Cisco UCS with MapR: Delivering Advanced Performance for Hadoop Workloads

Solution Brief
February 2013

Highlights

Optimized for Performance

- Cisco UCS® with MapR delivers an integrated Hadoop solution that is specifically engineered to handle the most demanding MapReduce and HBase workloads.

Highly Scalable Platform

- The Cisco Unified Computing System™ (Cisco UCS) provides a highly scalable platform that can be optimized and easily scaled for any size of Hadoop cluster.

Advanced Hadoop Distribution

- MapR brings cutting-edge innovation to make Hadoop easy, dependable, fast, and ready for all big data analytics.

Ease of Management

- Cisco UCS Manager provides unified, embedded management of all computing, networking, and storage-access resources.

Choice of Configurations

- The solution provides a choice of Cisco UCS configurations, letting organizations select performance and capacity as their needs dictate.

Enterprise-Class Support and Services

- Reference configurations from Cisco offer confidence and help accelerate implementation of successful Hadoop deployments with worldwide enterprise-class support from Cisco and MapR.

MapR on the Cisco Unified Computing System™ (Cisco UCS®) delivers a fully optimized Hadoop solution that provides lights-out data center capabilities and ease of use with superior performance for different classes of Hadoop applications.

Big data technology, and Apache Hadoop in particular, are finding use in an enormous number of applications and are being evaluated and adopted by enterprises of all kinds. As this important technology helps transform large volumes of data into actionable information, many organizations are struggling to deploy effective and reliable Hadoop infrastructure that performs and scales and is appropriate for mission-critical applications in the enterprise. Deployed as part of a comprehensive data center architecture, the Cisco UCS with MapR solution delivers a powerful and flexible infrastructure that increases business and IT agility, reduces total cost of ownership (TCO), and delivers exceptional return on investment (ROI) at scale, while fundamentally transforming the way that organizations do business with Hadoop technology.

MapR: A Complete Hadoop Platform

As the technology leader in Hadoop, the MapR distribution provides enterprise-class Hadoop solutions that are fast to develop and easy to administer. With significant investment in critical technologies, MapR offers the industry's most comprehensive Hadoop platform, fully optimized for performance scalability. MapR's distribution delivers more than a dozen tested and validated Hadoop software modules over a fortified data platform, offering exceptional ease of use, reliability, and performance for Hadoop solutions (Figure 1).

- **Ease of use:** The MapR 100 percent POSIX-compliant system allows users to access the Hadoop cluster through industry-standard APIs such as Network File Service (NFS), Open Database Connectivity (ODBC), Linux Pluggable Authentication Modules (PAM), and Representational State Transfer (REST). MapR also provides multitenancy, data-placement control, and hardware-level monitoring of the cluster.
- **Reliability:** The MapR distribution provides lights-out data center capabilities for Hadoop. Features include self-healing of the critical services that maintain the cluster nodes and jobs, snapshots that allow point-in-time recovery of data, mirroring that allows wide-area intercluster replication, and rolling upgrades that prevent service disruption.
- **Performance:** The MapR distribution is twice as fast as any other Hadoop distribution. To provide superior and exceptional perfo rmance over other Hadoop distributions, MapR uses an optimized shuffle algorithm, direct access to the disk, built-in compression, and code written in advanced C++ rather than Java. As a result, the MapR distribution provides the best hardware utilization when compared to any other distribution.

MapR technology innovations bring these capabilities onto a single data platform built for all big data analytics. The platform supports a wide range of applications that process structured and unstructured data stored in files as well as NoSQL databases. This single data platform for all Hadoop workloads further solidifies MapR's value proposition of providing the best ROI for hardware utilization.

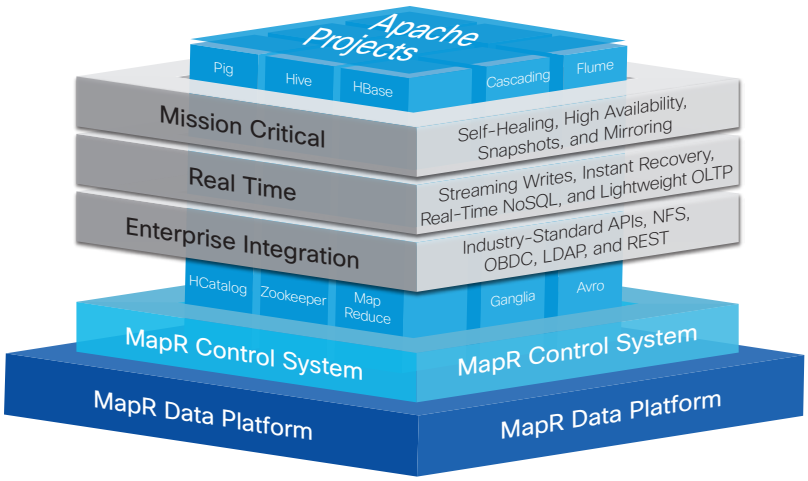

Figure 1. The MapR Hadoop Distribution Assists Mission-Critical Applications by Supporting the Entire Apache Hadoop Stack over a Fortified Data Platform

Cisco UCS with MapR Solution

The Cisco UCS solution for MapR is based on the Cisco® Common Platform Architecture (CPA) for Big Data. The Cisco CPA is a highly scalable architecture designed to meet a variety of scale-out application demands with transparent data and management integration capabilities built using the following components:

- **Cisco UCS 6200 Series Fabric Interconnects** provide high-bandwidth, low-latency connectivity for servers, with integrated, unified management provided for all connected devices by Cisco UCS Manager. Deployed in redundant pairs, Cisco fabric interconnects offer the full active-active redundancy, performance, and exceptional scalability needed to support the large number of nodes that are typical in clusters serving big data applications. Cisco UCS Manger enables rapid and consistent server configuration using service profiles, automating ongoing system maintenance activities such as firmware updates across the entire cluster as a single operation. Cisco UCS Manager also offers advanced monitoring with options to raise alarms and send notifications about the health of the entire cluster.

- **Cisco UCS 2200 Series Fabric Extenders** extend the network into each rack, acting as remote line cards for fabric interconnects and providing highly scalable and extremely cost-effective connectivity for a large number of nodes.
- **Cisco UCS C240 M3 Rack Servers** are designed for a wide range of computing, I/O, and storage-capacity demands in a compact two-rack-unit (2RU) design. Cisco UCS C240 M3 servers are powered by dual Intel® Xeon® processor E5-2600 series CPUs and support up to 768 GB of main memory (128 GB or 256 GB is typical for big data applications). These servers support a range of disk drive options as well as Cisco UCS virtual interface cards (VICs) optimized for high-bandwidth and low-latency cluster connectivity, with support for up to 256 virtual devices.

The solution is offered as reference architecture blueprints and as Cisco UCS SmartPlay solutions that can be purchased by ordering a single part number.

Reference Architecture

Available reference architecture blueprints offer a choice of high-performance and high-capacity options, selected according to the specific computing and storage requirements of the organization.

- **High-performance option:** The high-performance option offers a balance of computing power and I/O bandwidth optimized to achieve an excellent price-to-performance ratio. Equipped for performance, Cisco UCS C240 M3 Rack Servers are powered by two Intel Xeon E5-2665 processors (16 cores), with 256 GB of memory and 24 1-terabyte (TB) Small Form-Factor (SFF) disk drives.
- **High-capacity option:** The high-capacity option is optimized for low cost per terabyte and is built using Cisco UCS C240 M3 Rack Servers powered by two Intel Xeon E5-2640 processors (12 cores), with 128 GB of memory and 12 3-TB Large Form-Factor (LFF) disk drives.

Reference architectures are available in both single and multirack configurations, with considerable capacity built into the Cisco Unified Fabric provided in each rack.

Architectural Scalability

The single-rack configuration provides two fully redundant Cisco UCS 6248UP 48-Port Fabric Interconnects (to connect up to five racks) or two Cisco UCS 6296UP 96-Port Fabric Interconnects (to connect up to 10 racks and 160 servers), along with two Cisco Nexus 2232PP 10GE Fabric Extenders and 16 Cisco UCS C240 M3 Rack Servers (either high-performance or high-capacity CPU configurations). Multirack configurations include two Cisco Nexus 2232PP fabric extenders and 16 Cisco UCS C240 M3 servers for every additional rack.

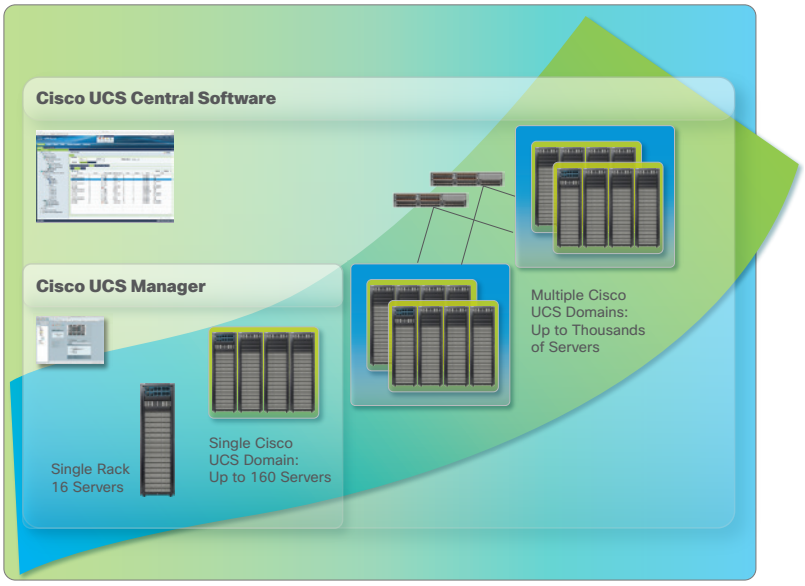

Figure 2. The Cisco UCS with MapR Solution Can Scale to Thousands of Servers

Each server in the configuration connects to the Cisco Unified Fabric through two active-active 10 Gigabit Ethernet links using a Cisco UCS VIC. Each high-performance rack can support up to 256 cores and 32 GBps (SATA) or 48 GBps (SAS) I/O bandwidth. Each high-capacity rack can support up to 576 TB of raw storage.

Up to 160 servers are supported in a single management domain with a pair of fabric interconnects. Scaling beyond 160 servers can be accomplished by interconnecting multiple Cisco UCS domains using Cisco Nexus® 6000 or 7000 Series Switches. With Cisco UCS Central Software, thousdands of servers and hundreds of petabytes (PB) of storage can be managed through a single interface with the same automation that Cisco UCS Manager provides (Figure 2).

Cisco SmartPlay Solutions

Both the high-performance and high-capacity options are available through the Cisco SmartPlay program (Table 1). With only a single part number to order, the program makes it easy to quickly deploy a powerful and secure big data environment without the expense or risk entailed in designing and building a custom solution.

Conclusion

Despite the compelling power of big data technology, deployment of successful, reliable, and high-performance infrastructure for Apache Hadoop can be daunting. The MapR Hadoop distribution provides critical technology advances to make Hadoop implementation easy, dependable, and fast for enterprise-ready deployments. MapR has made crucial investments to improve the underlying technology while maintaining the Apache Hadoop API and application compatibility for broad applicability. The combination of MapR and Cisco UCS brings the power of the MapR distribution to a dependable deployment model that can be implemented rapidly and customized for either high performance or high capacity using Cisco Unified Fabric and powerful and efficient Cisco UCS rack servers. Whether you are deploying a large data center or buying single racks through the Cisco SmartPlay program, the Cisco UCS with MapR solution can be sized to meet the challenges of Hadoop.

For More Information

- For more information about the Cisco SmartPlay program, please visit http://www.cisco.com/go/smartplay.
- For more information about Cisco UCS big data solutions, please visit http://www.cisco.com/go/bigdata.
- For more information about the Cisco CPA for Big Data, please visit http://blogs.cisco.com/datacenter/cpa/.

Table 1. Cisco SmartPlay Solutions Are Optimized for High Performance or High Capacity and Are Tested and Validated for Rapid Deployment

Base Rack Solution	Big Data High Capacity	Big Data High Performance
Part Number	UCS-EZ-BD-HC	UCS-EZ-BD-HP
Computing and Storage	16 Cisco UCS C240 M3 Rack Servers, each with: • 2 Intel Xeon processors E5-2640 at 2.5 GHz • 128 GB of memory • Cisco UCS P81E VIC • 12 LFF 3-TB 7.2K 3.5-inch SAS HDDs • LSI MegaRAID 9266-CV 8i card	16 Cisco UCS C240 M3 Rack Servers, each with: • 2 Intel Xeon processors E5-2690 at 2.9 GHz • 256 GB of memory • Cisco UCS P81E VIC • 24 SFF 1-TB 7.2K SFF SATA HDDs • LSI MegaRAID 9266-CV 8i card
Network	10-Gbps unified fabric supported by: • 2 Cisco UCS 6296UP 96-Port Fabric Interconnects • 2 Cisco Nexus 2232PP 10GE Fabric Extenders	

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices**.

LE-38203-00 02/13