

INFOSYS DATA WIZARD SOLUTION



Business imperatives

As a business, there are multiple reasons to migrate your enterprises data warehouses and data lakes. Migration enables you to optimize your data storage and compute costs, retire and modernize legacy appliances to cloud or leap forward your business and digital transformation journey

with next generation data platforms. Unlike transactional or operational data stores, migration of your enterprise data warehouses, and data lakes are fraught with many execution challenges. Irrespective of your reasons, it involves operational risks that require careful planning and execution. With the sheer number of data pipelines, ETL jobs,

configurations, schedules that feed data to them; and the array of meta stores, integrations, catalogs that enable their effective use, and the many downstream applications that depend on the data warehouses and data lakes, the migration exercise could quickly spiral out of control if not planned, managed and executed efficiently.

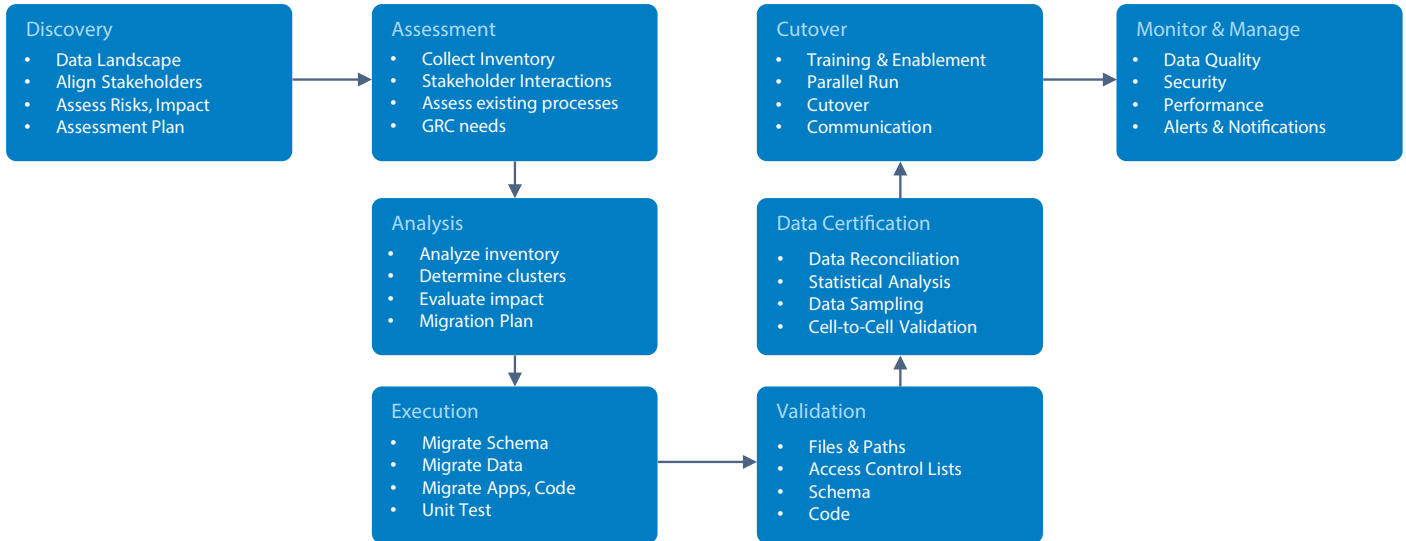


Figure 1. Data Migration Process

Migration, Made Easy with Infosys Data Wizard

Infosys Data Wizard can make data warehouse and data lake migrations seamless, manageable, and quick. It offers an intuitive graphical user interface that is designed to follow your execution approach through different phases of your migration – from project initiation to configuration, inventory collection, analysis, migration, validation, and all the way through to data certification, tracking, and management.

	INVENTORY COLLECTION	MIGRATION PLANNING	MIGRATION PROCESS	OPERATIONAL RISK	DATA INTEGRITY
CHALLENGES	Inventory collection is complex and require multiple scripts to be run to gather information at different levels of detail.	Data lake and Warehouses have multiple feeds and sinks that span across business lines and ought to be carefully planned and scheduled.	Complex migration workflows with better planning, scheduling, tracking, notification and co-ordination are the need of the hour.	Reduce operational risks associated with live migration of data when the systems are in production and in active use. Cutover to target system should be smooth.	Validation of migration is often error prone. Data assets must maintain their integrity during and after migration to target environment.
SOLUTIONS	Suite of scripts and automations to discover and gather asset dependencies, integrations and associated metadata.	Provision to adopt built-in recommendation engine for wave planning and ability to tailor per user needs and priorities.	Intuitive GUI and guidance-based workflows from project configuration, to execution, to validation and data certification.	Built-In Integration with 3 rd party migration execution engines with capability for both full and catchup data loads.	Robust data certification method based on multi-point validation framework to ensure integrity and quality of data.

Figure 2. Migration Challenges and Solutions enabled by Infosys Data Wizard

Process Guidance, and Automation.

Following the process need not be cumbersome anymore. Infosys Data Wizard can offer guidance and recommendations all the way in your migration, while offering full control to the migration architect to customize the approach, as and when needed.

Enable Best-in-class Partner solutions.

Infosys Data Wizard integrates best-in-class partner solutions to provide a unified wizard for planning and execution of your migrations. It offers minimal to zero downtime and impact to business, reduces migration cycle-time, and brings operational efficiencies.

Provides Choices and Flexibility.

All migrations are not planned and executed in the same way. We understand that different considerations and trade-offs need to be made when you migrate, be it with your choice of execution engines, partner solutions, and more. Infosys Data Wizard offers you the choices to pick what best fits your timelines, needs and budgets.

Data Certification.

Data integrity must be preserved at all costs, during and after migration. Infosys Data Wizard brings forth a suite of validations that range from simple reconciliation and digests, to statistical metrics and cell-to-cell validations. Again, the spectrum of validations to use are configurable and can be tailored to suit one's individual needs.

Case study

For a world leading retail chain and an early adopter of technology, the existing solution on Azure Data Lake Storage Gen 1

and Databricks needed to be migrated to Azure Data Lake Storage Gen 2 with GPU capability.

Business case for migration

Enterprises who were early adopters of Azure Data lake Storage (ADLS Gen 1) and moved their data estate to Azure cloud, had their data storage and compute primarily on Hadoop based eco-system. As the data volumes have grown multi-fold, the cost curves has significantly increased with security, scalability and performance challenges.

The advent of object storage has heralded a new wave of innovation in storage and Microsoft Azure has introduced Gen 2 of Azure Data Lake storage (ADLS Gen2). Azure Data Lake Storage Gen2 combines features from

Azure Data Lake Storage Gen1, such as file system semantics, directory, and file level security and scale with low-cost, tiered storage, high availability/disaster recovery capabilities from Azure Blob storage. It thus provides significant and immediate advantages with respect to Scale, Performance, Security, Manageability and Cost-effectiveness.

For existing customers of ADLS Gen1, no new features will be added to ADLS Gen1. Hence the migration to ADLS Gen2 becomes an absolute necessity for enterprises to optimize their cost curves and service increasing demands.



Benefits

- **Unified Migration Management:** Enables a unified solution for end-to-end management of migration process. Role-based personalized dashboards and reports to validate different stages of migration process.
- **Data Integrity & Certification:** Multi-point data validation framework to certify data integrity and quality.
- **40-50 % improvement in Productivity:** Provides guided automation and acceleration of migration steps.
- **Upto 40% improvement in Time to Market:** Enable significant reduction in window-time resulting from automation in migration and validation.

Business Challenge

Multiple in-flight business projects meant migration had to be performed with zero or minimal downtime. Over 2.5 petabytes of data and 250+ data pipelines, with a mix of batch and real time workloads had to be migrated.

Infosys Solution

- Infosys Data Wizard for end-to-end migration management was implemented
- Adopted a wave-based approach to migrate data
- Automated data validation strategy with Databricks on Microsoft Azure to validate at scale
- Performed code remediation of 250+ data pipelines
- Partnered with WANDisco for live data migration

Key Metrics

- Data volume processed – 2.4 Petabytes
- Data pipelines remediated – 265
- File throughput – 300 files per second, per node
- Data volume throughput – 5.6 Gib per second

Business Benefits

- Zero impact to in-flight business initiatives
- Enabled 30% effort savings in overall program execution

For more information, contact askus@infosys.com



© 2021 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.