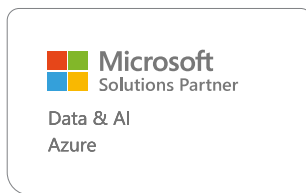


# Intelligent and automated legacy data workload modernization to Azure Databricks

Up to 95% automated migration of legacy data warehouses, ETL, Hadoop, analytics, BI, reporting, and Mainframe workloads



Azure Databricks offers a unified data analytics platform combining the best of data lake architectures and warehouses to deliver high-performance, scalable, and cost-efficient solutions.

Migrating legacy data workloads to Azure Databricks may present some challenges, such as ensuring business continuity and maintaining data integrity. Enterprises can face issues like incomplete documentation and complex workload dependencies. Leveraging robust automation tools and support for such migration can ultimately lead to a more efficient, scalable, and innovative data environment.

LeapLogic, Impetus' automated cloud accelerator, simplifies and de-risks the migration of legacy workloads to Azure Databricks. Achieving up to 95% automation across the four-step process, its intelligent engine converts diverse legacy code into Azure Databricks-native equivalents, ensuring 100% preservation of business logic and optimized performance.

## Key Benefits

Compared to traditional approaches, LeapLogic enables:

- **4x** faster transformation
- **1.5x** faster validation
- **2x** cheaper than manual migration
- **50%** migration acceleration
- **100%** risk compliance
- **100%** SLA adherence

## Watch LeapLogic in action

▶ Teradata to Databricks

▶ Oracle to Databricks

▶ Hadoop to Databricks

▶ Informatica to Databricks

▶ SAS to Databricks

# What's possible?

## For data warehouse

- Plan a phased migration instead of a 'big bang'
- Mitigate risks beforehand with no downtime
- Handle proprietary elements like BTEQs efficiently
- Ensure optimization to meet production SLAs
- Consider all workloads in totality
- Stabilize fast with a minimal parallel run period
- Cutover and retire your legacy data warehouse
- Strategize for people and processes

## For Hadoop

- Assess infrastructure and workload inventory
- Map infrastructure and workloads on Azure Databricks
- Assess TCO and forecast for Azure Databricks
- Detect and optimize patterns
- Navigate complexities and risks
- Migrate efficiently at scale
- Assure quality through technology mapping
- Save cost and time with automation
- Convert workloads to Azure Databricks-native equivalent
- Ensure validation equivalence and acceptance
- Provision through infrastructure as code

## For ETL

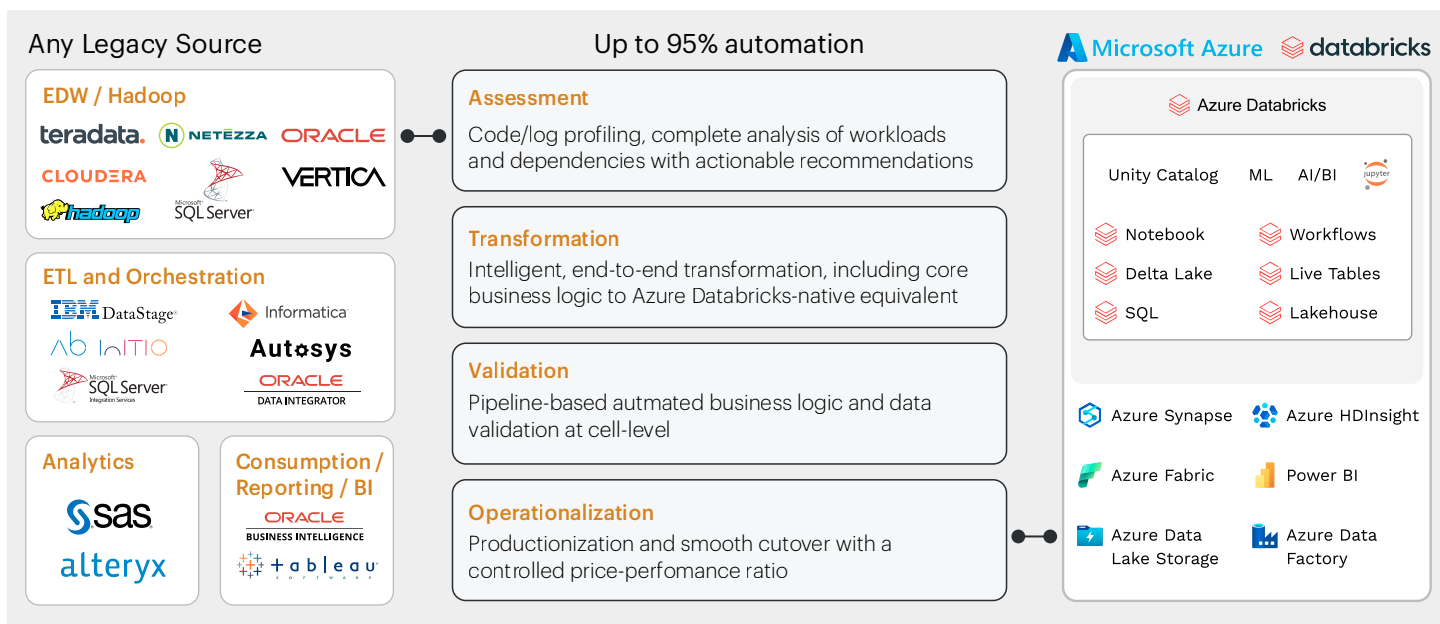
- Assess code complexity, usage patterns, etc.
- Identify and analyze complex interdependencies. For example, for **Informatica**, it can be from XML files to workflows and then to mappings and transformations. Similarly, for **DataStage**, identify and analyze all jobs and components for each script and job activity, sequencer, lookup, aggregator, transformer stage, join, etc.
- Transform core business logic to Azure Databricks-equivalent format
- Package for production-ready jobs
- Validate scripts for Azure Databricks-equivalence of all use cases

## For analytics and reporting workloads

- Assess usage patterns with automated accelerators. For example:
  - SAS ETL – mostly SQL + some SAS procedural
  - SAS procedural – mostly statistical
  - SAS advanced algorithms
- Convert code to Azure Databricks-native stacks
- Map the conversion target for each usage pattern
- Enable datasets and migrate them as cloud stores or access them via JDBC
- Use a staggered approach to convert and validate scripts for Azure Databricks-equivalence across all use cases
- Execute thorough integration testing on staging
- Execute on production

# Automated migration to Azure Databricks with LeapLogic

LeapLogic enables end-to-end modernization of workloads to an Azure Databricks-native stack in four steps:



## Step 1: Assessment

- Integrated assessment of jobs, DML, DDL, ETL, procedures, logs, scheduler scripts, analytics scripts, etc.
- Comprehensive inventory listing
- Workload prioritization as per the business use case
- Dependency analysis
- Advanced blueprint of Azure Databricks stack
- Optimization opportunities for schema, code, and orchestration
- Prescriptive recommendations
- Phased migration plan with timelines and cost estimates

Entity Name	Partition By	Cluster By	Bloom Filter Index	ZOrder Index	Split By	Sort By
ds_tbl_db_ss_text1	-	-	-	-	-	-
ds_tbl_db_web_sales	-	-	WS_SOLD_DATE_SKIWS...	WS_ITEM_SK	-	-
ds_tbl_db_store_sales	-	-	SS_SOLD_DATE_SKISS...	SS_ITEM_SK	-	-
Entity Name	Partition By	Cluster By	Bloom Filter Index	ZOrder Index	Split By	Sort By
Data Source	Entity Type	Volume	Primary Key	Unique Key	Transactional	-
ds_tbl_db	-	2	ss_item_sk,ss_ticket_num...	-	No	-
No. of Buckets	Frequency of Use	Tune Based on Workload	Tune Based on Table Sizes	Compaction (bin-packing)	-	-
-	2424	-	-	-	-	-
ds_tbl_db_inventory	-	-	INV_ITEM_SK	INV_DATE_SK	-	-
ds_tbl_db_s_promotion	-	prom_promotion_id	-	-	-	-
ds_tbl_db_catalog_sales	-	-	CS_SOLD_DATE_SKICS...	CS_ITEM_SK	-	-
ds_tbl_db_date_dim	-	-	D_DATE_SK	D_DATE_SKID_DATE_SK	-	-

Sample of schema optimization recommendations

## Step 2: Transformation

- Intelligent grammar engine supporting a variety of workloads and formats
- End-to-end transformation, packaging, and orchestration to Azure Databricks-native format
- Notebook-based inline editor for query optimizations
- Delivers a verified, executable package with performance SLAs met
- Extensible, repeatable, and verifiable methodology

## Step 4: Operationalization





- Productionization and go-live
- Infrastructure as code
- Execution using cloud-native orchestrators
- Automated DevOps, including CI/CD, etc.
- Target environment stabilization
- Smooth cutover

## Step 3: Validation

- Pipeline-based automated validation
- Sample or customer data-based code validation
- Auto-generation of reconciliation scripts
- Automated SQL/query and business-level validation
- Cell-to-cell validation reports
- Data type and entity-level matching

LeapLogic is built to efficiently assess, migrate, validate, and move complex workloads from legacy systems to Azure Databricks. Based on a sophisticated grammar engine that transforms the most complex functions, this automated conversion helps enterprises streamline their modernization journey with advanced analytics and business growth.

## Explore resources to support your transformation initiatives

-  30% performance improvement by converting Netezza and Informatica to Azure-Databricks stack
-  7 best practices to modernize data architecture on Databricks with LeapLogic
-  Revolutionize your data landscape: Automated Databricks modernization with LeapLogic
-  Maximize migration success with unique validation features of LeapLogic

To automate the transformation of legacy workloads to Azure Databricks, write to us at [info@leaplogic.io](mailto:info@leaplogic.io)

BOOK A DEMO 

**leaplogic** | an **IMPETUS** product

LeapLogic, an Impetus product, automates the end-to-end transformation of legacy data warehouse, ETL, BI/reporting, analytics, Mainframe, and Hadoop workloads to native cloud platforms for optimum cost and performance. Impetus Technologies helps enterprises solve the data, AI, and cloud puzzle by combining unmatched cloud and data engineering expertise. Recognized as an AWS Advanced Consulting Partner, Elite Databricks Consulting Partner, and Elite Snowflake Services Partner, Impetus has been the 'Partner of Choice' for several Fortune 500 enterprises for transforming their digital nuclei and driving unmatched innovation and growth for over a decade.

To learn more, visit [www.leaplogic.io](http://www.leaplogic.io) or [www.impetus.com](http://www.impetus.com). You can also write to us at [inquiry@impetus.com](mailto:inquiry@impetus.com).

© 2024 Impetus Technologies, Inc. All rights reserved. Product and company names mentioned herein may be trademarks of their respective companies. SEPT 2024