

A Big Data Analytics Environment Optimized for Compute and Storage

Data drives the modern organizations of the world. Making sense of this data, discovering the various patterns and revealing the unseen connections has become critical to business success. Organizations are using more data than ever before to make better decision and unlock new data driven revenue streams. Being able to handle the massive data growth and extract timely insights from this increasingly important asset is therefore a strategic imperative for enterprises today and has spurred an entire Big Data industry.

Hadoop has emerged as the platform of choice for Big Data Analytics. Hadoop Distributed File System (HDFS) as the defacto scale out file system for Hadoop has served to scale storage and compute needs for big data workloads across commodity hardware. However as we scale beyond the petabyte range, economics related to storing and archiving of this valuable asset has become a major concern for the big data industry.

The challenges that limit the overall adoption and scalability of traditional big data systems include:

- Need for moving data from source system's storage into HDFS cluster resulting in delayed analytics.
- Increased operational and capital cost of holding copies of data in different systems.
- Security risk of associated with data residing outside of enterprise IT security domain in shadow IT environments.
- Inability to scale storage nodes independent of compute in HDFS, leading to unnecessary resource cost. Specially for warm data/archive data that only uses storage and little to no compute resources.
- Reliance on data replication (factor of 3) for data protection at large scale which increases the overall cost of data storage.

CHALLENGES



Object-based storage systems overcome these limitations and are an attractive complement to HDFS giving a scalable and cost effective option for creating Big Data lakes.



SOLUTION FEATURES & BENEFITS

- Independent scaling of storage from expensive compute resources
- Efficient data durability, with erasure coding – double the usable capacity as compared to 3x replication
- In-place analytics directly from storage leading to faster insight and reduced infrastructure costs
- Support for multi-site automated DR and local storage access at remote locations
- Single namespace across locations eliminates management workload and complexity through a unified view of data
- Start small and easily expand solution with non-disruptive scaling
- Hybrid and MultiCloud interoperability with AWS, Google Cloud Platform, and Azure.

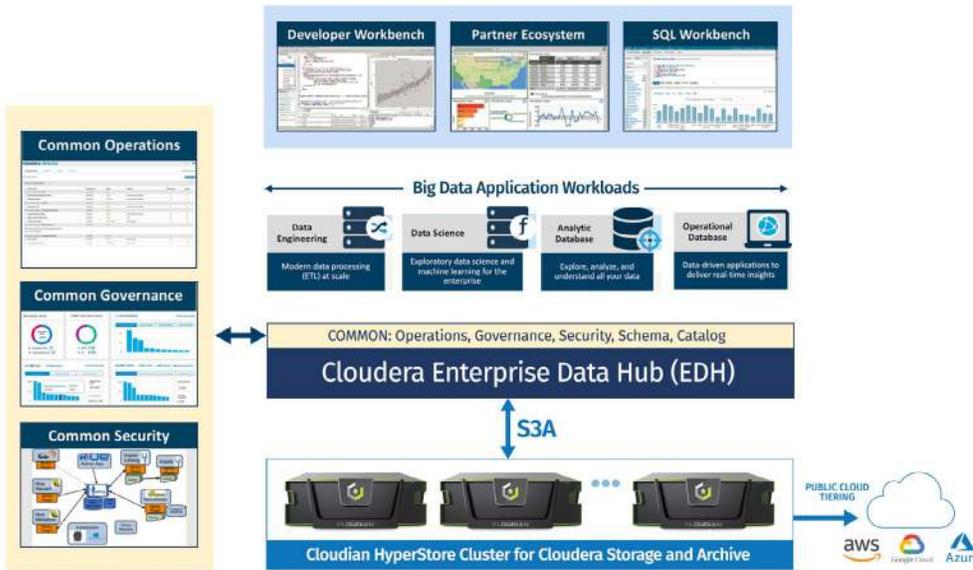
cloudera

Cloudera delivers the modern platform for machine learning and advanced analytics built on the latest open source technologies. The world's leading organizations trust Cloudera to help solve their most challenging business problems by efficiently capturing, storing, processing and analyzing vast amounts of data.

www.cloudera.com

Solution

Cloudera's Enterprise Data Hub (EDH) is a modern big data platform powered by Apache Hadoop at the core. It provides a central scalable, flexible, secure environment for handling workloads from batch, interactive, to real-time analytics. Cloudera HyperStore provides a Software Defined Storage platform that delivers industry leading scalability supporting 100s of petabytes (PBs) with via the S3 Restful API, with data integrity and protection.



Single platform solution for Big Data Applications scaling as needed to support more workloads, more users, and more data across all locations.

A Big Data lake built on Cloudera's HyperStore working on Cloudera's EDH via an S3A connector, provides a economical and scalable storage option, for Big Data workloads. This enables enterprises to,

- Create a tiered Big Data storage environment where hot data can reside within HDFS for immediate access, while HyperStore provides the storage/archive layer for warm data
- Analyze data in place without moving it from the source system into HDFS
- Scale storage independent of compute nodes, which is a huge cost saving benefit
- Take advantage of the associated metadata stored within HyperStore for accessing and searching through data under archive
- Use erasure coding instead of 3x replication for the HyperStore layer, getting increased storage density and doubling the usable storage capacity as compared to traditional HDFS

In addition HyperStore supports multi-site deployment for data protection and disaster recovery. HyperStore's policy-based replication ensures that key assets are automatically copied and available at two or more sites, providing failover instances if the primary site goes offline.

Learn More

Contact your Cloudera or Cloudera sales representatives to learn more about this solution offering and how you can benefit.

Cloudera:

www.cloudera.com

Cloudera:

www.cloudera.com

Cloudera, Inc.

177 Bovet Road, Suite 450

San Mateo, CA 94402

Tel: 1.650.227.2380

Email: info@cloudera.com

www.cloudera.com

©2018 Cloudera, Inc. Cloudera, the Cloudera logo, HyperFile, HyperScale, and HyperStore are registered trademarks or trademarks of Cloudera, Inc. All other trademarks are property of their respective holders. SB-CDA-0918