# Lenovo Big Data Validated Design for the MapR Converged Data Platform with Containers

Configuration reference number: BGDMR01XX81

**MapR converges operational and analytical workloads in a single platform**

**MapR brings unprecedented dependability, ease-of-use and world-record speed to Big Data**

**Uses powerful, versatile new Lenovo® ThinkSystem SR650 server**

**Leading security and reliability Innovative energy efficiency and performance**

**Dan Kangas (Lenovo)**

**Weixu Yang (Lenovo)**

**Ajay Dholakia (Lenovo)**

**James Sun (MapR)**

LENOVO PRESS

# Table of Contents

# 1 Introduction

This document describes the reference architecture for the Big Data Solution based on the MapR Converged Data Platform. It provides a predefined and optimized hardware infrastructure for the MapR Converged Enterprise Edition, the commercial edition of the MapR Converged Data Platform that supports enterprise-grade features for business-critical production deployments. This reference architecture provides the planning, design considerations, and best practices for implementing the MapR Converged Data Platform with Lenovo products.

The Lenovo and MapR teams worked together on this document, and the reference architecture that is described herein was validated by both Lenovo and MapR.

MapR brings the power of data analytics to the enterprise. MapR platform services are the core data handling capabilities of the MapR Converged Data Platform and consist of MapR-FS, MapR-DB, and MapR Data Science Refinery  The enterprise-friendly design provides a familiar set of file and data management services, including a global namespace, high availability (HA), data protection, self-healing clusters, access control, real-time performance, secure multi-tenancy, and management and monitoring.

On top of the platform services, MapR packages a broad set of Apache™ Hadoop® open source ecosystem projects that enable big data applications. The goal is to provide an open platform that lets you choose the right tool for the job. MapR tests and integrates open source ecosystem projects such as Apache Drill, Apache Hive™, Apache Pig, Apache HBase™ and Apache Mahout™, among others.

The predefined configuration provides a baseline configuration for a big data solution which can be modified based on the specific customer requirements, such as lower cost, improved performance, and increased reliability.

The intended audience of this document is IT professionals, technical architects, sales engineers, and consultants to assist in planning, designing and implementing the big data solution with Lenovo ThinkSystem hardware. It is assumed that you are familiar with Apache Hadoop components and capabilities.

# 2 Business problem and business value

This section describes the business problem that is associated with big data environments and the value that is offered by the MapR solution that uses Lenovo hardware.

## 2.1 Business problem

Data sets grow rapidly - in part because they are increasingly gathered by low cost and numerous information-sensing mobile devices, aerial devices (remote sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers and wireless sensor networks. The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s. In 2015 data consumed in the world was estimated to be an astronomical 8 zettabytes (1 zettabyte - 1 trillion terabytes) but by 2020 it's expected to have increased significantly again to be 40 zettabytes (ZB).

Big data spans the following dimensions:

- **Volume:** Big data comes in one size: enormous! Enterprises are awash with data, easily amassing terabytes to petabytes of information. For example, human generated data through social media is growing at a 10x rate and machine generated is growing at a 100x rate.
- **Velocity:** Often time-sensitive, big data must be used as it is streaming into the enterprise to maximize its value to the business. Data ingestion comprises both batch and real-time loading methods.
- **Variety:** Big data extends beyond structured data, including unstructured data of all varieties, such as text, audio, video, click streams, and log files.

Big data is more than a challenge; it is an opportunity to find insight into new and emerging types of data to make your business more agile. Big data also is an opportunity to answer questions that, in the past, were beyond reach. Until now, there was no effective way to harvest this opportunity.

## 2.2 Business value

MapR provides the industry's only converged data platform that uniquely allows applying analytical insights to operational processes in real time to create competitive advantage for our customers. MapR is a data platform that converges historically separate product segments/categories in order to enable extraordinary new value never before possible. The MapR Converged Data Platform ("MapR Platform" or "MapR") is powered by the industry's fastest, most reliable, secure, and open data infrastructure that dramatically lowers TCO and enables global real-time data applications.

The key benefits of MapR are to ensure customers' production success for Hadoop, Apache Spark™ and much more. MapR was engineered for the data center with IT operations in mind. MapR serves business-critical needs that cannot afford to lose data, must run on a 24x7 basis, require immediate recovery from node and site failures – all with a smaller data center footprint. MapR supports these capabilities for the broadest set of data applications from batch analytics to interactive querying and real-time streaming.

MapR deployed on Lenovo ThinkSystem servers with Lenovo networking components provides superior performance, reliability, and scalability. This 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.

# 3 Requirements

The functional and non-functional requirements for the MapR reference architecture are described in this section.

## 3.1 Functional requirements

A big data solution supports the following key functional requirements:

- Various application types, including batch and real-time analytics
- Industry-standard interfaces, so that existing applications can work
- Real-time streaming and processing of data
- Various data types and databases
- Various client interfaces
- Large volumes of data

## 3.2 Non-functional requirements

Customers require their big data solution to deliver business value without significant overhead. The following non-functional requirements are key:

- Simplified:
  - Ease of development
  - Easy management at scale
  - Advanced job management
  - Multi-tenancy for users, data, and applications
- Reliable:
  - Data protection with snapshot and mirroring
  - Automated self-healing
  - Insight into software/hardware health and issues
  - Mission-critical high availability (HA) and business continuity (99.999% uptime)
- Fast and scalable:
  - Real-time streaming capabilities
  - Superior NoSQL database agility
  - Web-scale file system performance
  - Linear scale-out on a distributed architecture
- Secure:
  - Strong authentication and authorization
  - Kerberos and broad user registry support
  - Data confidentiality and integrity
  - Comprehensive data security auditing

# 4 Architectural overview

This reference architecture provides validation for MapR 6.0 running Persistent Application Client Containers (PACC), MapR-DB, MapR-ES, and MapR Data Science Refinery on Lenovo ThinkSystem hardware platform. The MapR 6.0 Converged Data Platform integrates Hadoop, Spark, and Apache Drill with real-time database capabilities, global event streaming, and scalable enterprise storage to power a new generation of big data applications. The MapR Platform delivers enterprise grade security, reliability, and real-time performance while dramatically lowering both hardware and operational costs of your most important applications and data. This MapR Converged Data Platform solution is shown below.

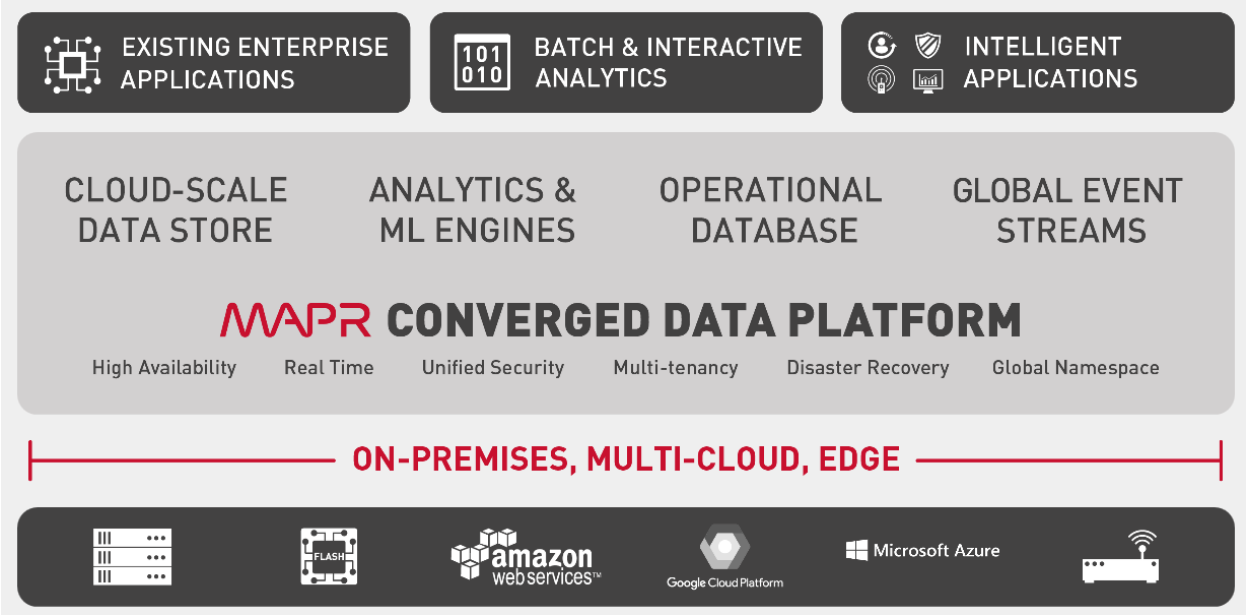



***Figure 1: The MapR Converged Data Platform***

## 4.1 MapR Persistent Application Client Containers (PACC)

MapR Persistent Application Client Containers (PACC) support containerization of existing and new applications by providing containers with persistent data access from anywhere. Typical containers have no scalable, out-of-box way to let applications persist their state on the cluster's distributed file system. All data written by containerized applications is typically lost when there is an application or hardware failure. Attempting to store data on local nodes, for example, forces IT departments to track down the data and move files when containers are redeployed. This defeats the ease-of-use benefits of containers.

With the introduction of MapR PACC, containerized applications can easily leverage all the MapR platform services including MapR-FS file system, MapR-DB data base, and Map-ES event streams as a persistent data store. MapR provides the pre-built Docker image for connecting the containerized application to all converged data platform services, including MapR-FS, MapR-DB, and MapR-ES. MapR PACC applications are flexible, can be shared and deployed across private MapR clusters as well as MapR nodes in the cloud.

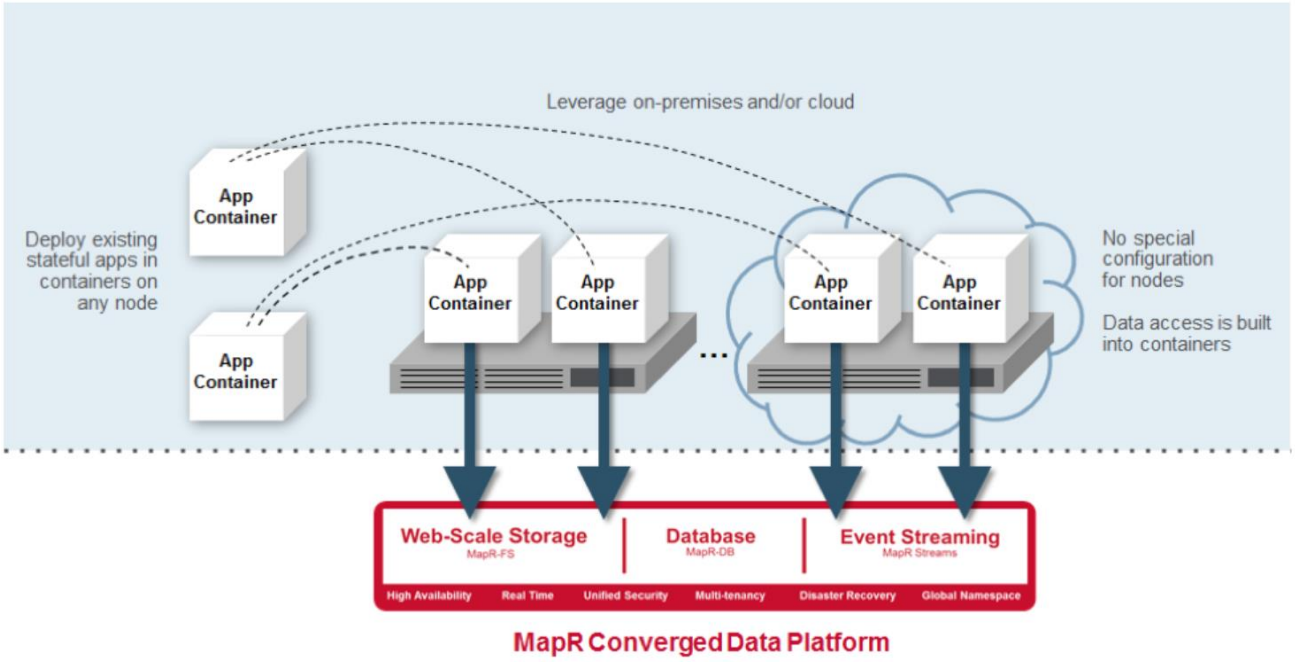



***Figure 2: MapR Persistent Application Client Containers (PACC)***

## 4.2 MapR-DB Database

MapR-DB is a high performance NoSQL (“Not Only SQL”) database management system built into the MapR Converged Data Platform. It is a highly scalable multi-model database that brings together operations and analytics, and real-time streaming and database workloads to enable a broader set of next-generation data-intensive applications in organizations.

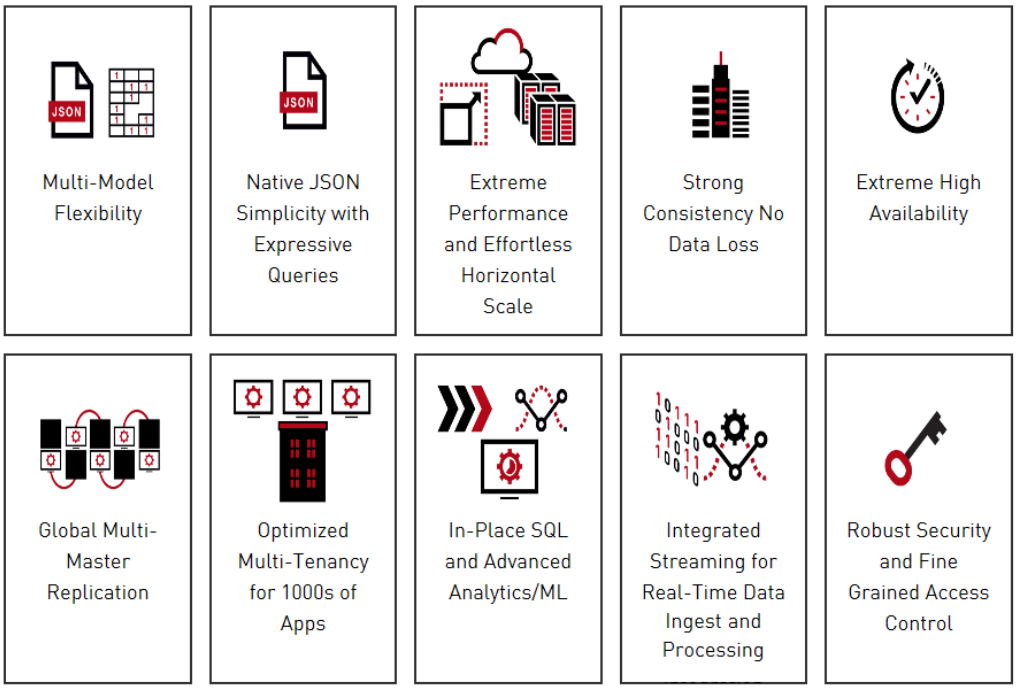



***Figure 3: MapR-DB High Performance NoSQL Database***

## 4.3 MapR-ES Event Stream

MapR-ES is the first big data-scale streaming system built into a converged data platform, and the only big data streaming system to support global event replication reliably at IoT scale. The MapR Converged Data Platform allows you to quickly and easily build breakthrough, reliable, real-time applications by providing:

- **Single cluster** for streams, file storage, database, and analytics.
- **Persistence** of streaming data, providing direct data access to batch and interactive frameworks, eliminating data movement.
- **Unified security** framework for data-in-motion and data-at-rest, with authentication, authorization, and encryption.
- **Utility-grade reliability** with self-healing and no single point-of-failure architecture.

## 4.4 MapR Data Science Refinery

**The** MapR Data Science Refinery is an easy-to-deploy and scalable data science offering with native platform access and superior out-of-the-box security. It allows for agile, containerized solutions that can scale to fit the needs of all types of data science teams. Within the MapR platform, Data Science Refinery offers support for popular open source tooling in a pre-configured offering that can be distributed to many data science teams across a multi-tenant environment.

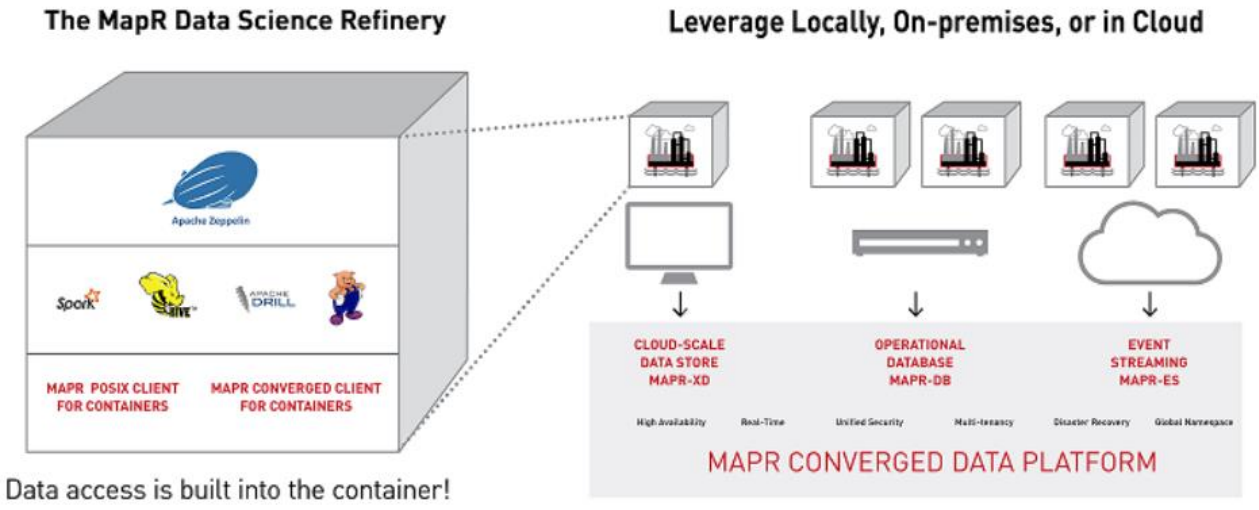



***Figure 4: MapR Data Science Refinery***

The MapR Data Science Refinery offers:

- **Access to All Platform Assets -** The MapR FUSE-based POSIX Client allows app servers, web servers, and other client nodes and apps to read and write data directly and securely to a MapR cluster like a Linux filesystem. In addition, connectors are provided for interacting with both MapR-DB & MapR-ES via Apache Spark connectors.

- **Superior Security -** MapR platform is secure-by-default, and Apache Zeppelin on MapR leverages and integrates with this security layer using the built-in capabilities provided by the MapR Persistent Application Container (PACC).
- **Extensibility -** Apache Zeppelin is paired with the Helium framework to offer pluggable visualization capabilities.
- **Simplified Deployment -** A pre-configured Docker container provides the ability to leverage MapR as a persistent data store. The Dockerfile is also available, allowing users to customize the image as needed to support specific application needs.

# 5 Component model

MapR supports dozens of open source projects and is committed to using industry-standard APIs to provide a frictionless method of developing and deploying exciting new applications that can meet the most stringent production runtime requirements.

## 5.1 Enterprise-Grade Platform Services

MapR platform services are the core data handling capabilities of the MapR Platform and consist of MapR-FS, MapR-DB, and MapR-ES. Its enterprise friendly design provides a familiar set of file and data management services, including a global namespace, high availability (HA), data protection, self-healing clusters, access control, real-time performance, secure multi-tenancy, and management and monitoring.

A few of the key attributes of the MapR converged platform are shown below:

- Linux operating systems
  Red Hat Linux is supported by this Lenovo ThinkSystem reference architecture.
- MapReduce
  MapR provides high performance for MapReduce operations on Hadoop and publishes performance benchmarking results. The MapR architecture is built in C/C++ and harnesses distributed metadata with an optimized shuffle process, enabling MapR to deliver consistent high performance. Both the classic MapReduce and YARN frameworks are supported by MapR.
- File-based applications
  MapR is a Portable Operating System Interface (POSIX) system that fully supports random read-write operations. By supporting industry-standard Network File System (NFS), users can mount a MapR cluster and run any file-based application, written in any language, directly on the data residing in the cluster. All standard tools in the enterprise including browsers, UNIX tools, spreadsheets, and scripts can access the cluster directly without any modifications.
- NoSQL Database
  MapR has removed the trade-offs that organizations face when looking to deploy a NoSQL solution. Specifically, MapR-DB delivers ease of use, reliability, and performance advantages for NoSQL applications. It provides native multi-model support for JSON document and wide column data models. MapR-DB also provides scalability, strong consistency, and continuous low latency with an architecture that eliminates compaction delays or background consistency corrections (“anti-entropy”).
- Stream processing
  MapR provides a simplified architecture for real-time stream computational engines such as Spark Streaming, Apache Flink®, and Apache Storm™. Streaming data feeds can be written directly to the MapR Platform for long-term storage and MapReduce processing.

## 5.2 Open Source Engines and Tools

MapR packages a broad set of Apache open source ecosystem projects as well as open source Kubernetes that enable big data applications. The goal is to provide an open platform that lets you choose the right tool for the job. MapR tests and integrates open source ecosystem projects such as Apache Drill, Apache Hive™,

Apache Pig, Apache HBase™ and Apache Mahout™, among others.

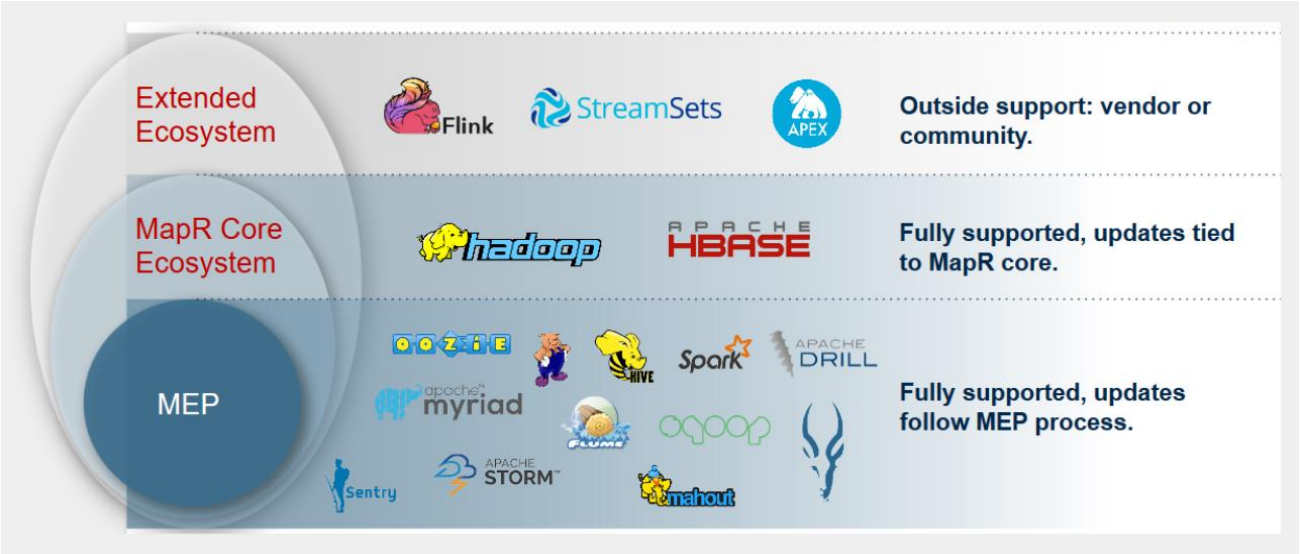



***Figure 5: MapR key components for open source engines and tools***

- Batch Processing
  Apache Spark provides fast and general engine for large-scale data processing. Apache Pig is a language and runtime for analyzing large data sets, consisting of a high-level. MapReduce/Yarn is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. Machine Learning
  Popular machine learning libraries are available in the MapR Platform which includes Apache Mahout and Spark MLlib.

- Machine Learning
  Popular machine learning libraries are available in the MapR Platform which includes Apache Mahout and Spark MLlib.

- SQL on Hadoop
  There are a number of applications that support SQL access against data contained in MapR. MapR is also leading the development of Apache Drill that brings American National Standards Institute (ANSI) SQL capabilities to Hadoop. Drill enables data analysts to perform self-exploratory tasks on any type of data stored in Hadoop, including JSON and Parquet files. Apache Hive is a data warehouse system for Hadoop that facilitates easy data summarization, ad hoc queries, and the analysis of large data sets stored in Hadoop compatible file systems. Spark SQL offers real-time in-memory processing to gain insights to your data.

- NoSQL
  Apache HBase is supported in addition to the core MapR-DB (HBase API compatible) to handle NoSQL workloads. The integrated NoSQL database, MapR-DB, is built on the core MapR Data Platform which set records on both the TeraSort and the MinuteSort benchmarks. Recently, MapR-DB ran over 30,000 batch output operations per second per node, and showed as much as an eleven-fold speed improvement over HBase. With its in-memory feature, MapR-DB can store a database in memory for additional performance gains.

- Streaming
  Spark Streaming support is available in addition to the core MapR-ES (Kafka API compatible) to handle event streaming for Internet of Things (IoT) data streaming on the edge devices.

- Data Integration and Access
  Hue, HttpFS, Sqoop, and Flume are all supported for efficiently ingesting and accessing data in a cluster.

- Apache Flume – A distributed, reliable, and highly available service for efficiently ingesting large amounts of data into a cluster.

- Sentry - Sentry is supported for enforcing fine grained role-based authorization to data and metadata stored on a Hadoop cluster.

- Workflow and Data Governance – Oozie is the workflow coordination manager which is supported with MapR.

- Provisioning and Coordination
  Apache ZooKeeper is distributed service for maintaining configuration information, providing distributed synchronization, and providing group service.

Other open source projects included with MapR:

- Kubernetes - The MapR Data Fabric includes a natively integrated Kubernetes volume driver to provide persistent storage volumes for access to any data located on-premises, across clouds, and to the edge. Stateful applications are easily deployed in containers for production use cases, machine learning pipelines, and multi-tenant use cases. By combining the power of the PACC and the Data Fabric for Kubernetes, MapR customers can scale storage and compute resource independently. Their custom applications can easily share the compute and storage resources effortlessly.

## 5.3 Commercial Engines and Applications

One of the key developer benefits of the MapR Platform is its basis on well-known, open APIs and interfaces. This enables commercial software vendors such as SAP, SAS, HP, and Cisco to easily deploy large scale applications onto the MapR Platform. It also means that even small teams of developers can create enterprise-grade software products by exploiting the built-in protections of the MapR Platform as well as mature commercial processing engines.

Lastly, details about the various MapR and Linux version operability can be found at:
http://maprdocs.mapr.com/home/InteropMatrix/r_os_matrix.html?hl=os,matrix

# 6 Operational model

This section describes the operational model for the MapR reference architecture.

To show the operational model for different sized customer environments, this reference architecture describes four different models for supporting different amounts of data. Throughout the document, these models are referred to as starter rack, half rack, full rack, and multi-rack configuration sizes. The full bill of material for building these models is available as a predefined configuration which can be used for ordering purposes or as a starting point for customizing the configuration.

A MapR deployment consists of cluster nodes, networking, power, and racks. The predefined configurations can be implemented as-is or modified based on specific customer requirements, such as lower cost, higher CPU performance, increased storage, or increased reliability. Key workload requirements such as the data growth rate, sizes of datasets, and data ingest patterns help in determining the proper configuration for a specific deployment. A best practice when a MapR cluster infrastructure is designed is to conduct the proof of concept testing by using representative data and workloads to ensure that the proposed design works.

## 6.1 Hardware description

This reference architecture uses Lenovo ThinkSystem SR650 (2U) servers, Lenovo RackSwitch G8052 and G8272 top of rack switches, and ThinkSystem NE10032 cross-rack switch. MapR clusters do not require unique 1U master node servers.

### 6.1.1 Lenovo ThinkSystem SR650

The Lenovo ThinkSystem SR650 is an ideal 2-socket 2U rack server for small businesses up to large enterprises that need industry-leading reliability, management, and security, as well as maximizing performance and flexibility for future growth. The SR650 server is particularly suited for big data applications due to its rich internal data storage, large internal memory and selection of high performance Intel processors. It is also designed to handle general workloads, such as databases, virtualization and cloud computing, virtual desktop infrastructure (VDI), enterprise applications, collaboration/email, and business analytics.

The SR650 server supports:

- Up to two Intel® Xeon® Scalable Processors
- Up to 3.0 TB 2666 MHz TruDDR4 memory (certain CPU part numbers required),
- Up to 24x 2.5-inch or 14x 3.5-inch drive bays with an extensive choice of NVMe PCIe SSDs, SAS/SATA SSDs, and SAS/SATA HDDs
- Flexible I/O Network expansion options with the LOM slot, the dedicated storage controller slot, and up to 6x PCIe slots

***Figure 6: Lenovo ThinkSystem SR650***

Combined with the Intel® Xeon® Scalable Processors (Bronze, Silver, Gold, and Platinum), the Lenovo SR650 server offers an even higher density of workloads and performance that lowers the total cost of ownership (TCO). Its pay-as-you-grow flexible design and great expansion capabilities solidify dependability for any kind of workload with minimal downtime.

The SR650 server provides high internal storage density in a 2U form factor with its impressive array of workload-optimized storage configurations. It also offers easy management and saves floor space and power consumption for most demanding use cases by consolidating storage and server into one system.

This reference architecture recommends the storage-rich ThinkSystem SR650 for the following reasons:

- **Storage capacity**: The nodes are storage-rich. Each of the 14 configured 3.5-inch drives has raw capacity up to 12 TB and each, providing for 168 TB of raw storage per node and over 3000 TB per rack.
- **Performance**: This hardware supports the latest Intel® Xeon® Scalable processors and TruDDR4 Memory.
- **Flexibility**: Server hardware uses embedded storage, which results in simple scalability (by adding nodes).
- **PCIe slots**: Up to 7 PCIe slots are available if rear disks are not used, and up to 3 PCIe slots if the Rear HDD kit is used. They can be used for network adapter redundancy and increased network throughput.
- **Higher power efficiency**: Titanium and Platinum redundant power supplies that can deliver 96% (Titanium) or 94% (Platinum) efficiency at 50% load.
- **Reliability:** Outstanding reliability, availability, and serviceability (RAS) improve the business environment and helps save operational costs

For more information, see the Lenovo ThinkSystem SR650 Product Guide:

https://lenovopress.com/lp0644-lenovo-thinksystem-sr650-server

### 6.1.2 Lenovo RackSwitch G8052

The Lenovo System Networking RackSwitch G8052 (as shown in Figure 7) is an Ethernet switch that is designed for the data center and provides a virtualized, cooler, and simpler network solution. The Lenovo RackSwitch G8052 offers up to 48 1 GbE ports and up to 4 10 GbE ports in a 1U footprint. The G8052 switch is always available for business-sensitive traffic by using redundant power supplies, fans, and numerous high-availability features.

***Figure 7: Lenovo RackSwitch G8052***

Lenovo RackSwitch G8052 has the following characteristics:

- Forty-eight 1 GbE RJ45 ports
- Four standard 10 GbE SFP+ ports
- Low 130W power rating and variable speed fans to reduce power consumption

For more information, see the RackSwitch G8052 Product Guide:

https://lenovopress.com/tips1270-lenovo-rackswitch-g8052

### 6.1.3 **Lenovo RackSwitch G8272**

Designed with top performance in mind, Lenovo RackSwitch G8272 is ideal for today's big data, cloud, and optimized workloads. The G8272 switch offers up to 72 10 Gb SFP+ ports in a 1U form factor and is expandable with four 40 Gb QSFP+ ports. It is an enterprise-class and full-featured data center switch that deliver line-rate, high-bandwidth switching, filtering, and traffic queuing without delaying data. Large data center grade buffers keep traffic moving. Redundant power and fans and numerous HA features equip the switches for business-sensitive traffic.

The G8272 switch (as shown in Figure 8) is ideal for latency-sensitive applications, such as client virtualization. It supports Lenovo Virtual Fabric to help clients reduce the number of I/O adapters to a single dual-port 10 Gb adapter, which helps reduce cost and complexity. The G8272 switch supports protocols, including Data Center Bridging/Converged Enhanced Ethernet (DCB/CEE) for support of FCoE and iSCSI and NAS.

***Figure 8: Lenovo RackSwitch G8272***

The enterprise-level Lenovo RackSwitch G8272 has the following characteristics:

- 48 x SFP+ 10GbE ports plus 6 x QSFP+ 40GbE ports
- Support up to 72 x 10Gb connections using break-out cables
- 1.44 Tbps non-blocking throughput with very low latency (~ 600 ns)
- Up to 72 1Gb/10Gb SFP+ ports
- OpenFlow enabled allows for easily created user-controlled virtual networks
- Virtual LAG (vLAG) and LACP for dual switch redundancy

For more information, see the RackSwitch G8272 Product Guide:

https://lenovopress.com/tips1267-lenovo-rackswitch-g8272

### 6.1.4 **Lenovo RackSwitch NE10032 - Cross-Rack Switch**

The Lenovo ThinkSystem NE10032 RackSwitch that uses 100 Gb QSFP28 and 40 Gb QSFP+ Ethernet technology is specifically designed for the data center. It is ideal for today's big data workload solutions and is an enterprise class Layer 2 and Layer 3 full featured switch that delivers line-rate, high-bandwidth switching, filtering and traffic queuing without delaying data. Large data center-grade buffers help keep traffic moving, while the hot-swap redundant power supplies and fans (along with numerous high-availability features) help provide high availability for business sensitive traffic.

The NE10032 RackSwitch has 32x QSFP+/QSFP28 ports that support 40 GbE and 100 GbE optical transceivers, active optical cables (AOCs), and direct attach copper (DAC) cables. It is an ideal cross-rack aggregation switch for use in a multi rack big data cluster.


***Figure 9: Lenovo ThinkSystem NE10032 cross-rack switch***

For further information on the NE10032 switch, visit this link:

https://lenovopress.com/lp0609-lenovo-thinksystem-ne10032-rackswitch

## 6.2 Cluster and Edge nodes

The MapR reference architecture is implemented on a set of server nodes that make up a cluster. A MapR cluster node contains the MapR Converged Data Platform services and performs the core batch and streaming data processing among other cluster related services. Edge nodes do not have MapR services loaded, and are free to run client software or analytics software such as the Data Science Refinery. Edge nodes may be connected to the same high speed data network as cluster nodes for maximum performance. Edge nodes can also be located outside of the MapR cluster and it's data network, but still accessible via lower speed remote networks.

Cluster and Edge Nodes use Lenovo ThinkSystem SR650 servers with locally attached storage. MapR runs well on a homogenous server environment with no need for different hardware configurations for data services. Server nodes can run three different types of services which are MapR data services, MapR management services and other optional MapR services.

Unlike other Hadoop distributions that require different server configurations for management nodes and data nodes, the MapR reference architecture requires only a single node hardware configuration. Each node is then configured to run one or more of the mentioned services.

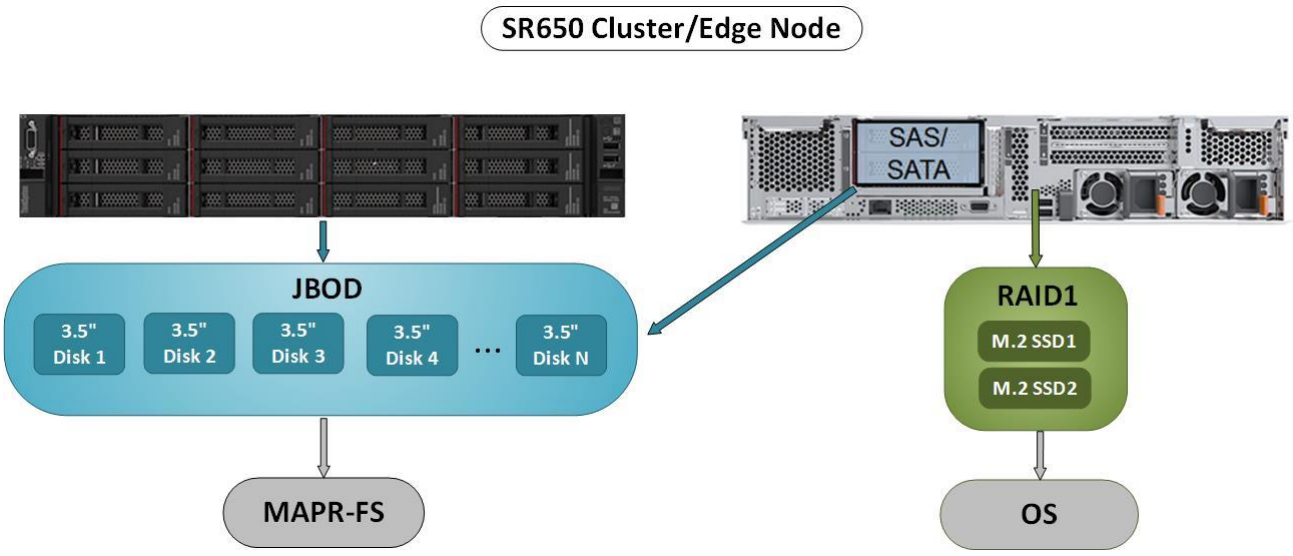
## 6.2.1 **Predefined Configuration Summary**

Table 1 lists the recommended system components for cluster nodes.

***Table 1: Node predefined configuration***

| Component | Cluster and Edge node configuration |
|---|---|
| Server | ThinkSystem SR650 |
| Processor | 2x Intel® Xeon® processors: 6130 Gold, 16-core 2.1Ghz |
| Memory - base | 256 GB: 8x 32GB 2666MHz RDIMM |
| Disk (OS) | Dual M.2 128GB or 480GB SSD with RAID1 |
| Disk (data) | 4 TB drives: 14x 4TB NL SAS 3.5 inch (56 TB total)<br>Alternate HDD capacities available:<br>6 TB drives; 14x 6TB NL SAS 3.5 inch (84 TB total)<br>8 TB drives: 14x 8TB NL SAS 3.5 inch (112 TB total)<br>10 TB drives: 14x 10TB NL SAS 3.5 inch (140 TB total)<br>12 TB drives: 14x 12TB NL SAS 3.5 inch (168TB total) |
| HDD controller | OS: M.2 RAID1 mirror enablement kit<br>MapR-FS: ThinkSystem 430-16i 12Gb HBA |
| Hardware storage protection | OS: RAID1<br>MapR-FS: None (JBOD). By default MapR-FS stripes data across multiple disks for data redundancy |
| Hardware management network adapter | Integrated XCC management controller - dedicated 1Gb or shared LAN port |
| Data network adapter | ThinkSystem 10Gb 4-port SFP+ LOM |

The Intel® Xeon® processor E5-2680 v4 is recommended to provide sufficient performance. A minimum of 256 GB of memory is recommended for most MapReduce workloads with 512 GB or more recommended for Drill, Spark, and memory-intensive MapReduce workloads.




## 6.2.2 **Storage Configuration**

Two sets of disks are used, one set of disks is for operating system and the other set of disks is for data. For the operating system disks, RAID 1 mirroring should be used.

Each node in the reference architecture has internal storage. External storage is not used in this reference architecture. Available data space assumes the use of MapR replication with three copies of the data, and 25% capacity reserved for efficient file system operation and to allow time to increase capacity if needed.

In situations where higher storage capacity is required, the main design approach could be to increase the amount of data disk space per node. Using 6 TB drives instead of 4 TB drives increases the total per node data disk capacity from 56 TB to 84 TB, a 50% increase. However, when increasing data disk capacity, there is a trade-off with the node's storage performance (IO throughput). For some workloads, increasing the amount of user data that is stored per node can *decrease* disk parallelism and negatively affect performance. Increasing drive size also affects rebuilding and repopulating the replicas if there is a disk or node failure. Higher density disks or nodes results in higher rebuild times. Drives that are larger than 4 TB are not recommended for best disk performance due to this balance of capacity versus performance. In this case, *higher capacity should be achieved by increasing the number of nodes in the cluster.*

For the case where higher IO throughout from each node is required, the data nodes can be configured with 24 2.5-inch SAS drives, (which may have less storage capacity than 3.5" drives) but much higher IO throughput due to parallelism from higher drive count and faster SAS technology .

For the HDD controller selection, just-a-bunch-of-disks (JBOD) is the best choice for a MapR cluster. It provides excellent performance and when combined with the default of 3x data replication, also provides significant protection against data loss. The use of RAID with data disks is discouraged because it reduces performance and the amount data that can be stored. Data nodes can be customized according to client needs.

Lenovo storage adapters provide a true JBOD configuration for best performance. In cases where only a RAID controller is available with no JBOD configuration, RAID0 may be configured but with a single HDD per RAID array. This will most closely emulate the JBOD configuration. Multiple HDDs in a single RAID0 array are not recommended nor required since a failure of a single HDD will cause all HDDs in that array to go off-line.

### 6.2.3 **Minimum Node Count**

For a Production environment and High Availability the absolute minimum number of nodes is 6 to accommodate separate CLDB and Zookeeper services on 3 nodes each. For a proof of concept (POC) or Non-Prod cluster, 5 nodes can be used by overlapping CLDB and ZK on a single common node. Reducing a cluster to less than 5 nodes moves the cluster away from a validation platform towards a sandbox environment and will encounter reduced work load capability and service warnings during operation.

Other Hadoop branded clusters which use master node and worker node types will specify a similar 5 node minimum using 3 worker nodes and 2 master nodes for POC purposes.

### 6.2.4 **Node Service Layout**

The location of various MapR services on certain nodes depends on the total number of nodes in the cluster. MapR is very flexible in its ability to use any node for any MapR service. Here we recommend three MapR service layout templates for small, medium and large deployments to address high availability (HA). For additional information on planning and installing a MapR converged data platform, reference this link: http://maprdocs.mapr.com/home/install.html.

**Note:** A primary consideration in the node layout is separating CLDB services from Zookeeper services for maximum node stability. The remaining services can be spread across remaining nodes evenly to avoid overloading a single node with excessive services.

**Single Rack Deployment**

In this scenario, we recommend using 3 CLDB services on nodes 1, 2, and 3; and, 2 web and resource manager services in nodes 1 and 2. Zookeeper services are spread across separate nodes 3, 4 and 5. All nodes in the cluster have the base services such as file server, node manager, spark, and NFS gateway services installed.

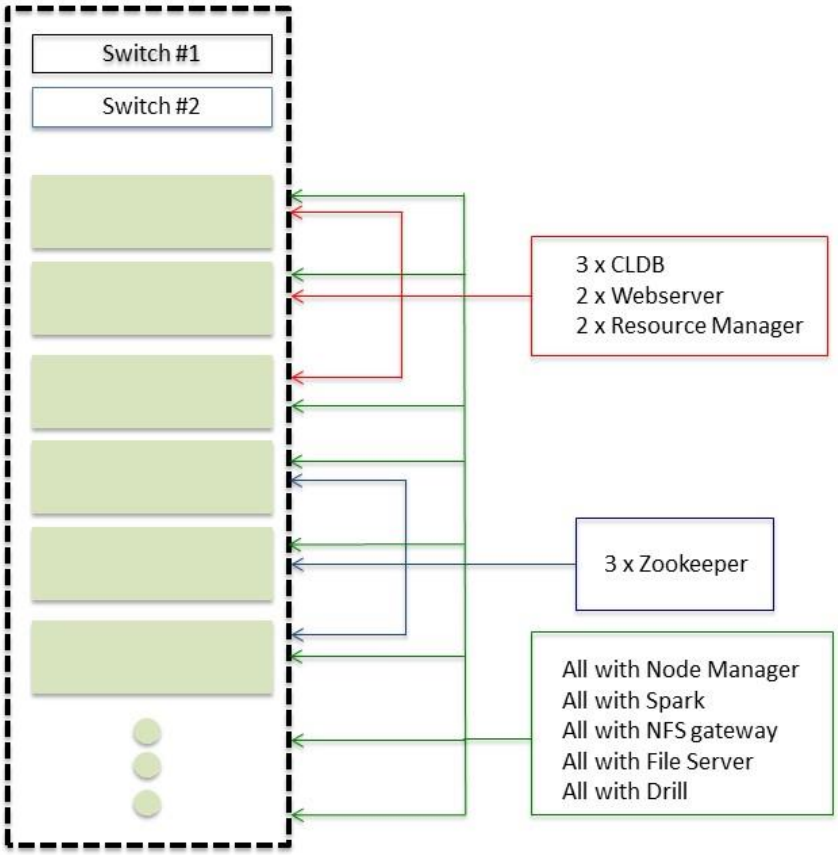



***Figure 10: Single Rack Service Layout***

**Medium Size Deployment (Two Racks)**

In this scenario, similar to single rack deployment, we recommend using 3 CLDB services spread across the 2 racks (1 is on a node in one rack and 2 others on nodes in the other rack). Two web and resource manager services are in node 1 and 2 in rack 1. Zookeeper services are spread across two racks also but on different nodes than CLDBs -. All nodes in the cluster have the base services such as file server, node manager, spark, and NFS gateway services installed.

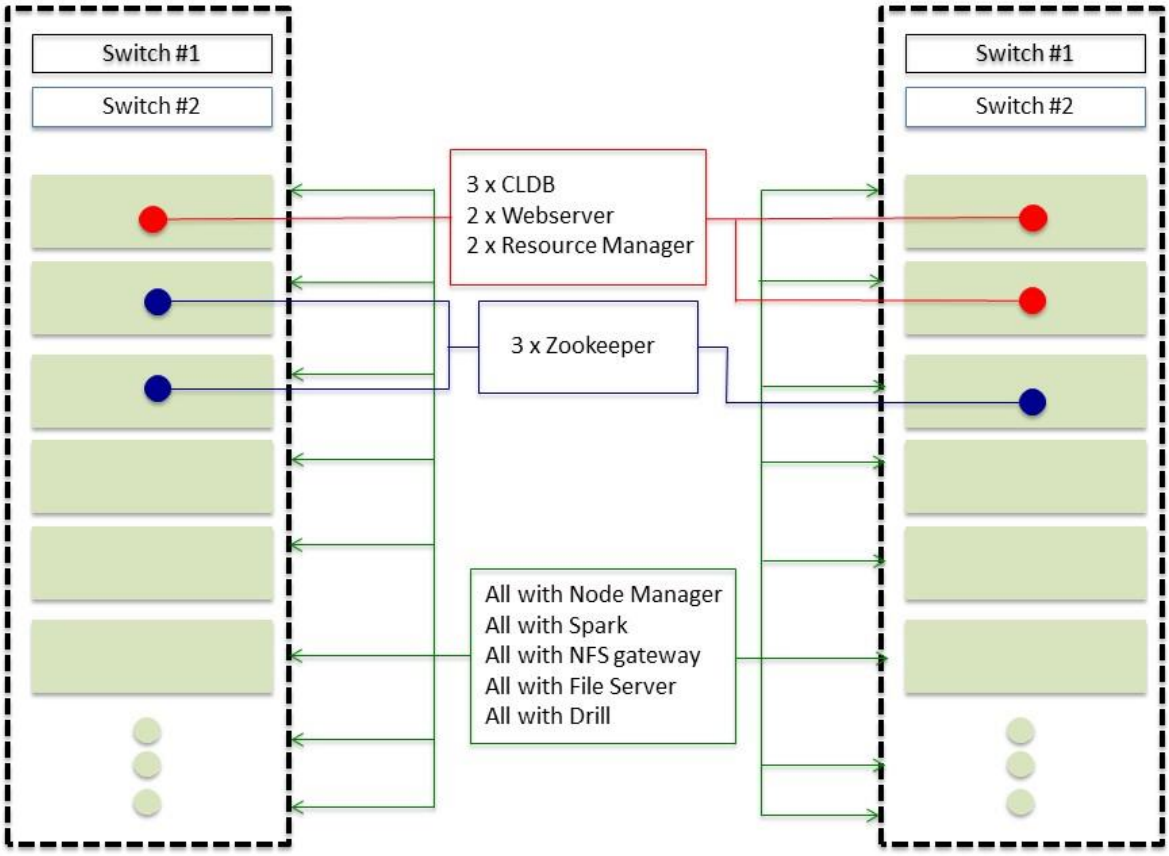



***Figure 11: MapR Two Rack Service Layout***

### Large Multi Rack Deployment

In this scenario, taking into consideration a rack failure, we recommend spreading the main services across racks using three CLDB, web and resource manager services in node 1 across the racks 1, 2 and 3. Zookeeper services are deployed in node 2 across racks 1, 2 and 3. All nodes in all racks in the cluster have the base services such as file server, node manager, spark, and NFS gateway services installed.

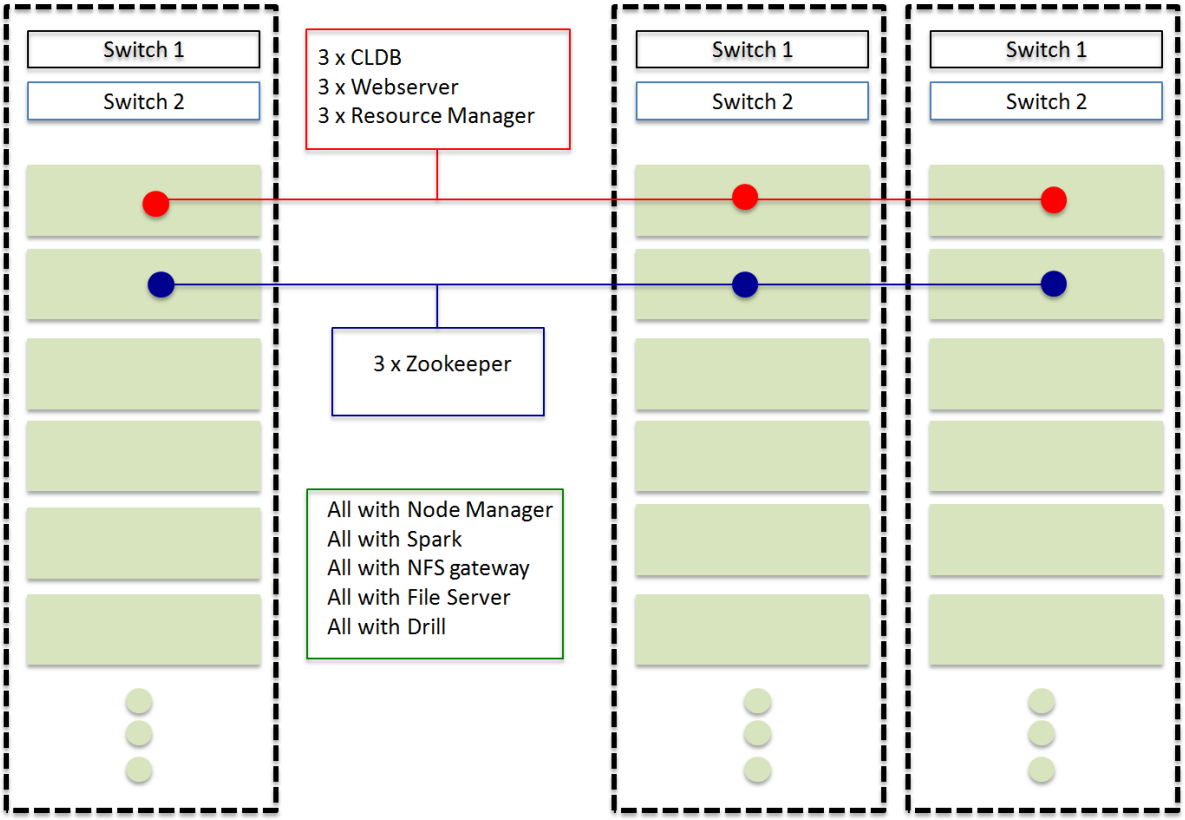



***Figure 12: MapR Large Multi Rack Service Layout***

## 6.3 Systems management

*Systems management* of a cluster includes hardware management, Operating System, and MapR cluster management. For MapR clusters, the MapR Control System (MCS) allows you to manage the cluster (including nodes, volumes, users, and alarms) through a comprehensive graphical user interface with all the functionality of the command-line or REST APIs.

*Hardware management* uses the Lenovo XClarity™ Administrator, which is a centralized resource management solution that reduces complexity, speeds up response and enhances the availability of Lenovo server systems and solutions. XClarity™ is used to install the OS onto new worker nodes; update firmware across the cluster nodes, record hardware alerts and report when repair actions are needed.

Figure 13 shows the Lenovo XClarity™ Administrator interface in which servers, storage, switches and other rack components are managed and status is shown on the dashboard. Lenovo XClarity™ Administrator is a virtual appliance that is quickly imported into a server-virtualized environment.

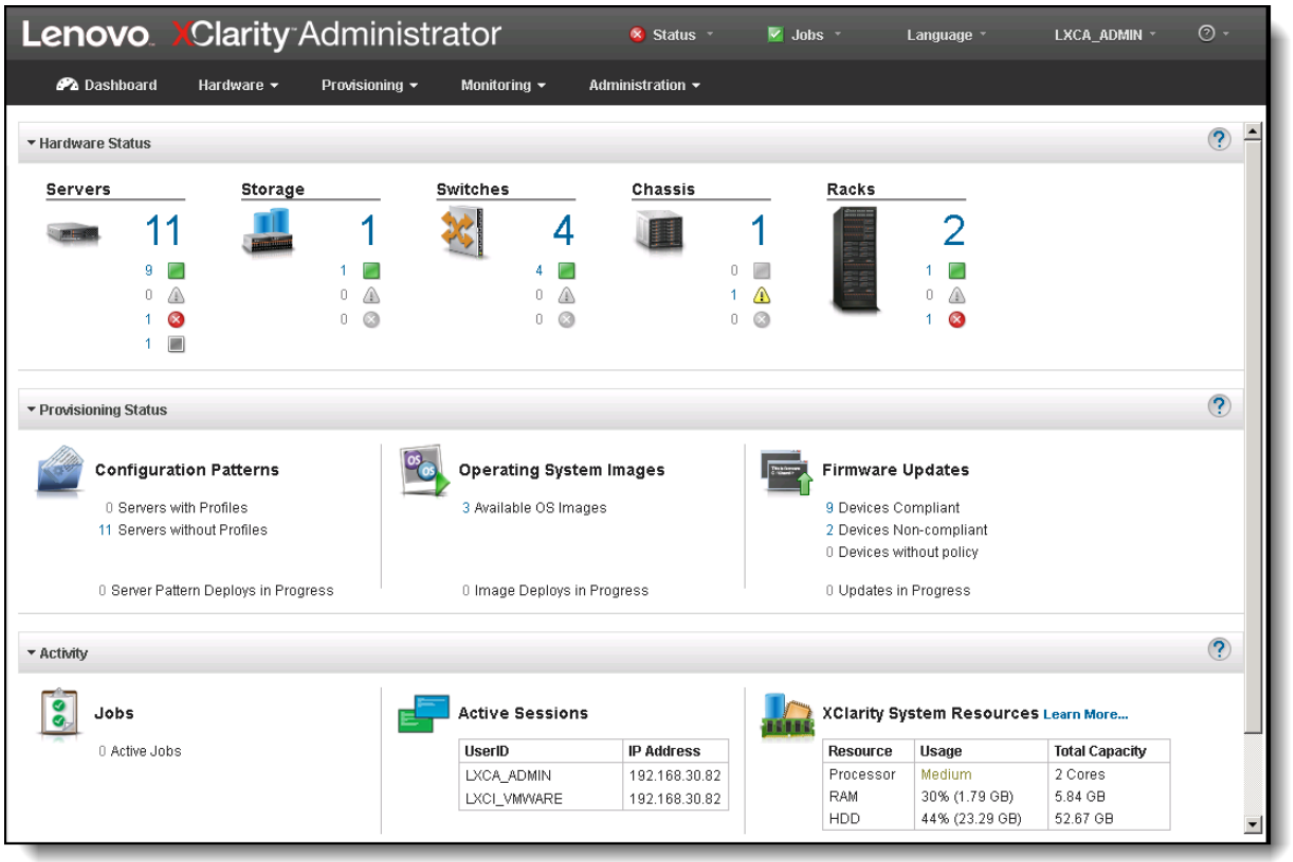



***Figure 13: XClarity™ Administrator interface***

In addition, xCAT provides a scalable distributed computing management and provisioning tool that provides a unified interface for hardware control, discovery and operating system deployment. It can be used to facilitate or automate the management of cluster nodes. For more information about xCAT, see “Resources” on page 46.

## 6.4 Networking

The reference architecture specifies two networks: a high-speed data network and a management network. Two types of top of rack switches are required; one 1Gb for out-of-band management and a pair of 10Gb for the data network with High Availability. See Figure 14 below.

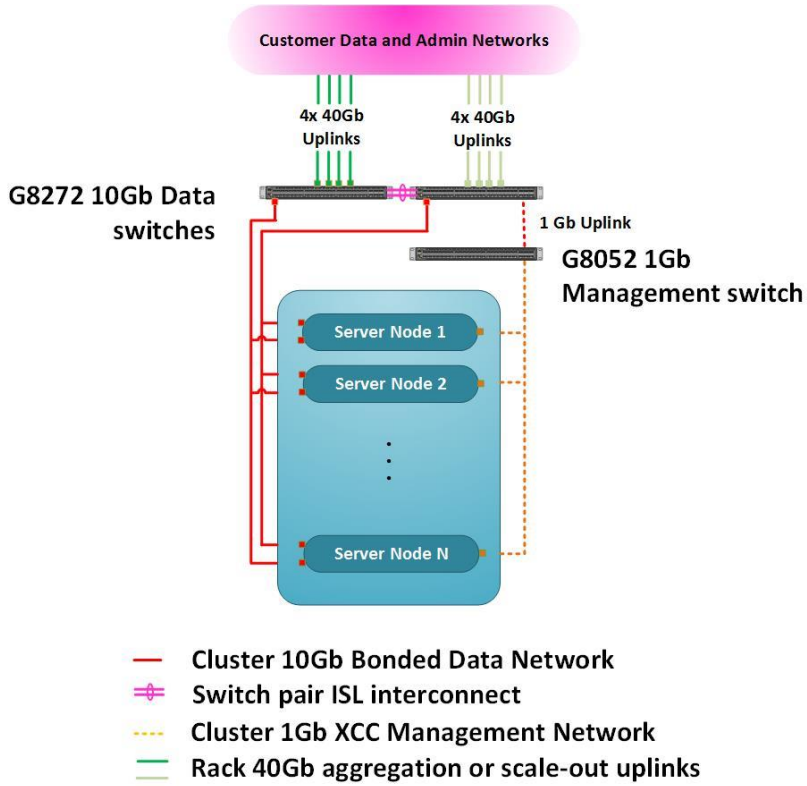



***Figure 14: MapR single rack network***

### 6.4.1 Data network

The data network creates a private cluster among multiple nodes and is used for high-speed data transfer across all the cluster, edge, and other nodes, and for ingesting data into the cluster. The MapR cluster typically connects to the customer's corporate data network. The recommended 10 GbE switch is the Lenovo RackSwitch™ G8272 that provides 48 10Gb Ethernet ports with 40Gb uplink ports.

The two 10GbE NIC ports of each node are link aggregated into a single bonded network connection. The two data switches are connected together as a Virtual Link Aggregation Group (vLAG) pair using LACP to provide the switch redundancy. Either G8272 switch can drop out of the network and the other G8272 continues transferring 10Gb traffic. The switch pairs are connected with dual 10Gb links called an ISL, which allows maintaining consistency between the two peer switches.

### 6.4.2 Hardware management network

The hardware management network is a 1GbE network for out-of-band hardware management. The recommended 1GbE switch is the Lenovo RackSwitch G8052 with 10Gb SFP+ uplink ports. Through the XClarity™ Controller management module (XCC) within the ThinkSystem SR650 servers, the out-of-band

network enables hardware-level management of cluster nodes, such as node deployment, UEFI firmware configuration, hardware failure status and remote power control of the nodes.

MapR functions have no dependency on the XCC management controller. The MapR Data and OS management networks can be shared with the XCC hardware management network, or can be separated via VLANs on the respective switches. The MapR cluster and hardware management networks are then typically connected directly to the customer's existing admin network to facilitate remote maintenance of the cluster.

### 6.4.3 **Multi-rack network**

The data network in the predefined reference architecture configuration consists of a single network topology. A rack consists of redundant G8272 access level 10Gb switches. Cluster nodes are connected with bonded 10Gb links (NIC teaming) for further redundancy to each server node. Additional racks can be added as needed for scale out. Beginning with the third rack a core switch for rack aggregation is used and the Lenovo NE10032 core switch with 40Gb and 100Gb uplinks is the best choice for this purpose.

Figure 15 shows a 2-rack configuration which is easily upgraded from single rack by adding the second rack and a LAG interconnect.

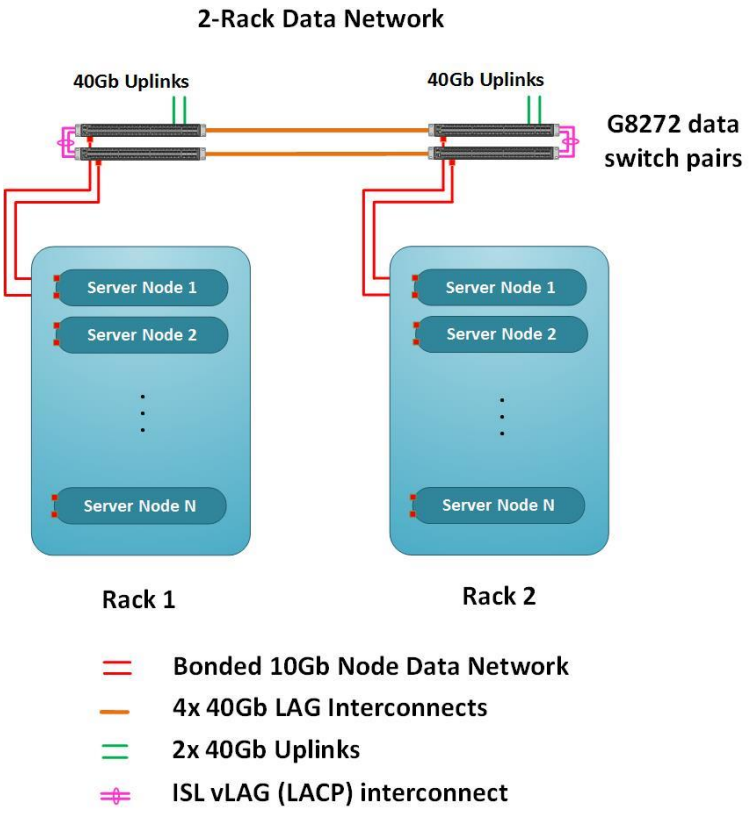



***Figure 15: MapR 2-rack network configuration***

Figure 16 shows how the network is configured when the MapR cluster contains 3 or more racks. The data network is connected across racks by four aggregated 40 GbE uplinks from each rack's G8272 switch to a core NE10032 switch. The 2-rack configuration can be upgraded to this 3-rack configuration as shown. Additional racks can be added with similar uplink connections to the NE10032 cross rack switch.

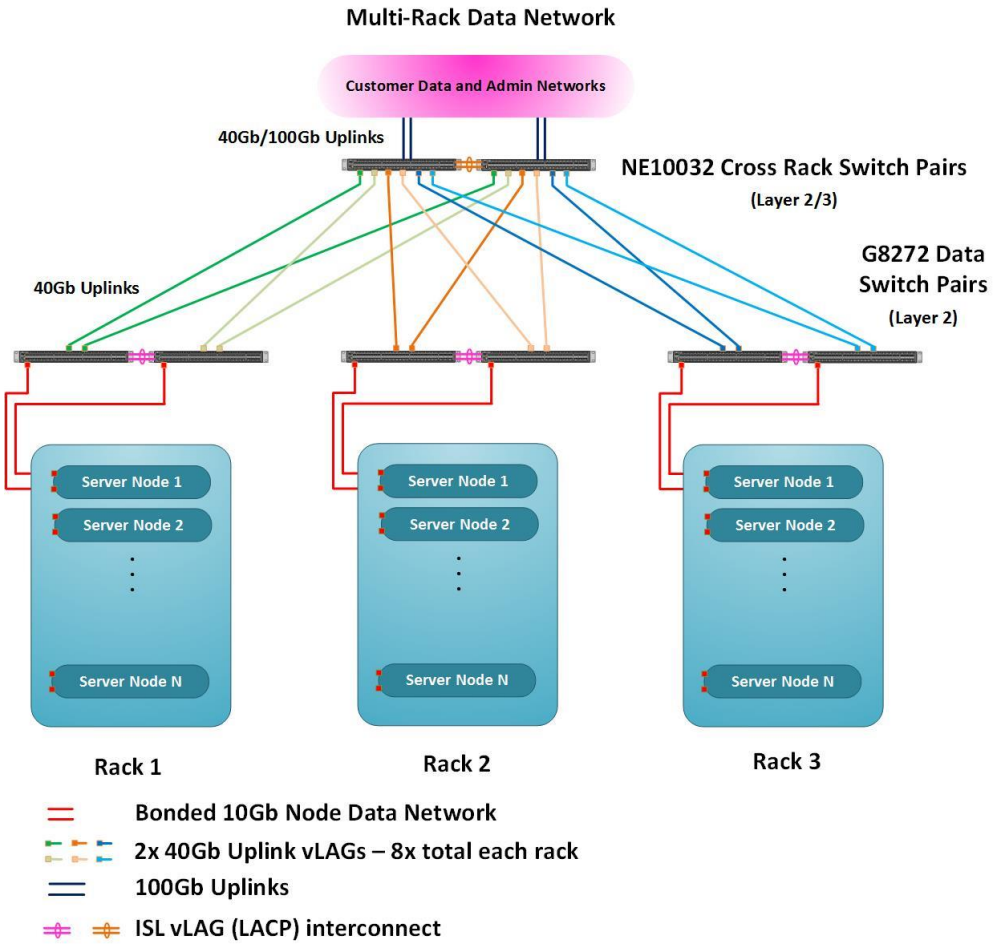



***Figure 16: MapR multi-rack rack network configuration***

Within each rack, the G8052 1Gb management switch can be configured to have two uplinks to the G8272 switch for propagating the management VLAN across cluster racks through the NE10032 cross-rack switch. Other cross rack network configurations are possible and may be required to meet the needs of specific deployments and to address clusters larger than three racks.

## 6.5 Predefined cluster configurations

The intent of the predefined configurations provided in this reference architecture is to ease initial sizing for customers and to show example starting points for four different-sized workloads: the starter rack, half rack, full rack, and a 3 rack multi-rack configuration. These consist of worker nodes, edge nodes, network switches, cabling, and rack hardware. Table 2 below, show the storage capacities and Figure 17 and Figure 18: Multi-rack MapR predefined configuration show graphics of the pre-defined configurations. The table lists the number of nodes and storage space available for data that each predefined configuration provides. Storage space is described in two ways: the total amount of ***raw storage space*** and the amount of ***usable space*** available for customer data. Usable data space assumes the use of MapR-FS replication with three copies of the data and 25% reserve working capacity. The estimates that are listed in the table are for uncompressed data. Compression rates can vary widely based on file contents and usable space must be calculated based on the specific compression rate used.

*Table 2: Storage Capacity of Predefined configurations without compression*

| Rack Storage Capacity in TB | Starter Rack (Non-Prod) | Half Rack | Full Rack | Multi-Rack |
|---|---|---|---|---|
| Raw storage (4TB) | 280 | 560 | 1064 | 3192 |
| Available data space (4TB) | 70 | 140 | 266 | 798 |
| Raw storage (6TB) | 420 | 840 | 1596 | 4788 |
| Available data space (6TB) | 105 | 210 | 399 | 1197 |
| Number of nodes | 5 | 10 | 19 | 57 |
| Number of Racks | 1 | 1 | 1 | 3 |

See *Estimating disk space* on page 29 for calculating node counts considering software compression.

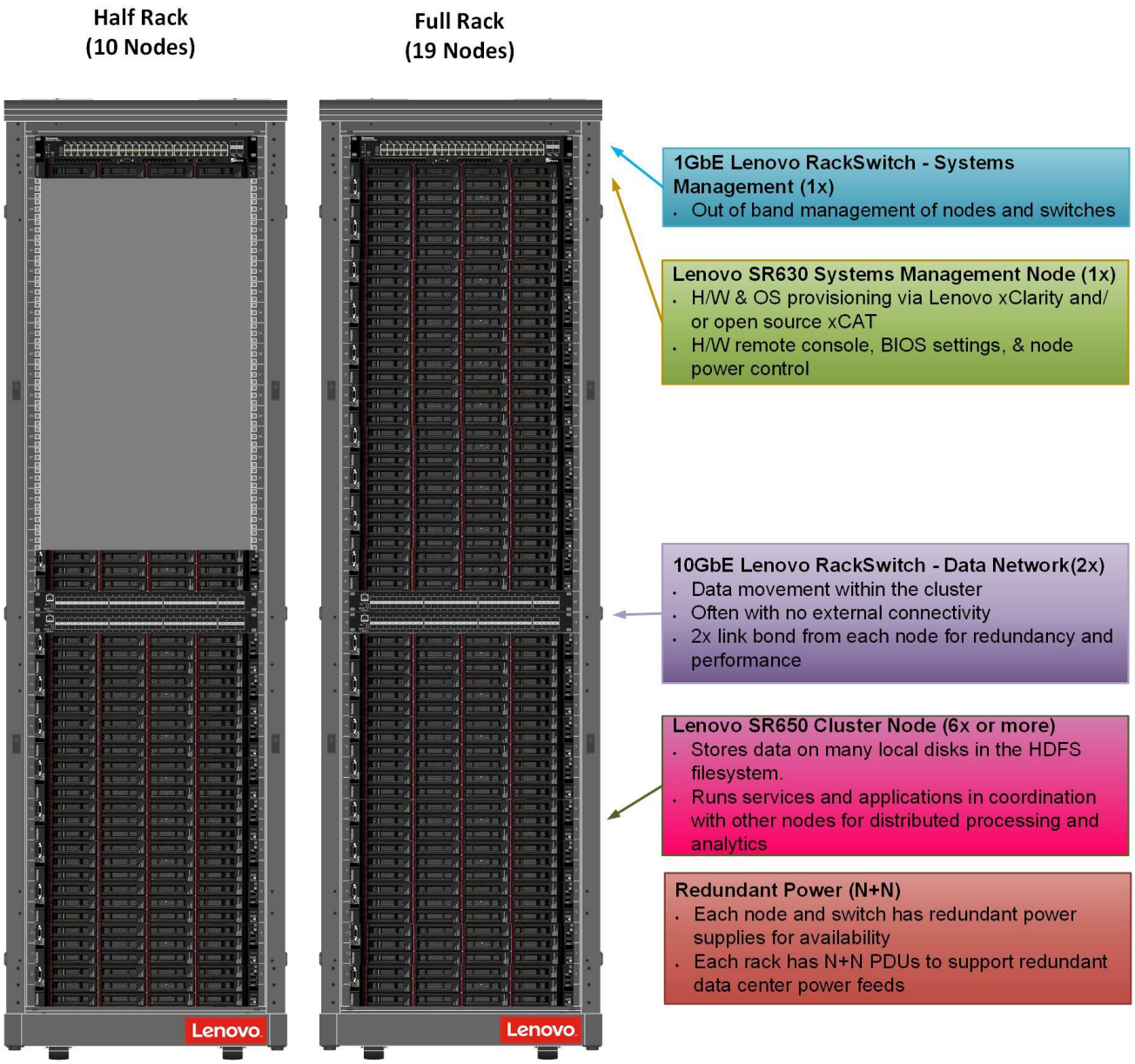



*Figure 17: Half rack and full rack MapR predefined configurations*

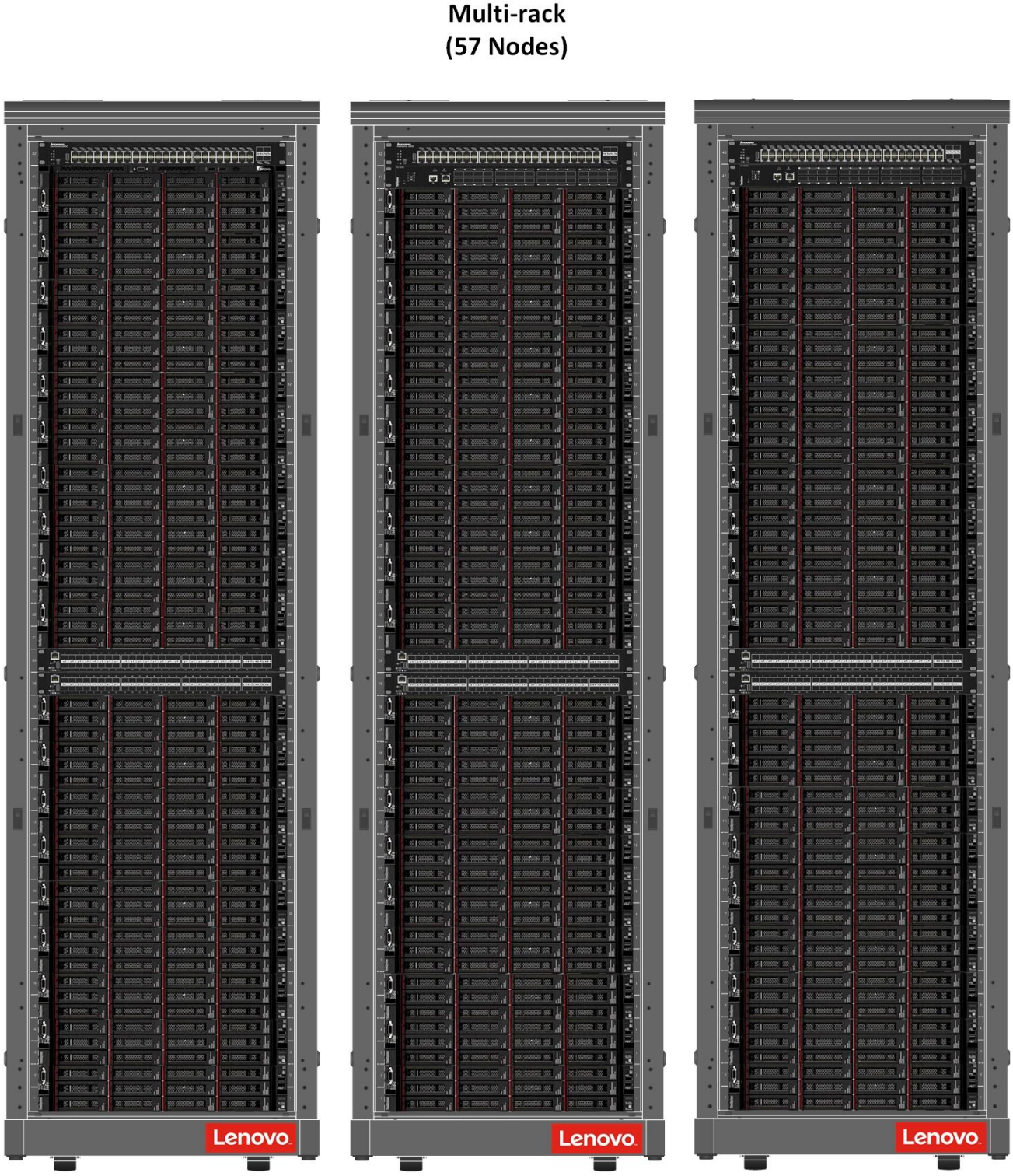



*Figure 18: Multi-rack MapR predefined configuration*

# 7 Deployment considerations

This section describes specific design choices for deploying the MapR solution.

## 7.1 Increasing cluster performance

There are two approaches that can be used to increase cluster performance: increasing node memory, and the use of a high-performance job scheduler and MapReduce framework. Often, improving performance comes at increased cost and you must consider the cost-to-benefit trade-offs of designing for higher performance.

In the MapR predefined configuration, node memory can be increased to 768 Gb with 24x 32GB RDIMMs, 1,536 GB using 24x 64GB LRDIMMs and up to 3,072 GB per node using 3DS RDIMMs and Intel processors that support 1.5TB each.

## 7.2 Designing for high ingest rates

Designing for high ingest rates is difficult. It is important to have a full characterization of the ingest patterns and volumes. The following questions provide guidance to key factors that affect the rates:

- On what days and at what times are the source systems available or not available for ingest?
- When a source system is available for ingest, what is the duration for which the system remains available?
- Do other factors affect the day, time and duration ingest constraints?
- When ingests occur, what is the average and maximum size of ingest that must be completed?
- What factors affect ingest size?
- What is the format of the source data (structured, semi-structured, or unstructured)? Are there any data transformation or cleansing requirements that must be achieved during ingest?

To increase the data ingest rates, consider the following points:

- Ingest data with MapReduce job, which helps to distribute the I/O load to different nodes across the cluster.
- Ingest when cluster load is not high, if possible.
- Compressing data is a good option in many cases, which reduces the I/O load to disk and network.
- Filter and reduce data in earlier stage saves more costs.

## 7.3 Designing for Storage Capacity and Performance

Selection of the HDD form factor, number of drives, and size of each drive can skew a worker node towards highest capacity or highest disk IO throughput.

### 7.3.1 Node Capacity

The 3.5" HDD form factor gives the maximum local storage ***capacity*** *for a node*. 12TB and larger HDDs are available and can be used to replace the 4TB HDDs used in this reference architecture to give a total of up to the 168 TBs per node. The 4TB HDD size provides the best balance of HDD capacity and performance per node. When increasing data disk capacity, some workloads may experience a decrease in disk parallelism, creating a bottleneck at that node which negatively affects performance. To increase capacity beyond the

4TB HDD size recommended in this reference architecture, the number of nodes in the cluster should be increased to maintain good I/O disk node performance

### 7.3.2 **Node Throughput**

The 2.5" HDD form factor gives the maximum local storage ***throughput*** for a node configuration. In cases where the maximum local storage throughout per node is required, the worker node can be configured with 24x 2.5-inch SAS drives. The 2.5-inch HDD has less total capacity per drive and gives less total capacity per node than the 3.5" form factor, but allows for higher parallel access to the drives - more data can be accessed simultaneously. The SR650 configuration using 2.5" and 3.5" HDDs is listed below as an example of maximum node capacity vs. parallel HDD connections for various drive sizes.

| HDD Form Factor | HDD size | Max. node storage capacity | Parallel HDD Connections |
|---|---|---|---|
| 3.5" HDDs, 14x HDDs | 10 TB Drive | 140 TB | 14 |
| | 8 TB Drive | 112 TB | 14 |
| 2.5" HDDs, 24x HDDs | 2.4 TB Drive | 57.6 TB | 24 |

Solid State Drives (SSDs) are also available in the 2.5" form factor for the SR650 with a higher capacity per drive than spinning HDDs, but at a significantly higher cost per drive.

In the 2.5" HDD configuration of the SR650, it is recommended to use 3 host bus adapters for maximum parallel throughput vs. a single host bus adapter.

## 7.4 Designing for in-memory processing with Apache Spark

Methods described in this reference architecture apply for general Spark considerations as well; however, there are additional considerations. Conceptually, Spark is similar in nature to high performance computing where analytics is an important workload.

It is important that memory capacity be carefully considered, as both the Spark execution and storage of intermediate results should reside fully in memory. This is to achieve maximum performance, although Spark will continue to execute with performance benefits even when an application doesn't fully fit within memory. In general, disk access for storage or caching is very costly to Spark processing. The memory capacity considerations are highly dependent on the application, so to get an estimate one can load in to cache an RDD (Resilient Distributed Dataset) of a desired dataset and monitor memory consumption. In summary, for Spark workloads with high execution and intermediate storage requirements, memory capacity is the primary consideration.

Additional considerations for memory configuration include the bandwidth and latency requirements. Applications with high transactional memory usage should focus on DIMM configurations that are balanced across the CPU memory controllers and their memory channels. The following table provides ideal worker node memory configurations for bandwidth/latency sensitive workloads.

***Table 3: Recommended memory configurations for 2-socket worker nodes***

| Capacity | DIMM Description | Quantity |
|---|---|---|
| 128GB | 16GB TruDDR4 Memory (1Rx4, 1.2V) 2666MHz RDIMM | 8 |
| 256GB | 32GB TruDDR4 Memory (2Rx4, 1.2V) 2666Mhz RDIMM | 8 |
| 384GB | 32GB TruDDR4 Memory (2Rx4, 1.2V) 2666Mhz RDIMM | 12 |
| 512GB | 32GB TruDDR4 Memory (2Rx4, 1.2V) 2666Mhz RDIMM | 16 |
| 768GB | 64GB TruDDR4 Memory (4Rx4, 1.2V) 2666MHz LRDIMM | 12 |
| 1,536GB | 64GB TruDDR4 Memory (4Rx4, 1.2V) 2666MHz LRDIMM | 24 |
| 3,072GB * | 128GB TruDDR4 Memory (8Rx4 1.2V) 2666Mhz 3DS RDIMM | 24 |
| | DIMM counts to be avoided: 2,6,10,14,18,20,22 | |

| Best | Better | Avoid |
|---|---|---|

*Notes: DIMM quantity is of the same part number (speed, size, rank, etc.)*
*Requires CPU part numbers that support 1.5TB of memory each.*

Some memory configurations are unbalanced and negatively affect memory interleaving by the memory controller. Although these DIMM configurations are supported by the hardware and will function, they should be avoided in favor of the higher performance configurations. Reference the *Intel Xeon Scalable Family Balanced Memory Configurations* white paper in the Resources section.

## 7.5 Processor and Network Considerations

Similarly, processor selection may vary based on the level of desired level of workload parallelism. For example, Apache recommends 2-3 tasks per CPU core. Large working sets of data can drive memory constraints, which can be alleviated through further increasing parallelism, resulting in smaller input sets per task. In this case, higher core counts can be beneficial. Naturally, the nature of the operations is considered, as they may be simple evaluations or complex algorithms.

The cluster data network using 10Gb bonded NIC interfaces connected with dual 10Gb network switches provides up to 20Gb of network connectivity between nodes in the cluster. The ThinkSystem 4-port 10Gb LAN on Motherboard (LOM) adapter is recommended in this reference architecture. Alternate 10Gb network adapters are available as well as Lenovo hardware for higher data rate networks.

***Table 4: Network adapters for cluster nodes***

| Code | Description |
|---|---|
| AT7S | Emulex VFA5.2 2x10 GbE SFP+ PCIe Adapter |
| AT7T | Emulex VFA5.2 2x10 GbE SFP+ PCIe Adapter and FCoE/iSCSI SW |
| ATPX | Intel X550-T2 Dual Port 10GBase-T Adapter |
| ATRN | Mellanox ConnectX-4 1x40GbE QSFP+ Adapter |
| AUAJ | Mellanox ConnectX-4 2x25GbE SFP28 Adapter |
| AUKN | ThinkSystem Emulex OCe1410B-NX PCIe 10Gb 4-port SFP+ Ethernet Adapter |
| AUKP | ThinkSystem Broadcom NX-E PCIe 10Gb 2-Port Base-T Ethernet Adapter |
| AUKS | ThinkSystem Broadcom NX-E PCIe 25GbE 1-Port SFP28 Ethernet Adapter |
| AUKX | ThinkSystem Intel X710-DA2 PCIe 10Gb 2-Port SFP+ Ethernet Adapter |
| B0WY | ThinkSystem Intel XXV710-DA2 PCIe 25Gb 2-Port SFP28 Ethernet Adapter |
| AUZV | ThinkSystem Broadcom NetXtreme PCIe 1GB 4-Port RJ45 Ethernet Adapter |
| AUZW | ThinkSystem I350-T4 PCIe 1Gb 4-Port RJ45 Ethernet Adapter |
| AUZX | ThinkSystem NetXtreme PCIe 1Gb 2-Port RJ45 Ethernet Adapter |

| | |
|---|---|
| AUZY | ThinkSystem I350-T2 PCIe 1Gb 2-Port RJ45 Ethernet Adapter |
| B21R | ThinkSystem QLogic QL41262 PCIe 25Gb 2-Port SFP28 Ethernet Adapter |

## 7.6 Estimating disk space

When you are estimating disk space within a cluster, consider the following points:

For improved fault tolerance and performance, MapR stripes data across the multiple cluster nodes. By default, the file system maintains three replicas.
Compression ratio is an important consideration in estimating disk space and can vary greatly based on file contents. If the customer's data compression ratio is unavailable, assume a compression ratio of 2.5:1.
To ensure efficient file system operation and to allow time to add more storage capacity to the cluster if necessary, reserve 25% of the total capacity of the cluster.

Assuming the default three replicas are maintained by the cluster, the raw data disk space and the required number of nodes can be estimated by using the following equations:

*Total raw data disk space = (User data, uncompressed) * (4 / compression ratio)*

*Total required worker nodes = (Total raw data disk space) / (Raw data disk per node)*

You should also consider future growth requirements when estimating disk space.

Based on these sizing principals, Table 5 shows an example for a cluster that must store 500 TB of uncompressed user data. The example shows that the cluster needs 800 TB of raw disk to support 500 TB of uncompressed data. The 800 TB is for data storage and does not include operating system disk space. A total of 15 nodes are required to support a deployment of this size.

*Total raw data disk space = 500TB * (4 / 2.5) = 500 * 1.6 = 800TB*

*Total required worker nodes = 800TB / (4TB * 14 drives) = 800TB / 56TB = 14.2 => 15 nodes*

***Table 5: Example of storage sizing with 4TB drives***

| Description | Value |
|---|---|
| Data storage size required (uncompressed) | 500 TB |
| Compression ratio | 2.5:1 |
| Size of compressed data | 200 TB |
| Storage multiplication factor | 4 |
| Raw data disk space needed for the cluster | 800 TB |
| Storage needed for MapR-FS 3x replication | 600 TB |
| Reserved storage for headroom (25% of 800TB) | 200 TB |
| Raw data disk per node (with 4TB drives * 14 drives) | 56 TB |
| Minimum number of nodes required (800/56) | 15 |

## 7.7 Scaling considerations

The MapR architecture is linearly scalable but it is important to note that some workloads might not scale completely linearly, so planning ahead for these items will help ease the effort.

When the capacity of the infrastructure is reached, the cluster can be scaled out by adding nodes. Typically, identically configured nodes are best to maintain the same ratio of storage and compute capabilities. A The

MapR cluster is scalable by adding additional SR650 Worker nodes, Edge nodes or network switches. As the capacity of a rack is reached, new racks can be added to the cluster.

When a MapR reference architecture implementation is designed, future scale out should be a key consideration in the initial design. There are two key aspects to consider: networking and management. These aspects are critical to cluster operation and become more complex as the cluster infrastructure grows.

The cross rack networking configuration that is shown in Figure 16 provides robust network interconnection of racks within the cluster. As racks are added, the predefined networking topology remains balanced and symmetrical. If there are plans to scale the cluster beyond one rack, a best practice is to initially design the cluster with multiple racks (even if the initial number of nodes fit within one rack). Starting with multiple racks can enforce proper network topology and prevent future re-configuration and hardware changes. As racks are added over time, multiple NE10032 switches might be required for greater scalability and balanced performance.

Also, as the number of nodes within the cluster increases, so do many cluster management tasks, such as updating node firmware or operating systems. Building a cluster management framework as part of the initial design and proactively considering challenges in managing a large cluster pays off significantly in the long run.

Proactive planning for future scale out and the development of cluster management framework as a part of initial cluster design provides a foundation for future growth that can minimize hardware reconfigurations and cluster management issues as the cluster grows.

## 7.8 Designing with Docker Containers

Building and deploying Docker containers is accomplished using MapR prebuilt Docker images and customizing them as needed via the dockerfile. The MapR runtime Docker image (PACC) is the interface between the application running in the container and the MapR cluster resources such as persistent MapR-FS file system storage.

### 7.8.1 PACC Container Overview

The MapR Persistent Application Client Container (PACC) is a Docker-based container image that includes a container-optimized MapR client. The PACC provides seamless access to MapR Converged Data Platform services, including MapR-FS, MapR-DB, and MapR-ES. The PACC makes it fast and easy to run containerized applications that access data in MapR.

Pre-built Docker container base images are available from Docker hub, as well as example MapR container applications on GitHub. This section provides an overview of designing with the MapR PACC while a full description is available from MapR at:

https://mapr.com/blog/getting-started-mapr-client-container/

A prerequisite to using the PACC is having Docker 1.12.5 or later version loaded on the nodes where the container is created and where the container is launched. Refer to this link for other requirements:

https://maprdocs.mapr.com/52/AdvancedInstallation/BeforeDeployingtheMapRPACC2.html

### 7.8.2 Creating Docker Containers with PACC

While a MapR-provided Docker image cannot be modified directly, one can build a custom image that is based on a MapR Persistent Application Client Container (PACC). The following example shows a custom Dockerfile that is used to create a new Docker image. In this example, an application has a JAR file that takes a producer as a parameter and runs a custom function.

The Docker files are configured to:

1. Install and configure the MapR Client: this is done by building the container from the MapR PACC image
2. Deploy the Java application: this is done by copying the Jar into the container
3. Run the Java application: simply call the java command.

The application will automatically inherit from the various components installed by the container: MapR-DB and MapR-ES Streams Client, POSIX Client for Containers, Java 8, etc.

MapR PACC images can be found on the Docker Hub here:
https://hub.docker.com/r/maprpartners/

**Creating a Custom Dockerfile**

The example Docker file below is defined with the following steps and is an example developing a MapR-ES aware application:
:

- Create a new directory /usr/share/mapr-apps/ to deploy the application
- Copy the application from maven target directory into this directory
- Copy the run.sh file to this folder and make it executable
- For the Web service the HTTP port 8080 is exposed
- The run.sh is automatically started using the CMD command

The run.sh script will create /apps/logs using a mount point and will start the Java application when the container starts.

```
FROM maprtech/pacc:5.2.0_2.0_centos7

# Create a directory for the MapR Application and copy the Application
RUN mkdir -p /usr/share/mapr-apps/

COPY ./target/sensor-service-1.0-SNAPSHOT.jar /usr/share/mapr-apps/sensor-service.jar
COPY run.sh /usr/share/mapr-apps/run.sh
RUN chmod +x /usr/share/mapr-apps/run.sh

CMD ["start", "/usr/share/mapr-apps/run.sh", "/apps/sensors:computer"]
```

#### Building a Custom Docker Image From the Dockerfile

Build a new custom image using the Docker build command. In this example, the files are located in the sensor-service directory and apache maven is being used for software management (mvn clean package command):

```
$ cd sensor-service
$ mvn clean package
$ docker build -t mapr-sensor-producer .
```

### 7.8.3 Configuring containers for MapR-DB and MapR-ES

Before launching a container that uses MapR-ES or MapR-DB, each must be created and configured.

#### Configuring MapR-ES for ingesting external data stream

For example, for MapR-ES create a new MapR Streams and new folder in the MapR File System with the following commands:

```
$ maprcli stream create -path /apps/sensors -produceperm p -consumeperm p -topicperm p
$ maprcli stream topic create -path /apps/sensors -topic computer
```

#### Configuring MapR-DB for storing data in tables

MapR-DB is a data store that uses OJAI documents stored in JSON tables. To create a JSON table in MapR-DB for use by the container, a table with the name myJasonTable is initialized as follows:

```
$ maprcli table create -path /tables/myJsonTable -tabletype json
```

### 7.8.4 Launching container applications for MapR-ES and MapR-DB

Run the container with the following command:

```
$ docker run -it -e MAPR_CLDB_HOSTS=192.168.99.18 -e MAPR_CLUSTER=my.cluster.com -e MAPR_CONTAINER
_USER=mapr --name producer -i -t mapr-sensor-producer
```

This command creates a new container based on the mapr-sensor-producer image that was just built. The command use the following mandatory variables:

- The name of the container is producer
- MAPR_CLDB_HOSTS: the list of CLDB hosts of your MapR cluster
- MAPR_CLUSTER: the name of the cluster
- MAPR_CONTAINER_USER : the user that will be used to run the application

These two variables are used to configure the MapR Client embedded in the container.
The Java application is automatically started by Docker, and you should see messages in the Kafka Console.

## 7.9 High Availability (HA) considerations

When a MapR cluster on ThinkSystem is implemented, consider availability requirements as part of the final hardware and software configuration. Typically, a standard Hadoop deployment has some high availability features, but MapR enhancements make it more highly available for mission-critical environments. MapR best practices provide significant protection against data loss. MapR ensures that failures are managed without causing an outage. There is redundancy that can be added to make a cluster even more reliable. Some consideration must be given to hardware and software redundancy.

### 7.9.1 Networking considerations

A second redundant switch can be added to ensure HA of the hardware management network. The hardware management network does not affect the availability of the MapR-FS or Hadoop functionality, but it might affect the management of the cluster; therefore, availability requirements must be considered.

MapR provides application-level Network Interface Card (NIC) bonding for higher throughput and high availability. Customers can either choose MapR application-level bonding or OS-level bonding and switch-based aggregation of some form matching the OS bonding configuration when using multiple NICs. Virtual Link Aggregation Groups (vLAG) can be used between redundant switches. If 1Gbps data network links are used, it is recommended that more than one is used per node to increase throughput.

### 7.9.2 Hardware availability considerations

With no single point of failure, redundancy in server hardware components is not required for MapR. MapR automatically and transparently handles hardware failure resulting in the loss of any node in the cluster running any data or management service. The default three-way MapR replication of data ensures that no data is lost because two additional replicas of data are maintained on other nodes in the cluster. MapReduce tasks from failed nodes are automatically started on other nodes in the cluster. Failure of a node running any management service is automatically and transparently recovered as described in the following services.

- All ZooKeeper services are available for read operations, with one acting as the leader for all writes. If the node running the leader fails, the remaining nodes will elect a new leader. Most commonly, three ZooKeeper instances are used to allow HA operations. In some large clusters, five ZooKeeper instances are used to allow fully HA operations even during maintenance windows that affect ZooKeeper instances. The number of instances of ZooKeeper services that must be run in a cluster depends on the cluster's high availability requirement, but it should always be an odd number. ZooKeeper requires a quorum of (N/2)+1 to elect a leader where N is the total number of ZooKeeper nodes. Running more than five ZooKeeper instances is not necessary.
- All CLDB services are available for read operations, with one acting as the write master. If the node running the master CLDB service goes down, another running CLDB will automatically become the master. A minimum of two instances is needed for high availability.
- One ResourceManager service is active. Other ResourceManager instances are configured but not running. If the active ResourceManager goes down, one of the configured instances automatically

takes over without requiring any job to restart. A minimum of two instances is needed for high availability.

- All NFS servers are active simultaneously and can present an HA NFS server to nodes external to the cluster. To do this, specify the virtual IP addresses for two or more NFS servers for NFS high availability. Additionally, use round-robin Domain Name System (DNS) across multiple virtual IP addresses for load balancing in addition to high availability. For NFS access from within the cluster, NFS servers should be run on all nodes in the cluster and each node should mount its local NFS server.
- The MapR web server can run on any node in the cluster to run the MapR Control System. The web server also provides a REST interface to all MapR management and monitoring functions. For HA, multiple active web servers can be run with users connecting to any web server for cluster management and monitoring. Note that even with no web server running, all monitoring and management capabilities are available using the MapR command line interface.
- Within racks, switches and nodes have redundant power feeds with each power feed connected from a separate PDU.

### 7.9.3 **Storage availability**

RAID data disk configuration is not necessary and should be avoided in MapR clusters. The use of RAID causes a negative impact on performance. MapR provides automated setup and management of storage pools. The three-way replication provided by MapR-FS provides higher durability than RAID configurations because multiple node failures might not compromise data integrity.

If the default 3x replication is not sufficient for availability requirements, the replication factor can be increased on a file, volume, or cluster basis. Replication levels higher than 5 are not normally used. Mirroring of MapR volumes within a single cluster can be used to achieve very high replication levels for higher durability or for higher read bandwidth. Mirrors can be used between clusters as well. MapR efficiently mirrors by only copying changes to the mirror. Mirrors are useful for load balancing or disaster recovery.

MapR also provides manual or scheduled snapshots of volumes to protect against human error and programming defects. Snapshots are useful for rollback to a known data set.

Lenovo storage adapters provide a true JBOD configuration for best performance. In cases where only a RAID controller is available with no JBOD configuration, RAID0 may be configured but with a single HDD per RAID array. This will most closely emulate the JBOD configuration. Multiple HDDs in a single RAID0 array are not recommended nor required since a failure of a single HDD will cause all HDDs in that array to go off-line.

### 7.9.4 **Software availability considerations**

Operating system availability is provided by using RAID1 mirrored drives for the operating system.

The MapR Platform is unique because it was designed with a “no NameNode” architecture for high availability. MapR is designed with no single point of failure and no single bottleneck for data access. With MapR, the file metadata is replicated, distributed, and persistent, so that there is no data loss or downtime even in the face of multiple disk or node failures.

The MapR ResourceManager HA improves recovery time objectives and provides for a self-healing cluster. Upon failure, the MapR ResourceManager automatically restarts on another node in the cluster. NodeManagers can automatically pause and then reconnect to the new ResourceManager. Any currently running jobs or tasks continue without losing any progress or failing.

You can easily set up a pool of NFS nodes with HA and failover using virtual IP addresses. If one node fails, the virtual IP addresses will be automatically reassigned to the next NFS node in the pool.

It is also common to place an NFS server on every node where NFS access to the cluster is needed.

## 7.10 Migration considerations

If migration of data or applications to MapR is required, you must consider the type and amount of data to be migrated and the source of the data being migrated. Most data type can be migrated, but you must understand the migration requirements to verify viability. Standard Hadoop tools such as *distcp* (distributed copy) can be used to migrate data from other Hadoop distributions. For data in a POSIX file system, you need to NFS or FUSE client mount the MapR cluster and use standard Linux commands to copy the files into the MapR cluster. Either Sqoop or database import/export tools with MapR NFS can be used to move data between databases and MapR.

You also need to consider whether applications must be modified to use Hadoop functionality. With the MapR read/write file system that can be mounted by a standard NFS client, your applications continue to work with no code changes, so you can completely avoid the significant effort of rewriting applications to conform to Hadoop APIs.

## 7.11 Planning and Installation Tips

MapR provides substantial guidance for planning and installing a MapR cluster at this link:
https://maprdocs.mapr.com/home/AdvancedInstallation/PlanningtheCluster.html
Additional Tips noted in this reference architecture are listed below:

- Planning -On Production clusters, CLDB services should be on different nodes than Zookeeper services. This requires a minimum of 6 nodes in a High Availability (HA) production cluster.
- Planning -Proof of Concept starter racks for high work loading with less than 5 nodes are not recommended. The service count per node will increase and negatively impact node stability.
- Planning & OS Provisioning - MapR recommends at least 128GB for subdirectory /opt/root. During OS provisioning of the nodes, the root partition where /opt/root is located will need to be 128GB plus other required overhead else the installer posts a warning (yellow flag during Verification phase).
- OS Provisioning: Ensure Transparent Huge Pages (THP) are Disabled for best file system performance; rhel7 installer will default to Enabled.
- OS Provisioning - Set the data network interface MTU to 9000 for higher performance
- MapR Installation - In Production clusters, CLDB services should be installed on 3 nodes minimum for best reliability during a single failed CLDB node. MapR Installer will force an odd number of CLDB nodes.

- MapR Installation - Disk Stripe Width. MapR filesystem maintains redundant storage via striping data across multiple drives called drive pools. Disk write performance can be increased by increasing disk stripe width from the default of 3, up to 7 or 8, but will require longer pool rebuild times when a disk fails.
- MapR Installation - For this reference architecture the following changes were made to the Installer's default layout to meet MapR guidelines:
    1. Created a new CONTROL2 node group consisting of 3 nodes for the Zookeeper service. These Zookeeper services were re-assigned from the original 3 nodes where CLDB services were also loaded.
    2. Created a new HIVE node group to move services off the MASTER node group which was flagged in Red for having too many services (greater than 23). The following services were re-assigned to HIVE node group: Hive Server2, Hive Metastore, Hive WebHCat, MySQL.

# 8 Analytics Demo with Data Science Refinery

This reference architecture demonstrates the analytical capabilities of the MapR Converged Data Platform with Data Science Refinery. Using pre-built containers based on PACC the MapR Data Science Refinery Demo showcases how containers are used on Lenovo bare-metal hardware to perform real-time Twitter analytics using MapR-DB, MapR-ES, and MapR Data Science Refinery. These containers ingested a live stream of data via a Twitter developer API into a MapR cluster deployed on the Lenovo SR650 served as an edge node.

The flow diagram below illustrates how raw data in the form of tweets are ingested into the cluster via MapR-ES event streaming, formatted and stored in a table via MapR-DB, and the Data Science Refinery analytics and graphical presentation via a Zeppelin notebook.

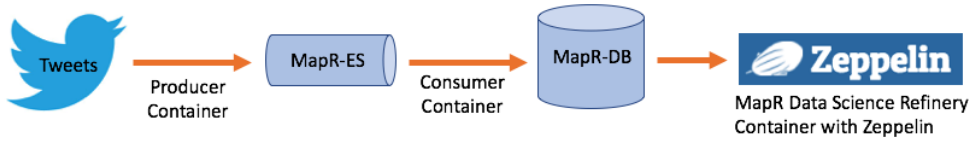



***Figure 19: Ingest and analytics flow diagram***

For the ingest phase, launching the Producer Container begins receiving raw tweet data into the cluster via MapR-ES as shown in the below screen shot.

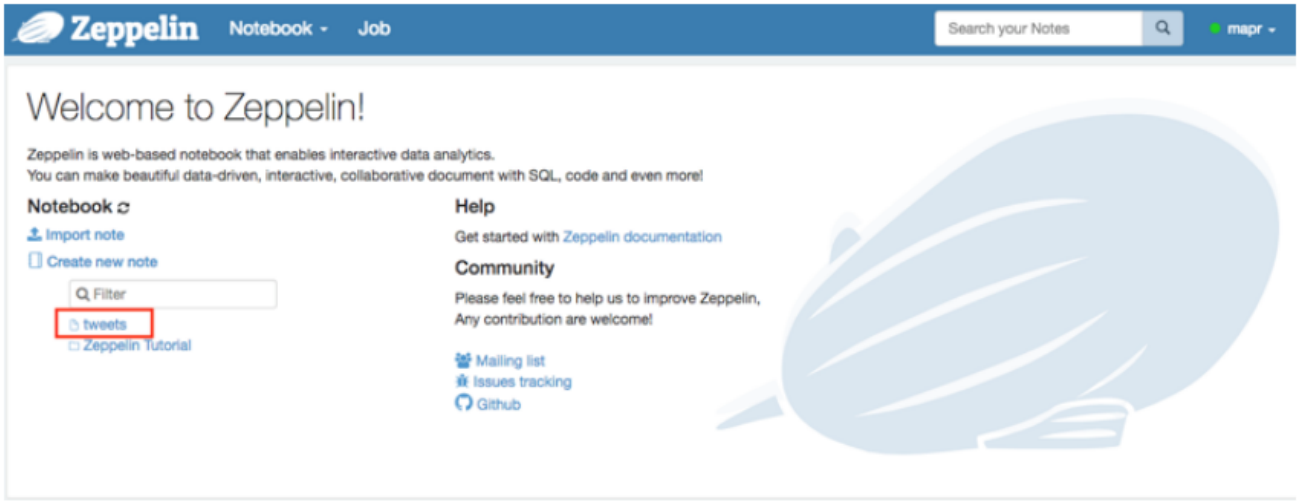

***Figure 20: Ingested tweet stream***

This unformatted data stream is formatted into individual tweets via the Consumer Container and MapR-DB, and stored in the database as a table for subsequent queries and analytics.


***Figure 21: Formatted stream stored in MapR-DB***

With the DSR Container running Data Science Repository to perform analytics and the Config-zeppelin Container to visualize the results in a web browser, the stored tweet data can be analyzed and studied.

In this demo, a web browser on the edge node (for example, IP address at port 9995; https://10.0.0.10:9995) is used to access Zeppelin where the "tweets" notebook has been created.



In the tweets notebook, the following charts with the corresponding Data Science Refinery and Drill commands show preconfigured analytics. The charts below show that over 400,000 tweets were ingested and stored in MapR-DB for analysis. The Top Ten twitter users measured by their follower count (followers_count) is analyzed from the stored data and presented as a pie chart.

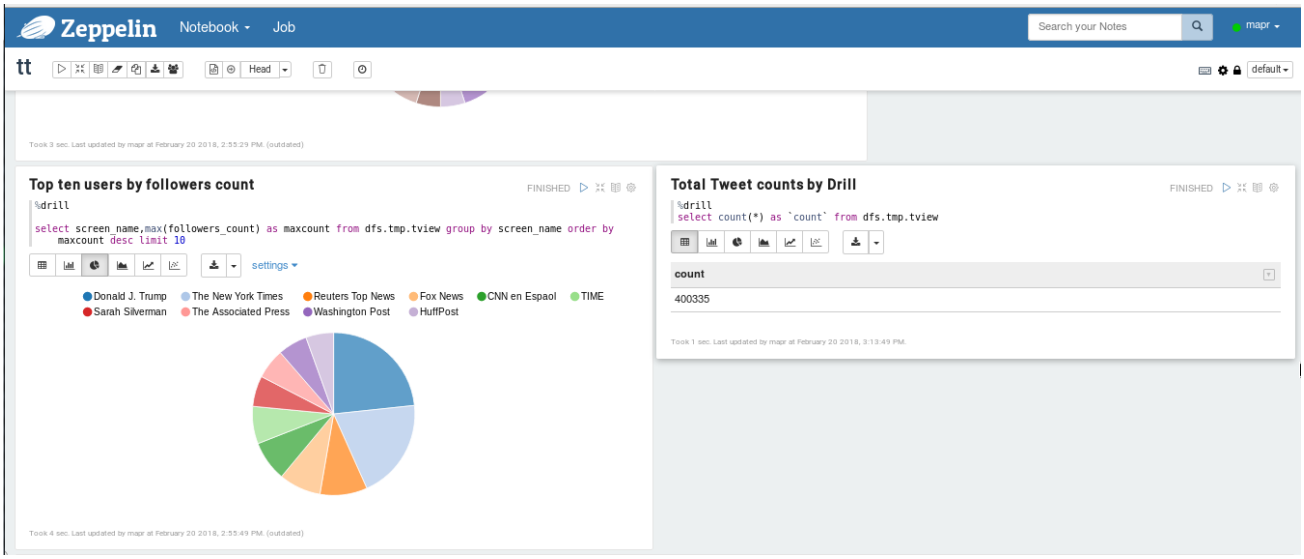



***Figure 22: Data Science Refinery Drill analysis and total tweet count***

In this example analysis, tweets from top ten states with subjects regarding tax reform were queried and presented in a pie chart format. In addition, the total tweet count by hour for this query is presented in the hourly bar chart.

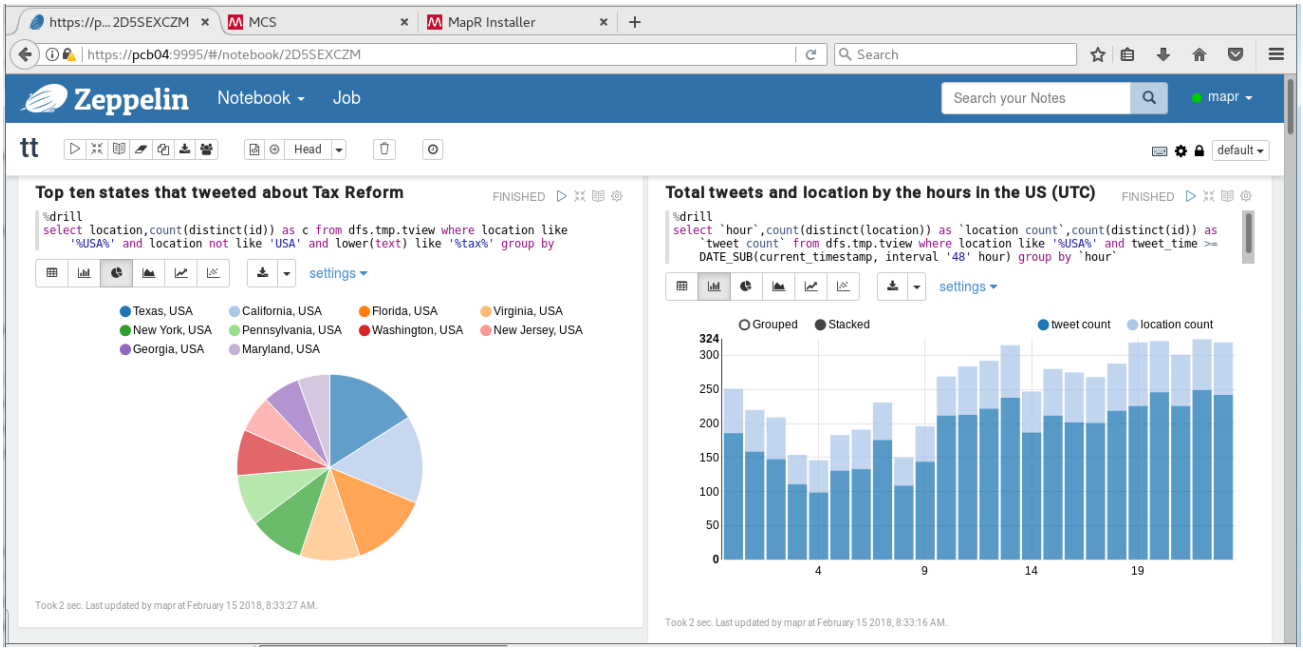



***Figure 23: Data Science Refinery - Drill query and count by hour***

To capture additional tweets with different sets of keywords, the 1-run.producer script can be modified to

launch several producers at the same time for a multi-stream ingest and higher data ingestion rate. It is very easy to scale up and down the container deployment to fit the analytical requirements.

The Zeppelin notebook is a primary tool for teams of data scientists to share and use various tools (i.e. Drill, Spark, and Hive) to extract data from a MapR cluster, then perform and visualize the data analysis with Data Science Refinery. For more information about MapR Data Science Refinery, refer to this URL:

https://mapr.com/products/data-science-refinery/

# 9 Predefined Configurations (Bill of Material)

This appendix includes the Bill of Materials (BOMs) for different configurations of hardware for the Big Data Solution for MapR deployments. There are sections for worker/edge nodes and networking.

The BOM includes the part numbers, component descriptions and quantities. Table 2 on page15 lists the quantity of each component defined in each of the predefined configurations.

This appendix is not meant to be exhaustive and must be verified with the ordering configuration tools. Any discussion of pricing, support and maintenance options is outside the scope of this document.

This BOM information is for the United States; part numbers and descriptions can vary in other countries. Other sample configurations are available from your Lenovo sales team. Components are subject to change without notice.

## 9.1 Cluster & Edge Nodes

Table 6 lists the BOM for the standard cluster and edge node.

***Table 6***: ***Worker node***

| Code | Description | Qty |
|---|---|---|
| 7X05CTO1WW | -SB- ThinkSystem SR650 - 1yr Warranty | 1 |
| AURC | ThinkSystem SR550/SR590/SR650 x16/x8(or x16) PCIe FH Riser 2 Kit | 1 |
| B0MK | Enable TPM 2.0 | 1 |
| AUPW | ThinkSystem XClarity Controller Standard to Enterprise Upgrade | 1 |
| AUR9 | ThinkSystem SR650/SR550/SR590 3.5" SATA/SAS 12-Bay Backplane | 1 |
| AUMV | ThinkSystem M.2 with Mirroring Enablement Kit | 1 |
| AUU6 | ThinkSystem 3.5" 4TB 7.2K SAS 12Gb Hot Swap 512n HDD | 14 |
| AVWF | ThinkSystem 1100W (230V/115V) Platinum Hot-Swap Power Supply | 2 |
| 5977 | Select Storage devices - no configured RAID required | 1 |
| AXCA | ThinkSystem Toolless Slide Rail | 1 |
| AUKK | ThinkSystem 10Gb 4-port SFP+ LOM | 1 |
| AUNM | ThinkSystem 430-16i SAS/SATA 12Gb HBA | 1 |
| A484 | Populate Rear Drives | 1 |
| AUVW | ThinkSystem SR650 3.5" Chassis with 8 or 12 bays | 1 |
| AWEN | Intel Xeon Gold 6130 16C 125W 2.1GHz Processor | 2 |
| AURZ | ThinkSystem SR590/SR650 Rear HDD Kit | 1 |
| B11V | -SB- ThinkSystem M.2 5100 480GB SATA 6Gbps Non-Hot Swap SSD | 2 |
| AUND | ThinkSystem 32GB TruDDR4 2666 MHz (2Rx4 1.2V) RDIMM | 8 |
| AURD | ThinkSystem 2U left EIA Latch Standard | 1 |
| 6570 | 2.0m, 13A/100-250V, C13 to C14 Jumper Cord | 2 |
| 2306 | Integration >1U Component | 1 |
| AUQB | Lenovo ThinkSystem Mainstream MB - 2U | 1 |
| AURS | Lenovo ThinkSystem Memory Dummy | 16 |
| AURP | Lenovo ThinkSystem 2U 2FH Riser Bracket | 1 |
| AURR | ThinkSystem M3.5 Screw for Riser 2x2pcs and SR530/550/558/570/590 | 2 |

| | | |
|---|---|---|
| AUSA | Lenovo ThinkSystem M3.5" Screw for EIA | 8 |
| AVWK | ThinkSystem EIA Plate with Lenovo Logo | 1 |
| AWF9 | ThinkSystem Response time Service Label LI | 1 |
| AWFF | ThinkSystem SR650 WW Lenovo LPK | 1 |
| AURM | ThinkSystem SR550/SR650/SR590 Right EIA Latch with FIO | 1 |
| B0ML | Feature Enable TPM on MB | 1 |
| B173 | Companion Part XClarity Controller, Enterprise Upgrade in Factory | 1 |
| A102 | Advanced Grouping | 1 |
| A2HP | Configuration ID 01 | 1 |
| 8971 | Integrate in manufacturing | 1 |
| AUTJ | Lenovo ThinkSystem Label Kit | 1 |
| AUSE | Lenovo ThinkSystem 2U CPU Entry Heatsink | 2 |
| AUSG | Lenovo ThinkSystem 2U Cyborg 6038 Fan module | 1 |
| AUSS | MS 12x3.5" HDD BP Cable Kit | 1 |
| 9206 | No Generic Preload Specify | 1 |
| AUT8 | ThinkSystem 1100W RDN PSU Caution Label | 1 |
| AUTS | ThinkSystem 2U 12 3.5"HDD Conf HDD sequence Label | 1 |
| AVJ2 | ThinkSystem 4R CPU HS Clip | 2 |
| AUT1 | ThinkSystem SR650 Lenovo Agency Label | 1 |
| AUSZ | ThinkSystem SR650 Service Label LI | 1 |
| AUTD | ThinkSystem SR650 model number Label | 1 |
| AUTQ | ThinkSystem small Lenovo Label for 24x2.5"/12x3.5"/10x2.5" | 1 |
| AUTA | XCC Network Access Label | 1 |

## 9.2 Systems Management Node

Table 7 lists the BOM for the Systems Management Node.

***Table 7: Systems Management Node***

| Code | Description | Qty |
|---|---|---|
| 7X01CTO1WW | -SB- ThinkSystem SR630 - 1yr Warranty | 1 |
| AUWC | ThinkSystem SR530/SR570/SR630 x8/x16 PCIe LP+LP Riser 1 Kit | 1 |
| B0MK | Enable TPM 2.0 | 1 |
| AUPW | ThinkSystem XClarity Controller Standard to Enterprise Upgrade | 1 |
| AUWB | ThinkSystem SR530/SR630/SR570 2.5" SATA/SAS 8-Bay Backplane | 1 |
| AUMV | ThinkSystem M.2 with Mirroring Enablement Kit | 1 |
| AVWA | ThinkSystem 750W (230/115V) Platinum Hot-Swap Power Supply | 2 |
| 5977 | Select Storage devices - no configured RAID required | 1 |
| AXCA | ThinkSystem Toolless Slide Rail | 1 |
| AUKK | ThinkSystem 10Gb 4-port SFP+ LOM | 1 |
| AUNG | ThinkSystem RAID 530-8i PCIe 12Gb Adapter | 1 |
| AUWQ | Lenovo ThinkSystem 1U LP+LP BF Riser Bracket | 1 |
| AUW0 | ThinkSystem SR630 2.5" Chassis with 8 Bays | 1 |
| AWEH | Intel Xeon Bronze 3106 8C 85W 1.7GHz Processor | 1 |

| | | |
|---|---|---|
| AUUV | ThinkSystem M.2 CV3 128GB SATA 6Gbps Non-Hot Swap SSD | 2 |
| AUNB | ThinkSystem 16GB TruDDR4 2666 MHz (1Rx4 1.2V) RDIMM | 1 |
| 6570 | 2.0m, 13A/100-250V, C13 to C14 Jumper Cord | 2 |
| 2305 | Integration 1U Component | 1 |
| AUS6 | Lenovo ThinkSystem 1U height CPU HS Dummy | 1 |
| AURR | ThinkSystem M3.5 Screw for Riser 2x2pcs and SR530/550/558/570/590 | 2 |
| AULP | ThinkSystem 1U CPU Heatsink | 1 |
| AVWJ | ThinkSystem 750W Platinum RDN PSU Caution Label | 1 |
| AUWF | Lenovo ThinkSystem Super Cap Holder Dummy | 1 |
| AVKJ | ThinkSystem 2x2 Quad Bay Gen4 2.5" HDD Filler | 1 |
| AUWK | Lenovo ThinkSystem 4056 Fan Dummy | 1 |
| AUWG | Lenovo ThinkSystem 1U VGA Filler | 1 |
| AUWL | Lenovo ThinkSystem 1U LP Riser Dummy | 1 |
| AVWK | ThinkSystem EIA Plate with Lenovo Logo | 1 |
| AWF9 | ThinkSystem Response time Service Label LI | 1 |
| AUX4 | MS 1U Service Label LI | 1 |
| AUX3 | ThinkSystem SR630 Model Number Label | 1 |
| AUWX | 8x2.5" HDD BP Cable Kit | 1 |
| AWGE | ThinkSystem SR630 WW Lenovo LPK | 1 |
| AUW3 | Lenovo ThinkSystem Mainstream MB - 1U | 1 |
| B0ML | Feature Enable TPM on MB | 1 |
| B173 | Companion Part for XClarity Controller Standard to Enterprise Upgrade in | 1 |
| 8971 | Integrate in manufacturing | 1 |
| AUTJ | Lenovo ThinkSystem  Label Kit | 1 |
| 9206 | No Generic Preload Specify | 1 |
| AVEN | ThinkSystem 1x1 2.5" HDD Filler | 4 |
| AVJ2 | ThinkSystem 4R CPU HS Clip | 1 |
| AUTC | ThinkSystem SR630 Lenovo Agency Label | 1 |
| AUTV | ThinkSystem large Label for non-24x2.5"/12x3.5"/10x2.5" | 1 |
| AUTA | XCC Network Access Label | 1 |

## 9.3 Management network switch

Table 8 lists the BOM for the Management/Administration network switch.

***Table 8***: ***Management/Administration network switch***

| Code | Description | Qty |
|---|---|---|
| 7159HC1 | Lenovo RackSwitch G8052 (Rear to Front) | 1 |
| ASY2 | Lenovo RackSwitch G8052 (Rear to Front) | 1 |
| A3KR | Air Inlet Duct for 442 mm RackSwitch | 1 |
| A3KP | Adjustable 19" 4 Post Rail Kit | 1 |
| 6201 | 1.5m, 10A/100-250V, C13 to IEC 320-C14 Rack Power Cable | 2 |
| 2305 | Integration 1U Component | 1 |

## 9.4 Data network switch

Table 9 lists the BOM for the data network switch.

***Table 9: Data network switch***

| Code | Description | Qty |
|---|---|---|
| 7159HCW | Lenovo RackSwitch G8272 (Rear to Front) | 2 |
| ASRD | Lenovo RackSwitch G8272 (Rear to Front) | 2 |
| ASTN | Air Inlet Duct for 487 mm RackSwitch | 2 |
| 6201 | 1.5m, 10A/100-250V, C13 to IEC 320-C14 Rack Power Cable | 4 |
| A3KP | Adjustable 19" 4 Post Rail Kit | 2 |
| 2305 | Integration 1U Component | 2 |
| 3792 | 1.5m Yellow Cat5e Cable | 2 |

## 9.5 Rack

Table 10 lists the BOM for the rack.

***Table 10: Rack***

| Code | Description | Qty |
|---|---|---|
| 1410HPB | Scalable Infrastructure 42U 1100mm Enterprise V2 Dynamic Rack | 1 |
| A2M8 | Scalable Infrastructure 42U 1100mm Enterprise V2 Dynamic Rack | 1 |
| 5895 | 1U 12 C13 Switched and Monitored 60A 3 Phase PDU | 3 |
| 2202 | Cluster 1350 Ship Group | 1 |
| 2304 | Integration Prep | 1 |
| 2310 | Solution Specific Test | 1 |
| AU8K | LeROM Validation | 1 |
| B1EQ | Network Verification | 1 |
| 5AS7A02047 | Hardware Installation Server (Business Hours) | 1 |
| 4275 | 5U black plastic filler panel | 4 |

For unused rack space, consider the use of blank plastic filter panels for the rack to better direct cool air flow.

Four PDU should be used for the half rack configuration and six PDUs for a full rack.

## 9.6 Cables

Table 11 lists the BOM for the cables, for each node. Quantities depend on total number of nodes in a rack

***Table 11: Cables***

| Code | Description | Qty |
|---|---|---|
| AT2S | -SB- Lenovo 3m Active DAC SFP+ Cables | * |
| A3RG | 0.5m Passive DAC SFP+ Cable | * |
| A51N | 1.5m Passive DAC SFP+ Cable | * |
| 3792 | 1.5m Yellow Cat5e Cable | * |
| A51P | 2m Passive DAC SFP+ Cable | * |
| 3793 | 3m Yellow Cat5e Cable | * |

# 10 Acknowledgements

This reference architecture has benefited very much from the detailed and careful review provided by:

**Lenovo business review**

- Prasad Venkatachar – Sr. Solutions Product Manager

**MapR technical review**

- Dale Kim, Sr. Director, Product Marketing

**MapR business review**

- Tom Scurlock, Worldwide Alliances and Channels at MapR Technologies
- Angie Chatham, Director of Global Alliances and Channels

# 11 Resources

Additional information is available at these links:

Lenovo ThinkSystem SR650 (MapR Cluster Node):

- Lenovo Press product guide: https://lenovopress.com/lp0644

Lenovo RackSwitch G8052 (1GbE Switch):

- Lenovo Press product guide: https://lenovopress.com/tips1270

Lenovo RackSwitch G8272 (10GbE Switch):

- Lenovo Press product guide: https://lenovopress.com/tips1267

Lenovo ThinkSystem NE10032 (40GbE/100GbE Switch):

- Lenovo Press product guide: https://lenovopress.com/lp0609

Intel Xeon Scalable Family Balanced Memory Configurations whitepaper:

- https://lenovopress.com/lp0742-intel-xeon-scalable-family-balanced-memory-configurations

Lenovo XClarity Administrator:

- Lenovo Press product guide: https://lenovopress.com/tips1200

MapR:

- MapR main website: www.mapr.com
- MapR Persistent Application container (PACC): https://mapr.com/products/persistent-application-client-container/
- MapR Data Science Refinery: MapR Community Data Science Refinery
- MapR-ES: https://mapr.com/products/mapr-streams/
- MapR-DB: https://mapr.com/products/mapr-db/
- MapR products: www.mapr.com/products
- MapR editions overview: www.mapr.com/products/mapr-distribution-editions
- MapR architecture overview: www.mapr.com/why-hadoop/why-mapr/architecture-matters
- MapR blogs: www.mapr.com/blog
- MapR Resources: www.mapr.com/resources
- MapR documentation: maprdocs.mapr.com
- MapR technical training: www.mapr.com/training
- MapR getting started: mapr.com/products/hadoop-download/

Open source software:

- Hadoop: hadoop.apache.org
- Pig: pig.apache.org
- Cascading:www.cascading.org
- Spark:spark.apache.org
- Apache Tez:tez.apache.org
- Mahout: mahout.apache.org
- Hive: hive.apache.org
- Drill:drill.apache.org
- Sqoop: sqoop.apache.org
- Flume: flume.apache.org
- Hue:gethue.com

- Sentry: sentry.incubator.apache.org
- Oozie: oozie.apache.org
- ZooKeeper: zookeeper.apache.org
- Sahara:wiki.openstack.org/wiki/Sahara

Other resources:

- xCat: xcat.org

# 12 Document history

| | | |
|---|---|---|
| Version 1.0 | 3/29/2018 | Inital publish for MapR 6.0 on SR650 |

# Trademarks and special notices



References in this document to Lenovo products or services do not imply that Lenovo intends to make them available in every country.

Lenovo, the Lenovo logo, and XClarity are trademarks of Lenovo.

Intel, the Intel logos, and Xeon are trademarks of Intel Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

Information is provided "AS IS" without warranty of any kind.

All customer examples described are presented as illustrations of how those customers have used Lenovo products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Information concerning non-Lenovo products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by Lenovo. Sources for non-Lenovo list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. Lenovo has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-Lenovo products. Questions on the capability of non-Lenovo products should be addressed to the supplier of those products.

All statements regarding Lenovo future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. Contact your local Lenovo office or Lenovo authorized reseller for the full text of the specific Statement of Direction.

Some information addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in Lenovo product announcements. The information is presented here to communicate Lenovo’s current investment and development activities as a good faith effort to help with our customers' future planning.

Performance is based on measurements and projections using standard Lenovo benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

Photographs shown are of engineering prototypes. Changes may be incorporated in production models.

Any references in this information to non-Lenovo websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this Lenovo product and use of those websites is at your own risk.