The (new) Azure cloud data migration playbook

How to accelerate time to value in cloud migrations

There's a better way to move to cloud

Many large enterprises leverage Azure cloud services for their data processing requirements to take advantage of scalability, agility, and cost optimization. Yet many enterprises still run their business–critical data estate on premises and are afraid to modernize and migrate their on–prem Hadoop environment to Azure as it presents many unknowns, from the cost of running on cloud, supportability of their application version, performance, to more. Contrary to popular belief, data modernization and migration from on–premises Hadoop to the cloud doesn't have to be hard — a shift in thinking is all it takes.

Today, most companies approach cloud migration with a single lens, which is "application first," and where workloads are migrated based on the importance of sequencing of applications. In this modern day and age, though, large enterprises have to be very agile and constantly change how they do business to stay relevant or come up with new products and offerings to capture bigger market share. Given the advancement in technology that allows us to migrate petabytes of data to the cloud quickly and without disruption, customers can benefit from insights, trends, and patterns using AI, machine learning (ML), and cloud elasticity faster than ever before — enabling enterprises to deliver results and value immediately.

This Playbook answers key questions about what's driving cloud migrations, why getting to the cloud fast matters, and why traditional approaches aren't built for modern business; it provides various approaches to accelerate the time it takes to get data into the cloud so your business can start reaping the benefits faster. We also include real–world examples of companies realizing the success of a data–first approach. Finally, we show you how to get started with the right data–migration strategy.

4 triggers are driving many Azure cloud migrations

Organizations are moving data to the cloud for many reasons but four stand out:

Costs. The cost of ownership becomes too high to maintain and scale on–premises data capabilities.

License expiration. Support for Hadoop ends, causing organizations to either upgrade to a new version or pay for support.

Innovation. Data leaders need to deliver applications and innovate faster than Hadoop allows.

End of support. Version sunsetting is creating opportunities to modernize while migrating to the cloud.

Hadoop Migration Triggers

- Cost of ownership
- License Expiration
- Performance and Autoscaling
- Better VM types
- High availability/disaster recovery
- Compliance
- Application, innovation, delivering applications faster
- End of support

Activating data in the cloud matters now more than ever

By moving to the cloud, organizations gain:

- Faster insights from AI and ML
- More agile storage capabilities
- Improved disaster recovery options

- Scalable storage and analytic capabilities
- Flexibility for future data strategies

"Companies have collected rich datasets over decades but are not leveraging this data to transform and create new applications, which could generate new revenue streams or optimize the way they do business."

– Amit Agrawal, *Chief Architect, National CSA Team, Microsoft*

Legacy approaches aren't built for modern migrations

Legacy technology such as bulk transfer devices and DistCp were designed for static data transfer copy rather than for moving active and changing data from on-premises storage to the cloud.

Manual migration tools create additional costs and frustration, including:

- Business disruption and downtime
- Additional IT resources needed
- Loss of data and data inconsistencies
- Custom code development and maintenance

By avoiding these pitfalls, your IT team can focus on business-critical projects.

60% of large-scale data migrations exceed time expectations, budgets, or both.

Biggest impacts to cost of prior Hadoop migrations

IT resources required	54%
Complexities of handling changing data	54%
Custom code development and maintenance	53%
Business disruption	32%
Loss of data or data inconsistencies	16%
Workload application/code migration	15%
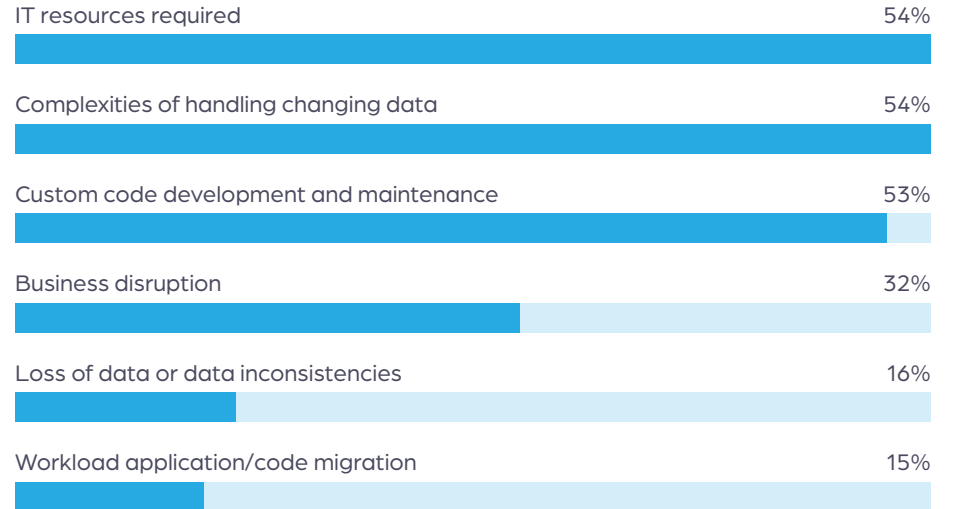
Source: DBTA Hadoop to Cloud Migration Benchmark Report

A framework for faster migrations to Microsoft Azure

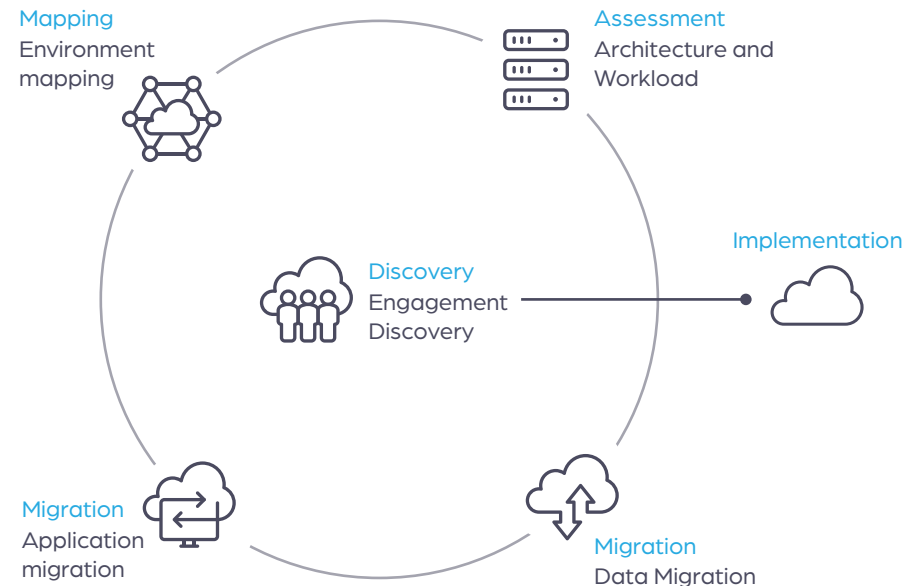Business conditions mandate different approaches based on the business need or use case, but you can start your data migration at any one of these five stages and quickly deliver business value, monetize the data, and innovate faster.

1. Discover your business needs by conducting an overall assessment of the project.

2. Map the right services and predictable TCO.

3. Assess your architecture and workloads to understand your application–readiness for migration and target environment.

4. Migrate your data using a data–first approach to quickly gain insight from existing data estates and make transformative decisions.

5. Migrate workloads and applications to help with agility and gaining confidence on running critical workloads on the public cloud.

Mapping
Environment mapping

Assessment
Architecture and Workload

Implementation

Discovery
Engagement Discovery

Migration
Application migration

Migration
Data Migration

Smart organizations use a data–first strategy

What is a data–first strategy?

A data–first strategy means moving as much of your live data into the cloud as fast as possible to take advantage of cloud–scale storage, analytics, and new capabilities.

Why is it smart to move your data first?

This approach provides flexibility for the workload and application migration while avoiding big–bang approaches.
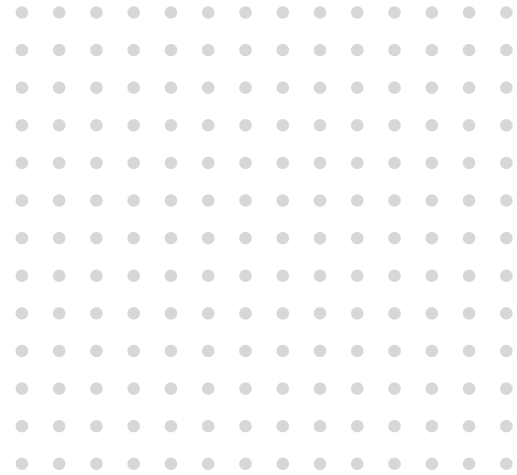
Organizations can realize a much faster ROI on their overall cloud migrations while the existing on–premises production workloads continue to run unaffected.

Requirements for data–first success

A data–first strategy requires meeting four key criteria:

- Efficiently handling migration of any volume of data, from terabytes to multiple petabytes

- Supporting frequent and actively changing transaction data from thousands to hundreds of thousands of events per second

- Requiring zero downtime of current source and production systems

- Validating successful transfer of all data, not just the initial migration

Case Study

Daimler AG moved 110 TB of critical data to Microsoft Azure with zero downtime

- Gained 5 times the storage capacity at 30% lower cost

- Modulated the migration link between on–premises and cloud to share the connection

- Guaranteed data consistency, real–time availability, and continuous replication, even with changing data

"The on–premises solution would have been nothing more than a dead end for us, as it would have been too inflexible as well as very expensive."

Guido Vetter, Head of Corporate Center of Excellence Advanced Analytics and Big Data, Daimler AG

How to get started with a data–first approach

Successful data leaders are embracing data migration, modernization, ML, and AI in the cloud. Choosing the right tools and partners will enable your live, business–critical data to flow freely — even during petabyte–scale migration — and will prepare you to build a modern platform that can leverage this data for competitive advantage.

Here are four steps to help get you started:

1 **Clarify the underlying drivers for the migration and map your approach to those drivers.**

2 **Proactively seek ways to move datasets early in the process to pull forward data science in the cloud.**

3 **Explore options to automate the migration and ensure that the data landed in the target cloud environment.**

4 **Take advantage of Microsoft's Migration Program to quickly execute a proof of concept that starts delivering value.**

About Cirata

Welcome to Cirata — a new company with over 45 patents and 15+ years of data science expertise in rapidly integrating high value datasets to leading cloud platforms for game changing AI activation and analytics insights.

We accelerate data–driven revenue growth by automating data transfer and integration to modern cloud analytics and AI platforms without downtime or disruption.

For more information on Cirata, visit www.cirata.com.

5000 Executive Parkway, Suite 270, San Ramon, CA 94583
US +1 877 (926–3472)   EMEA +44 (0) 114 3039985
APAC +61 2 8211 0620   All other +1 925 380 1728

10 | © 2023 Cirata Inc. | All rights reserved

www.cirata.com

eb–azcdm–230914