

    + Provides potential opportunity to abort if necessary

    + Not having to maintain two system for extended periods

    + Risk of "Load and Explode" mitigated by practice cycles

    - Requires a big preparation effort (expensive)

This will be the focus
of this presentation

Introduction: Data Migrations and Wire Walks

August 7, 1974, 7:15 AM

Philippe Petit walked between the newly built WTC Twin Towers
- A one-time event
- Do or die
- No net

In many ways similar to
a data conversion

Preparation is the root of success

Conversion – Project Overview

Timeline	Period 01	Period 02	Period 03	Period 04	Period 05	Period 06	Period 07	Period 08	Period 09	Period 10	Period 11	Period 12	Period 13	Period 14	Period 15	Period 16	Period 17	Period 18
Conversion Preparation		Staffing & Planning																
		Data Conversion Approach																
		Data Quality Specifications, Data Field Mapping Specifications																
			Data Cleanup															
			Create Conversion Programs (extract/transform/load)															
			Data Conversion Signoff Criteria															
			Plan for Hardware Readiness															
			Build Conversion Infrastructure															
			Interface Modification & Testing															
		Plan for Training																
Conversion Tasks									Practice Conversion 1									
											Practice Conversion 2							
													Practice Cutover				
				Initiate training					Training with Company Data									
									Refine extracts & Conversion Practice									
									Test and Polish up Interface Functionality									
														Actual Cut-Over & Go Live				
															Stabilize and Operate			

Preparation & Foundations:
Conversion Specifications & Project Planning

Conversion Specification Documents are the foundation

Project Management Preparation

- Team Formation
- Data Conversion Approach (DCA)
- Constraint and Risk Mitigation

Stakeholder and Data Preparation

- Data Quality Specifications (DQS)
- Data Field Mapping (DFM)
- Interface Register
- Data Conversion Signoff (DCS)

Project Management Preparation

Project Preparation: Identify Project Stakeholders

Project
- Sponsor & Project Board
- Project Manager
- Technical Architect / System Administrator

Business
- Division business data migration lead(s)
- Business Stakeholders – Data Owners (will assist in the data cleanup cycle)
- New System Subject Matter Experts
- Training for new system

IT
- Technical Architect / System Administrator
- Interface Team Lead
- ETL Lead and Subject Matter Experts
- Data Migration Scheduler (will scheduled, manage and track data as it is extracted, cleansed and uploaded)

Project Preparation: Data Conversion Approach (DCA) Components

1) Time & Scope
- Timeline / Gantt Chart
- Business Scope (e.g. Business units)
- Data Scope (e.g. volumes)
- Constraint and Risk Mitigation
- Security Scope (e.g. medical records, corporate secrets)
- Conversion Effort Management
  - conversion analysis and programming
  - hiring outside resources and / or contractors

2) Human Resources
Create a Stakeholder map & RACI Chart
Core Team
- Sponsor, Board, Project Management
- Technical staff

Business Team
- Key resources
- Training staff

3) Technical Resources
Hardware & technical
- Hardware infrastructure
  - Computing platform
  - Storage
  - Networking
- Software licenses
- Document agreed-upon customizations

4) Budget Resources
- Staff, Hardware
- Software Modification Effort/Cost
- Data Cleanup Effort
- Extract / Transform / Load (ETL)
- Scheduling and Migration
- Business Training Budget

Project Management Preparation: Risk Mitigation

Business

- Strained resources (availability of Business Experts)
- Data Loss, Data Corruption or Unwanted Data
- Extended downtime
- Target Application Parameterization

Technical

- Semantics (right, complete data landing in the wrong field)
- Target System Stability
- Orchestration - occurs when the processes of data migration are not performed in order

Almost all the conversion risks can be significantly mitigated with three practice conversion cycles before going live

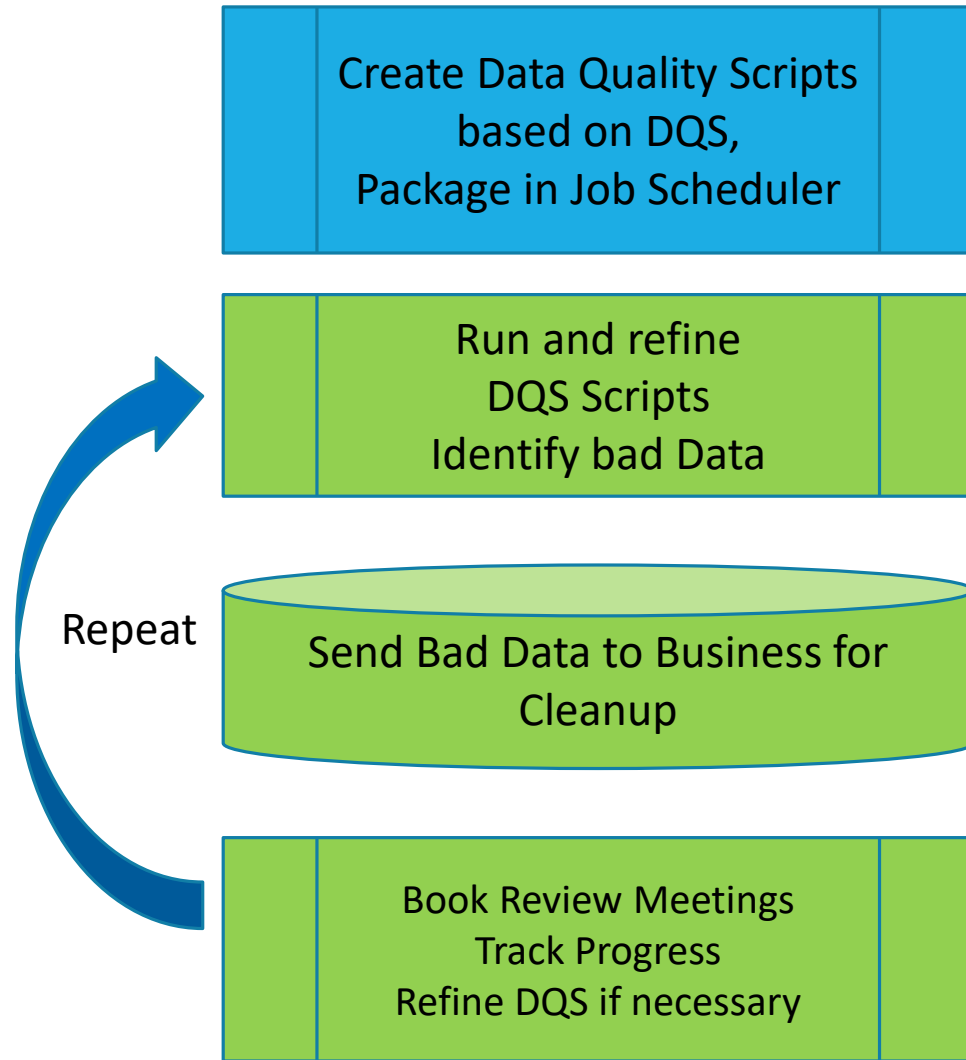Stakeholder and Data Preparation – Focus on the 4 "Anchor" documents:

- Data Quality Specifications (DQS)

- Data Field Mapping (DFM)

- Interface Register

- Data Conversion Signoff (DCS)

Data Quality Specification (DQS): Define the meaning of "Good Data"

For each Data Set
- Book DQS Workshop with the Business - Bring File/Field Description & any available base knowledge – Don't come empty-handed!
- Document what constitutes "good data"
- Document what "bad data" to eliminate
- **Speak in Business terms. Bring examples:**

- **Accuracy**: Can it be validated? Are there duplicates or near-duplicates? Does it follow the rules?
- **Completeness**: Does it provide all the information required? What data may be missing?
- **Compliance**: Does it comply with regulatory standards?
- **Consistency**: Is it consistent and easily to convert with little or no exceptions?
- **Data Types**: e.g. numeric data in a legacy Alpha field, open for errors
- **Integrity**: Does it have a coherent, logical structure? Is there orphaned data e.g. PO without Vendor)
- **Order**: Is the data in the right place?
- **Relevance**: Is it relevant to its intended purpose? How much of that data do we really need?
- **Timeliness and Value**: Is it current? Up to date? Worth converting?

These rules will determine the code necessary to identify "Bad Data"

Data Quality Standards (DQS) Usage -> The Data Cleanup Cycle

Create Data Quality Scripts based on DQS, Package in Job Scheduler

Run and refine DQS Scripts Identify bad Data

Repeat

Send Bad Data to Business for Cleanup

Book Review Meetings Track Progress Refine DQS if necessary

Objective: Identify non-compliant data

- Who: Business Users, Legacy System Experts, ETL Technical staff
- What: SQL Scripts that will be packaged into CL's and Scheduler jobs
- When: DQS Jobs will be run on weekly basis, results will be fed back to the Business

Objective: Eliminate non-compliant data

- DQS Script will identify the non-compliant data.
- The Business will then have to find a way to fix it in the legacy system
- Refine DQS and DQS Scripts if new bad data is discovered

Repeat cycle until Go Live

- Data Cleanup is iterative (new data is created every day!)

Data Field Mapping (DFM)

Build First Specification

Get an early start – best effort based on experience

Document mappings , rules and processes

Validate with The Business

Book DFM Workshops with the Business

Review mappings & Transformation processes & rules

Get signoff on all from the Business

Source	Source Data Element	Destination	Target Data Element	Transformation/ Cleansing Rules	Success/ Signoff Criteria	Notes
Legacy Customer Master	Customer Number	New Customer Master	Customer ID	Only customers with transactions in the last 3 years	Get verification by the Business	New Customer ID will start at number 10000 Track old Customer to New Customer in a separate table
Legacy Customer Master	Customer Name	New Customer Master	Customer Name Split into - First Name - Last Name	Split Name into - First Name - Last Name - Ensure no duplicates or near-duplicates	Get verification by the Business	Review data to check if the same name appears in different addresses as well

Conversion Code Scheduling : Consistency and Repeatability

Group each set of instructions - Data Cleanup and ETL (using Data Field Mapping) - into a repeatable package
- Procedures must be repeatable with consistency, so that each run can be compared

The sequence of Scheduler Groups must be themselves run in a consistent order, documented in the DCA
- Data Cleanup (recurring)
- Setup Data
- Master Data
- Transaction Data
- Interface Data

For each job scheduler group run, record results for comparison and trending purposes
- The quantity of data processed for each extract
- The duration of each extract
- The success and failures, as per the pre-established success criteria (quantities, dollar amounts, etc.)
- Go through the Data Conversion Signoff process – Improve DCS documents if any gaps are discovered

Interface Register

Requirements

Interfaces often involve outside stakeholders and physical infrastructure (low control)
- Transition planning has to start early

Establish Base List of
– Data Interfaces
– Stakeholders
– Characteristics
– Effort & Lead time required for transition and testing

Organize Data Conversion workshops by interface with
– Legacy System Experts (Data Owners)
– New System Experts
– IT Conversion Analysts

Aim:

Map Interfaces from legacy to new system

Establish task list for each interface
– Decommission?
– Modify?
– Replace?
– How much effort is necessary?

Key Deliverables:

Document transition Approach for each interface
– Old vs. New
– Task list to transition

Success & Signoff Criteria

Data Conversion Signoff (DCS)

Deliverable

– List of tests that will effectively validate the conversion results

Aggregate Reconciliation Data Tests

• Have all the records been migrated?

• Do the totals match between the old and the new systems?

• Run equivalent jobs on old system and new system and compare results

• Run duration comparisons

Balancing Between Inter-related Systems

• Ensure balance with accounting systems are all sync'd up

  – Vendors with open balances in the Accounts Payable Ledger

  – A cycle count (physical) should be completed prior to loading inventory positions

Individual Data Tests

• Review Conversion Utility reports – e.g. SAP exception reports

• Key data sample validations e.g. high-value customers

Usage

– FOR EACH CONVERSION CYCLE:
  -> Each DCS to get formal Business
  -> Each data group must get a signoff

– Store each e-mail response in a purpose-built sub-directory

Stakeholder and Data Preparation:
Create a Data Conversion Signoff (DCS) Tracker

Create a sub-directory and tracker to store DCS Signoff documents

For each Data set signoff, get a signed acknowledgement

– Each of data set must be signed off explicitly with an e-mail message by the Data Owners

– Each e-mail message will be stored in a DCS Signoff subdirectory
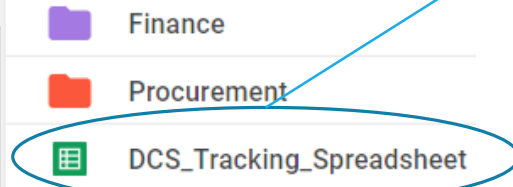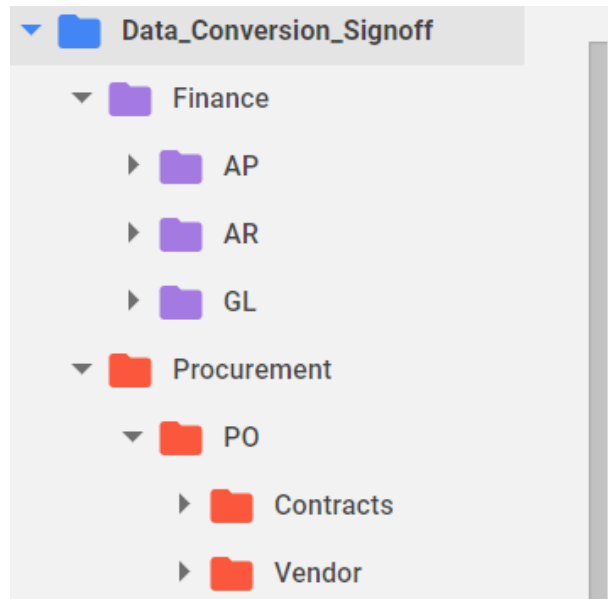
Data Conversion Signoff							
General Category	SubCategory 1	SubCategory 2	Conversion Cycle	Approved by (Business Stakeholder)	Signoff Date	Signoff Review Notes	Signoff Document link
Finance	AR		1	Jimmy Page	8/5/2017	Good	Link
Finance	AP		1	Pete Townshend			
Finance	GL		1	Steven Tyler			
Procurement	PO	Contracts	1	Roger Waters	8/7/2017	Review mapping	Link
Procurement	PO	Vendor	1	Jon Anderson	8/2/2017	Good	Link
Inventory	BOM		1	Jon Anderson	8/2/2017	Good	Link
Inventory	Inventory Transactions		1	Brian May	8/6/2017	Good but need to update DCS	Link
Inventory	Item Master		1	Roger Waters	8/7/2017	Review mapping	Link
Inventory	Purchasing		1	Peter Gabriel	8/4/2017	Missing Fields	Link

Folder structure:
- Data_Conversion_Signoff
  - Finance
    - AP
    - AR
    - GL
  - Procurement
    - PO
      - Contracts
      - Vendor

- Finance
- Procurement
- DCS_Tracking_Spreadsheet

Hardware, Licenses & Infrastructure

"The WTC Walk" Infrastructure: Hardware Matters!

Grounding Wires are essential to the wire stability over long distances

They are typically anchored to the ground – in the case of the WTC Walk, they were anchored horizontally
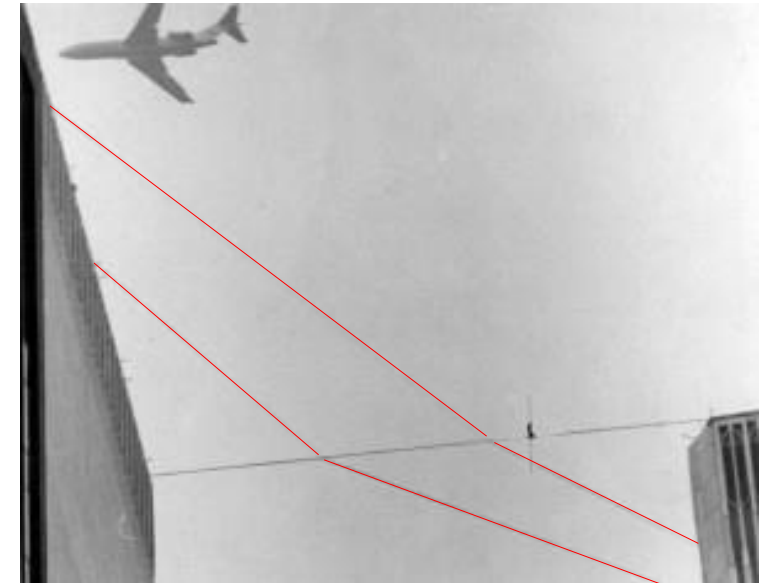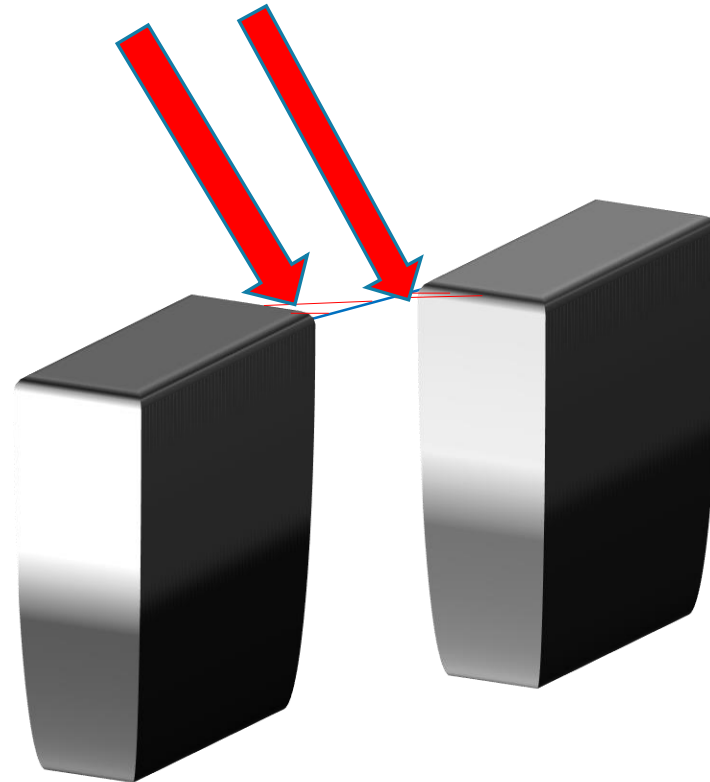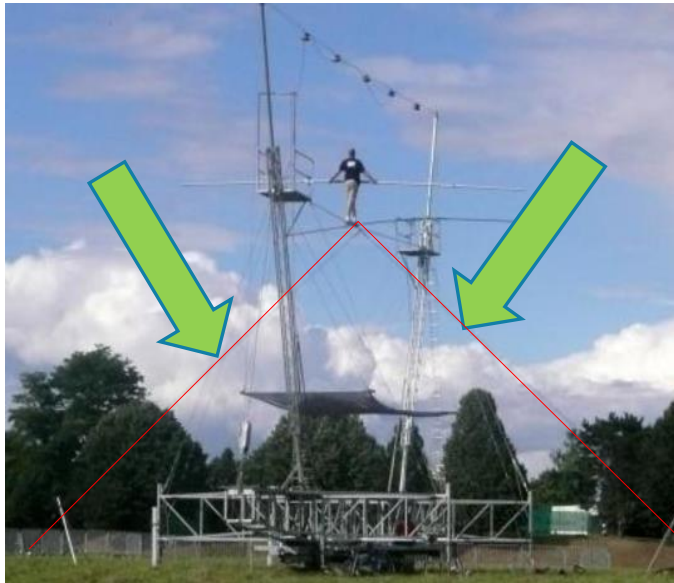
Hardware and Infrastructure

Key Principle: To Protect Production, to minimize downtime due to conversion

Hardware Requirements:

Determine allowable downtime (refer to Migration Plan & Schedule) – should be part of the DCA

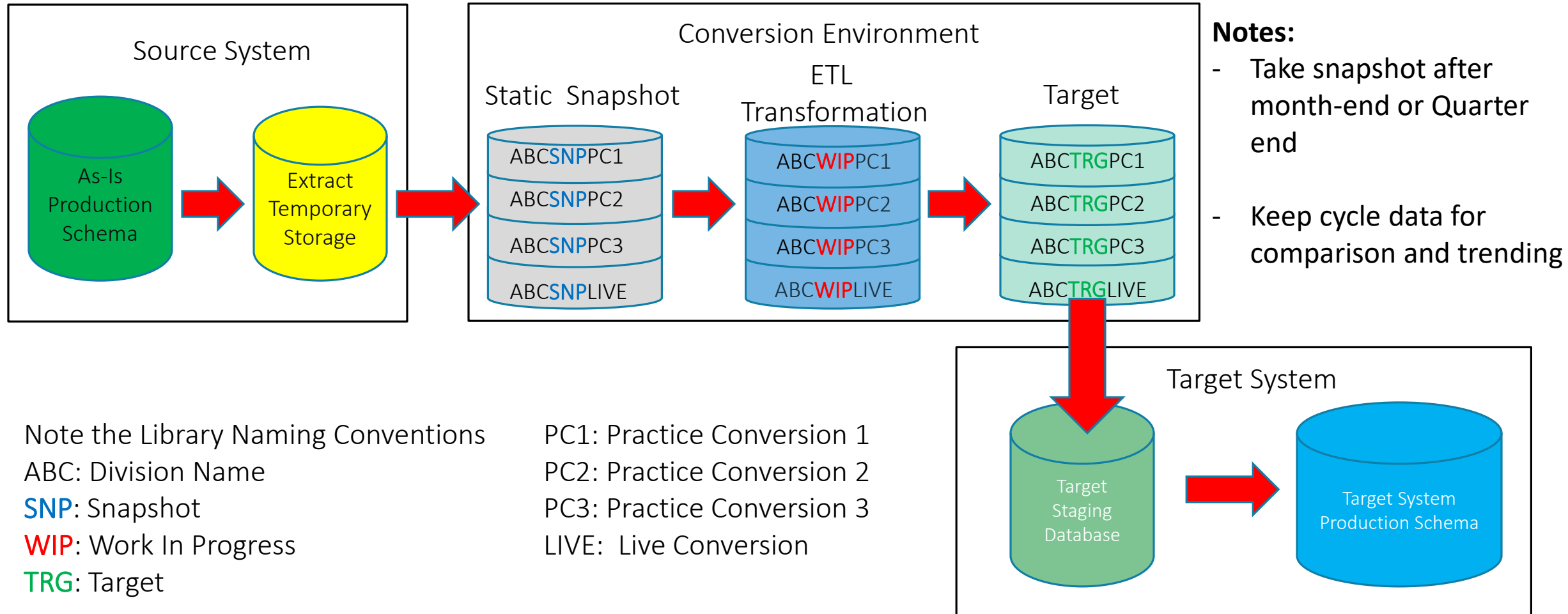From this, determine Hardware Requirements

- Network bandwidth – particularly critical if Business is shifting to Cloud Computing
- Storage Capacity, including temporary staging capacity requirements
- CPU Capacity

Validation:

- Validate hardware requirements early on with vendor
- There will be 3 practice conversion cycles

      Make use of those to validate hardware and sizing in test conversion cycles

Hardware and Infrastructure
Sample Conversion Data Migration Path

Source System

As-Is Production Schema → Extract Temporary Storage

Conversion Environment

Static Snapshot

- ABCSNPPC1
- ABCSNPPC2
- ABCSNPPC3
- ABCSNPLIVE

ETL Transformation

- ABCWIPPC1
- ABCWIPPC2
- ABCWIPPC3
- ABCWIPLIVE

Target

- ABCTRGPC1
- ABCTRGPC2
- ABCTRGPC3
- ABCTRGLIVE

Notes:
- Take snapshot after month-end or Quarter end
- Keep cycle data for comparison and trending

Target System

Target Staging Database → Target System Production Schema

Note the Library Naming Conventions
ABC: Division Name
SNP: Snapshot
WIP: Work In Progress
TRG: Target

PC1: Practice Conversion 1
PC2: Practice Conversion 2
PC3: Practice Conversion 3
LIVE: Live Conversion

Migration Practice Cycles and Execution

Practice Cycles and Actual Conversion

Preparation is far enough along – Prepare Data Conversion Cycles

Timeline	Period 01	Period 02	Period 03	Period 04	Period 05	Period 06	Period 07	Period 08	Period 09	Period 10	Period 11	Period 12	Period 13	Period 14	Period 15	Period 16	Period 17	Period 18
Conversion Preparation	Staffing & Planning	Data Conversion Approach																
		Data Quality Specifications, Data Field Mapping Specifications																
			Data Cleanup															
			Create Conversion Programs (extract/transform/load)															
				Data Conversion Signoff Criteria														
				Plan for Hardware Readiness														
				Build Conversion Infrastructure														
				Interface Modification & Testing														
		Plan for Training																
Conversion Tasks									Practice Conversion 1		Practice Conversion 2		Practice Cutover					
				Initiate training					Training with Company Data									
									Refine extracts & Conversion Practice									
									Test and Polish up Interface Functionality									
														Actual Cut-Over & Go Live				
															Stabilize and Operate			

Three Practice Cycles: The key to Minimize Risks
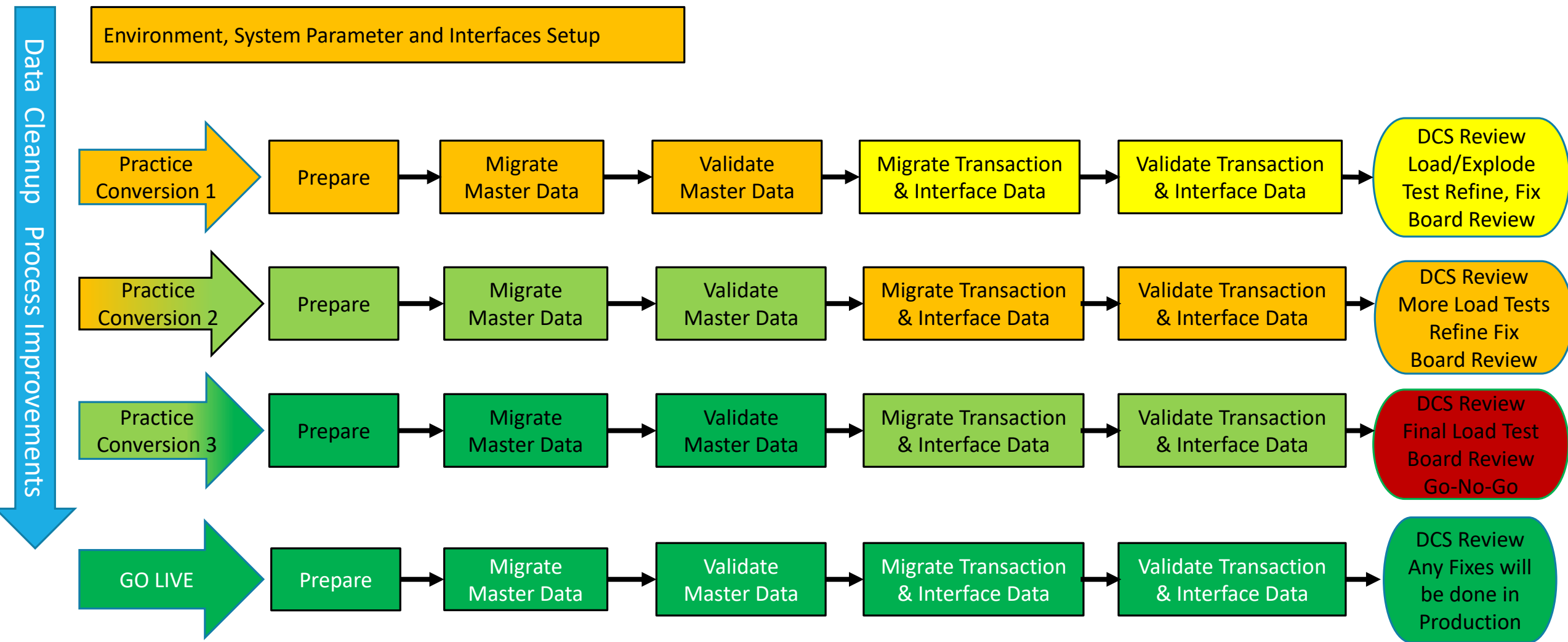
Goal: To ensure NO SURPRISES WITH THE LIVE CONVERSION

For each conversion practice, record:
- Time Durations and Data Volumes
- Verification Checklist Results against DCS
- Scheduling Notes – correct any mistake regarding the order of each data load
- Load tests on the target system –> Did the target system work as expected?
- Document any surprises and/or learnings to take into account in the next cycle

With each cycle: Fix Mistakes, Document learnings, Sharpen Signoff Documents
- Data Conversion Signoff (DCS)
- Data Field Mapping (DFM) and Interface Register
- Data Quality Specifications (DQS)
- ETL and Scheduling
- Interface Application

Three Practice Cycles: Discover/Resolve Issues Early

Environment, System Parameter and Interfaces Setup

Data Cleanup Process Improvements

Practice Conversion 1
Prepare → Migrate Master Data → Validate Master Data → Migrate Transaction & Interface Data → Validate Transaction & Interface Data → DCS Review Load/Explode Test Refine, Fix Board Review

Practice Conversion 2
Prepare → Migrate Master Data → Validate Master Data → Migrate Transaction & Interface Data → Validate Transaction & Interface Data → DCS Review More Load Tests Refine Fix Board Review

Practice Conversion 3
Prepare → Migrate Master Data → Validate Master Data → Migrate Transaction & Interface Data → Validate Transaction & Interface Data → DCS Review Final Load Test Board Review Go-No-Go

GO LIVE
Prepare → Migrate Master Data → Validate Master Data → Migrate Transaction & Interface Data → Validate Transaction & Interface Data → DCS Review Any Fixes will be done in Production

"The WTC Walk" Cycle 1: Two Trees, a Bow and a Fishing Line

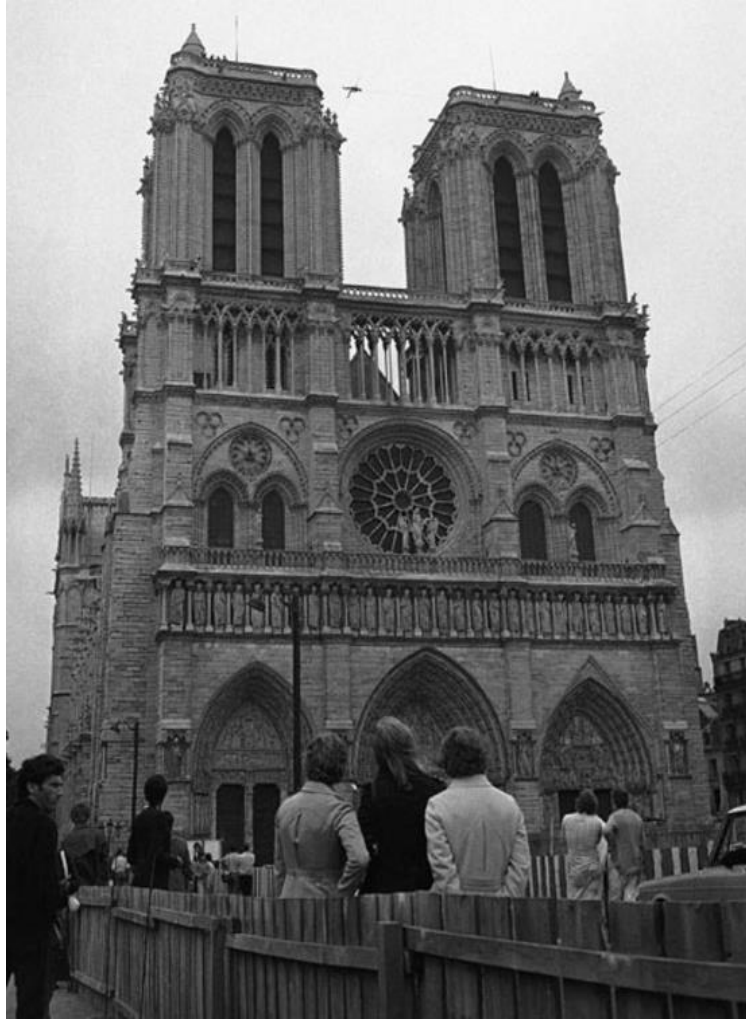First Practice Cycle: Simulate the distance between Twin Towers: 140 feet
- Draw a string between two Trees, 140 feet apart with a bow, arrow and fishing line
- Walk successfully on a wire for 140 feet

- Practice arrow shot to get the wire from one tower to the other
- Must get the wire walk right!

Technical Archery

"The Walk" Cycle 2: 1971 Notre Dame Cathedral

"The WTC Walk" Cycle 3: 1973 Sydney Harbour Bridge
The last practice run, the last chance to iron out any lingering kinks

Going Live!

The point of 3 Practice cycles: Meet the goal of NO SURPRISES

No surprises: Cycle Signoff's

- DCS are ready to check (been run before!)
- Converted data is within tolerance for each data set

No Surprises: Final Conversion

- The data should be clean
- ETL Code should run flawlessly
- Data Scheduling times are now predictable down to a bracket of 5 minutes
- Interfaces migration will be done as per schedule

No Surprises: Open for Business The next Day!

- In effect, no surprises

Recap :

Key to Data Migration Success: Planning, Practice and Consistency in Execution

Plan and Guide the migration effort with strong foundations

- Data Conversion Approach (DCA) – the overall approach
- Constraints and Risks Mitigation Document
- Data Quality Specifications (DQS)
- Data Field Mapping (DFM)
- Interface Register
- Data Conversion Signoff (DCS)

Practice Cycles Enable confident execution

- To Mitigate risks & provide predictable results
- To Reduce uncertainty / Increase Business confidence

Document learnings

- Have a look-back, correct any unforeseen issues immediately
- Close the project, formally release resources

Questions

Email: dambrine@tylogix.com