MAKE YOUR DATA FUTURE-READY WITH INFOSYS DATA WAREHOUSE OPTIMIZATION
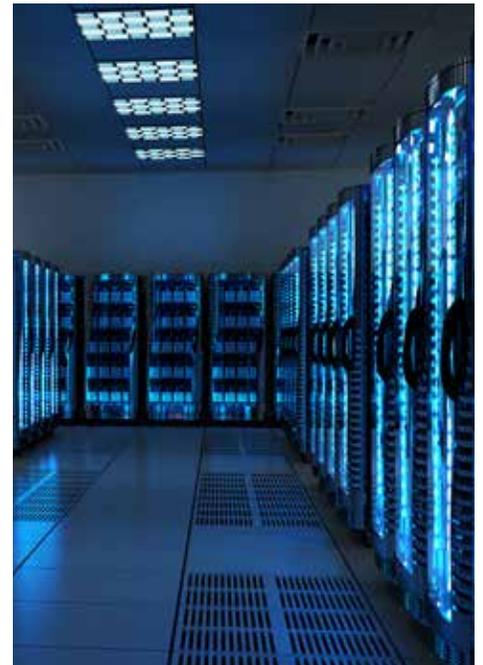
"Data is expected to grow to 40 zettabytes and machine-generated data is projected to increase 15 times by 2020 as quoted by a premier global market intelligence firm."

Advances in automation and technology have been generating huge volumes of data. This will continue to grow exponentially in the future. Big Data comprises nontraditional and digital data, as well as traditional data generated from sources such as customer relationship management (CRM), enterprise resource planning (ERP), transactional systems, etc. To ensure business success, enterprises must harness big data to gain meaningful business insights and make accurate decisions in real time. However,

companies struggle to address some key big data challenges such as:

- How do we store TB / PB of data efficiently?
- How can we process these large data sets for actionable insights?
- How do we visualize and consume insights quickly to stay competitive?

While appliances such as Teradata and Netteza offer scalability and high computational power compared to traditional relational database management systems (RDBMS), the ratio of cost to performance remains high. Today, large enterprises are looking to big data platforms to address these challenges, reduce operational costs, and improve the efficiency of their data warehouses.

Infosys solution and offerings

Infosys has a comprehensive Data Warehouse Optimization Solution that leverages popular architecture and methodologies to meet client-specific needs:

- **ETL / MDM offloading**

    Offloads data processing workloads onto Hadoop to improve performance and reduce data processing cycle times

- **Augmented data warehouses**

    Offloads high volume storage and processing to Hadoop and delivers ready-to-consume results to a traditional data store

- **Big data warehouses**

    These are built on Hadoop and enable data consumption

- **Data archival**

    Data is archived on Hadoop to reduce storage cost and meet the compliances around online data access

Our implementations leverage a combination of the above methods. This has yielded benefits, such as improving processing speed by ten times and enabling 50%–70% cost reduction.

The Infosys solution is supported by end-to-end service offerings that ensure continuous support and a successful migration.

Service Offering	Description
Consulting services	Current state assessment, architecture and roadmap definition, tool evaluation, and business case preparation
Migration services	Migration from traditional data warehouses to Hadoop
Implementation services	Design, develop, and validate the implementation of big data solutions

Infosys assets

Big data assessment framework
Framework for assessment of applications, considering business and technical factors, to derive its fitment for migration from existing DWBI technologies

Reference architecture patterns
Reference architecture patterns for ETL offloading, augmented DW, big data warehouses, and data archival developed based on implementation experience, which help in accelerated implementation of big data platform

Big data migration workbench
Infosys Migration workbench can accelerate your migration from a traditional warehouse to a big data warehouse. It automates table migration, data migration, and data validation

Augmented DW playbook
Compilation of Infosys knowledge and experience gained from various client engagements as well as other knowledge sources such as – vendor alliances, trainings, proof-of-concepts, white papers, and conferences

Data Integration Toolkit
Infosys Data Integration Toolkit has a set of tools and frameworks that accelerate the development cycle involved in building ETL processes on Hadoop. It has tools to handle data ingestion, data processing, data validation, and ETL workflow control and security
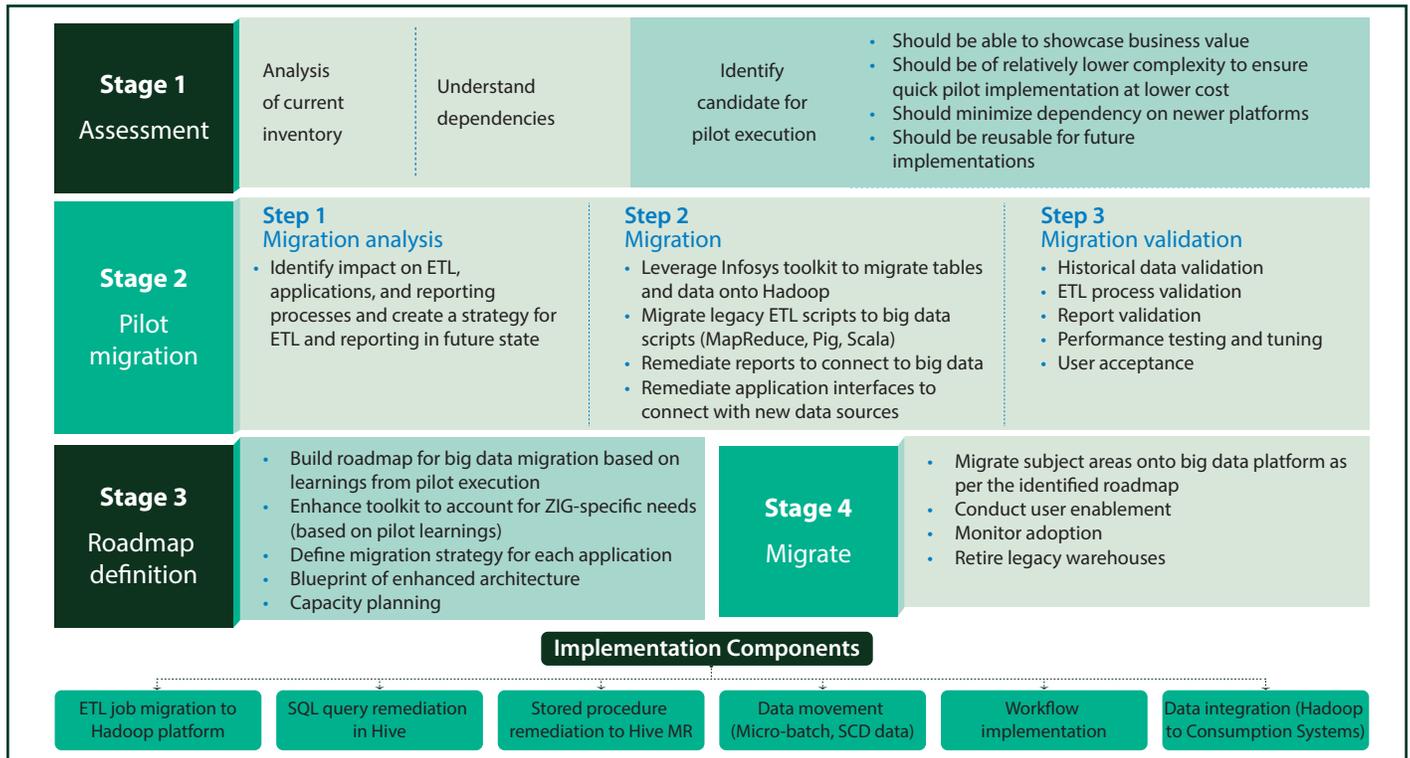
MDM on Hadoop toolkit
Infosys Consumer Identify Solution addresses this challenge by offloading the heavy lifting involved in customer record matching to a big data platform

Infosys approach

Infosys handles migration to the optimized DW platform using a phased approach to minimize risk and negative impact to business processes.

Stage 1 — Assessment

Analysis of current inventory	Understand dependencies	Identify candidate for pilot execution	• Should be able to showcase business value • Should be of relatively lower complexity to ensure quick pilot implementation at lower cost • Should minimize dependency on newer platforms • Should be reusable for future implementations

Stage 2 — Pilot migration

Step 1 — Migration analysis
• Identify impact on ETL, applications, and reporting processes and create a strategy for ETL and reporting in future state

Step 2 — Migration
• Leverage Infosys toolkit to migrate tables and data onto Hadoop
• Migrate legacy ETL scripts to big data scripts (MapReduce, Pig, Scala)
• Remediate reports to connect to big data
• Remediate application interfaces to connect with new data sources

Step 3 — Migration validation
• Historical data validation
• ETL process validation
• Report validation
• Performance testing and tuning
• User acceptance

Stage 3 — Roadmap definition
• Build roadmap for big data migration based on learnings from pilot execution
• Enhance toolkit to account for ZIG-specific needs (based on pilot learnings)
• Define migration strategy for each application
• Blueprint of enhanced architecture
• Capacity planning

Stage 4 — Migrate
• Migrate subject areas onto big data platform as per the identified roadmap
• Conduct user enablement
• Monitor adoption
• Retire legacy warehouses

Implementation Components

ETL job migration to Hadoop platform	SQL query remediation in Hive	Stored procedure remediation to Hive MR	Data movement (Micro-batch, SCD data)	Workflow implementation	Data integration (Hadoop to Consumption Systems)

Success Stories

Infosys DW Optimization solution has delivered significant benefits to a variety of clients:

- **Shortened load time from two hours to ten minutes and reduced storage costs by 83%** for a consumer electronics giant by implementing big data augmented warehouse
- **Reduced storage costs by US$6.6 million and improved performance by ten times** for a leading financial major
- **Processed information residing in 1.3 billion device-related and over 1 billion quotes-related records within four hours** (the expected time as per the service level agreement was within six hours) by leveraging Hadoop for a US-based network giant

Related reading: Infosys Big Data Migration Workbench: **infy.com/migration-workbench**

For more information, contact askus@infosys.com

Infosys®
Navigate your next