A Hadoop Distributed System (HDFS) gives you clustered storage that federates many data nodes into a single pool where both compute and storage are co-located. By 2023, 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 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 computer and storage creates an inefficient balance and utilization of resources, and becomes very costly with today’s storage capacity demands.

**Hitachi Vantara at a Glance**

Your data is the key to new revenue, better customer experiences and lower costs. With technology and expertise, Hitachi Vantara drives data to meaningful outcomes.
Hadoop gets very expensive as demands for storage capacity increase.

Data Optimizer for Hadoop 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 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 for Hadoop 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 Data Optimizer for Hadoop

Reserve Hadoop for active data while less-frequency accessed data is store on HCP.

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.

Intelligent Hadoop Data Management

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

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

Scale Compute and Storage Independently

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

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.

Seamless Access to Hadoop Data with no disruptions

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

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.

Outcomes

- 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.
Purpose-Built for HDFS

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

Data Optimizer for Hadoop 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.

Optimize Hadoop Resource Consumption and Utilization

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

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.

Avoid Data Protection Headaches

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

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.

A Secure, Flexible and Cost-Effective Storage Solution

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

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.

“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
**SOLUTION PROFILE**

**Lumada DataOps: Data Optimizer for Hadoop**

**Lumada Data Ops**

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

Open and modular, Lumada DataOps delivers AI-driven automation and collaboration, and includes:

- **Data Integration and Analytics**
- **Data Catalog**
- **Data Optimizer for Hadoop**

Lumada DataOps 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 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 Data Optimizer for Hadoop video.
- **Visit** the Lumada DataOps page.
- **Calculate** your Hadoop cost savings.
- **Learn** how Hitachi Content Platform makes your data securely available, anywhere, anytime.

---

**WE ARE HITACHI VANTARA**

Hitachi Vantara solves digital challenges by guiding you from what’s now to what’s next. Our unmatched industrial and digital capabilities benefit both business and society.

---

**Hitachi Vantara**

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
HitachiVantara.com | community.hitachiVantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
hitachiVantara.com/contact

© Hitachi Vantara LLC 2022. All Rights Reserved. HITACHI and Lumada are trademarks or registered trademarks of Hitachi, Ltd. Pentaho is a trademark or registered trademark of Hitachi Vantara LLC.