Lenovo Big Data Reference Architecture for MapR Distribution including Apache Hadoop

Easy to implement hardware, software and services for analyzing big data

Architecture Highlights

- Provides guidance for deploying Lenovo systems with a MapR 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 projects are not all alike. What works for one company may not work for another. Instead of re-inventing the wheel to apply business analytics to big data, it makes business sense to take advantage of what has already been proven.

Lenovo Big Data Reference Architecture for MapR Distribution including Apache Hadoop leverages the MapR container architecture to store metadata and provide a reliable service distributed across the entire cluster. It uses Hadoop's MapReduce framework to tackle very large data sets spread across many nodes.

This Lenovo reference architecture is certified by MapR 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 helps your team to be up and running quickly.

MapR deployed on Lenovo servers and networking components provides a superior performance, reliability and scalability. The reference 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 flexibility in meeting varying big data analytics requirements.
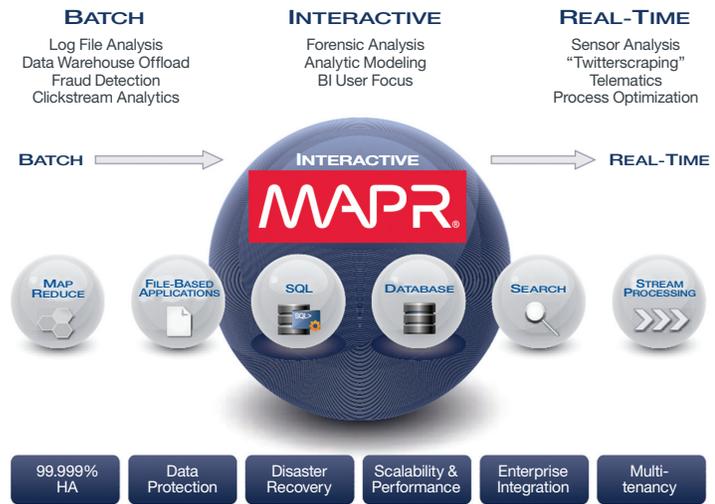
The Lenovo System x M5 servers, like the powerful two-socket x3650 M5, enhance performance and reduce power consumption of big data clusters. Purpose built for big data workloads, the 2U two-socket x3650 M5 BD server supports industry leading data storage capacity, the latest Intel Xeon E5-2600 v2 high performance compute processing power, flash storage options, and energy efficient features. The core MapR 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 MapR 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 MapR 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.

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 MapR management services and adapted from the standard Hadoop distribution which places the management services on separate servers. The hardware management platform is addressed with the Lenovo XClarity Administrator.
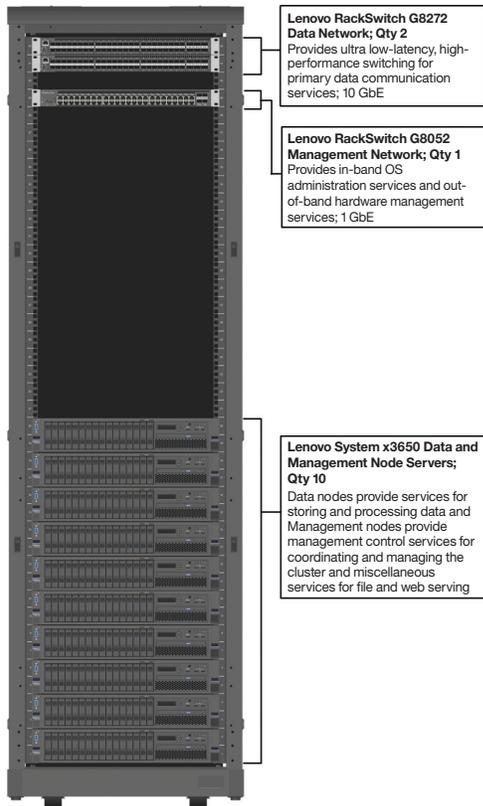
Lenovo XClarity is an agentless centralized resource management solution that that is aimed at reducing complexity, speeding response, and enhancing the availability of Lenovo server systems. The solution seamlessly integrates into Lenovo M5 rack servers. Through an uncluttered, dashboard-driven GUI, 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 medium configuration for this reference architecture. This architecture also offers small, large and very large 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 Networking Switch

1 x Management Networking Switch
  1 x Lenovo RackSwitch G8052 Networking Switch

10 x Data & 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
  2 x 1.2 TB NL SATA 3.5 inch, OS Disks
  14 x 4 TB NL SATA 3.5 inch, Data Disks

Lenovo RackSwitch G8272 Data Network; Qty 2
Provides ultra low-latency, high-performance switching for primary data communication services; 10 GbE

Lenovo RackSwitch G8052 Management Network; Qty 1
Provides in-band OS administration services and out-of-band hardware management services; 1 GbE

Lenovo System x3650 Data and Management Node Servers; Qty 10
Data nodes provide services for storing and processing data and Management nodes provide management control services for coordinating and managing the cluster and miscellaneous services for file and web serving

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 Lenovo Big Data Reference Architecture for MapR Distribution including Apache Hadoop, contact your Lenovo Business Partner or visit **lenovo.com**/systems/solutions.

[1] #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).

[2] #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

LYS03179-USEN-00