

# Lenovo Big Data Reference Architecture for Cloudera Distribution for Hadoop

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



## Architecture Highlights

- Provides guidance for deploying Lenovo systems with a Cloudera distribution for Hadoop to coordinate the processing of the data across a massively parallel environment.
- Includes the latest data center equipment available such as the Lenovo x3650 M5 and x3550 M5 and Lenovo RackSwitch Ethernet switches and Lenovo XClarity.
- Supports entry through high-end configurations and the ability to easily scale as the use of big data grows

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. Today, Cloudera uses the latest big data technologies such as the massive map-reduce scale-out capabilities of Hadoop to open the door to a world of possibilities.

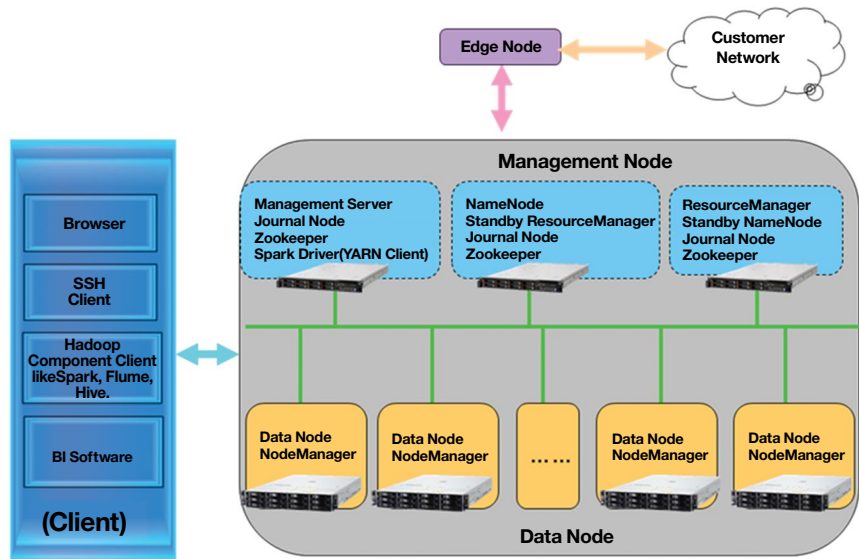
This Lenovo Big Data RA for Cloudera Distribution for Hadoop is certified by Cloudera and provides a thoroughly tested and integrated solution that combines the benefits of leading-edge technologies with mature, enterprise-ready features. Starting with a preconfigured hardware platform that is Cloudera-certified helps your team to be up and running analytics quickly.

Cloudera allows organizations to run large-scale, distributed analytics jobs on clusters of cost-effective server hardware. This infrastructure can be leveraged to tackle very large data sets by breaking up the data into “chunks” and coordinating the processing of the data across a massively parallel environment.

Cloudera deployed on this Lenovo reference architecture provides superior performance, reliability, and scalability. This architecture supports entry through high-end configurations and the ability to easily scale as the use of big data grows. A choice of infrastructure components provides the flexibility to meet broad range of big data analytics requirements

**Why Lenovo and Cloudera**

Lenovo and Cloudera have collaborated to deliver successful, integrated solutions to customers worldwide. These customers recognize Lenovo’s leadership in the worldwide server market,<sup>1,2</sup> and Cloudera’s leadership in Apache Hadoop as one of the largest contributors to this open source ecosystem.



Sample design of a Big Data environment using Cloudera

The Lenovo M5 servers, like the powerful two-socket x3650 M5 BD and x3550 M5, enhance performance and reduce power consumption of big data clusters. Purpose built for big data workloads, the 2U two-socket x3650 M5 BD server supports industry leading data storage capacity, the latest Intel Xeon E5-2600 v2 high performance compute processing power, flash storage options, and energy efficient features. The core Cloudera reference architecture leverages this server as a data node for scale-out clusters.

The versatile, two-socket 1U x3550 M5 rack server delivers outstanding performance as a big data management node in the Cloudera reference architecture. This compact, easy-to-use server features a pay-as-you-grow design to help lower costs and manage risks.

The data network is a private cluster data interconnect among nodes used for data access, moving data across nodes within a cluster, and ingesting data into the Cloudera cluster. While a 1 Gb Ethernet switch is sufficient for some workloads, a 10 Gb Ethernet switch can provide extra I/O bandwidth for added performance. The recommended 10GbE switch is the Lenovo RackSwitch G8272.

Regarding Storage, each server node in the reference architecture has an internal, directly attached storage. External storage is not utilized in this reference architecture. In situations where higher storage capacity is required, the design approach followed in this reference architecture is to increase the amount of data disk space per node or increase the number of nodes in the cluster.

# ARCHITECTURE BRIEF

## BIG DATA

For system management, this reference architecture includes one platform for hardware management and another for cluster system management. The mechanism for cluster systems management is done via Cloudera Manager and adapted from the standard Hadoop distribution which places the management services on separate servers. Because the Management Node hosts important and high-memory functions, it is important that it is a powerful and fast server.

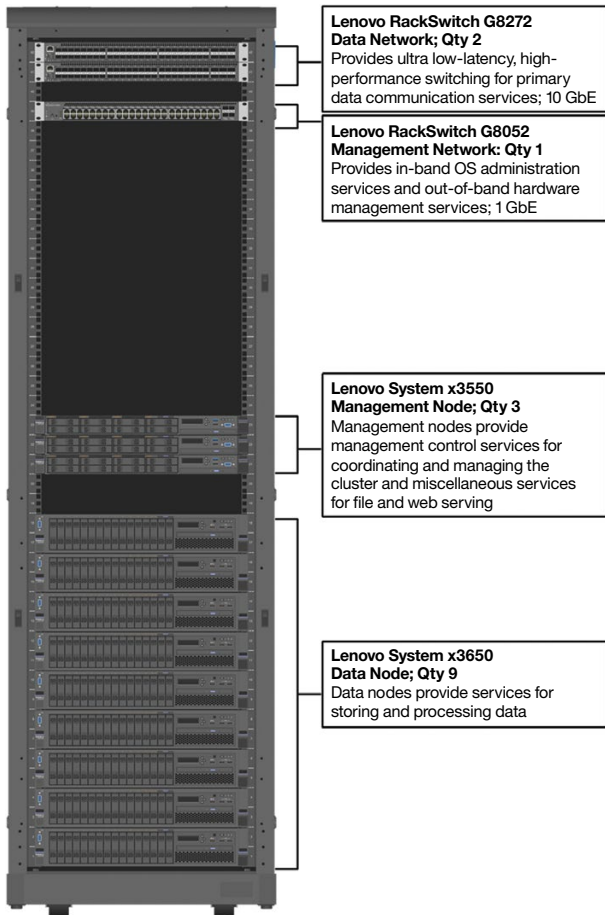
Hardware management is addressed with the Lenovo XClarity Administrator. Lenovo XClarity is an agentless centralized resource management solution that is aimed at reducing complexity, speeding response, and enhancing the availability of Lenovo server systems. The solution seamlessly integrates into Lenovo M5 rack servers. Through an uncluttered, dashboard-driven GUI, XClarity provides automated discovery, monitoring, firmware updates, pattern-based configuration management, hypervisor operating system deployments. Lenovo XClarity also features extensive REST APIs that provide deep visibility and control via higher-level cloud orchestration and service management software tools.

The example below shows the advanced configuration for this reference architecture. This architecture also offers Enhanced and Advanced configurations. Configurations may also be customized to best suit the workloads running in your environment. To accelerate time to value, Lenovo has service offerings and expertise to implement this reference architecture and to accommodate customization.

The following is an abbreviated Bill of Materials for the standard configuration:

### Assemblies and Components

- 2 x Data Networking Switches
  - 1 x Lenovo RackSwitch G8272 Network Switch
- 1 x Management Networking Switches
  - 1 x Lenovo RackSwitch G8052 Networking Switch
- 3 x Management Node Servers
  - 1 x Lenovo System x3650 M5 Rack Server
  - 2 x E5-2630 v3 (10C, 2.30 GHz)
  - 8 x 16 GB of system memory
  - 3 x 1 TB NL SATA 3.5 inch
- 9 x Data Node Servers
  - 1 x Lenovo System x3650 M5 Rack Servers
  - 2 x E5-2630 v3 (10C, 2.30 GHz)
  - 8 x 16 GB of system memory
  - 2 x 1.2 TB NL SATA 3.5 inch, OS Disks
- 14 x 4 TB NL SATA 3.5 inch, Data Disks



Sample Configuration: Lenovo Big Data Reference Architecture for Cloudera Distribution for Hadoop

## Why Lenovo

Lenovo is the leading provider of x86 systems for the data center. The portfolio includes rack, tower, blade, dense and converged systems, and supports enterprise class performance, reliability and security. Lenovo also offers a full range of networking, storage, software and solutions, and comprehensive services supporting business needs throughout the IT lifecycle.

## For more information

To learn more about the Big Data RA for Cloudera Distribution for Hadoop, contact your Lenovo Business Partner or visit [www.lenovo.com/systems/solutions](http://www.lenovo.com/systems/solutions)

<sup>1</sup> #1 x86 Reliability; ITIC 2014-2015 Global Server Hardware, Server OS Reliability Report, May 2014 (latest report); <http://www.lenovo.com/images/products/system-x/pdfs/analyst-reports/XSL03126USEN.PDF>. (2015 survey results pending).

<sup>2</sup> #1 Customer Satisfaction; TBR Customer Satisfaction Survey, May 2015 (latest report); [http://www.lenovo.com/images/products/system-x/pdfs/white-papers/tbr\\_x86servers\\_top\\_csat\\_1q15\\_wp.pdf](http://www.lenovo.com/images/products/system-x/pdfs/white-papers/tbr_x86servers_top_csat_1q15_wp.pdf)



© 2015 Lenovo. All rights reserved.

**Availability:** Offers, prices, specifications and availability may change without notice. Lenovo is not responsible for photographic or typographic errors. **Warranty:** For a copy of applicable warranties, write to: Warranty Information, 500 Park Offices Drive, RTP, NC, 27709, Attn: Dept. ZPYA/B600. 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. Visit [www.lenovo.com/lenovo/us/en/safecomp.html](http://www.lenovo.com/lenovo/us/en/safecomp.html) periodically for the latest information on safe and effective computing.