

Start your engines...

Data migrations

Process, Risk
and Audit
Areas:

Presented by:
Rajesh
Ramachandran,
CISA, CRISC,
CAPM

Agenda

- Data migration definition. Data conversion in Data migration
- Need for Data Migration projects
- Different types of data migration
- Types of Data Migration in my experience in financial institutions
- Roles in DM - PA, CA, BA, PM, Programmer and others
- Processes or steps in Data Migration Projects
- ETL in Data migration. Migration Architecture
- Difference between DM and DW.
- Risks involved in Data Migration.
- Risk Mitigation.
- Testing in Data Migrations
- Different types of Audit controls such as application controls, file controls, field controls.
- Auditing and Data Migration - A discussion

Definitions

- **Data migration** is the process of transferring data between storage types, formats, or computer systems. It is a key consideration for any system implementation, upgrade, or consolidation.
- **Data migration** is usually performed programmatically to achieve an *automated migration*, freeing up human resources from tedious tasks.

Definitions Continued...

- Data migration occurs for a variety of reasons, including server or storage equipment replacements, maintenance or upgrades, application migration, website consolidation and data center relocation.

Definitions continued

- **Data conversion** is required if source and target systems utilize different:
 - ▶ Field Formats and sizes
 - File or database structure(Eg: relational database, flat files, virtual storage
 - access method or VSAM)
 - coding schemes - (ASCII vs EBCEDIC)
 - Hardware and or OS platforms
- The goal is to convert existing data into new required format, coding and structure while preserving the meaning and integrity of data.

Need for Data Migration

- Changing business models
- Ever changing regulations
- Constant technological progress
- Corporate Events

▶ Require that companies change their business applications from time to time. This necessitates as a side effect that data be migrated as well to the new applications.

▶ So Application replacement as form of IT maintenance is contingent on underlying data migration. But data migrations are an everlasting although infrequently performed discipline.

▶ According to surveys the success of data migration projects within budget and time is less than 20%. This calls for a stringent data migration process with well defined quality assurance measures.

Categories of Data Migration

► The major categories are:

- **Storage Migration** - A business may choose to rationalize the physical media to take advantage of more efficient storage technologies. This will result in having to move physical blocks of data from one tape or disk to another, often using virtualization techniques. The data format and content itself will not usually be changed in the process and can normally be achieved with minimal or no impact to the layers above.
- **Database Migration** - It may be necessary to move from one database vendor to another, or to upgrade the version of database software being used. The latter case is less likely to require a physical data migration, but this can happen with major upgrades. In these cases a physical transformation process may be required since the underlying data format can change significantly. This may or may not affect behavior in the applications layer, depending largely on whether the data manipulation language or protocol has changed - but modern applications are written to be agnostic to the database technology so that a change from Sybase, MySQL, DB2 or SQL Server to Oracle should only require a testing cycle to be confident that both functional and non-functional performance has not been adversely affected.

Categories of Data Migration Continued:

- **Application migration** - Changing application vendor - for instance a new CRM or ERP platform - will inevitably involve substantial transformation as almost every application or suite operates on its own specific data model and also interacts with other applications and systems within the enterprise application integration environment. Furthermore, to allow the application to be sold to the widest possible market, commercial off-the-shelf packages are generally configured for each customer using metadata. Application programming interfaces (APIs) may be supplied by vendors to protect the integrity of the data they have to handle.
- **Business Process Migration** - Business processes operate through a combination of human and application systems actions, often orchestrated by business process management tools. When these change they can require the movement of data from one store, database or application to another to reflect the changes to the organization and information about customers, products and operations. Examples of such migration drivers are mergers and acquisitions, business optimization and reorganization to capture new markets or respond to competitive threat.

Categories of Data Migration continued:

- The first two categories of migration are usually routine operational activities that the IT department takes care of without the involvement of the rest of the business. The last two categories directly affect the operational users of processes and applications, are necessarily complex, and delivering them without significant business downtime can be challenging. A highly adaptive approach, concurrent synchronization, a business-oriented audit capability and clear visibility of the migration for stakeholders are likely to be key requirements in such migrations.

Type of Data Migrations in my experience in financial institutions

- **Denovo** - A financial institution or a business organization that is new to banking. Here only metadata and hardware provisions are required. There is no data to migrate. Extensive training on the new processor product and process needed along with possibly training in banking.
- **Conversion** - A bank new to the processor. Data Migration is required. Extensive training on the new processor product and process is also needed.
- **Conversion Merger** - A bank already on the processor acquires another bank. So more data is added to existing set up. The acquiring bank staff are already familiar with target processor.
- **Merger - merger** - Two or more banks already on processor merge as a single bank.
- **Partial data migrations** - Only certain branches or business units move to a different bank.
- **Deconversions(full or partial)** - Existing bank decides to go to a new processor. Data has to be transferred securely , archived and cleaned as per retention standards.

Typical roles in Data Migration

- Programmer or developer – They help with data conversion programs, tool creation and/or report generation or migration or conversion modules.
- Programmer Analyst – They are responsible for the technical solution and implementation of the data migration project. They interact with programmers, conversion analysts, users, project managers and other technical groups. They understand the entire process or parts of processes from a technical view point. They also code programs, reports, test the process and design the conversion architecture.
- Conversion Analyst - They usually have domain knowledge but should be knowledgeable on the product. They are the business analyst for the conversion process and interact with users as well as Programmer analyst and Project manager.
- Technical consultant – Infrastructure sizing and system integrations.
- Program or Project Manager – One who is responsible for the complete delivery of the Migration project – technical, business, infrastructure, user requirements etc. and is the face of senior management
- Other roles the migration team interacts with are Development Business analyst, Application Developers, Data architect, product architect or manager, DBA, change management, account management, configuration management, support and maintenance, command center, information security, compliance, clients etc.

Processes or steps in Data Migration Project

- Tape cut or Data cut – The source data is made available for use by the Extraction process.
- Data Discovery – This step helps in determining the current state of incoming data. In house utilities are used to report on the data.
- Product Definition – In the first step of transform process the source data is mapped to target products. Where no suitable products or values exist the gaps are resolved through discussion with end user or old processor.
- Data Transformation.
- Multiple transform and load runs for test cuts.
- Readiness review cut or Data Load cut – One penultimate tape cut is used to test the operational readiness of the bank on the new processor. This is called readiness review. All operations are simulated from isolate UAT areas with user access. For banks already on processor just a data load is done if needed. This includes sometimes a pilot phase as well.

Processes/steps in Data Migration Project cont.

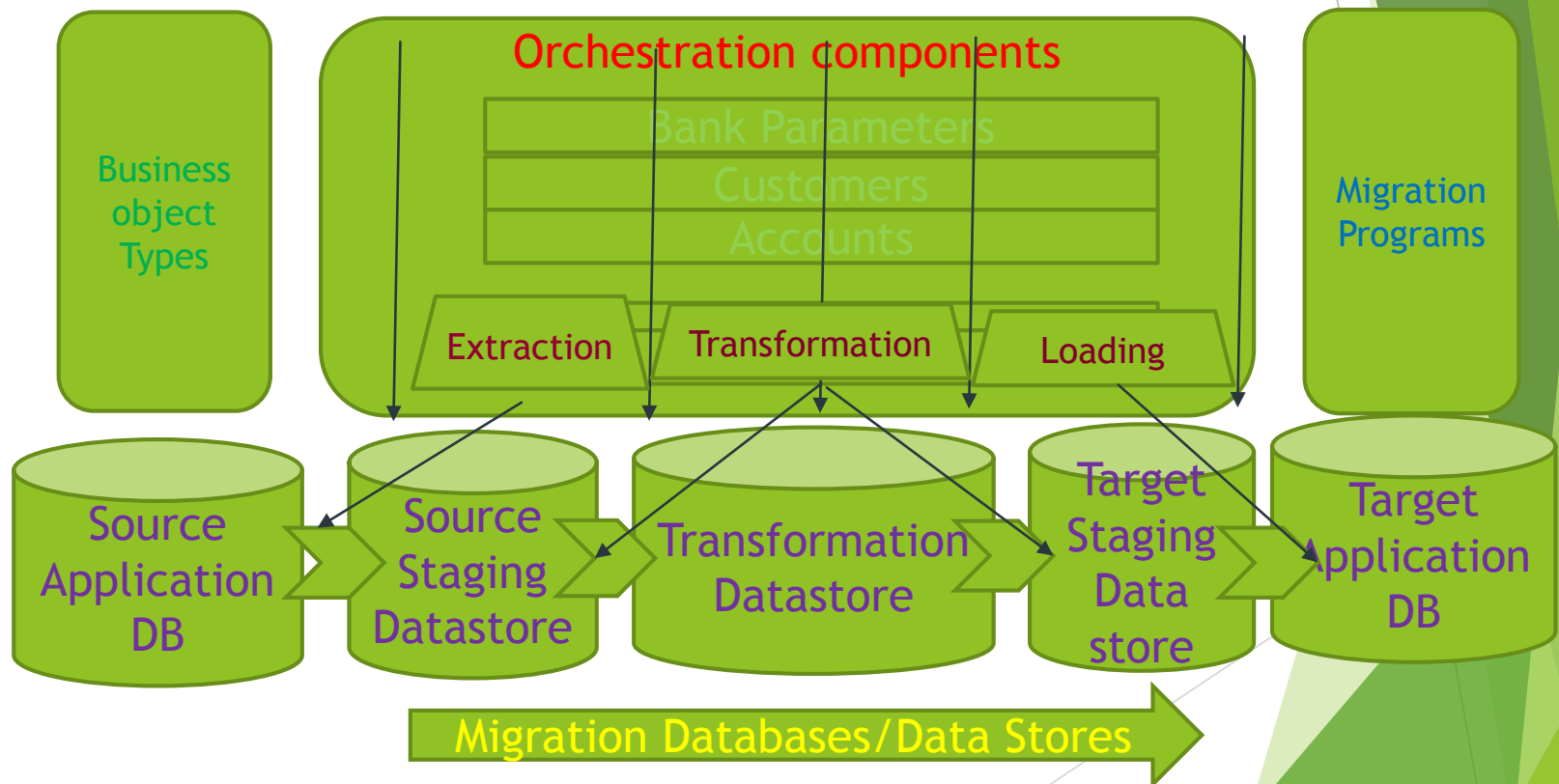
- Customer collateral Distribution.
- Application Interfaces and customer customization development.
- Operational readiness – A parallel process to data migration that occurs is to set up or update the infrastructure (server, storage, network, processes and IT teams at processor and bank branches. This should factor in the growth plans for the bank.
- Training – The bank's personnel are trained on the new product as needed.
- Live cut with Implementation – The implementation is done usually on a weekend and loaded to production system during the published outage window. This could be in form of parallel run cut over, phased cut over or abrupt cut (big bang) over as required by the application(s).
- Post Implementation Support for a pre-determined duration.
- Handover to Support followed by project closure.

ETL in Data Migration

- **Data extraction** is the act or process of retrieving data out of (could be poorly structured) data sources. For example the source could be reports, non RDBMS files like IMS or VSAM or text files. Usually utilities or programs have to be applied to get them ready for further use like analysis and transformation
- **Data Transformation** converts a set of data values from the data format of a source data system into the data format of a destination data system. This can be divided into two steps:
 - Data mapping that maps data elements from the source data system to the destination data system and captures any transformation that must occur.
 - Code Generation that creates the actual transformation program. This includes data cleansing, data quality and data conversion as well.
- **Data Loading** is the process of loading the migrating data into the final target (database, more specifically, operational data store. Database utilities, application insert and update programs are used. Data Delete programs also should be developed as part of this for data back out procedures if needed.
 - For test cut these processes are repeated multiple times till the required data quality is achieved. Usually in these runs the actual End of Day jobs are run in test to ensure process in production would process successfully. The data gets loaded to test databases.

Data Migration Architecture

- Even though data migration projects implement programs, scripts and transformation rules, nevertheless they rest upon a generic architecture. Then there is an orchestration component i.e. order of running the program flow.



Difference between Data Migration and Data Warehouse

- ▶ There is a difference between data migration and data integration activities. Data migration is a project (part of a larger program like core banking migration) by means of which data will be moved or copied from one environment to another, and removed or decommissioned in the source. During the migration (which can take place over months or even years), data can flow in multiple directions, and there may be multiple migrations taking place simultaneously. The Extract, Transform and Load actions will be necessary, although the means of achieving these may not be those traditionally associated with the ETL acronym.
- ▶ Data integration, by contrast, is a permanent part of the IT architecture, and is responsible for the way data flows between the various applications and data stores - and is a process rather than a project activity. Standard ETL technologies designed to supply data from operational systems to data warehouses would fit within the latter category.

Risks in Data Migration projects

- ▶ The risks particularly associated with data migrations are:
 - Disruption of routine operations.
 - Conflicts and contention between legacy and migrated operations.
 - Violation of security and confidentiality of data
 - Data inconsistencies and loss of data integrity during the migration process

- ▶ At a deeper level we can classify them under:
 - Business risks - Profitability, Regulators and reputation
 - IT Management risks - Target Application stability, Data or information loss, Delays, Migration abort, Extended downtime and Project Budget overruns
 - Data Migration Process risks - Subprogram risks such as completeness, semantics, corruption, stability. Program risks such as Execution time and orchestration. Infrastructure risks such as Dimensioning and Interference risks an New Business Application risks such as Parameterization risks.

Risk Mitigation

The usual risk mitigations steps should be:

- Data migration risks should be tackle as part of the core banking implementation program.
- Plan in a detailed manner the data migration including emergency communication among stakeholders for exception situations.
- Use appropriate methodologies including a scalable architecture.
- Use appropriate tools
- Importantly have a fall back scenario i.e. process and programs to reverse the migration or some portions of it or select updates and deleted post migration. Sometimes these programs/processes may need to be developed as part of the project.
- Test base risk mitigation

Testing in data migrations

- ▶ The following testing techniques are usually employed in a Data Migration project and this falls under two categories:
 - Data Validation - This includes;
 - Completeness and Type Correspondence test
 - Semantics
 - Appearance
 - Processability
 - Integrability
 - Migration Run tests;
 - Partial Migration Run test
 - Full Migration Run test
- A future slide shows mapping of these testing techniques to Data Migration Process risks.

Test Based Risk Mitigation

Risk Group	Risk	Risk Mitigation
Subprogram	Stability	Full migration Run Test Partial Migration
	Corruption	Appearance Test Processability Test Integration Test
	Semantics	Appearance Test Processability Test Integration Test
	Completeness	Run Completeness and Type Correspondence Test

Test Based Risk Mitigation

Risk Group	Risk	Rik Mitigation
Programming Risk	Execution Risk	Full Migration Run Tests
	Orchestration Risks	Full Migration Run Tests Partial Migration Run Tests
	Semantics	Full Migration Run Tests Partial Migration Run Tests
Infrastructure	Dimensioning	Full Migration Run Tests Partial Migration Run Tests
	Interference	Operation Risk. No Testing
Target Application	Parameterization	Appearance Test Processability Test Integration Test Completeness and Type Correspondence Test

Audit Controls

- ▶ The controls involve a combination among:
 - Input/origination controls
 - Batch controls and balancing - Total monetary amount, Total items, Total documents, hash totals
 - Error reporting and handling - Handling transactions
 - Data validation
 - Processing controls
 - Data file controls
 - Output controls
 - Application controls.
 - Field controls.

File Audit checks or controls

- Before and after image reporting
- Maintenance error reporting and handling
- Source documentation retention
- Internal and External Labelling
- Version usage
- Data file security
- One-for-one checking
- Transaction file checking
- File updating and maintenance authorization
- Data transfer accuracy
- ▶ Data files or database tables fall in the categories as follows:
- ▶ System control parameters, standing data, Master data/Balance Data, Transaction files.

Field Audit Checks and/or controls

- Sequence check
- Limit check
- Range check
- Validity check
- Reasonableness check
- Table look ups
- Existence check
- Key verification
- Completeness Check
- Duplicate check
- Logical relationship check

Application controls

- Document review - SDLC , Functional and Technical design, Program changes, User manuals, Technical reference
- Flow of transaction through the system
- User performance procedures - Separation of duties, Balancing, Error control and correction, Distribution of reports etc.
- Data integrity testing - Relational and referential integrity
- ACID test for online applications
- Testing application - Depending on the complexity of data and/or applications one or a combination of these among others could be used: They are snapshots, Mapping, Tracing and tagging, test data deck, parallel operation, Integrate test facility, Parallel simulation, Transaction selection programs, Embedded audit data collection etc.

Audit and Data Migration- A Discussion

How do we apply the general audit controls to a data migration project or process.

The auditor should pick and choose from the general controls to the checks available in the Data Migration process itself apart from General Project Management Controls.

Data Migration naturally lends itself to self audit at each stage intrinsic to the process. The results are obvious when they go wrong.

Questions?

Thank you!

Rajesh Ramachandran

Mobile – 9320391219

Email – rajesh9172@gmail.com

Acknowledgements:

Wikipedia

1.CISA and CRISC review manuals

*2.Article on Risks in Core Banking Migration by
K.K..Mookhey in Mumbai Chapter e-journal 2014-15
Issue 2*

*3.Testing and Quality Assurance in Data Migration
projects by Florian Matthes, Christophe Schulz and
Klaus Haller - Presentation at the 27th IEEE
International Conference on Software Maintenance*