



Getting Started with Hadoop Migration and Modernization - Q&A

Most companies use Hadoop for big data analytical workloads. The problem is, on-premises Hadoop deployments have failed to deliver business value after it is implemented. Over time, the high cost of operations and poor performance places a limitation on an organization's ability to be agile. As a result, data platform teams are looking to modernize their Hadoop workloads to the data lakehouse.

In this video, learn about:

- Use cases for modernizing Hadoop workloads
- How the data lakehouse solves the inefficiencies of on-premises Hadoop
- Success stories from organizations that have modernized Hadoop with the data lakehouse on Dremio



Kamran Hussain

Field Solution Architect
Manager

Dremio



Alex Merced

Developer Advocate

Dremio



Tony Truong

Sr. Product Marketing

Dremio

Webinar Q&A

1. What are my options to deploy Dremio if I have HDFS on-prem?

Answer: (a) Leave the data in HDFS, connect Dremio to it and start consuming it as if it were a table. (b) Move the data to modern object storage (MinIO or ECS on-prem) or Cloud Data Lake storage like S3, ADLSgen2, Google storage. Dremio delivers sub-second performance and gives business users self-service access to all their data, regardless of where it is stored.

- 2. This webinar sums up what we are doing. My ETL pipelines are hard to keep up with because data is copied everywhere. Would I be able to join data in Postgres with object storage? I'm hoping this will reduce the amount of pipeline maintenance?**

Answer: Yes, Dremio can connect to relational and nosql databases in addition to object storage. There is no need to copy data, or build cubes to accelerate the queries. Dremio's Lakehouse engine allows consuming data from many sources very simple and fast. You can build virtual datasets on top of disparate data sources, unifying the experience for the business users. You can also create self-service analytics using Dremio's semantic layer, thus reducing the need for additional software costs.

- 3. New feature - Dremio to Dremio connector. Interesting use case possibility?**

Answer: Many of Dremio's customers have multiple Dremio Clusters for different lines of business. Some of the time the datasets have to be shared across different business units. The new Dremio to Dremio connector allows creating Dremio as a source and consuming data from one cluster to another.