

Lenovo Big Data Solutions for Cloudera Data Platform

Enabling analytics and machine learning on growing business data

The Big Data Challenge

Big data is more than a challenge. It is an opportunity to find new insights in data to make your business more agile and to answer questions that were previously beyond reach. To open the door to a world of possibilities Cloudera employs the latest big data technologies to address critical business value drivers – growing business, connecting products and services, and protecting business.

Apache Hadoop and Apache Spark are open source software frameworks that are used to reliably manage and analyze large volumes of structured and unstructured data. Cloudera enhances this technology to withstand the demands of your enterprise, adding management, security, governance, and analytics features. The result is that you get an enterprise-ready solution for complex, large-scale analytics.

The Lenovo Big Data Solutions for Cloudera Data Platform (CDP) provide a predefined and optimized hardware infrastructure for CDP Private Cloud, a hybrid cloud version of CDP that seamlessly connects on-premises environments to public clouds with consistent, built-in security and governance. These solutions enable analysis of large data sets easily and quickly through a massively parallel processing environment, and provide exceptional reliability, scalability and flexibility. Entry through high-end configurations are supported along with the ability to easily scale as enterprise use of big data grows.

Cloudera Data Platform

CDP is an enterprise analytics and management platform, enabling ingestion, management, and delivery of any analytics workload from Edge to AI. It provides enterprise grade security and governance, and self-service access to integrated, multi-function analytics on centrally managed and secured business data. CDP allows you to meet the exponential demand for analytics and machine learning services with a petabyte-scale hybrid data architecture, delivering faster time to value and supporting critical workloads at scale.

CDP gives you complete visibility into all your data. The CDP control plane allows you to manage the data, infrastructure, analytics, and analytic workloads across hybrid and multi-cloud environments, providing consistent security and governance across the entire data lifecycle.

Highlights

- Deliver analytics and machine learning services to react faster to changing business requirements
- Meet the growing demand for analytics and machine learning services with a scalable data architecture
- Consistently and easily enforce security and governance policies across hybrid and multi-cloud deployments to ensure regulatory compliance
- Invest in a platform powered by open source, ensuring continual and rapid innovation to address evolving business requirements

One Data Platform. Many Applications.

Cloudera Data Platform provides a consistent experience across Public Cloud, Multi-Cloud, and Private Cloud deployments.

CDP Private Cloud provides a disaggregation of compute and storage, and allows independent scaling of compute and storage clusters. Through the use of containerized applications deployed on Kubernetes, CDP Private Cloud brings both agility and predictable performance to analytic applications. The three main benefits of CDP Private Cloud are:

- **Simplified multitenancy and isolation** - The containerized deployment of applications in CDP Private Cloud ensures that each application is sufficiently isolated and can run independently from others on the same Kubernetes infrastructure. Such a deployment also helps in independently upgrading applications based on your requirements. In addition, all these applications can share a common Data Lake instance.
- **Simplified deployment of applications** - CDP Private Cloud deployment ensures a much faster deployment of applications with a shared Data Lake compared to monolithic clusters where separate copies of security and governance data would be required for each separate application. In situations where you need to provision applications on an arbitrary basis, for example, to deploy test applications; CDP Private Cloud enables you to rapidly perform such deployments.
- **Better utilization of infrastructure** - CDP Private Cloud enables you to provision resources in real time when deploying applications. In addition, the ability to scale or suspend applications on a need basis in CDP Private Cloud ensures that your on-premises infrastructure is utilized optimally.

Powered by Lenovo

Lenovo Solutions for Cloudera Data Platform and earlier versions provide an optimized hardware infrastructure designed for high performance and scalability, handling the Big Data analytics and machine learning requirements of your business today and in the future.

All Lenovo ThinkSystem and ThinkAgile servers are high performance systems, consistently holding numerous world performance benchmarks. Engineered for always-on productivity, ThinkSystem and ThinkAgile servers are consistently ranked high in x86 server customer satisfaction and #1 in x86 server reliability¹.

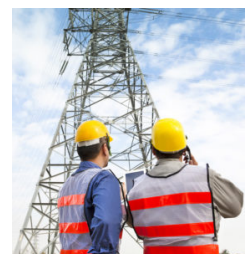
Complementary Offerings

Connecting the clustered server environment in these solutions can be easily accomplished with network switches. The recommended offerings for these solutions are 10GbE switches.

Lenovo XClarity™ Administrator is a centralized resource management solution that is aimed at reducing complexity, speeding response, and enhancing the availability of Lenovo server systems and solutions. It captures industry-leading proactive platform alerts, enabling administrators to migrate workloads or replace failing components without incurring downtime.

Customer Spotlight

A utility needed a unified and more holistic view of their company data - both legacy database and unstructured data. They consolidated their data with Cloudera on Lenovo ThinkSystem SR650 servers, and are now implementing predictive analytics on power consumption utilization as well as fraud detection.



Tying It All Together

In today's rapidly-changing technology environment, empowering your data center transformation isn't just a necessity—it's also a journey. Regardless of your current environment, Lenovo Services is a true business partner that will take you from where you are, to where you want to be. At every stage, you'll get our expertise and services to help you:

- **Drive Digital Transformation.** You'll get the best architectures suited to your unique needs, along with our industry insights, expert guidance, and hands-on experience.
- **Foster Innovation.** Free up your internal resources to focus on initiatives that grow your business.
- **Simplify Your Support Experience.** Gain a trusted partner who understands your systems and solutions to fully support and optimize your data center.

Why Lenovo

Lenovo is a leading provider of data center infrastructure solutions. We partner with you to identify, design, install and support the solution that best ensures your organization's needs are met throughout the IT lifecycle. Lenovo complements a portfolio of leading x86 infrastructure with a full range of storage, software, and comprehensive services that provides excellent performance, reliability, and security for your IT environment from the edge to the cloud.

For More Information

To learn more about Lenovo solutions for Cloudera Data Platform, contact your Lenovo sales representative or Business Partner or visit: www.lenovo.com/systems

Reference Architecture:

[Lenovo Big Data Reference Design for Cloudera Data Platform on ThinkSystem Servers](#)

¹ ITIC reliability study, <https://lenovopress.com/lp1117-itic-reliability-study>

Trademarks: Lenovo, the Lenovo logo, Lenovo Services, ThinkAgile, ThinkSystem, and XClarity® are trademarks or registered trademarks of Lenovo. Other company, product, or service names may be trademarks or service marks of others.

BDGCL02SB04

Lenovo solutions for Cloudera Data Platform provide flexibility, scalability and high performance at a cost-effective price



CLouDERA