MapR Enterprise Edition & Enterprise Database Edition

Reference Architecture

A PSSC Labs Reference Architecture Guide
June 2015

20432 North Sea Circle, Lake Forest, CA 92630 * www.pssclabs.com * tel: 949-380-7288 * fax: 949-380-9788

PSSC LABS

Introduction

PSSC Labs continues to bring innovative compute server and cluster platforms to market. Focusing on specific applications where performance and reliability are critical, highlights PSSC Labs strengths. The Apace Hadoop data framework requires substantial compute and storage capabilities coupled tightly together. MapR Enterprise Edition & Enterprise Database Edition adds high availability, enhanced performance, and extensive data management capabilities.

PSSC Labs is the first manufacturer to design a server platform specifically for Hadoop. Introducing the World's highest density, lowest power consuming, Enterprise Ready Big Data server platform designed specifically for Hadoop workloads. The Big Data BD12000 Enterprise Server offers the absolute highest possible compute and storage density combined with high performance Data IO throughput. PSSC Labs already delivers the Big Data BD12000 for small POC to large production clusters spanning hundreds or thousands of nodes.

Key Features:

- MapR Enterprise 4.1 Certified
- Reduce Data Center Footprint By 50%
- Reduce Power Consumption By 40%
- Nearly 50% Greater Data IO Rates
- Patent Pending Tool Free Maintenance

Technical Specifications:

- Up to 12 x 3.5" SATAIII or 14 x SSDs in 1U Rack Space
  - 72 TB using twelve x 6 TB Hard Drives
- Supports UP to 2 x Intel® Xeon® E5 Series Processors
- Supports Up to 256 GB ECC Enterprise Memory
- 10 GigE, 40 GigE and Infiniband Network Support
- Redundant Power Supply
- Redundant Operating System Hard Drives
- Red Hat®, CentOS, Ubuntu, MS Windows® Compatible

"The Big Data BD 12000 platform offers a unique, highly optimized system design with a direct data path for disk I/O for up to 12 SATA/SAS devices in a 1u Rackmount chassis. This hardware explicitly lacks a raid controller and a backplane, as its intended use is for Hadoop."

Jim Scott
Director Enterprise Strategy & Architecture

20432 North Sea Circle, Lake Forest, CA 92630 * www.pssclabs.com * tel: 949-380-7288 * fax: 949-380-9788

Other Big Data Server Platforms

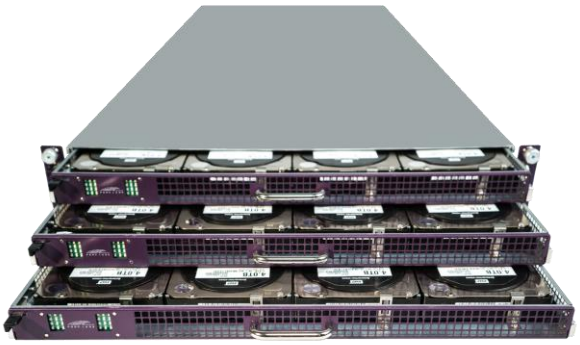
In addition to our Big Data BD 12000 server platform, PSSC Labs offers additional servers that may better compliment MapR workloads. Below are two additional server options that contain 24 x hard drives. Depending on the use case, MapR does recommend 24 x hard drive server configurations.

SureStore U2000SD:

- Up to 26 x 2.5" SSD or SAS Hard Drives
- 2U Rackmount Chassis with Redundant Power
- Supports UP to 2 x Intel® Xeon® E5 Series Processors
- Supports Up to 512 GB ECC Enterprise Memory
- 10 GigE, 40 GigE and Infiniband Network Support
- Redundant Power Supply
- Redundant Operating System Hard Drives
- Red Hat®, CentOS, Ubuntu, MS Windows® Compatible

Big Data BD 24000:

- **Up to 24 x 3.5" , SATA III, SSD or SAS Hard Drives**
- 2U Rackmount Chassis with Redundant Power
- Supports UP to 2 x Intel® Xeon® E5 Series Processors
- Supports Up to 512 GB ECC Enterprise Memory
- 10 GigE, 40 GigE and Infiniband Network Support
- Redundant Power Supply
- Redundant Operating System Hard Drives
- Red Hat®, CentOS, Ubuntu, MS Windows® Compatible



PSSC LABS

BigData BD12000 Sample Configurations

Every organization or use case requires different computing needs. The Big Data BD12000 offers the greatest flexibility possible. Below are three different proposed architectures for different workloads: high density storage, high computational requirements and a balanced configuration.

Big Data BD12000 High Density

- 48 TB Total Storage
- 12 Xeon E5 Processor Cores
- 64 GB ECC Memory
- 2 x 10GigE Network Ports
- 2 x GigE Network Ports
- Remote IPMI Management
- CentOS Linux OS
- Power Draw Estimate
  215 Watt Idle / 285 Watt Full Load

Big Data BD12000 High Compute

- 24 TB Total Storage
- 20 Xeon E5 Processor Cores
- 256 GB ECC Memory
- 2 x 10GigE Network Ports
- 2 x GigE Network Ports
- Remote IPMI Management
- CentOS Linux OS
- Power Draw Estimate
  265 Watt Idle / 355 Watt Full Load

Big Data BD12000 Balanced

- 36 TB Total Storage
- 16 Xeon E5 Processor Cores
- 128 GB ECC Memory
- 2 x 10GigE Network Ports
- 2 x GigE Network Ports
- Remote IPMI Management
- CentOS Linux OS
- Power Draw Estimate
  230 Watt Idle / 320 Watt Full Load

A Sample of Organizations Currently Using PSSC Labs for Hadoop.

collective

OPOWER

Integral Ad Science

OpenX

cardlytics®

pulsepoint™

20432 North Sea Circle, Lake Forest, CA 92630 * www.pssclabs.com * tel: 949-380-7288 * fax: 949-380-9788

Big Data RAX : MapR Enterprise Edition Validated Turn Key Cluster Solutions

PSSC Labs offers a complete, turn-key cluster that is ready to run on delivery. With over 2000+ Computing & Big Data Clusters delivered to date, PSSC Labs understands everything that is necessary for a successful deployment. All necessary hardware including servers, network equipment, power and infrastructure are included. PSSC Labs Cluster Engineers preconfigure network, storage, operating system and BIOS settings to the end user's specifications. MapR Enterprise Edition can be installed at PSSC Labs factory. The final step is the running of sample data sets to ensure proper functionality and performance. Each Big Data RAX Turn Key Cluster is custom configured for specific needs and budget.

Below is an overview of each different server platform PSSC Labs offers for MapR® Enterprise Edition turn-key deployments. Depending on the complexity of the environment, some software resources can be installed on different server platforms.

MapR-OOP 12000 DATA NODE		
Tech Specs	**Key Features**	**Software Resource**
○ 1U High Density Form Factor ○ 2 x Intel® Xeon® E5 Processors ○ 12 x SATAIII or SAS Hard Drives or 14 x SSDs ○ 12 TB to 72 TB Storage Capacity ○ 32 GB to 256 GB ECC Memory ○ 2 x GigE Network Adapters ○ Optional 10 GigE, 40 GigE, Infiniband Support ○ Dedicated IPMI / iKVM	○ Enterprise Platform ○ Redundant Power Supply ○ Improved Data IO Throughput ○ 40% Reduction in Power Consumption ○ 2 x the Density of Standard Server ○ Flexible Configuration Options ○ 3 Year Warranty Included (24 x 7 x 365 NBD Available)	○ YARN/MRv1 Daemons ○ MapR-Fileserver Daemons ○ MapR-DB ○ Zookeeper ○ Apache Drill Daemon ○ Solr Servers

SURESTORE U2000 MANAGEMENT NODE (ONLY RECOMMEDNED FOR 50+ NODE CLUSTERS)		
Tech Specs	**Key Features**	**Software Resources**
○ 2 x Intel® Xeon® E5 Processors ○ 1 TB to 24 TB SATA III, SAS, SSD Hard Drives ○ 32 GB to 256 GB ECC Memory ○ 2 x GigE Network Adapters ○ Optional 10GigE, 40GigE, Infiniband Support ○ Dedicated IPMI / iKVM	○ Enterprise Platform ○ Redundant Power Supply ○ Redundant Storage ○ Raid Levels 0,1,5,6,10,50 ○ Flexible Configuration Options ○ 3 Year Warranty Included (24 x 7 x 365 NBD Available)	○ MapR-MCS management ○ CLDB ○ JobTracker ○ Standby JobTracker ○ Zookeeper ○ Oozie

PSSC LABS

SURESTORE U1000 EDGE NODE		
Tech Specs	**Key Features**	**Resources**
○ 2 x Intel Xeon E5 Processors ○ 1 TB to 12 TB SATAIII, SAS, SSD Hard Drives ○ 32 GB to 256 GB ECC Memory ○ 2 x GigE Network Adapters ○ Optional 10GigE, 40GigE, Infiniband Support ○ Dedicated IPMI / iKVM	○ Enterprise Platform ○ Redundant Power Supply ○ Redundant Storage ○ Raid Levels 0,1,5,6,10,50 ○ Flexible Configuration Options ○ 3 Year Warranty Included (24 x 7 x 365 NBD Available)	○ Hive command line client ○ Hadoop command line client ○ Drill command line client ○ Flume Agents ○ Hue Server ○ MapR-FS Posix client

MapR-Rax Turn-Key Cluster Sample Configurations

MapR Rax 300

- 300 TB Total Storage
- 6 Data Nodes
- 36 Xeon E5 Processor Cores
- 196 GB ECC Memory
- 10GigE Network Backplane
- Remote IPMI Management
- CentOS Linux OS
- MapR Installation Service
- MapR Validation Service

MapR Rax 500

- 500 TB Total Storage
- 1 Edge Node
- 10 Data Nodes
- 160 Xeon E5 Processor Cores
- 1280 GB ECC Memory
- 10GigE Network Backplane
- Remote IPMI Management
- CentOS Linux OS
- MapR Installation Service
- MapR Validation Service
- Rack & Roll Service

MapR Rax 1500

- 1500 TB Total Storage
- 1 Edge Node
- 30 Data Nodes
- 480 Xeon E5 Processor Cores
- 3840 GB ECC Memory
- 10GigE Network Backplane
- Remote IPMI Management
- CentOS Linux OS
- MapR Installation Service
- MapR Validation Service
- Rack & Roll Service

"We believe strongly in our ability to deliver the highest performance, highest reliability server platforms to MapR Enterprise end users. Our experience delivering clusters ranging from several hundred TBs to several dozen PBs ensures a successful MapR Enterprise deployment."

Larry Lesser
PSSC Labs, CTO

20432 North Sea Circle, Lake Forest, CA 92630 * www.pssclabs.com * tel: 949-380-7288 * fax: 949-380-9788

PSSC LABS

Total Cost of Ownership Comparison

PSSC Labs goal is to offer solutions with the absolute lowest total cost of ownership. The below chart compares different server manufacturer's solution for a 1 Petabyte (raw) Hadoop environment. PSSC Labs Big Data BD 1000 requires 50% less rack space and consumes 40% less power. Coupled with MapR's built-in transparent compression, this maximizes usable capacity and cost per usable TB for data within your environment.

	PSSC Labs Big Data RAX 1000 for 1PB Total Storage Space	Dell Configuration for 1PB Total Storage	HP Configuration for 1PB Total Storage	Cisco Configuration for 1PB Total Storage
Required Data Center Footprint	Single x 42U Rack	Two x 42U Rack	Two x 42U Rack	Two x 42U Rack
Power Consumption Estimate*	4300 Watts Total @ Idle 5500 Watts Total @ Load	5800 Watts Total @ Idle 8500 Watts Total @ Load	6000 Watts Total @ Idle 8800 Watts Total @ Load	5700 Watts Total @ Idle 8700 Watts Total @ Load
Required Power Circuits	Single x 30 Amp / 208V / Single Phase	Two x 30 Amp / 208V / Single Phase	Two x 30 Amp / 208V / Single Phase	Two x 30 Amp / 208V / Single Phase
Pre-installation and Validation of Cloudera Enterprise 5 at Factory	Yes. MapR Enterprise Edition preinstalled and tested.	No. Additional services and fees required.	No. Additional services and fees required.	No. Additional services and fees required.
Onsite Physical Installation	Yes. Cluster arrives pre-racked, cabled and labeled.	No. Additional services and fees required.	No. Additional services and fees required.	No. Additional services and fees required.
Cluster Management Training	Yes.	No. Additional services and fees required.	No. Additional services and fees required.	No. Additional services and fees required.
Dedicated Remote Monitoring Capabilities	Yes. IPMI 2.0 Network Standard	Yes.	Yes.	Yes.
Hardware Warranty	3 Year NBD Service Available.	3 Year NBD Service Available.	3 Year NBD Service Available.	3 Year NBD Service Available.

*Dell, HP and Cicso power estimates based on manufacturers website power draw estimates.