

Cisco UCS Common Platform Architecture for Big Data with MapR: Delivering Advanced Performance for Hadoop Workloads



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

HIGHLIGHTS
<p>Optimized for Performance</p> <ul style="list-style-type: none"> The Cisco UCS CPA for Big Data with MapR delivers an integrated Hadoop solution that is specifically engineered to handle the most demanding big data workloads.
<p>Highly Scalable Platform</p> <ul style="list-style-type: none"> Cisco UCS CPA provides a highly scalable platform that can be optimized and easily scaled for any size of Hadoop cluster.
<p>Advanced Hadoop Distribution</p> <ul style="list-style-type: none"> MapR brings the latest architectural innovations to make Hadoop easy, dependable, and fast, helping ensure big data production success.
<p>Ease of Management</p> <ul style="list-style-type: none"> Cisco UCS Manager provides unified, embedded management of all computing, networking, and storage-access resources.
<p>Choice of Configurations</p> <ul style="list-style-type: none"> The solution provides a choice of Cisco UCS configurations, letting organizations select performance and capacity as their needs dictate.
<p>Enterprise-Class Support and Services</p> <ul style="list-style-type: none"> Reference configurations from Cisco offer confidence and help accelerate implementation of successful Hadoop deployments with worldwide enterprise-class support from Cisco and MapR.

Big data technology, and Apache Hadoop in particular, is finding use in an enormous number of applications, and enterprises of all kinds are evaluating and adopting it. Because 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 CPA 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 organizations do business with Hadoop technology.

The Cisco UCS CPA for Big Data improves both performance and capacity, featuring the Intel Xeon E5-2600 v2 family of processors, industry-leading storage density, and a transparent cache acceleration for Big Data.

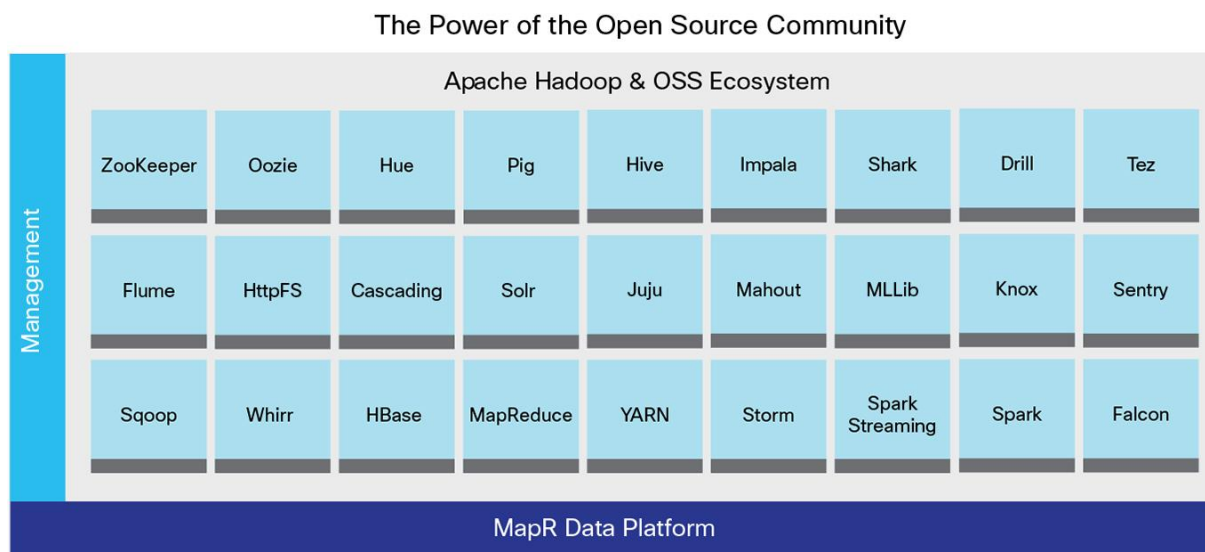
MapR: A Complete Hadoop Platform

As the technology leader in Hadoop, the MapR Distribution provides an enterprise-class, high-performance Hadoop solution that is fast to develop and easy to administer. With significant investment in architectural innovations, MapR 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 Portable Operating System Interface (POSIX)-compliant system allows users to access the Hadoop cluster through industry-standard application programming interfaces (APIs) such as Network File System (NFS), Open Database Connectivity (ODBC), Linux Pluggable Authentication Modules (PAM), and Representational State Transfer (REST). MapR also provides multi-tenancy, data-placement control, and hardware-level monitoring of the cluster.
- **Reliability:** MapR 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 consistent point-in-time recovery of data, mirroring that allows wide-area intercluster replication, and rolling upgrades that prevent service disruption during software upgrades.
- **Performance:** MapR is twice as fast as any other Hadoop distribution. To provide superior and exceptional performance 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, MapR distribution provides the best hardware usage 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 the MapR value proposition of providing the best ROI for hardware use.

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



Cisco UCS CPA with MapR Solution

The Cisco UCS CPA with MapR solution is based on the Cisco[®] Common Platform Architecture for Big Data. With its fabric-based infrastructure, Cisco UCS CPA for Big Data delivers exceptional performance, capacity, management simplicity, and scale to help customers derive value from the most challenging big data deployments.

The Cisco UCS 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 Cisco UCS Manager providing integrated, unified management for all connected devices. 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 Manager 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 v2 series CPUs, and they support up to 768 GB of main memory (128 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 CPA SmartPlay Solution Paks that you can purchase 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.

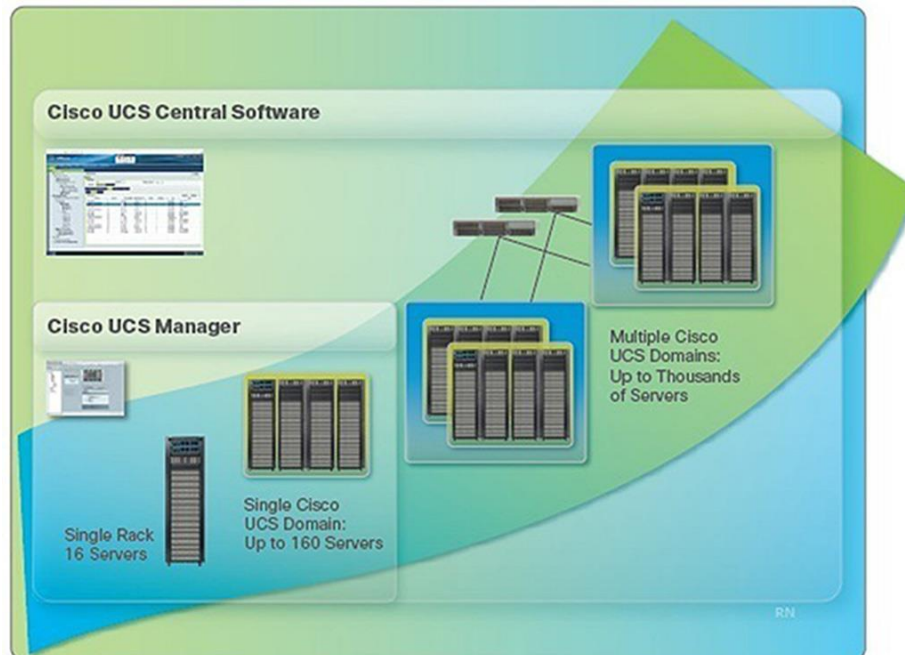
- **Performance and capacity balanced option:** This configuration provides an excellent balance of computing power and storage capacity for Hadoop and NoSQL databases. It supports up to 32 Gbps of I/O bandwidth and 384 terabytes (TB) of unformatted storage and scales up to 10 racks without additional switches.
- **Capacity optimized option:** This option provides a high -capacity storage configuration for storage-intensive Hadoop deployments. The configuration supports up to 768 TB of unformatted storage per rack for a total of 7.68 petabytes (PB) when scaled to a 10-rack configuration.
- **Capacity optimized with flash memory option:** This solution accelerates performance with a transparent, high-performance flash -memory cache powered by LSI Nytro MegaRAID technology. The card has 200 GB of flash memory that you can use as a transparent cache tier for hard-disk drives and operating system images, freeing all 12 hard-disk drives for data. It offers 768 TB of unformatted storage and 3.12 TB of flash memory per rack, for a total of 7.68 PB and 31.25 TB of flash memory per domain.

Reference architectures are available in both single- and multi-rack configurations. Multi-rack configurations include two Cisco Nexus 2232PP fabric extenders and 16 Cisco UCS C240 M3 Rack-Mount Servers for every additional rack.

Architectural Scalability

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

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



Cisco SmartPlay Solution Paks

Solution paks 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. The solution scales by adding servers as needed.

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 Distribution for Hadoop provides critical technology advances to make Hadoop implementation easy, dependable, and fast for production 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 CPA brings the power of MapR to a dependable deployment model that you can implement rapidly and customize 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, you can size the Cisco UCS CPA with MapR solution to meet the challenges of Hadoop.

Table 1: Cisco SmartPlay Solution Paks Are Optimized for Performance or Capacity or a Blend of Each and Are Tested and Validated for Rapid Deployment

	Performance Optimized (UCS-SL-CPA2-P)	Performance and Capacity Balanced (UCS-SL-CPA2-PC)	Capacity Optimized (UCS-SL-CPA2-C)	Capacity Optimized with Flash Memory (UCS-SL-CPA2-CF)
Connectivity	<ul style="list-style-type: none"> • 2 Cisco UCS 6248UP 48- Port Fabric Interconnects • 2 Cisco Nexus 2232PP 10 GE Fabric Extenders 	<ul style="list-style-type: none"> • 2 Cisco UCS 6296UP 96- Port Fabric Interconnects • 2 Cisco Nexus 2232PP 10 GE Fabric Extenders 	<ul style="list-style-type: none"> • 2 Cisco UCS 6296UP 96- Port Fabric Interconnects • 2 Cisco Nexus 2232PP 10 GE Fabric Extenders 	<ul style="list-style-type: none"> • 2 Cisco UCS 6296UP 96- Port Fabric Interconnects • 2 Cisco Nexus 2232PP 10 GE Fabric Extenders
Management	<ul style="list-style-type: none"> • Cisco UCS Manager 	<ul style="list-style-type: none"> • Cisco UCS Manager 	<ul style="list-style-type: none"> • Cisco UCS Manager 	<ul style="list-style-type: none"> • Cisco UCS Manager
Servers	8 Cisco UCS C240 M3 Rack Servers, each with: <ul style="list-style-type: none"> • 2 Intel Xeon processors E5-2680 v2 • 256 GB of memory • LSI MegaRaid 9271CV8i card • 24 900-GB 10K SFF SAS drives (168 TB total) 	16 Cisco UCS C240 M3 Rack Servers, each with: <ul style="list-style-type: none"> • 2 Intel Xeon processors E5-2660 v2 • 256 GB of memory • LSI MegaRaid 9271CV8i card • 24 1TB 7.2K SFF SAS drives (384 TB total) 	16 Cisco UCS C240 M3 Rack Servers, each with: <ul style="list-style-type: none"> • 2 Intel Xeon processors E5-2640 v2 • 28 GB of memory • LSI MegaRaid 9271CV8i card • 12 4TB 7.2K LFF SAS drives (768 TB total) 	16 Cisco UCS C240 M3 Rack Servers, each with: <ul style="list-style-type: none"> • Intel Xeon processors E5-2660 v2 • 128 GB of memory • Cisco UCS Nytro MegaRAID 200-GB Controller • 12 4TB 7.2K LFF SAS drives (768 TB total)

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 Cisco CPA for Big Data, please visit <http://blogs.cisco.com/datacenter/cpa/>.
- For more information about MapR and Cisco UCS CPA with MapR, please visit <http://www.mapr.com/cisco>.



© 2014 Cisco and/or its affiliates. All rights reserved. Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)

© 2014 MapR Technologies. All rights reserved. Apache Hadoop, HBase and Hadoop are trademarks of the Apache Software Foundation and not affiliated with MapR Technologies.

C22-731120-00 04/14