

Lenovo Big Data Validated Design

for Cloudera Enterprise

Lenovo™

Easy to implement hardware, software and services for big data analytics architecture

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 with the massive map-reduce scale-out capabilities of Hadoop MapReduce and the faster in-memory processing capability unleashed with Apache Spark.

This document describes the reference architecture for Cloudera Enterprise on bare-metal and on VMware vSphere with locally attached storage. It provides a predefined and optimized hardware infrastructure for the Cloudera Enterprise, a distribution of Apache Hadoop and Apache Spark with enterprise-ready capabilities from Cloudera. This reference architecture provides the planning, design considerations, and best practices for implementing Cloudera Enterprise with Lenovo products.

Starting with a preconfigured hardware platform that is Cloudera-certified helps your team to be up and running analytics applications quickly with the following key big data benefits:

- Gain new business insights with an engineered platform to store, manage and process data at scale built upon the foundation of a reliable and flexible data management architecture
- Enjoy a significant reduction in complexity and an increase in value and performance from optimized Cloudera Enterprise clusters of varying sizes on Lenovo's infrastructure

Cloudera allows organizations to run large-scale, distributed analytics on clusters of cost-effective Lenovo data center hardware. This solution enables analysis of large data sets easily and quickly through a massively parallel processing environment. Deploying on this Lenovo validated design yields a big data solution with exceptional performance, reliability, scalability and flexibility. This architecture supports entry size through high-end configurations and the ability to easily scale as an enterprise use of big data grows. A choice of infrastructure components provides the flexibility to meet a broad range of big data analytics requirements.

To enable faster time to deployment this solution may be shipped as a pre-integrated system with all servers, switches, and storage equipment installed in a rack that is wired with networking and power cables. Additionally, Lenovo offers a single support contact for this integrated hardware configuration, enabling customers to avoid call routing and delayed support.

Highlights

- Deploy with confidence using configurations that are predefined and pretested
- A Cloudera certified solution on CDH5.12 utilizing Lenovo ThinkSystem servers & networking equipment
- Provides VMware virtualization option for up to 4 VMs per server; Avoids any data locality issues by employing Hadoop virtualization extensions
- Features Cloudera Impala enabling administrators to run SQL analytics within a Big Data environment

Lenovo ThinkSystem SR650 & ThinkSystem SR630 servers for a high capacity, high performance and cost effective Big Data solution



Lenovo ThinkSystem SR650 server, ThinkSystem SR630 server and ThinkSystem G8272 networking switch

Support for Apache Hadoop and Apache 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 offers support for both batch and streaming analytics. Take advantage of Apache Spark's distributed in-memory storage for high performance processing across a variety of use cases including batch processing, real-time streaming and advanced modeling and analytics.

Configuration Details

The Lenovo ThinkSystem SR650 and SR630 servers serve as the data nodes and management nodes for this solution, respectively. These 2-socket high performance servers are energy efficient and enable the deployment of large cost effective big data clusters. Both servers feature the latest Intel 6130 processors and support industry leading high IOP flash storage options. The ThinkSystem SR650 high capacity server utilizes on-board storage, eliminating the need for SAN attached storage to support robust Hadoop and Spark clusters.

This big data solution also features the Lenovo ThinkSystem G8272 10Gb and G7028 1Gb Ethernet switches to support the data network and management network, respectively.

Why Lenovo

Lenovo is a leading provider of x86 servers for the data center. Featuring rack, tower, blade, dense and converged systems, the Lenovo server portfolio provides excellent performance, reliability and security. Lenovo also offers a full range of networking, storage, software, solutions, and comprehensive services supporting business needs throughout the IT lifecycle. With options for planning, deployment, and support, Lenovo offers expertise and services needed to deliver better service-level agreements and generate greater end-user satisfaction.

For More Information

To learn more about the Lenovo Big Data Validated Design for Cloudera Enterprise, contact your Lenovo Business Partner or visit: www.lenovo.com/systems/solutions

© 2017 Lenovo. All rights reserved.

Availability: Offers, prices, specifications and availability may change without notice. Lenovo is not responsible for photographic or typographical errors. **Warranty:** For a copy of applicable warranties, write to: Lenovo Warranty Information, 1009 Think Place, Morrisville, NC, 27560, Lenovo makes no representation or warranty regarding third party products or services. **Trademarks:** Lenovo, the Lenovo logo, System x, ThinkServer are trademarks or registered trademarks of Lenovo. Microsoft and Windows are registered trademarks of Microsoft Corporation. Intel, the Intel logo, Xeon and Xeon Inside are registered trademarks of Intel Corporation in the U.S. and other countries. Other company, product, and service names may be trademarks or service marks of others.

CRN: BGDCL01XX72

11/2017