Cirata Data Migrator

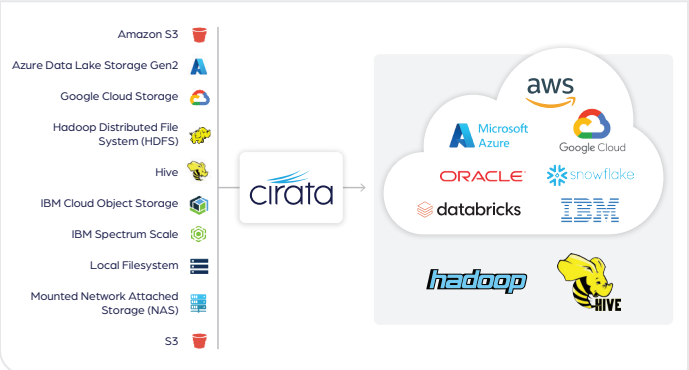
Automated data transfer across on-premises and cloud with no business disruption

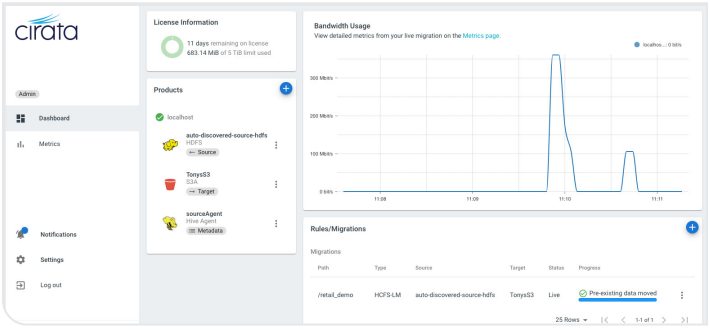
Organizations are modernizing their on-premises data architectures to more innovative solutions, both in the cloud to utilize analytics and AI applications, and to develop more robust DR environments across on-premises and cloud. With large datasets, transfer challenges can result from the amount of data under active change, and the potential for disrupting existing business critical operations.

Product overview

Data Migrator is an enterprise-class data replication software platform that keeps data consistent across distributed environments. It is an ideal solution for data lake migration and modernization, establishment of disaster recovery implementations with near-zero recovery point objectives, and support for hybrid and multi-region deployments.

The solution is noninvasive and moves big data sets with a single pass through the source storage, eliminating the overhead of repeated scans, while also supporting near-real-time replication of any ongoing changes with zero disruption to current production systems.





Key benefits

Flexibility

Data Migrator supports all leading cloud platforms as well as last mile metadata transformation to the most common analytic formats, providing users the ability to use their preferred target cloud and analytics technologies immediately.

Business continuity

Data Migrator continuously transfers data even as it is actively changing, without imposing production system downtime or business disruption, and simplifying the overall data transfer process.

Scalability

Data Migrator can accommodate data transfer at any scale. Organizations can begin with small terabyte transfers and scale up to multi-petabyte deployments using the same software product.

Cost & risk avoidance

Data Migrator is fully automated and does not require any custom code development or changes to existing applications. This minimizes the need for IT resource involvement and reduces overall project costs and risks.

Data Migrator capabilities

- **Latest enhancements:** Apache Iceberg support, Apache Ozone live migration, live CEPH migration, checksum verification, metadata verification and repair.
- **Quick deployment and operation:** Data Migrator is installed on your chosen source host(s). Deployment can be performed in minutes without impacting current operations, so users can begin moving data immediately.
- **Complete and continuous data transfer:** Existing datasets can be moved with a single pass through the source storage, eliminating the overhead associated with multiple scans, while also supporting continuous replication of any ongoing changes from source to target with zero disruption to current production systems.
- **Support for data and metadata migration:** Data Migrator supports the transfer of unstructured datasets as well as metadata stored in structured Hive tables. Data Migrator transforms metadata from the source metastore format to various supported metadata targets including Databricks, Snowflake, AWS Glue and others.
- **Support for multiple sources and targets:** Data Migrator supports a variety of sources including Hadoop File System, IBM Spectrum Scale, Network Attached Storages, and cloud object stores (AWS S3, ADLS Gen2, GCP, IBM COS, etc) and migrates data to target on-premises file systems like HDFS or cloud storages such as AWS S3, ADLS Gen2, and GCP. Along with data, Data Migrator also migrates Hive metadata to targets such as Databricks and Snowflake, and transforms the metadata for immediate use with migrated data.
- **Data transfers at any scale:** Data Migrator supports the transfer of datasets at any scale, from terabytes to multi-petabytes, without impacting current production environments. Horizontal scaling capabilities allow users to scale their data transfer capacity by configuring multiple transfer agents to maximize the productivity of available bandwidth.
- **Configurability and control:** Users can configure data transfer jobs to meet their organization's specific needs. This includes standard configuration options such as defining sources, targets, and data to be migrated, as well as advanced capabilities such as migration prioritization, path mapping, and network bandwidth management controls.
- **Browser-based user interface:** Users can leverage the Cirata user interface (UI), a browser-based UI that allows them to manage the complete data transfer process from a single management console.
- **Programmatic interface:** Data transfers can also be managed through a comprehensive and intuitive command-line interface or using the self-documenting REST API to integrate the solution with other programs as needed.
- **Data transfer verification:** Data transfer verification scans both source and target environments to ensure data fidelity and validate the success of all data transfers. Notifications can be used to specify the status of transfer verifications and receive the results by email.
- **Metrics and monitoring:** Data Migrator keeps you updated on data transfer jobs, with health and status metrics providing estimates for data transfer completion, email notifications and real-time insights on usage for hands-off operation.
- **Data Target Match:** In an event when unplanned issues occur that cause data to become inconsistent or fail to migrate, Data Migrator's Target Match feature identifies and transfers the missing data, ensuring users maintain consistent and correct data at all times.

Use cases

Data modernization
Shift away from legacy data technologies and siloed or underutilized datasets to more advanced and capable data platforms, typically in the cloud, that enable advanced analytics, AI, faster decision making, and more flexible and elastic storage and compute to unlock the full value from the data.

Disaster recovery
Maintain a current replica of actively used data in another location (either cloud or on-premises) for failover purposes in case the primary production environment becomes unavailable. Providing the ability to replicate the data in near-real-time is critical to meet any near-zero RTO (recovery time objective) and RPO (recovery point objective) requirements.

Hybrid and multi-cloud
Implement flexible architectures that maintain data in hybrid environments, which can include on-premises, cloud, multi-cloud, and intercloud deployments. As a result, organizations are able to utilize the cloud and data architecture of their choosing so they can:

- Leverage best in-class capabilities
- Improve availability
- Increase regional coverage

