

Lenovo Big Data Solutions for Cloudera Enterprise

Provide secure remote access to employees

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.

The Lenovo Big Data Solutions for Cloudera Enterprise provide a predefined and optimized hardware infrastructure for Cloudera Enterprise, a distribution of Apache Hadoop and Apache Spark with enterprise-ready capabilities from Cloudera. This document describes validated designs in compute-intensive, storage-intensive, streaming data and private cloud environments. The compute-intensive solution highlights Cloudera on bare-metal and on VMware virtualized servers with locally attached storage, while the storage-intensive solution highlights Cloudera with software-defined direct-attached storage. The streaming data solution allows businesses to quickly identify threats and opportunities from real-time streaming and machine learning applications, and the private cloud solution utilizes the Cazena SaaS (Software as a Service) Data Lakes platform for cloud-like capabilities. These solutions are detailed in reference architectures that provide the planning, design considerations, and best practices for implementing Cloudera Enterprise with Lenovo products.

Lenovo[™]

Highlights

- Derive new insights with an optimized infrastructure that stores, manages and processes data at scale
- Get up and running quickly with Cloudera-certified solutions that are designed to suit various capacity and scalability needs
- Deploy with confidence using predefined and pretested configurations for compute-intensive, storage-intensive, streaming data and private cloud environments

One Data Platform. Many Applications.

Cloudera provides users access to petabytes of diverse data and engines to process and query data, as well as develop and serve models quickly. The platform also provides several layers of fine-grained security and complete auditability for companies to prevent unauthorized data access and demonstrate accountability for actions taken.

Cloudera provides a shared data experience by bringing your data warehouse, data science, data engineering, and operational database workloads together on a single, integrated data platform. It enables diverse analytic processes to operate against a shared data catalog that preserves business context like security and governance policies, and makes it easier for IT to set and enforce policies while providing business access to self-service analytics.

Cloudera Enterprise can read direct from and write direct to cloud object stores like Amazon S3 and Azure Data Lake (ADLS) as well as on-premises storage environments, or HDFS and Kudu on IaaS. This provides flexibility to work on the data that you want wherever it lives, with zero copies or moves. Cloudera can also run on any compute resource for ultimate deployment flexibility - users can self-service via PaaS offering, or opt for more configurability and management via IaaS, private cloud, or on-premises.

Support for Apache Hadoop and Spark

Both Apache Hadoop and Apache Spark are fully integrated components of the Lenovo Big Data Validated Design for Cloudera Enterprise. Apache Hadoop with MapReduce provides an excellent solution for batch processing. Apache Spark's distributed in-memory storage delivers high performance processing - ideal for real-time streaming and advanced modeling and analytics.

Benefit from Lenovo Expertise

Starting with a preconfigured hardware platform that is Cloudera-certified helps your team get up and running quickly. Cloudera allows organizations to run large-scale, distributed analytics on clusters of cost-effective Lenovo data center hardware. These solutions enable analysis of large data sets easily and quickly through a massively parallel processing environment.

Deploying on one of the Lenovo validated designs yields a big data solution with exceptional performance, reliability, scalability and flexibility. These solutions support entry size through high-end configurations and the ability to easily scale as enterprise use of big data grows. A choice of infrastructure components provides the flexibility to meet a broad range of big data analytics requirements.

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.







Solution Details

Lenovo offers four solutions for Cloudera:

- Compute-intensive environments** can utilize Lenovo ThinkSystem SR650 servers running Cloudera on bare metal or on a virtualized platform with VMware vSphere. The SR650 utilizes on-board storage, eliminating the need for SAN attached storage to support robust Hadoop and Spark clusters. Lenovo ThinkSystem SR630 servers act as management nodes. An alternate solution are the 1-socket ThinkSystem SR655 servers as data nodes with ThinkSystem SR635 as management nodes.
- Storage-intensive environments** can utilize the Lenovo ThinkSystem SD530 servers with ThinkSystem D3284 JBOD storage, lowering the overall cost per TB of storage with just a small impact to overall performance. This configuration provides the flexibility to adjust compute and storage levels separately to directly match the requirements of the business.
- Businesses with data streams from financial markets, retail transactions, IoT sensors, and other sources can utilize the **streaming data** configuration. ThinkSystem SR650 servers running Cloudera on bare metal with ThinkSystem SR630 servers as management nodes provide a real-time streaming analytics solution to quickly identify threats and opportunities.
- Customers looking to implement **Cloudera in a cloud model** can utilize Cazena SaaS Data Lakes on Lenovo ThinkAgile HX. This solution is ideal for cloud migrations, new Cloudera deployments, expanding existing Cloudera clusters to the cloud or hybrid data architectures that span cloud and on-premises clusters.

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.

Lenovo Cloudera Big Data Solutions			
Internal/Direct Attach Storage	ThinkSystem SR650/SR630 SR655/SR635	 SR650/SR655	 SR630/SR635
Disaggregated Storage	ThinkSystem SD530	 SD530	
Cloud Model Deployment	ThinkAgile HX	 ThinkAgile HX	

Complementary Offerings

Connecting the servers to the storage 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.



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.

When Cloudera is implemented on ThinkAgile HX Series, Lenovo Services guarantees a superior service experience. Providing outstanding value while supporting your uptime requirements, we offer choices to match your workload requirements ranging from base warranty extensions to same-day committed repair, as well as hard drive retention, installation and customized service options. Investing in Lenovo Services guarantees genuine Lenovo quality parts, as well as reliable and consistent service from highly skilled, trained and certified technicians, with access to our global remote and field support teams.

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 Citrix Virtual Apps and Desktops, contact your Lenovo sales representative or Business Partner or visit: www.lenovo.com/systems/solutions

Other Resources:

Reference Architectures -

[Lenovo Big Data Validated Design for Cloudera Enterprise](#)

[Lenovo Big Data Validated Design for ThinkSystem SR655 and SR635](#)

[Lenovo Big Data Validated Design for Cloudera Streaming Analytics](#)

[Lenovo Big Data Validated Design for Cloudera on ThinkAgile HX with Cazena SaaS Data Lakes](#)

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



¹ ITIC reliability study, <https://lenovopress.com/ip1117-itic-reliability-study>

Trademarks: Lenovo, the Lenovo logo, Lenovo Services, ThinkAgile, ThinkSystem, and XClarity® are trademarks or registered trademarks of Lenovo. Azure® is a trademark of Microsoft Corporation in the United States, other countries, or both. Other company, product, or service names may be trademarks or service marks of others.

BDGCL02SB94