

Automating Hadoop Migration to the Cloud

As enterprises migrate Hadoop to the cloud they're faced with a unique set of challenges. Massive volumes of data need to be migrated with limited resources. Data must be continually synchronized to avoid business disruption. How can you ensure successful migration?

Traditional approaches for migrating Hadoop to the cloud are unable to achieve the speed and scale required of today's enterprise:

- **Manual hand-coding** is resource-intensive and results in delays and cost overruns, requires specialized resources, and is prone to error.
- **Lift-and-shift**, one-and-done migration is not feasible with the size and complexity of Hadoop environments, especially if you consider enterprises with multi-petabyte data volumes.
- **Data-transfer solutions** like AWS Snowball and Google Transfer Service are available but do not provide capabilities like active replication and synchronization while transitioning applications, which are critical for successful migrations.
- **Singular point solutions** are unable to manage the data siloed across business units even though data is co-located in Hadoop Data Lakes.

Gartner research shows these traditional approaches to migration require significantly more time, specialized resources, and ultimately higher cost, whereas choosing automation solutions ensures successful long-term outcomes. For migrations, it's not only the tool but also the strategy around the tool that determines success.

Rethinking Hadoop Migration through Automation: Infoworks Replicator

Infoworks pioneered end-to-end automation for seamless Hadoop migration, enabling large-scale Hadoop data lakes to be migrated to the cloud 3x faster and at 1/3 the cost of traditional approaches.

Infoworks Replicator provides rapid, high-performance migration of Hadoop data and metadata to the cloud. The solution is provisioned to empower business teams to migrate to the cloud object stores without being experts in the big data (Hadoop) or cloud technology space.

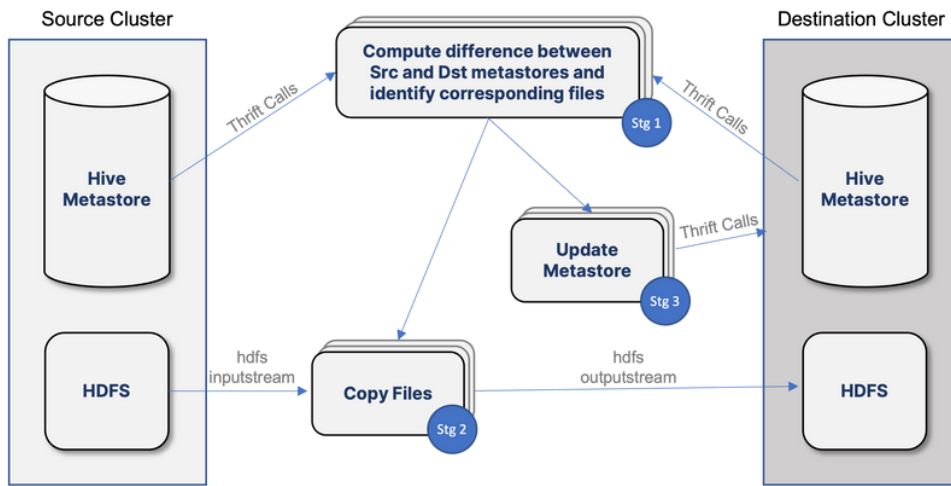
Infoworks Replicator migrates Hadoop to the cloud 3x faster at 1/3 the cost of traditional approaches

- **Automation**
Powers faster migrations and reduces resource requirements to speed time to value.
 - **Synchronization**
Continuous operation and synchronization ensure zero business disruption.
 - **Scalability**
Migrate petabytes of data and metadata to any cloud seamlessly.
 - **Extensibility**
Automate your cloud data platform post-migration to accelerate deployment of new analytics use cases.
-

How It Works

Replicator data migration is facilitated with multi-stage jobs:

- Determine the difference between source and target
- Calculate what the payload would be for the current load
- Copy across the data with user-defined parallelism, maintaining bandwidth thresholds
- Update the metastore
- Perform data validation to ensure that data copies are intact with cryptographic checksums of the files moved.



The Infoworks Difference

Infoworks Replicator provides a comprehensive solution to simplify, automate, and accelerate migration of Hadoop data and metadata to the cloud.

- **End-to-end automation** enables acceleration of data migration. Users replicate data from on-premise to the cloud at the click of a button with no-code automated data replication. Specialized data engineering resources are not required.
- **Automatic and incremental synchronization** of data and metadata.
- **Continuous Incremental replication** with high-speed, distributed difference computation across endpoints detects incremental changes to synchronize data.
- **Enterprise-grade scalability** leverages existing Hadoop clusters for data plane and independently scale-out performance. In a real-world use case, 1 PB of data is replicated in 24 hours.
- **Robust API endpoints** allow users to manage data validation pipelines in bulk.

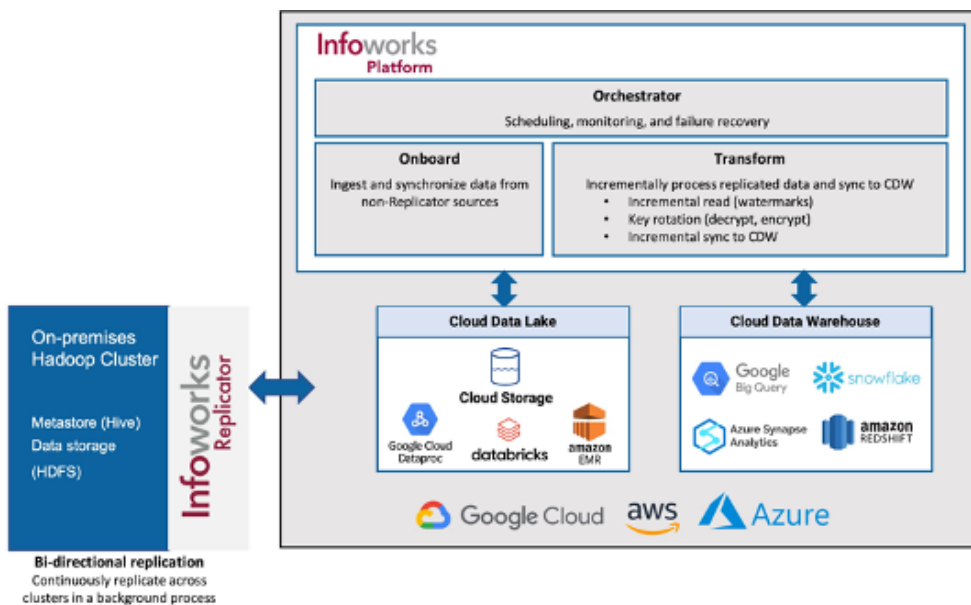
Automation solutions will reduce data management manual tasks by more than 45%.

GARTNER

Modernizing Your Cloud Data Operations

Beyond replicating data, Infoworks accelerates new analytics use cases in the cloud by up to 3x by modernizing cloud data operations with Infoworks Platform.

Infoworks Replicator is extensible to Infoworks Platform enabling onboarding, preparation, and operationalization of data from over 200 sources in hybrid multi-cloud environments. Infoworks modernizes cloud data operations by connecting directly to the data sources, enabling enterprises to decommission on-premises data lakes - realizing significant cost savings and productivity gains.



Simplify, Automate, Accelerate Hadoop Migration

Migrating Hadoop to the cloud at scale is a massive undertaking. Traditional approaches are unable to achieve expected results. Infoworks solves the challenges of migrating Hadoop to the cloud, enabling 3x faster migration at 1/3 the cost and with fewer specialized resources required. Find out more at

Learn More

Infoworks Replicator Introduction: www.infoworks.io/infoworks-replicator/
 Infoworks Replicator Technical Overview: infoworks.learnupon.com/store
 Infoworks Platform Introduction: www.infoworks.io/products/platform/

About Infoworks

Infoworks pioneered end-to-end automation to migrate to the cloud and modernize cloud data operations better, faster, and at lower cost. Some of the world's largest enterprises rely on Infoworks to realize the true value of their most strategic asset - their data.

LEARN MORE QUICK LINKS

[Infoworks Replicator Introduction](#)

[Infoworks Replicator Technical Overview](#)

[Infoworks Platform Introduction](#)