

# Lenovo Big Data Validated Design

For Cloudera Enterprise with VMware

Lenovo™

**Achieve greater efficiency and flexibility with a virtualized big data cluster**

## 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 faster in-memory analytics using Apache Spark and the massive scale-out capabilities possible with Apache Hadoop. Lenovo big data solutions incorporating Cloudera Enterprise software enable organizations to run batch and streaming analytics on large distributed clusters of cost-effective scale out server and networking hardware from Lenovo.

Deployed in a VMware vSphere virtualized server environment, the Lenovo Big Data Validated Design for Cloudera Enterprise with VMware delivers a certified solution for both Cloudera Hadoop and Spark environments. Employing standard VMware ESXi VMs, this virtualized solution helps organizations derive insights from both structured and unstructured data.

Starting with a preconfigured hardware platform that is Cloudera-certified, this solution helps your team be up and running quickly with the latest analytics applications. 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 built upon the Lenovo x3650 M5 server and Lenovo x3550 M5 server.

## Virtualizing Big Data Cluster Facilitates Greater Dynamic Scalability

This Lenovo Big Data Validated Design for Cloudera Enterprise with VMware provides a thoroughly tested and integrated solution which combines the benefits of leading-edge technologies with mature, enterprise-ready features. Combining VMware vCenter with Cloudera Enterprise and Lenovo servers, networking equipment and services, Lenovo can deliver a virtualized big data cluster that enables businesses to:

- Achieve greater efficiency and flexibility for managing a big data cluster in a virtualized environment
- Derive business insights from data with an engineered platform to store, manage and process data at scale
- Deploy a Cloudera certified solution that is designed to suit various capacity and scalability needs
- Harness solution elasticity enabled by the ability to grow or shrink the cluster based on workload demands.

### 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

## Lenovo data center products for high capacity, high performance virtualized platforms



Lenovo System X3650 M5 Server,  
Lenovo System X3550 M5 Server and  
Lenovo G8272 10Gb Ethernet Switch



## Architectural Overview

At the core of this architecture reside the Lenovo System x3650 M5 rack server and System x3550 M5 rack server. These rack based servers deliver the performance and stability required for business-critical big data workloads. Each of these servers are extremely powerful, flexible and support large DDR4 memory arrays.

This Lenovo solution employs several nodal configurations for the System x3650 M5 and System x3550 M5 servers. The various nodal options support a wide variety of VM quantities and performance levels. By having a number of options to choose from allows administrators to select the configuration most appropriate for their sizing and performance requirements. These servers host virtualized Cloudera Enterprise data nodes and management nodes. But in this solution these nodes are deployed on VMs rather than the usual bare metal platforms employed in big data clusters. Deploying a Cloudera cluster on VMware vSphere enables this solution to achieve greater provisioning and operational flexibility than typically achieved by deploying a Cloudera cluster on bare metal.

This big data solution also features the Lenovo G8272 10Gb Ethernet switch to support the data network. It is an enterprise class Layer 2 and Layer 3 full featured switch that delivers line-rate, high bandwidth switching, filtering, and traffic queuing without impeding data traffic.

## 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 with VMware, contact your Lenovo Business Partner or visit:

<http://www3.lenovo.com/us/en/data-center/solutions/c/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: BGDCL02XX72

06/2017