Cisco UCS with the Intel Distribution for Apache Hadoop Software

Highlights

Optimized for Performance

· The Cisco Unified Computing System™ (Cisco UCS®) with the Intel® Distribution for Apache Hadoop software integrates feature-enhanced software with Cisco UCS servers based on Intelligent Intel® Xeon® processors to propel performance of the most challenging MapReduce and HBase workloads.

Ease of Deployment

· Cisco UCS Manager and the Intel Manager for Apache Hadoop software automate server infrastructure deployment and scaling, reducing risk of configuration errors that can cause downtime.

Robust Manageability

· The solution provides a single point of management for up to thousands of servers along with their network infrastructure.

Integration with Enterprise Applications

· Big data and enterprise applications can coexist in the same system, sharing high-bandwidth connectivity so that analytic results can be quickly put to use.

Architectural Scalability

· The solution is designed to grow to its maximum scale without the need for complex layers of switching infrastructure.

Enterprise-Class Support

· Intel provides technical support and professional services for the Intel Distribution. Cisco provides support and services for Cisco UCS.

The Cisco Unified Computing System™ (Cisco UCS®) with the Intel® Distribution for Apache Hadoop software uses the power of hardware-enhanced software to deliver performance, capacity, and security for enterprise-class Hadoop deployments.

Cisco and Intel have a long history of collaboration and innovation that was first demonstrated with the announcement of the Cisco Unified Computing System in 2009. In their long-term collaboration, the two companies have worked together to design and deliver the next generation of open standards-based architectures for enterprises, now extending to big data. The solution combines the Intel Distribution for Apache Hadoop software with the Cisco® Common Platform Architecture (CPA) for Big Data. The result is an enterprise-class solution that delivers performance and capacity while reducing risk and accelerating deployment.

The Rise of Big Data Technology

Big data technology, and Apache Hadoop in particular, is finding use in an enormous number of applications and is being evaluated and adopted by enterprises of all sizes. 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. Many of the challenges arise from the friction between the rapid pace of change inherent in open-source software and the need for enterprise-class performance, reliability, and support.

A Unique Solution from Industry Leaders

Cisco UCS with the Intel Distribution for Apache Hadoop software was development by the two companies to help reduce the time and risk of Hadoop deployment by enhancing features and controlling the release cycle and then optimizing the resulting software for outstanding performance and scalability when it is run on the Cisco CPA. With its enterprise-class support, the solution is a customer-centered platform that can be rapidly deployed, scaled on demand, and secured. The solution has the performance and reliability that organizations need to support their enterprise applications.

Cisco UCS with the Intel Distribution for Apache Hadoop software features:

- **Powerful computing infrastructure:** Cisco UCS servers are powered by the Intel® Xeon® processor E5 family, the core of a flexible and efficient data center that meets diverse business needs. This family of processors is designed to deliver versatility, with an outstanding combination of performance, built-in capabilities, and cost effectiveness. With these processors, I/O latency is dramatically reduced with Intel Integrated I/O, which helps eliminate
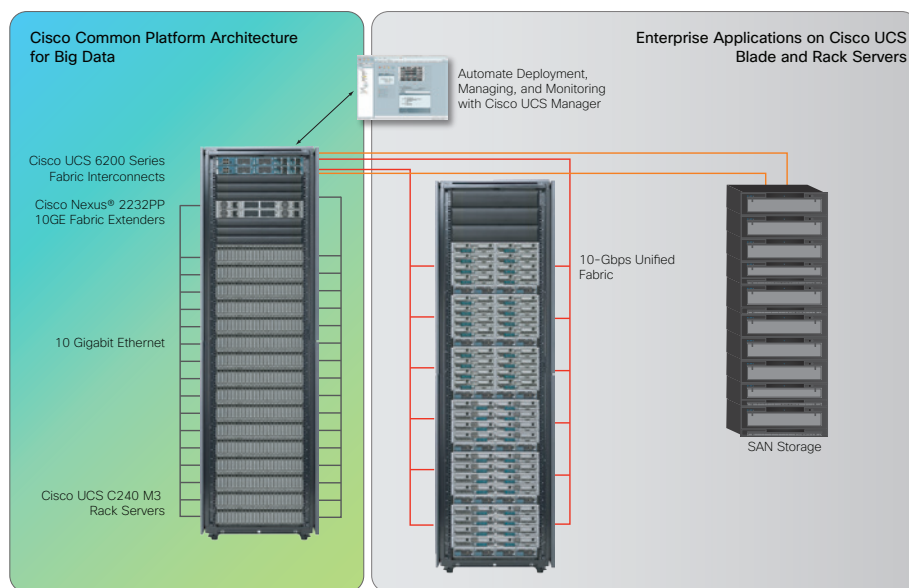
Figure 1. Cisco CPA for Big Data Integrates with Enterprise Applications in a Single Management Domain

data bottlenecks, streamline operations, and increase agility. Complementing the processing power of these servers is the massive storage capacity of Cisco UCS C240 M3 Rack Servers. The servers offer up to 24 Small Form-Factor (SFF) disk drives in the performance-optimized configuration or 12 Large Form-Factor (LFF) disk drives in the capacity-optimized configuration.

- **High-performance unified fabric:** The solution's low-latency, lossless 10-Gbps unified fabric is fully redundant. Through its active-active configuration, the fabric delivers high performance and scalability for up to 160 servers in a single switching domain and thousands of servers in a single management domain.

- **Ease of deployment:** Cisco UCS is the first unified system built from the beginning so that every aspect of server personality, configuration, and connectivity is set on demand, through Cisco UCS Manager. Through the powerful concept of Cisco service profiles, the Hadoop cluster's servers can be configured rapidly and automatically without the risk of configuration drift that can lead to errors that cause downtime. Unified management in Cisco UCS enables greater agility and more rapid deployment.

- **Robust manageability:** Big data environments can consist of hundreds of servers, resulting in immense management complexity. Cisco UCS provides a single point of management for the entire

system: for both blade servers supporting enterprise applications and rack servers supporting big data applications. With the system's self-aware, self-integrating infrastructure, IT departments can proactively monitor the system and reduce operating costs.

- **Integration with enterprise applications:** Big data environments need high-speed connectivity to transfer results to enterprise applications. The Cisco solution can host the Intel Distribution for Apache Hadoop software and enterprise applications from vendors including Microsoft, Oracle, and SAP in the same management and connectivity domains, further simplifying data center management (Figure 1).

- **Architectural scalability:** The system is designed with logically centralized connectivity management that is physically distributed across the racks and blade chassis that house big data and enterprise applications. After the initial system is established, it is designed to grow to maximum size without the need to add any new switching components or redesign the system's connectivity in any way. The solution can be deployed a rack at a time, with the initial rack hosting the system's fabric interconnects (described later in this document). Subsequent racks use Cisco fabric extenders, low-cost, low-power-consumption devices that bring the unified fabric to each server in the rack with no additional points of management.

- **Enterprise service and support:** Enterprises using Apache Hadoop to help with business-critical decisions want to know that the vendors providing the solution have the expertise to help them quickly proceed through the initial design, deployment, and testing. They also need to have confidence that they will receive timely and professional support if a critical component fails. One of the factors that makes this solution unique is the collaboration between Cisco and Intel support to make Cisco UCS with the Intel Distribution for Apache Hadoop software a fully supported, enterprise-class solution.

Intel Distribution for Apache Hadoop Software

The Intel Distribution for Apache Hadoop software is a controlled distribution based on the Apache Hadoop software, with feature enhancements, performance optimizations, and security options that are responsible for the solution's enterprise quality. The Intel Distribution for Apache Hadoop software includes (Figure 2):

- **Intel Manager:** The Intel Manager for Apache Hadoop software streamlines Hadoop cluster configuration, management, and resource monitoring. This powerful, easy-to-use, web-based tool allows IT departments to focus critical resources and expertise on deriving business value from the Hadoop environment rather than worrying about the details of cluster management. The Intel Manager for Apache Hadoop software provides installation and configuration features, wizard-based cluster management, proactive cluster health checks, monitoring and
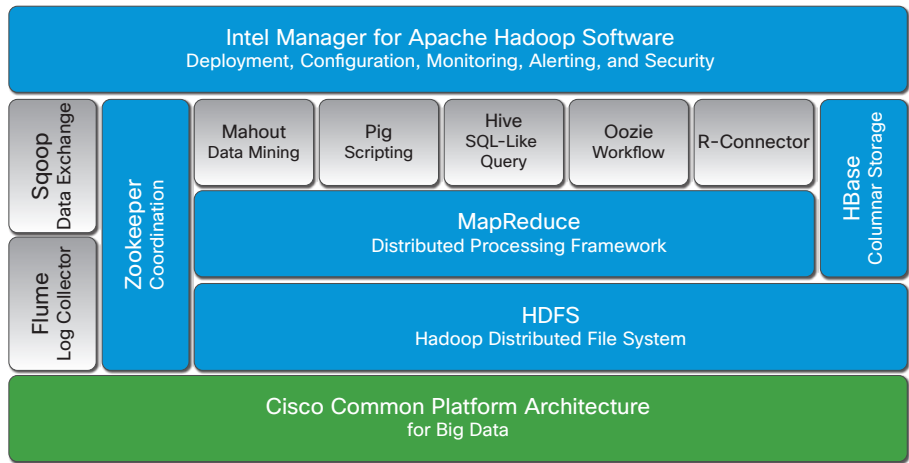
Figure 2. The Solution Combines the Intel Distribution with the Cisco CPA

logging, and secure authentication and authorization.

- **Hadoop Data Storage Framework (HDFS):** HDFS is a distributed, scalable, and portable file system that stores data about the cluster nodes. The Intel Distribution for Apache Hadoop software includes compression and encryption for enhanced security and performance.

- **Data Processing Framework (MapReduce):** This massively parallel computing framework is inspired by Google's MapReduce documents. The Intel Distribution for Apache Hadoop software includes dynamic replication capabilities that intelligently increases and decreases the number of data replicas according to workload characteristics.

- **Real-Time Query Processing Framework:** This component includes HBase, a scalable, distributed, columnar data storage system for large tables and the Hive data warehouse infrastructure for ad-hoc query processing. The Intel Distribution for Apache Hadoop software includes extensions to support big tables across geographically distributed data centers as well as feature additions that improve HBase and Hive performance.

Cisco CPA for Big Data

Cisco UCS with the Intel Distribution for Apache Hadoop software is optimized for high performance on the Cisco Common Platform Architecture 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.

The Cisco CPA is built using Cisco UCS, the first truly unified data center platform that combines industry-standard, x86-architecture servers with networking and storage access in a single system. Cisco UCS is smart infrastructure that is automatically configured through integrated, model-based management to simplify and accelerate deployment of enterprise-class applications and services running in bare-metal, virtualized, and cloud-computing environments. Benefits of the Intel Distribution for Apache Hadoop software available only from Cisco include the capability to unify both big data and enterprise applications in the same centralized management domain.

The Cisco CPA is built using the following components:

- **Cisco UCS 6200 Series Fabric Interconnects** establish a single point of connectivity and management for the entire system. The fabric interconnects provide high-bandwidth, low-latency connectivity for servers, with integrated, unified management for all connected devices provided 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 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 Nexus 2200 Series Fabric Extenders** bring the system's unified fabric to each rack, establishing a physically distributed but logically centralized network infrastructure. These low-cost, low-power-consumption devices act as remote line cards for the fabric interconnects, providing connectivity without adding the cost and management complexity that top-of-rack switches would require. The result is highly scalable and 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 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 that are configured on demand through Cisco UCS Manager.

Choice of Configuration

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

A single-rack configuration provides two fully redundant Cisco UCS 6200 Series 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.

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 I/O bandwidth. Each high-capacity rack can support up to 576 TB of raw storage.

Massive Scalability

The Cisco CPA supports the massive scalability that big data environments demand. Up to 160 servers are supported in a single switching
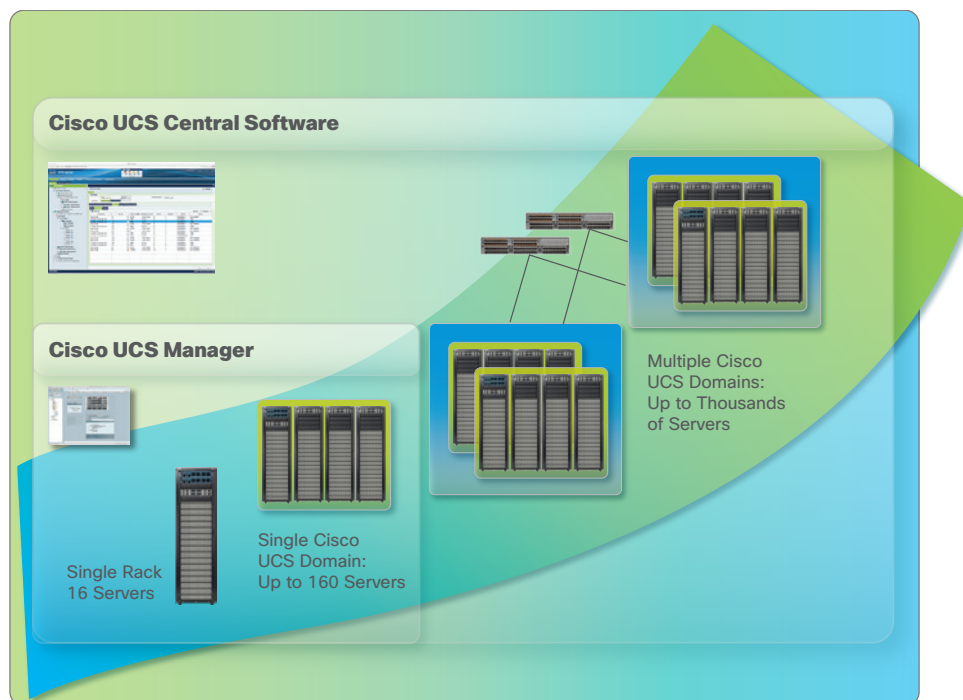
Figure 3. Cisco UCS with the Intel Distribution Can Scale to Thousands of Servers

domain with a pair of Cisco 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, thousands 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 3).

Cisco SmartPlay Configurations

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

Big data technology is becoming compelling for business organizations of all sizes. But although organizations want software that can meet mission-critical needs, they are understandably concerned about the risk and stability of unsupported open-source software.

Cisco UCS with the Intel Distribution for Apache Hadoop software provides critical technology enhancements that allow organizations to easily and safely deploy big data applications in enterprise environments. The combination of the Intel Distribution for Apache Hadoop software and Cisco UCS joins the power of big data with 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. Enterprise-class services can help with design, deployment, and testing, and organizations can continue to rely on these services through controlled and supported releases.

Whether you are deploying a large data center or buying single racks through the Cisco SmartPlay program, Cisco UCS with the Intel Distribution for Apache Hadoop software can be scaled to meet the challenges of any size of organization.

For More Information

- For more information about the collaboration between Cisco and Intel, please visit http://www.cisco.com/go/intel.

- For more information about Cisco UCS, please visit http://www.cisco.com/go/ucs.

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 VIC 1225 • 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 E5-2665 at 2.4 GHz • 256 GB of memory • Cisco UCS VIC 1225 • 24 SFF 1-TB 7.2K SFF SATA HDDs • LSI MegaRAID 9266-CV 8i card
Performance and Capacity per Rack	192 cores, 16 GBps I/O bandwidth, 576 TB storage capacity (raw) 720 TB (typical user storage capacity, 3-way replicated and compressed)	256 cores, 32 GBps I/O bandwidth, 384 TB storage capacity (raw) or 480 TB (typical user storage capacity, 3-way replicated and compressed)
Network	10-Gbps unified fabric supported by: • 2 Cisco UCS 6296UP 96-Port Fabric Interconnects (supports up to 160 servers) • 2 Cisco Nexus 2232PP 10GE Fabric Extenders	

- For more information about the Cisco SmartPlay program, please visit http://www.cisco.com/go/smartplay.

- For more information about Cisco CPA for Big Data, please visit http://blogs.cisco.com/datacenter/cpa/.

CISCO