

Data Migrator for Azure

Microsoft's preferred and native solution for Hadoop-to-Azure data migrations with zero business disruption and reduced business risk

Increasingly, organizations are migrating what they have implemented on-premises using Hadoop, Spark, and other technologies to more innovative solutions in the Azure cloud. Yet unlike building cloud-native applications from scratch, organizations that want to move from their existing Hadoop deployments to new solutions in Azure are faced with significant challenges. These challenges include the scale of the data that needs to be migrated, the amount of data changes occurring (ingests, updates, deletes), and the critical nature of these environments, which can't be disrupted.

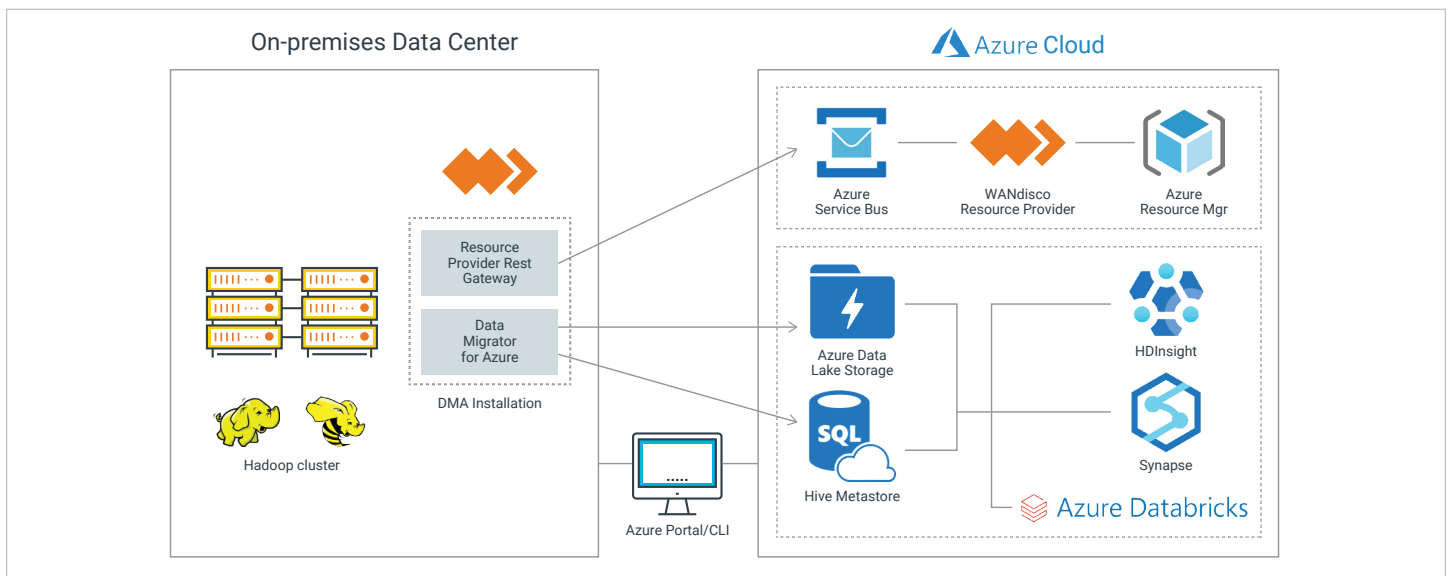
We make data migration non-disruptive

WANdisco's unique approach to migrating data at scale without disrupting the use of those datasets while an organization adopts cloud infrastructure and services has been a critically important answer to these challenges.

Data Migrator for Azure is a native Azure service that enables users to migrate petabyte-scale Hadoop data and Hive metadata to the Azure cloud with zero application downtime and zero risk of data loss, even while the source data is under active change. With Data Migrator for Azure, you can deploy and manage your data lake migrations using the same Azure management experience you enjoy today through the Azure Portal and Azure CLI.

We reduce business risks

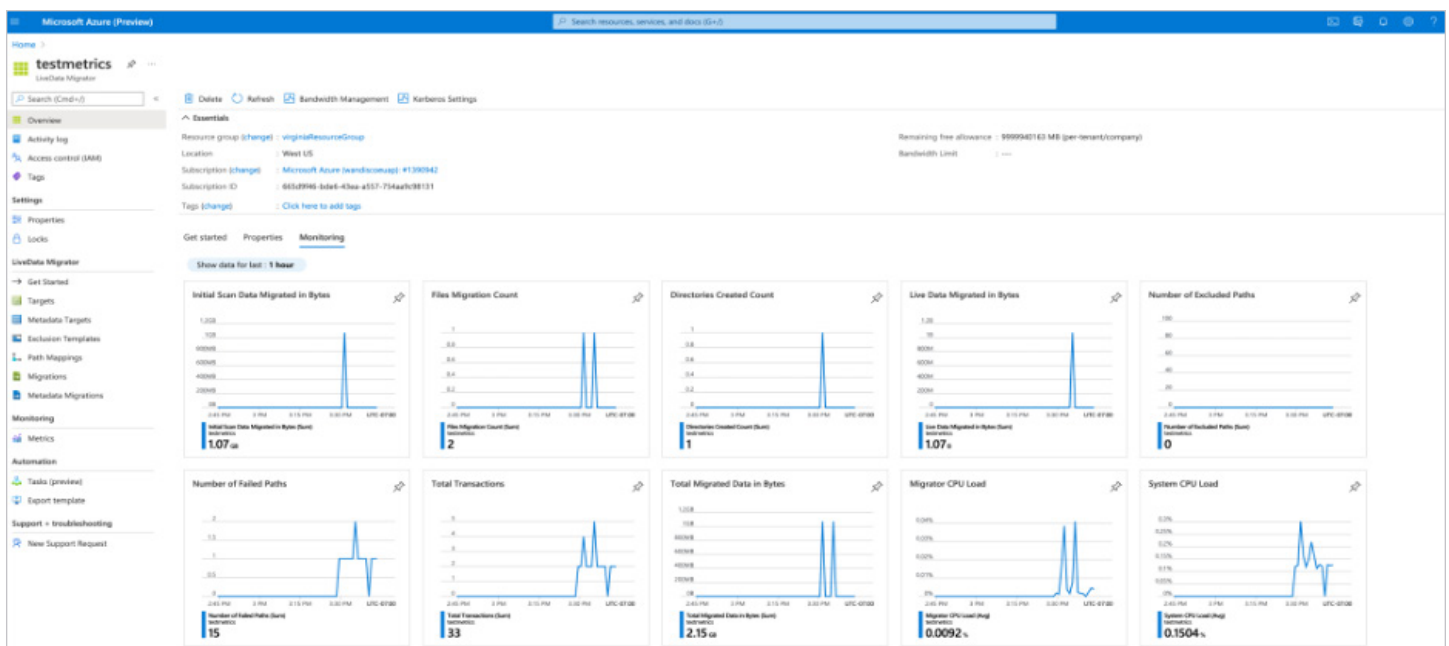
According to [Bloor Research](#), more than 80% of all data migration projects run over time and budget. Migrating large data volumes with traditional approaches, such as transfer devices or DistCp (distributed copy), requires disrupting the operation of on-premises applications and doesn't cater to data that are modified or created during migration. Reconciliation at scale is costly and does not guarantee a completely consistent data outcome. The overhead required to achieve non-disruptive, no-downtime big data migration is significant due to repeated scans, systems out of sync, and manual intervention for anticipated failures and interruptions. WANdisco Data Migrator for Azure minimizes these business risks and enables successful data migration with no business disruption and best time-to-value.



DMA Deployment Architecture

Data Migrator for Azure capabilities

- Core Service within Microsoft Azure:** Deep integration with Azure resources enables Data Migrator for Azure to be deployed at the same time as other native Azure services and with an equivalent user experience.
- Support for Native Azure security and manageability:** Data Migrator for Azure leverages Azure features such as Role-Based Access Control, Active Directory, Azure Policy enforcement, and Activity Log integration.
- Billing integration:** Customers are billed through Azure, eliminating the need for you to add a new vendor contract or require additional vendor approvals.
- Quick deployment and operation:** The Data Migrator for Azure resource can be created directly from the Azure portal. The Data Migrator for Azure service is installed on an edge node of your Hadoop cluster. Deployment can be performed in minutes without impacting current operations, so users can begin migrations immediately.
- Complete and continuous migration:** Migration of the selected data sets are performed with a single pass through the source storage system, eliminating the overhead of repeated scans while also supporting continuous migration of any ongoing changes from source to target, with zero disruption to current production systems.
- Hadoop data and Hive metadata migration:** Data Migrator for Azure supports migration of HDFS data and Hive metadata to Azure Data Lake Storage (ADLS) Gen2, as well as optional transformation to Delta Lake format on Azure Databricks.
- Migration at any scale:** Data Migrator for Azure migrates data sets at any scale, from terabytes to multi-petabytes, without impact to current production environments. Begin risk-free for small migrations and scale up to multi-petabyte initiatives without needing any additional installation requirements.
- Azure Portal and CLI Data Extension:** Users can manage the full data migration directly from the Azure portal. Additionally, Data Migrator for Azure can be configured and operated from the Azure CLI.
- Configurability and control:** You will have the ability to configure the migrations to meet your specific needs. Data Migrator for Azure includes standard configuration — such as defining sources, targets, and data to be migrated or excluded — as well as advanced capabilities such as path mapping, scheduling and network bandwidth management controls.
- Metrics and monitoring:** Data Migrator for Azure enables hands-off operations by providing information to keep you updated on the migration jobs, from health and status metrics that provide estimates for migration completion, files transferred over time, excluded paths, items that failed to transfer, as well as other real-time insights regarding usage.



Azure Portal: deploy, manage and monitor your data migrations

Data Migrator for Azure business benefits

This turnkey solution enables automated data lake migration with zero business disruption, minimized risk, and best time-to-value.

Business continuity

- No need for downtime of on-premises production clusters
- Zero changes to source applications
- Promotes “data first” approach to cloud migration

Complete and continuous migration

- Data migration with single pass of source storage
- Ongoing migration of any subsequent data changes
- Ensures zero data loss of source data and changes

Cost avoidance / IT efficiency

- Fully-automated migration minimizes need for IT resource involvement
- Zero custom code development or maintenance
- Faster time-to-value and adoption of AI and ML

Use Cases

Cloud data migration

Fully-automated data migration with zero disruption allows your users and systems to continue operating while migration is underway. Migrate to Azure Data Lake Storage, the only cloud storage service purpose-built for big data analytics.

Hybrid data estate

Maintain on-premises assets for as long as necessary. Organizations can maintain critical on-premises applications while expanding their investment and innovation in the cloud.

Cloud-burst

Enables enterprises to access compute and storage in the cloud when required, rather than building out on-premises infrastructure for peak capacity.

Cloud analytics

Once in ADLS, the data is available to Azure analytics services such as HDInsight, Synapse, and Azure Databricks. Optionally transform the data and metadata directly to the Delta Lake format on Databricks so that it becomes immediately available in Databricks’ preferred form.

Disaster Recovery

Enterprises that wish to decommission costly on-premises environments that sit idle and are only maintained for regulatory purposes can utilize the cloud as the new secure repository.

WANDISCO is the data activation company. WANDISCO makes petabyte and exabyte-scale data trapped in on-premises data lakes and IoT/ Edge/ file systems actionable by delivering cloud data migrations in real time with no business disruption or data loss. With the fastest time to business value and outcomes, zero downtime, and zero data loss, WANDISCO’s solutions keep geographically dispersed data at any scale consistent between on-premises or edge computing environments and cloud platforms allowing businesses to operate seamlessly in a hybrid or multi-cloud environment. WANDISCO has over a hundred customers and significant go-to-market partnerships with Microsoft Azure, Amazon Web Services, Google Cloud, Oracle, Databricks, and others as well as OEM relationships with IBM and Alibaba. For more information on WANDISCO, visit www.wandisco.com.

5000 Executive Parkway, Suite 270
San Ramon, CA 94583

www.wandisco.com

Talk to one of our specialists today

US +1 877 WANDISCO (926-3472)
EMEA +44 (0) 114 3039985
APAC +61 2 8211 0620
All other +1 925 380 1728

Join us online to access our extensive [resource library](#) and view our webinars.

Follow us to stay in touch

