

# ENTERPRISE DATA LAKE: ELIMINATE SILOS, SIMPLIFY AI ACCESS

All on-premises: Secure, optimized, multi-purpose



**Key Highlights**

- **Optimize** data storage and increase availability and scalability, while reducing costs.
- **Craft** multi-purpose data pipelines to connect data stores and analytics workloads from the Edge to AI.
- **Open and secure** data to allow access in a completely compliant manner.

How can you adopt AI-driven storage practices, modernize your data lake access and lower the cost of storing your data? Read on to find out the IBM Storage and Cloudera approach.

### Optimizing, not abandoning on-premises

While mass data storage is becoming increasingly cheap, compute is expensive and ephemeral. By separating compute and storage, data teams can easily and economically scale storage capacity to match rapidly growing datasets while only scaling distributed computation as required by their big data processing needs. Separation of compute and storage allows database software increased availability and scalability, and can dramatically reduce costs.

[Cloudera Data Platform \(CDP\)](#) is built with a cloud-native hybrid cloud architecture, to manage and secure the data lifecycle across private and public clouds. CDP's architecture not only enables the separation of compute and storage, but also maintains data state and context. The architecture simplifies management of metadata, metrics, audit, lineage, permissions, filters and attributes, while delivering powerful elastic performance for the data lifecycle in any cloud and data center.

By decoupling storage and compute environments, we offer customers increased flexibility by only adding storage or compute capacity as needed. This results in cost savings from the reduction in the data footprint at the customer site and the ability to do in-place analytics. Storage capacity can also be increased.

With a parallel architecture like [IBM Spectrum Scale](#), no single metadata node can become a bottleneck. Every node in the cluster can serve both data and metadata, enabling a single IBM Spectrum Scale file system to store billions of files. This architecture enables clients to grow their CDP environments seamlessly as the data grows.

#### About IBM Storage

IBM Storage for Data and AI simplified your infrastructure with optimized efficiency to drive faster results that are massively scalable and globally available from collect to organize to analyze and infuse. Learn how you can help your customers connect data across any architecture to deliver insights quickly and give them the edge to outthink and outperform the competition.

#### About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises. Learn more at [cloudera.com](https://cloudera.com)

### Crafting multi-purpose data pipelines

Connect your data stores and analytics workloads, no matter where they reside, to maximize and optimize. Whether you store on-premises, in the cloud, or analyze at the edge, pull together and consolidate everything on one platform.

Create complex, multi-purpose data pipelines with no code using [Cloudera DataFlow \(CDF\)](#). Reduce data integration development time, powered by [Apache NiFi](#).

IBM Spectrum Scale separates compute from storage and optimizes data placement to lower cost of CAPEX for the IT budget. For data users, IBM Spectrum Scale can grow online to YottaBytes in a single, accessible namespace, lowering OPEX costs and manual or time-consuming data movement.

### Opening and securing data as a commodity

Allowing access to all of your data, no matter where it resides or what the type, in one secure platform is a must. Security and governance is the key and a single data fabric through [Cloudera Data Platform-Data Center](#) on IBM Storage solutions is a simple solution. The [Elastic Storage System \(ESS\) 3000](#) or the [ESS 5000](#) together or individually offer the most optimized storage solutions for running analytics workloads.

To support the security and regulatory compliance requirements, IBM Spectrum Scale offers Federal Information Processing Standard (FIPS) compliant data encryption for secure data at rest, policy-based tiering/ILM, cold data compression, disaster recovery, snapshots, and backup and secure erase. Combined with [Apache Atlas](#) and [Apache Ranger](#) on CDP and [IBM Spectrum Discover](#), data users are provided additional data governance options and the ability to define and manage security policies.

### Find out more

Visit [cloudera.com/IBM](https://cloudera.com/IBM) or [ibm.com/it-infrastructure/storage/ai-infrastructure](https://ibm.com/it-infrastructure/storage/ai-infrastructure) for more information.