Lumada DataOps Suite: Lumada Data Optimizer for Hadoop

Optimize Cost and Utilization for Your Hadoop Environment

SOLUTION PROFILE

By 2025, the global datasphere will grow more than five times its current size. As the pace of digital transformation accelerates, maximizing the value of your data is key to gaining a competitive advantage. Hadoop can help you unlock the value hidden in all of this data. However, trying to keep pace with this astounding data growth has become very complex and expensive.

A Hadoop Distributed File System (HDFS) gives you clustered storage that federates many data nodes into a single pool where both compute and storage are co-located. As clusters fill with aged and inactive data, you must scale your storage capacity. Traditional storage scaling in Hadoop requires that you also scale compute. Having to simultaneously add compute and storage creates an inefficient balance and utilization of resources, and becomes very costly with today's storage capacity demands.

Hadoop Gets Very Expensive as Demands for Storage Capacity Increase

Lumada Data Optimizer for Hadoop (Data Optimizer) is an intelligent data tiering solution that reduces operating costs and gives you seamless access to HDFS data for real-time analytics with Hitachi Content Platform (HCP).

Automated Data Tiering for Reduced Costs and Seamless Access to Hadoop Data for Real-Time Analytics

With Lumada Data Optimizer for Hadoop, you can independently scale storage and compute for greater flexibility and resource utilization. Configuring volumes is quite simple. And you can leverage native Hadoop functionality – such as storage types, storage policies, and the Mover service: Use these to automatically tier older, infrequently accessed data into HCP, a cost-effective and industry-leading object storage solution. This approach optimizes resource utilization by reserving Hadoop nodes for active data while storing less-frequently accessed data on HCP.

Unlike offloading data with S3A, which removes files from HDFS, Data Optimizer integrates with HDFS to free up capacity and ensure that your data always remains securely accessible through HDFS. By dynamically tiering data between HDFS and HCP, you maintain seamless access to all your data, all the time.

> **Reduce Hadoop costs with an intelligent data tiering solution for seamless HDFS access.**

Reduce Hadoop Costs by Seamlessly Tiering HDFS Data to HCP

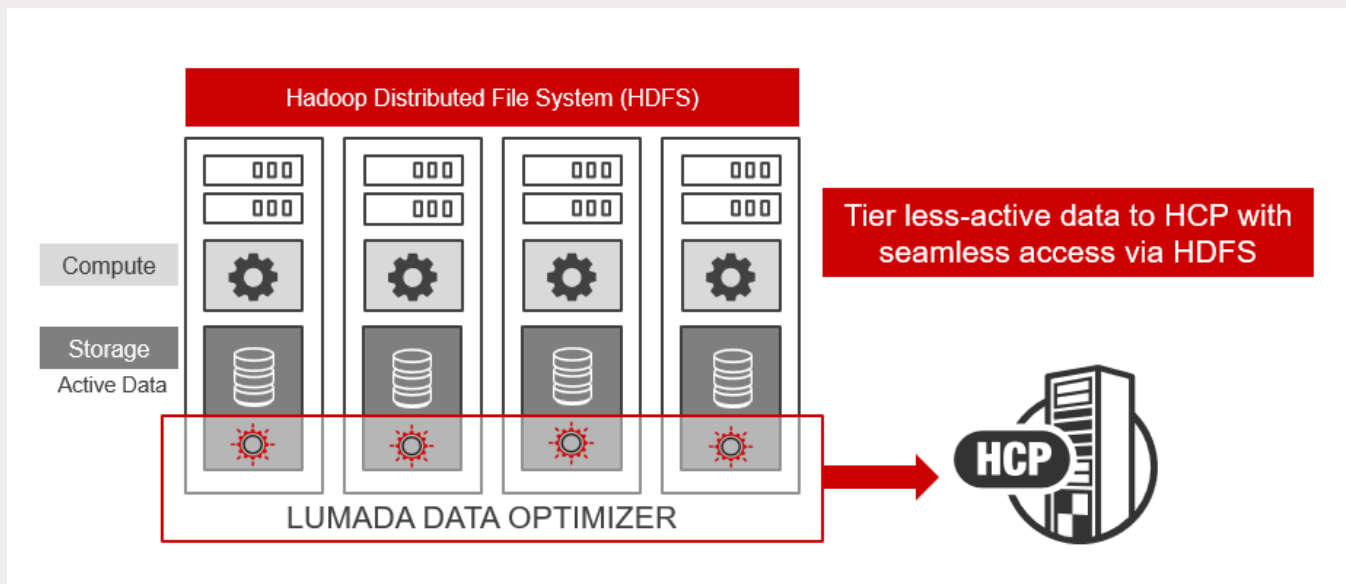**Lower Hadoop Costs With Lumada Data Optimizer for Hadoop**

Automatic policy-based tiering of Hadoop data to the most economic storage resource ensures that data continues to be utilized seamlessly and efficiently. Store data optimally for cost versus frequency of use.

Reserve Hadoop for active data while less-frequently accessed data is stored on HCP.

Seamless Access to Hadoop Data With No Disruptions

Data Optimizer automatically tiers data between HDFS and HCP to ensure that your data remains always accessible without having to alter data paths and application configurations.

Tier Hadoop data to and from HCP, a cost-effective storage platform for long-term data retention.

- Reduce Hadoop costs.
- Independently scale compute and storage resources.
- Seamlessly access tiered data via HDFS with no disruption.
- Massive scalability and more efficient resource utilization.

Scale Compute and Storage Independently

Improve utilization as you scale Hadoop. Add Hadoop nodes on demand, and dynamically grow storage with HCP to satisfy your data retention needs. Avoid scaling Hadoop compute nodes just to accommodate petabytes of cold data storage.

Gain massive data scalability with HCP for more efficient Hadoop resource utilization.

Intelligent Hadoop Data Management

Data Optimizer streamlines data management with policy-based metadata collection of HDFS to power the best decisions on how to optimally handle and store your Hadoop data.

Improve efficiency and gain better data insights to fuel the best business decisions.

Data Optimizer Is Purpose-Built for HDFS

Data Optimizer integrates with Hadoop and operates as an HDFS volume to move HDFS data to and from HCP. Since the files never leave HDFS, capacity is freed with no disruption and data continues to be accessed seamlessly via HDFS.

Gain seamless and secure HDFS access to Hadoop data that's been optimally tiered to HCP.

Optimize Hadoop Resource Consumption and Utilization

Hadoop maintains three copies of data for redundancy and availability, which consumes additional storage and compute resources. Tiering less-active data from HDFS to HCP consumes less storage and compute and reserves valuable Hadoop resources for your most active data.

Improve utilization and reduce costs by eliminating Hadoop's triple replication of inactive data.

Avoid Data Protection Headaches

With fifteen 9's data durability, erasure coding, replication, configurable redundancy and automatic repair and versioning, HCP is a resilient and self-protecting storage platform that makes data recovery easy.

Data moved to HCP doesn't require additional protection and consumes up to 40% less capacity.

A Secure, Flexible and Cost-Effective Storage Solution

HCP provides a long-term cloud object storage platform that is massively scalable. It makes compliance and data recovery easy with data mobility, AES-256 encryption, policy-based tiering to public clouds, robust data protection and up to fifteen 9s availability.

Rely on scalable, economical and highly available HCP storage for Hadoop data.

> "
> Decoupling compute and storage is proving to be useful in big data deployments. It provides increased resource utilization, increased flexibility, and lower costs.
> "

Ritu Jyoti
Program Vice President,
Artificial Intelligence Strategies

IDC

Source: "Five Benefits of Decoupling Compute and Storage for Big Data Deployments," Ritu Jyoti, Program Vice President, Artificial Intelligence Strategies, IDC

Lumada DataOps Suite

Lumada DataOps Suite provides intelligent data management for digital innovation through advanced insights based on trusted data.

The suite of products is open and modular to deliver AI-driven automation and collaboration, and includes:

- Lumada Analytics

- Lumada Data Integration

- Lumada Data Catalog

- Lumada Data Optimizer for Hadoop, and

- Lumada Edge Intelligence

Lumada is built with Pentaho technology that includes Pentaho Business Analytics and Pentaho Data Integration.

Next Steps

Check out the resources below to learn about how Lumada DataOps Suite enables better business and operational insights by improving data access with an intelligent data foundation that accelerates data discovery and automates management.

Reduce costs with an intelligent data tiering solution for Hadoop.

Watch the Lumada Data Optimizer for Hadoop video.

Visit the Lumada DataOps Suite Page.

Calculate your Hadoop Cost Savings.

Learn how Hitachi Content Platform makes your data securely available, anywhere, anytime.

Are you ready to make your data more agile and visible from edge to cloud with **Lumada DataOps Suite**?

We Are Hitachi Vantara

DataOps is the data practice for the AI era, connecting data consumers with data creators to accelerate collaboration and digital innovation. We are analytics, industrial expertise, technology and outcomes rolled into one great solution partner. Get Your DataOps Advantage.

Hitachi Vantara

SP-292-C MCoE November 2020