

WHITE PAPER



How to Ensure Data Quality During Data Migration



Naju D Mohan, *Delivery Manager*

Infosys
be more

Introduction

In today's business world, change is the only constant and changes make their appearances in various forms, some of which are:

- **Mergers and acquisitions**
- **New compliance requirements**
- **New Package implementations**
- **Migration to new technologies such as the cloud**
- **Big data programs**

Being data driven, the business has to upgrade and keep its intelligence up-to-date to realize the benefits of these changes. So in short, all these changes result in data migrations.

Who is the Primary Owner?

Most of the time, it is assumed that data migration is an IT problem 2. All the visible changes and the actions lie with the IT team, so the business moves on, putting the entire burden of data migration management on the IT team.

Mergers and acquisitions and compliance requirements clearly stand out as having its origin with the business team. So does the decision to implement a CRM, loyalty or HR package with its beginning at the business department. The need to optimize operating costs and make intelligent decisions and act in real-time, leads the business to migrate to cloud and embark on big data programs. But the onus of the migration management, often, lies with the IT team.

It must be clearly understood that any data migration without the business leading the program has a high rate of failure. Business has to not just care about data migration but command it.



Why Such a High Failure Rate for Data Migration Programs?

According to Gartner, 83% of the data migration programs fail to meet expectations, running over time and budget¹.

Some key reasons for this are:

1. Poor Understanding About Data Migration Complexity

- The focus on data migration is lost in the excitement of the new package implementation, migration to cloud or big data initiatives
- Most often, it is assumed that data fits one-one into the new system
- The whole attention is on the implementation of the new business processes with less or almost no focus on data migration

2. Lack of Proper Attention to Data

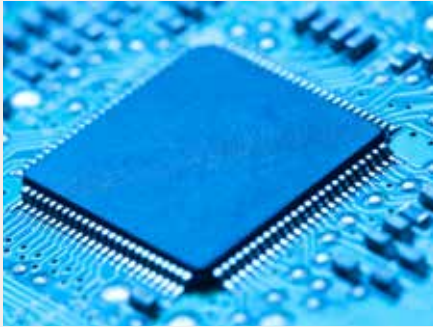
- Lack of data governance and proper tools for data migration can impact the quality of data loaded into the new system
- Mergers and acquisitions can introduce new data sources and diverse data formats
- Huge volumes of data may force us to overlook whether the data is still relevant for the business

3. Late Identification of Risks

- Poor data quality of the source systems and lack of documentation or inaccurate data models would be identified late in the migration cycle
- Lack of clarity on the job flows and data integrity relationship across source systems would cause data load failures

Is There a Right Validation Strategy for Data Migration?

An innovative data migration test strategy is critical to the success of the change initiatives undertaken by the business. The test strategy should be prepared in close collaboration with the business team as they are a vital stakeholder, who initiated the change resulting in data migration. The two principal components which should be considered as part of the test strategy are:



1. Risk-Based Testing

The data volumes involved in data migration projects emphasize the need for risk-based testing to provide optimum test coverage with the least risk of failure. Master test strategy can be created by ensuring proactive analysis with business and third-parties. Tables can be prioritized and bucketed based on the business criticality and sensitivity of data. Composite key agreed with the business can be used to select sample rows for validation in tables with billions of rows.

2. Data Compliance Testing

It is very important that the quality assurance (QA) team is aware of the business requirements that necessitated data migration, because the change would have been to meet new government regulations or compliance requirements. The test strategy must have a separate section to validate the data for meeting all compliance regulations and standards such as Basel II, Sarbanes-Oxley (SOX), etc.

A Test Approach Giving Proper Attention to Data

Data migration, as mentioned earlier, is often a by-product of a major initiative undertaken by the company. So in a majority of scenarios, there would be an existing application which was performing the same functionality. It is suitable to adopt a parallel testing approach which would save effort spent to understand the system functionality. The testing can be done in parallel with development in sprints, following an agile approach to avoid the risk of failure at the last moment.

1. Metadata Validation

Data migration testing considers information that describes the location of each source such as the database name, filename, table name, field or column name, and the characteristics of each column, such as its length and type, etc. as part of metadata. Metadata validation must be done before the

actual data content is validated, which helps in early identification of defects which could be repeated across several rows of data.

2. Data Reconciliation

Use automated data comparison techniques and tools for column to column data comparison. There could be duplicate data in legacy systems and it has to be validated that this is merged and exists as a single entity in the migrated system. Sometimes the destination data stores do not support the data types from the source and hence the storage of data in such columns have to be validated for truncation and precision. There could be new fields in the destination data store and it has to be validated that these fields are filled with values as per the business rule for the entity.

Benefits

A well thought-out data migration validation strategy helps to make the data migration highly predictable and paves the way for a first-time right release. Regular business involvement helps to maintain the testing focus on critical business requirements. A successful implementation of the shift-left approach in the migration test strategy helps identify defects early and save cost.



Case Study: Re-Platforming of Existing HP NEOVIEW Data Warehouse to Teradata

The Client

One of the largest super market chains in the United Kingdom which offers online shopping, DVD rentals, financial services, and multiple store locations.

The Objectives

- To complete the re-platform of HP Neoview to Teradata and re-platform associated services before HP discontinued support to Neoview
- To migrate the existing IT business services currently operating against a Neoview data warehouse onto a Teradata warehouse with minimal disruption
- To improve the performance of current Ab-Initio ETL batch processes and reporting services using Microstrategy, SAS, Pentaho, and Touchpoint

The QA Solution

The validation strategy was devised to ensure that the project delivered a like-for-like, 'lift-and-shift'. This project had environmental challenges and dependencies throughout the entire execution cycle. SIT phase overcame all the challenges by devising strategies that departed from traditional testing approach in terms of flexibility and agility. The testing team maintained close collaboration with the development and infrastructure teams while maintaining their independent reporting structure. The approach was to maximize defect capture within the constraints placed on test execution.

It was planned to have individual tracks tested independently on static

environment and then have an end-to-end SIT, where all the applications / tracks are integrated. Testing was always focused on migrating key business functions on priority such as sales transaction management, merchandise and range planning, demand management, inventory management, price and promotion management, etc.

The Benefits

- 15% reduction in effort through automation using in-house tools
- 100% satisfaction in test output through flexibility and transparency in every testing activity achieved through statistical models to define acceptance baseline

End Notes

1. Gartner, "Risks and Challenges in Data Migrations and Conversions," February 2009
2. <https://www.hds.com/go/cost-efficiency/pdf/white-paper-reducing-costs-and-risks-for-data-migrations.pdf>

For more information, contact askus@infosys.com



© 2017 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.