iDSS ENABLED AGILE DATA MIGRATION

Abstract

For enterprises, data is a strategic asset and data migration a critical activity. However, the challenges in maintaining data quality, consistency and privacy make it difficult for enterprises to pursue profitable data management strategies. This paper is a take on the criticality of data migration and how Infosys Data Services Suite reduces the time and risk associated with the same.

Infosys®
Navigate your next

Table of Contents

Business Case for iDSS

Modern enterprises are treating their data as a strategic asset. A well-executed data strategy can help identify opportunities to reduce cost, gain deeper insights, mitigate risk and serve stakeholders better.

As enterprises grow, there is a need for data migration, in other words, to move data across applications and systems. This is primarily due to organic growth, mergers & acquisitions, off-shoring, partnerships & alliances, and expansion into new markets.

Enterprises also face challenges in keeping data consistent across heterogeneous systems with different data formats. These systems require complex data mapping rules that need to be updated to keep up with data variety. Modern enterprises also need to adhere to global regulatory norms. This mix of constraints and risks makes it difficult for enterprises to pursue profitable data management strategies.

Traditionally, data migration life cycle

consists of sequential phases associated with one or more processes. It has the following standard activities:

- Source data analysis
- Data extraction
- Data cleansing & enrichment
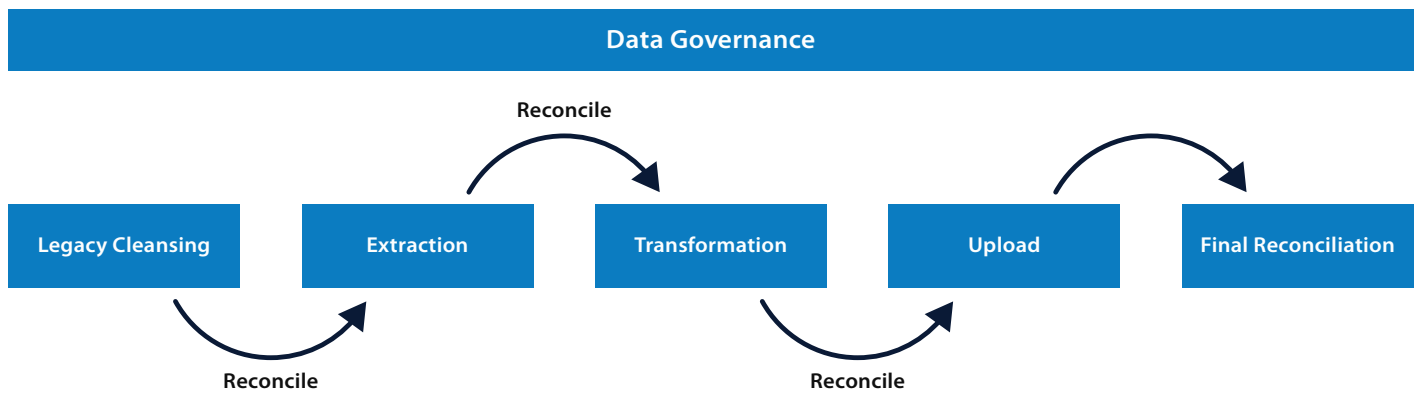- Transformation
- Load
- Reconciliation

Data Governance

Figure 1: Typical Data Migration Flow

Infosys Data Services Suite (iDSS) helps enterprises with this transformation and significantly reduces the time and risk of moving legacy databases. iDSS is

the Infosys proprietary end-to-end Data Management solution that addresses all data migration needs for structured and unstructured data. iDSS addresses

data quality assessment, data quality enrichment, data extraction, business transformation and data loading needs, in a cost effective and efficient manner.

Business Case for Agile

A detailed description of the agile methodology is out of scope for this document. However, we will briefly discuss the key concepts and illustrate the relevance and approach for agile in typical data migration projects.

Agile software method adheres to the following Agile principles:

• Frequent interactions with SMEs, individuals and shared consensus over processes and tools

• Working software (**Minimum Viable Product**) with minimal and relevant documentation

• Quick response to changing and evolving requirements

Agile is an iterative incremental framework and emphasizes on close working relationships between the business and the project team. Each iteration or sprints may be of 4-6 weeks duration and delivers an evolving requirement. Frequent Refactoring helps in refining the initial deliverable as bigger deliverables are built on top of each user story.

A team through multiple sprints completes each iteration – with each sprint covering full migration lifecycle that includes planning, requirements analysis, design, transformation & data loading, unit testing, and acceptance testing. These multiple sprints are necessary to release a product, to add a feature, or to complete an entire project.

The project plan is created at the following 3 levels:

Project Level: Quantifies the entire project size using Quick Function Point analysis or use case estimation technique. Skill, resourcing effort, complexity and risks are then considered to arrive at the duration and size.

Release Level: Breaks down the project into multiple user stories, and subsequently into prioritized user stories. These form the Product Burndown, i.e. one or multiple user stories are delivered through multiple sprints or building blocks. The prioritized user story identifies the critical and important ones that need to be selected upfront. Since the Product Burndown consists of multiple sprints, Sprint Burndown is introduced to indicate the rate at which each of these sprints are completed. Sprint burndown consists of the consolidated user stories and the subsequent 'velocity', rate at which user stories are delivered per sprint, as they are progressively completed. Technically, the burndown indicates the remaining user stories against the time left for the sprint (out of approx. 4-6 weeks per sprint duration).

Sprint Level: Each sprint consists of one or multiple user stories – with each user story accomplishing one entity or application or data type migration life cycle. The migration activities can be for a subset or the whole of master data, reference data or transactional data, depending on the horizontal and vertical partitioning strategy being adopted.

Each user story is further sub-divided into multiple individual tasks for the lifecycle of the migration activity such as mapping table, extraction, transformation and loading to target table with actual effort (in hours) for each. Ideally, each user story is completed, approximately, in a week. Sprint 0 is usually the discovery sprint, and subsequent Sprint 1 to Sprint 'n' are where the application level migration activities are accomplished. Estimation of user stories is done with Story Point estimation techniques like T-shirt sizing or Fibonacci series user story estimation.

Requirement Prioritization and Elaboration

An important leading question is on User story writing, and how it is done. The user stories need to be Detailed, Estimable, Emergent and Prioritized (DEEP) at all times. User story elaboration should follow the **INVEST** principle as a rough guide (acronym details as below):

I – Independent

N – Negotiable (to arrive at exact specifications)

V – Valuable to end users

E – Estimable in terms of effort both at individual and rounded up

S – Small enough to be completed in a week or 2. Otherwise, they need to be sub-divided into smaller logical user stories

T – Testable i.e. should have a DONE or ACCEPTANCE criteria

DONE criteria is an acceptance test parameter of a User story that is easily verifiable and can be independently validated. E.g. For migration of reference data from Siebel to CRM application as a User story, below is the DONE or

Acceptance criteria:

- Data Model of target database is created
- Data load to target application
- Unit testing, QA and SIT testing completed

Prioritization of the consolidated user stories is the logical next step. The Prioritization is done using **MoSCoW** approach (Must Have, Should Have, Could Have and Maybe Have).

Key Agile Migration Stages for Delivery Planning

In this section we will take a look at the key stages in a CRM System Migration.

Sprint 0 is the discovery phase, where requirement workshops are conducted for gathering, understanding and prioritizing requirements, mainly during the initiation phase User Stories on which stakeholders have enough clarity.

The actual architecture to be followed for the migration activity is refined in this phase.
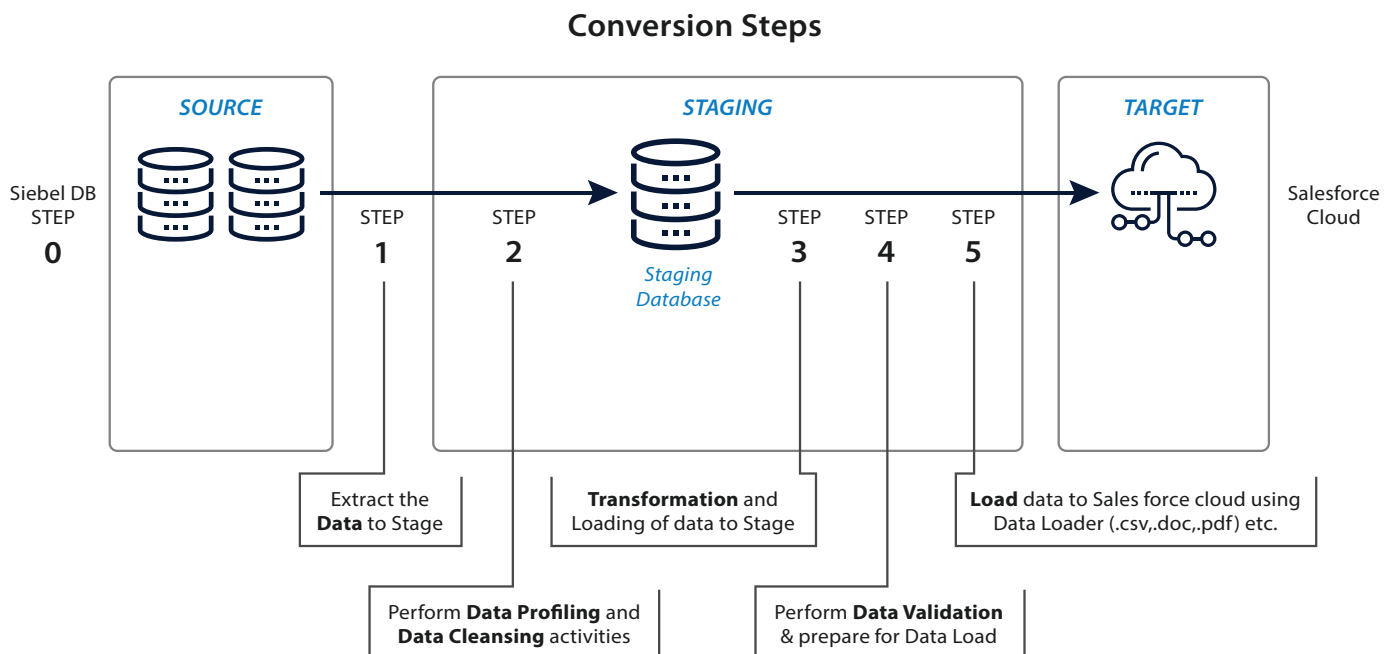
Conversion Steps

Figure 2: *Sample Data Migration Architecture and Approach*

Sprint 0 identifies the key entities as follows and identifies the relevant business rules for creating the High Level Design (HLD):
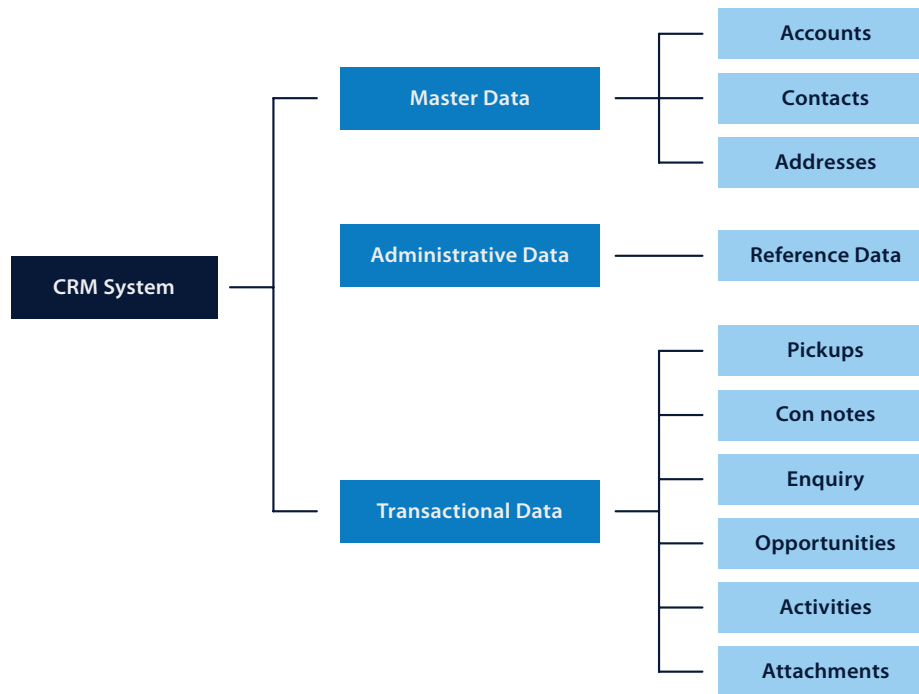

Figure 3: Key Entities in Source System

Sprint 1: This will deliver the source to target migration of chosen Reference Data entities (User Story 1...n) in the normal lifecycle sequence i.e. Requirements, Design, Build, Test and Acceptance.

Sprint 2 onwards: Subsequent sprints will deliver the same source to target migration iteratively for each chosen set of entities via User Story n+1...z.

Care needs to be taken to identify the independent and dependent entities linked by Primary –Foreign key relationship.

This is because the independent entities need to be taken up in earlier Sprints, and only then should the dependent entities be clubbed in subsequent Sprints. Sample is illustrated below:

Reference Data		
Sequence#	**Entity**	**Comments**
1	LOV	No dependency(deployed part of development of code)
2	RVM(Responsibility View Matrix)	No dependency
3	Organization	No dependency
4	Division	No dependency
5	Position	Division is to be created prior to Position
6	Employee	Position is to be created prior to Employee
7	Products	Only products targeted for September release(Dangerous goods)
8	Zipcode	Provided as part of seed data(Excel upload in SF)
9	State Model	Provided as part of seed data(Excel upload in SF)
10	PDQ	Can be discussed if this needs to be migrated
11	Holiday Calender	This needs to be created as is, but planning as a configuration item

Figure 4: Entities and Dependency Identification

Design and Approach for iDSS Level Agile Data Migration

This section outlines iDSS approach for data migration, key technical constituents, design and deployment approaches along with installation and transition to operation. It assumes only data migration in scope – and not database and schema migration, which is a separate discussion
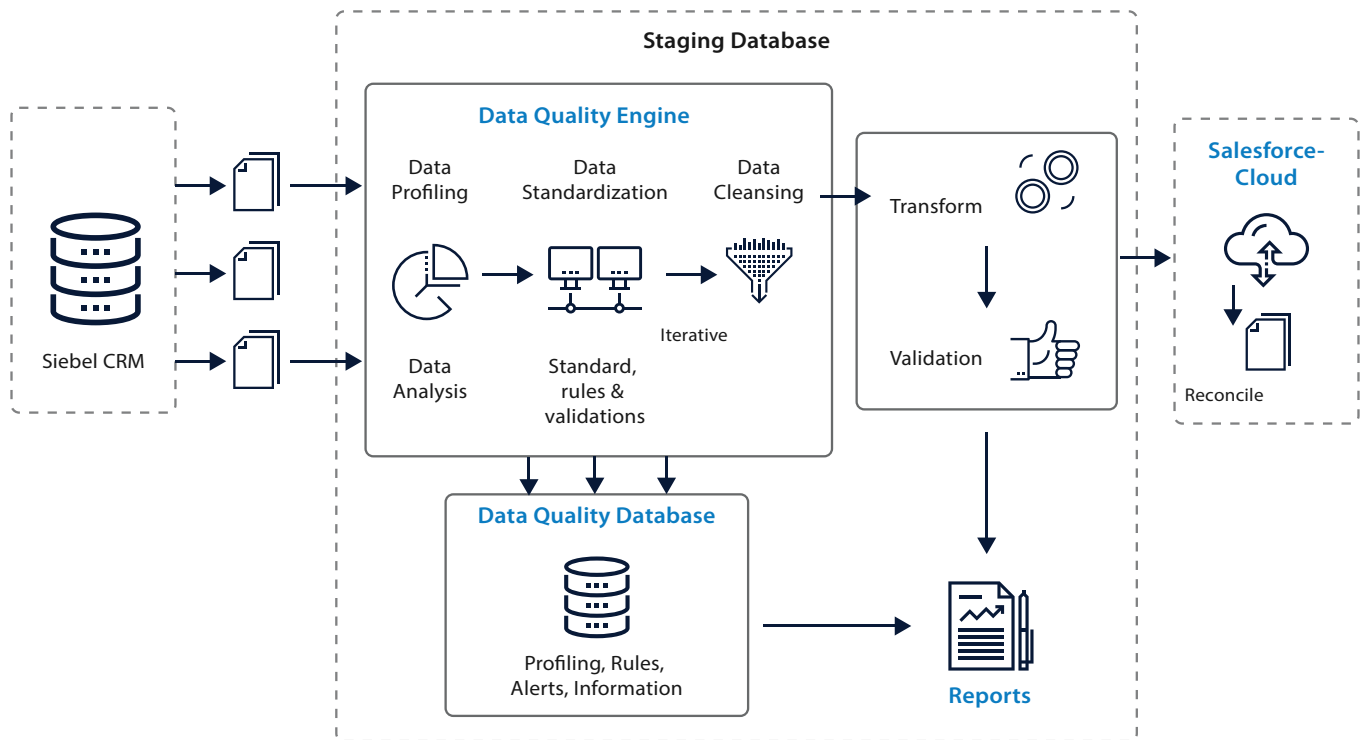

Figure 5: *iDSS Technical Architecture Schematic*

For each entity, end to end lifecycle of migration user story will consist of the following stages:

Dev Environment:

- Sample data extraction through iDSS tool and mapping exercise
- Generate output format files and upload into target through data loader

Aim: End-to-end data flow.

iDSS Advantages for Migration:

- **Metadata Extractor**: Automatically extracts system metadata definition as table and column definitions, data types, constraints, indexes & dependencies from source databases
- **Data Extractor:** Auto-extracts data using tables, views and queries on pre-defined templates and filtered based on business requirements
- **Mapping Builder:** Source (staged) data is transformed and mapped based on target table structures

iDSS Advantages for Data Quality:

- Data profiling and validation against iterative re-usable Business rule sets
- Automated and master data management using rules driven cleansing
- Automated data quality processing and regression testing for duplicate analysis and grouping strategy to enable Match and Merge

SIT environment:

- Loading 25 to 30% of production data and Functional Testing
- Identifying all entities that needs migration and apply basic rules to filter the data set
- Assessing data quality by validating the completeness, relevance and reliability

iDSS Advantages:

- **Task Developer:** Helps create different ETL tasks e.g. Data loading, notifications, file movement etc. and provisions them into workflow
- **Job Scheduler:** Sequence tasks based on dependency in workflow to enable monitoring over simultaneous migrations
- **Data Loader:** Loading the transformed data from source(s) to target databases using native import utilities and deployment scripts
- **Reconciliation Reporting:** Field level reconciliation reports for data load statistics and analyze / validate migration quality and completeness. This helps in corrective and rectification actions leading to faster sign – off.

Load/ Mock environment:

- Loading production data and analyzing performance of the load

- Modifying iDSS scripts if any performance improvements are required

- Performing validations and sharing deviation reports to business for further action

iDSS Advantages:

- **Automated reporting** in multiple file customizable and canned report formats along with push notifications to business stakeholders. This helps in corrective and rectification actions leading to faster sign – off.

- **Mapping** and **Conversion** followed by automated **load ready files** creation

- **Load ready files** can be accessed and reviewed in the integrated platform by multiple stakeholders simultaneously with credentials and access permissions customized for specific roles. Actions like sign – off and review followed by delegate within iDSS platform enables continuous improvement

UAT or Validation testing and verification:

- End to End Functional Testing

- Incremental data load testing

- Pre-go live activities

iDSS Advantages:

- **Enhanced Governance:** iDSS platform enables traceability and transparency with maintained history and audit logs for continuous governance

- **Pre-fabricated Migration Health Check Reports/Data Reconciliation Reports** helps to minimize manual testing

Production Load:

iDSS Advantages:

- **Fast load utilities** using database native loaders enable high migration performance during cut-over

- **Data visualization, correction and reprocessing capabilities** for quick turn-around-time to process rejected data

- **Rollback** and **Re-start** features enable quick course correction during cutover

- **Automated reporting** in multiple file customizable and canned report formats along with push notifications to business stakeholders helps in faster turn around and sign offs

Business Value Offered

With iDSS enabled Agile data migration framework, Infosys has delivered substantial benefits to multiple clients over the past 10 years.

Process related improvements:

- Eliminates various intermediate steps and establishes a seamless platform for managing all master data entities

- Makes UI usage flexible for searching, reviewing, and modifying the master data as required

- Provides one single view of the complete hierarchy for client data

- Provides dashboard for the key data profiling statistics to the client

stakeholders for better monitoring

- Enables easy segregation of work by enabling a central repository server allowing team members to divide multiple tasks among them

- Makes all the previous developments such as rules and transformations reusable

- Comprehensive reporting and visibility enables significant improvements on management reporting with the help of a dashboard which provides a global view of data management issues

Key tangible benefits:

- Manual effort reduction by 75%

- Increased productivity by 20% and reduced cost of quality

- Order completion cycle time reduced by 75%

Data related improvements:

- **Uniqueness:** More than 98% unique entities discovered with the help of de-duplication analysis

- **Accuracy:** More than 98% cleansed and correct entities post business rules' usage

- **Consistency/completeness:** More than 99% of source data present in output data set (excluding any in process profiling activities)

Conclusion

Agile design considerations mandate a faster turn around and fail fast approach. The earlier the failures can be highlighted; the faster rectifications can be planned. In data migration, the challenges increase manifold due to huge and critical data

volumes and history that is increasingly being treasured as a gold mine for artificial intelligence and machine learning analytics.

iDSS helps ensure data migration integrity as well as support incremental migration

for regular update and synchronizing of source and target databases. The automated approach taken by iDSS enables a zero-error migration strategy with minimal manual intervention and project management overheads.

About the Author

Tushar Subhra Das is a Senior Business Data Analyst with over 10 years of experience in Data Migration and Governance. He has worked with Australia based Insurance and Logistics clients for application migration, MDM and Data Quality and process governance. In his current role, Tushar is responsible for APAC deployments and enhancements including product developments for iDSS as the next generation industry standard data management platform.

Infosys®

Navigate your next

For more information, contact askus@infosys.com

Infosys.com | NYSE: INFY

Stay Connected   SlideShare