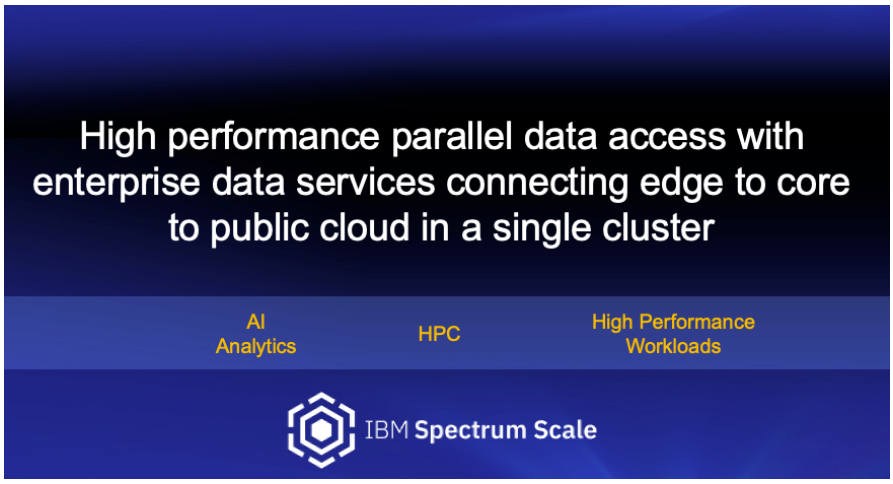
Optimize Cloudera analytics with IBM Spectrum Scale

Increase productivity and lower costs and increase SLAs

Highlights

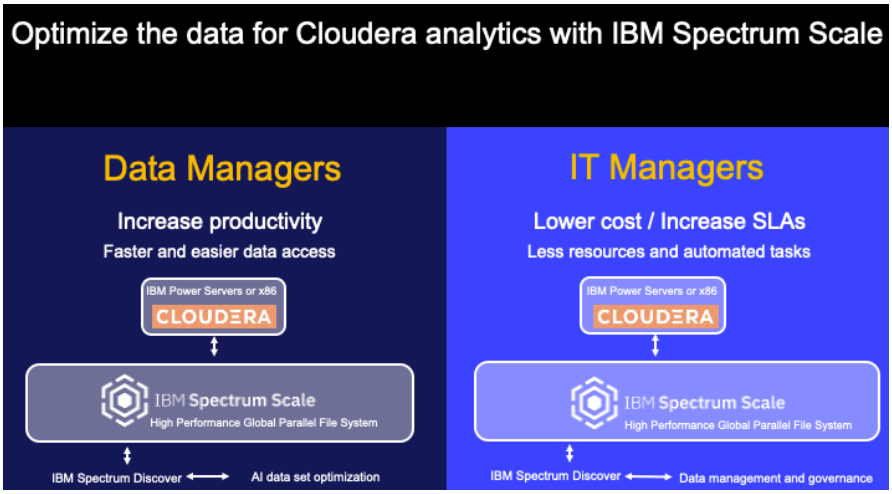
- Spectrum Scale certified with CDP Private Cloud Base
- Ease of management: single global data lake
- Economics: optimize cost and performance
- Robust: data availability, integrity and security built-in
- Global parallel performance: remove bottlenecks



High-performance storage

AI and analytics platforms require the harmonious marriage of a software platform and an infrastructure platform that is optimized for the large amounts of data that is an integral part of the solution. Many times data becomes problematic as performance bottlenecks persist and data is not located at the right place at the right time. Data can also become expensive very fast as multiple copies of data are often used and policy-based archiving is nonexistent. Data remains the key component for big data workloads but is often just an afterthought and considered a commodity component. The storage infrastructure must provide a flexible and scalable set of services to enable the best experience for advanced analytics and AI applications. The infrastructure platform must deliver dynamic and high-performance access to

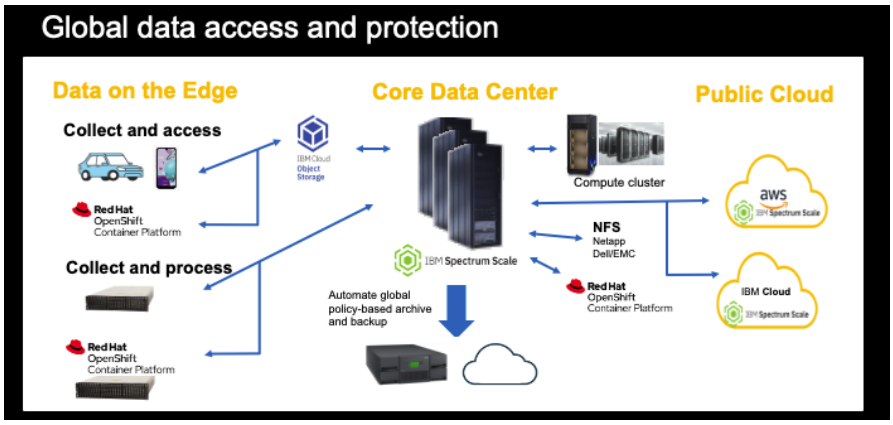
data and storage resources that are cost effective.



IBM and Cloudera

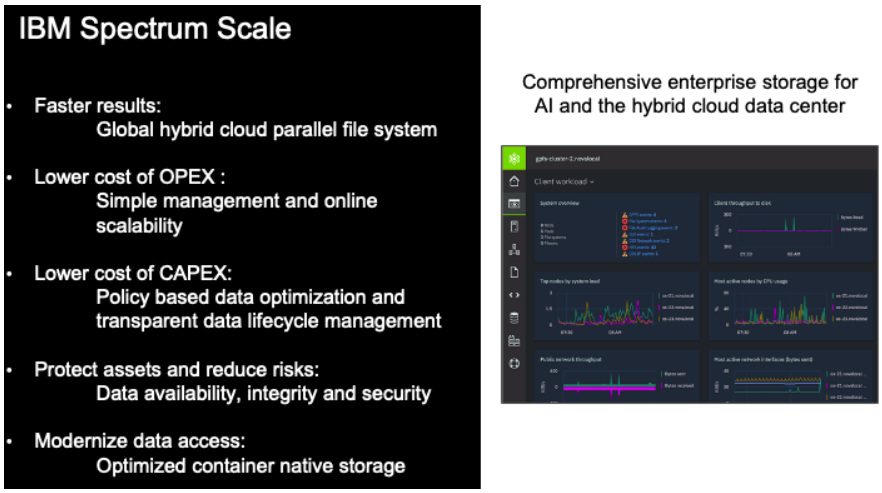
The partnership of Cloudera and IBM Storage brings these capabilities together in a solution that unleashes the insights in big data to power the digital enterprise. In 2019, Cloudera released Cloudera Data Platform (CDP) Data Center (now CDP Private CloudBase), the most comprehensive on-premises platform for integrated analytics from the Edge to AI –spanning ingest, processing, analysis, experimentation and deployment. It combines the best of Cloudera Enterprise Data Hub and HDP Enterprise Plus, a fusion of the latest and greatest open source data management and analytics technologies, integrated to work together, and optimized for deployment within the data center. Now Cloudera has certified Spectrum Scale for use with CDP Private Cloud Base.

IBM Spectrum Scale and the Elastic Storage Systems (ESS) building blocks provide an integrated storage system running IBM Spectrum Scale software with high-performance access and a global parallel file system with enterprise data services including optimized data placement. Clients leveraging Cloudera on IBM Storage and ESS can realize up to 60% reduction of storage infrastructure due to the data lifecycle management and data policies of IBM Spectrum Scale along with the ability to keep all the data on a single global hybrid cloud storage system.



An AI information architecture

IBM Spectrum Scale is a software-defined storage system based on a global parallel file system architecture that provides a high-performance file, object and optimized HDFS interface that supports the HDFS APIs. IBM Spectrum Scale becomes the storage layer for your CDP environment as an alternative to native Hadoop Distributed File System (HDFS). Support for HDFS APIs enables in-place analytics on enterprise storage instead of copying data from enterprise storage to analytics silos. In-place analytics not only eliminates duplication of data but also avoids the problems of running analytics on stale data. Spectrum Scale provides a global federated storage that can span from edge to the core data center to the public cloud. This means data users and line of business can leverage the vast amounts of data in one location for their analytics and AI workflows. Support for POSIX access enables super-fast ingest. In addition, Spectrum Scale provides a scalable environment to YBs of capacity and enterprise data services to optimize the data placement and lower cost of data capacity. Data administrators appreciate all the automated data optimization that makes their job easier with the efficiency of policy-based archiving and security that is built into the system. Shared storage to CDP allows for de-coupling of compute and storage to enable optimized and lower-cost configurations.



IBM Spectrum Scale

The value of Spectrum Scale for Cloudera is simple: High-performance parallel data access with enterprise data services connecting edge to core to public cloud in a single cluster. This makes Spectrum Scale second to none for AI/Analytics, HPC, and High-Performance Workloads.

Global data access

Why IBM?

IBM Storage for Data and AI is more than storage products or even storage solutions. It consists of a storage strategy to help customers on their journey to AI and the hybrid cloud data center. IBM continues to drive leadership for scalable high-performance workloads as well as efficient, secure, scalable, capacity storage for file and object-based solutions. Our products provide an enhanced strategy for AI and the hybrid cloud. We provide a foundation for the future for the edge, the core data center and the public cloud including Kubernetes containers and the Red Hat OpenShift platform. IBM Storage for Data and AI lowers complexity and cost with increased integration to an AI information architecture for the hybrid cloud that can be infused to the entire organization. Our message is easy “Storage Made Simple for Data and AI”.

Next steps

→ IBM Spectrum Scale Datasheet

→ IBM ESS Datasheet

For more information

Visit our solutions web page:
https://www.ibm.com/it-infrastructure/storage/ai-infrastructure

© Copyright IBM Corporation 2022.

IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at https://www.ibm.com/legal/us/en/copytrade.shtml, and select third party trademarks that might be referenced in this document is available at https://www.ibm.com/legal/us/en/copytrade.shtml#section_4.

This document contains information pertaining to the following IBM products which are trademarks and/or registered trademarks of IBM Corporation:
IBM® Spectrum Scale

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.