



# Cisco UCS Integrated Infrastructure for Big Data with Cloudera Enterprise



## Highlights



### Ease of deployment

- Cisco UCS® Manager automates deployment and scaling processes, helping reduce the risk of configuration errors that can cause downtime.



### Scalability for big data workloads

- Cisco UCS Integrated Infrastructure for Big Data solutions offer linear scalability and help simplify essential operations for single-rack and multiple-rack deployments.



### Comprehensive integrated infrastructure

- Cisco UCS Integrated Infrastructure for Big Data solutions include computing, storage, connectivity, and unified management.



### Complete security and data protection

- Cisco UCS Integrated Infrastructure for big data with Cloudera offers complete data protection with built-in support for Kerberos authentication, enterprise-class access controls, and data encryption.



### Simplified management

- Faster and easier big data infrastructure deployment: Cisco UCS Director Express for Big Data offers one-click provisioning, installation, and configuration.
- Cloudera Manager is a holistic interface that provides end-to-end system management and granular visibility into and control over every part of an enterprise data hub.



### Scale to thousands of servers with Cisco® Application Centric Infrastructure (Cisco ACI™)

- With Cisco ACI, you can easily scale the cluster to thousands of nodes. Cisco ACI treats the network as a single entity rather than a collection of switches.

## Cisco and Cloudera Deliver End-To-End Data Management and Governance for Businesses Looking to Tame Big Data

Big data is now a significant element in many industries. Big data is gaining significant traction across multiple industries. Organizations heavily using big data technologies include healthcare, education, and energy along with financial, utility, advertising, retail, public sector and manufacturing. But how do you put all this information to work for you?

The size and complexity of the analytics required to process big data can overpower the computing capabilities of traditional databases. That's why Apache Hadoop was developed. Hadoop can handle large volumes of data—both structured and unstructured—much more efficiently than traditional enterprise data warehouses. Because Hadoop is open source and designed to run on industry standard hardware, the cost savings are significant. Overall savings increase as organizations' data volumes grow.

Cisco offers an industry leading solution for Enterprise Hadoop deployments.

## Big Data Challenges Facing Enterprises Today

Traditional relational database systems (RDBMS) cannot cost-effectively scale to process the large amounts of data routinely being collected and processed by organizations today. Big data technology, Hadoop software in particular, are finding use in an enormous number of applications and are being evaluated and adopted by enterprises of all sizes that are confronted with large amounts of data. 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, security, reliability, and support.

## Hadoop, the Solution for Big Data

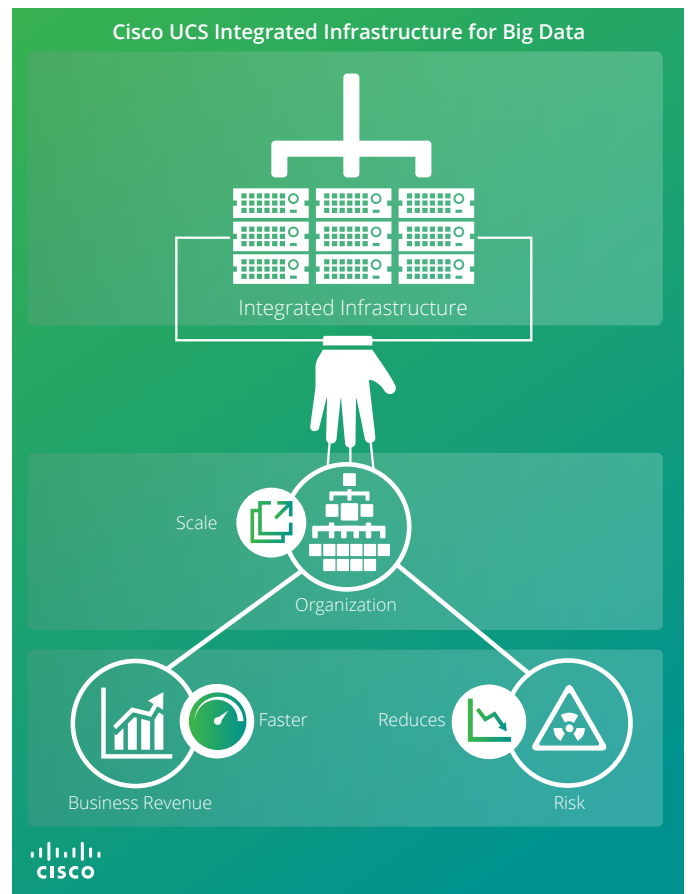
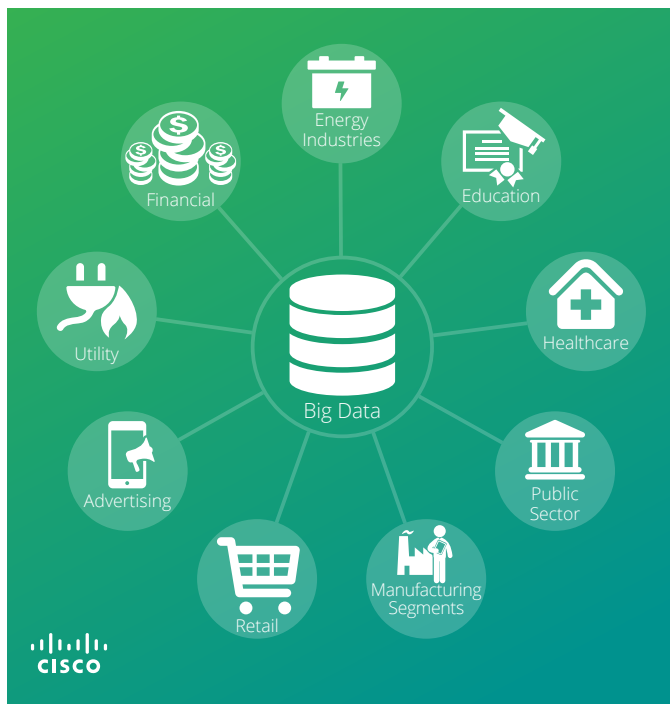
Chief among the big data technologies adopted by enterprises is Hadoop, an open source distributed data management platform. Hadoop was originally developed to address the overpowering size and complexity of the analytics required by enterprises in the era of big data.

Hadoop can handle large volumes of data—both structured and unstructured—at a more scalable and economical fashion.

## Cisco UCS Integrated Infrastructure for Big Data and Cloudera Enterprise

Cisco UCS® Integrated Infrastructure for Big Data includes computing, storage, connectivity, and unified management

capabilities to help companies manage the dramatic increase in data they must cope with today. It is built on Cisco UCS infrastructure using Cisco UCS 6200 Series Fabric Interconnects, Cisco Nexus® 2200 platform fabric extenders, and Cisco UCS C-Series Rack Servers. Installed in pairs, the fabric interconnects offer redundant, active-active connectivity and embedded management using Cisco UCS Manager.

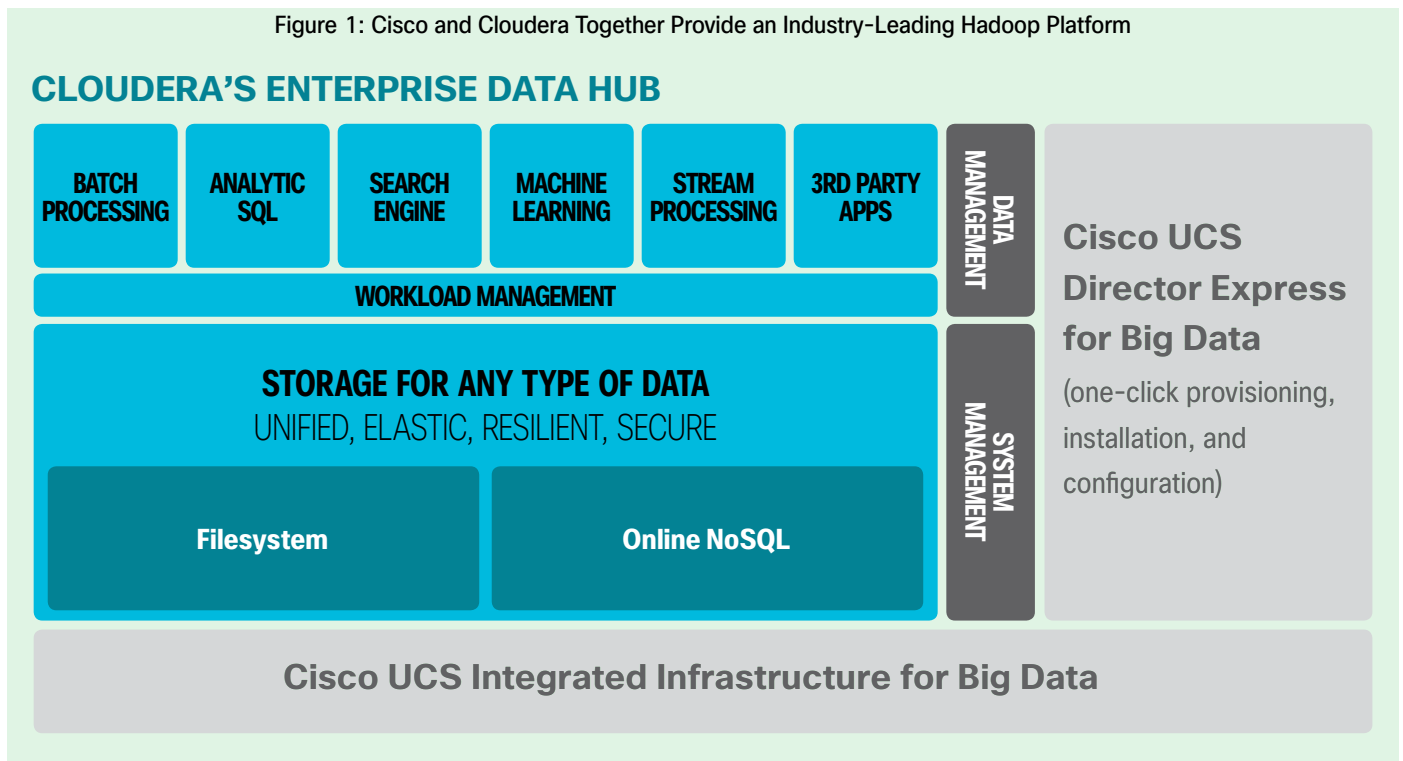


Cloudera is the leading provider of enterprise-ready big data software and services. Cloudera Enterprise includes the market-leading open source Hadoop distribution (CDH), a sophisticated administration and management tool (Cloudera Manager), native end-to-end data governance solution, comprehensive security toolsets, and technical support.

Together, Cisco and Cloudera provide organizations with enterprise-ready data management platform, as well as management integration with an enterprise application ecosystem. They seamlessly combine to provide a uniquely capable, industry-leading architectural platform for Hadoop-based applications (Figure 1).

The Cisco Unified Computing System™ (Cisco UCS) solution for Cloudera is based on Cisco UCS Integrated Infrastructure for Big Data, a highly scalable architecture that includes computing, storage, connectivity, and unified management capabilities and is designed to meet a variety of scale-out application demands.

Figure 1: Cisco and Cloudera Together Provide an Industry-Leading Hadoop Platform



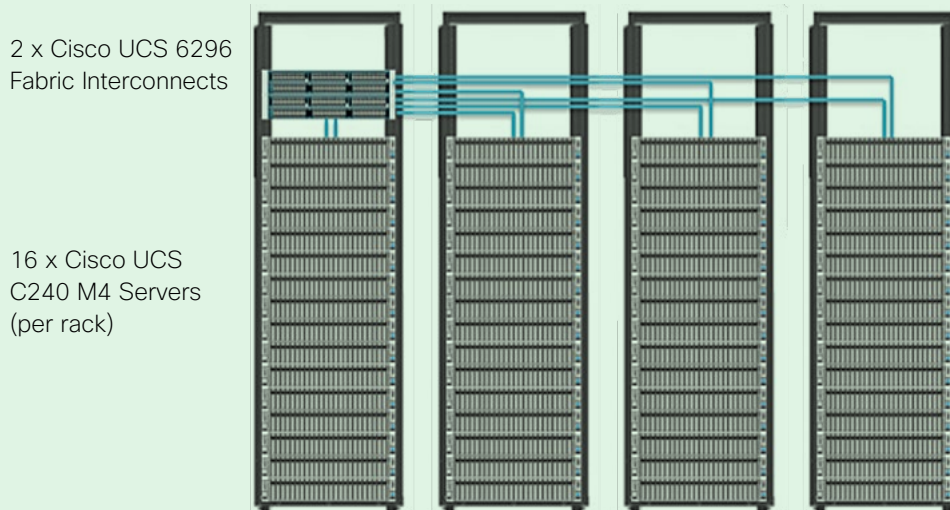
This architecture provides transparent data integration and management integration using the following components (Figure 2):

- Cisco UCS 6200 platform fabric interconnects: The interconnects establish a single point of connectivity and management for the entire system. They 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, the interconnects offer the full active-active redundancy, performance, and exceptional scalability needed to support the large number of nodes that are typical in clusters that serve big data applications. The 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. It also offers advanced monitoring with options to raise alarms and send notifications about the health of the entire cluster.
- Cisco UCS C240 M4 Rack Server: The rack server supports a wide range of computing, I/O, and storage-capacity demands in a compact design. The server is based on the Intel® Xeon® processor E5-2600 v3 series with 12-Gbps SAS throughput, delivering significant performance and efficiency gains over the previous generation of servers. The server uses dual Intel Xeon processor E5-2600 v3 series CPUs and supports up to 768 GB of main memory (128 or 256 GB is typical for big data applications) and a range of disk drive and SSD options. Twenty-four small-form-factor (SFF) disk drives are supported in the performance-optimized option, and 12 large-form-factor (LFF) disk drives are supported in the capacity-optimized option, along with two 1 Gigabit Ethernet embedded LAN-on-motherboard (LOM) ports.
- The Cisco UCS Virtual Interface Card (VIC) 1227 is designed for the M4 generation of Cisco UCS C-Series Rack Servers. The VIC is 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.
- Cisco UCS Director Express for Big Data provides a single-touch solution that automates deployment of Hadoop on Cisco UCS Integrated Infrastructure for Big Data. It also provides a single management pane across both physical infrastructure and Hadoop software. All elements of the infrastructure are handled automatically with little need for user input.

### End-to-End Administration for Hadoop

Cloudera Manager is the industry's first and most sophisticated management

Figure 2: Cisco UCS Integrated Infrastructure for Big Data: A 64-Node Cluster



application for Apache Hadoop and the enterprise data hub. Cloudera Manager sets the standard for enterprise deployment by delivering granular visibility into and control over every part of the data hub – empowering operators to improve performance, enhance quality of service, increase compliance and reduce administrative costs.

Cloudera Manager is designed to make administration of your enterprise data hub simple and straightforward, at any scale. With Cloudera Manager, you can easily deploy and centrally operate the complete Big Data stack. The application automates the installation process, reducing deployment time from weeks to minutes; gives you a cluster-wide, real-time view of nodes and services running; provides a single, central console to enact configuration changes across your cluster; and incorporates a full range of

reporting and diagnostic tools to help you optimize performance and utilization.

### Compliance-Ready Data Governance for Cloudera Enterprise

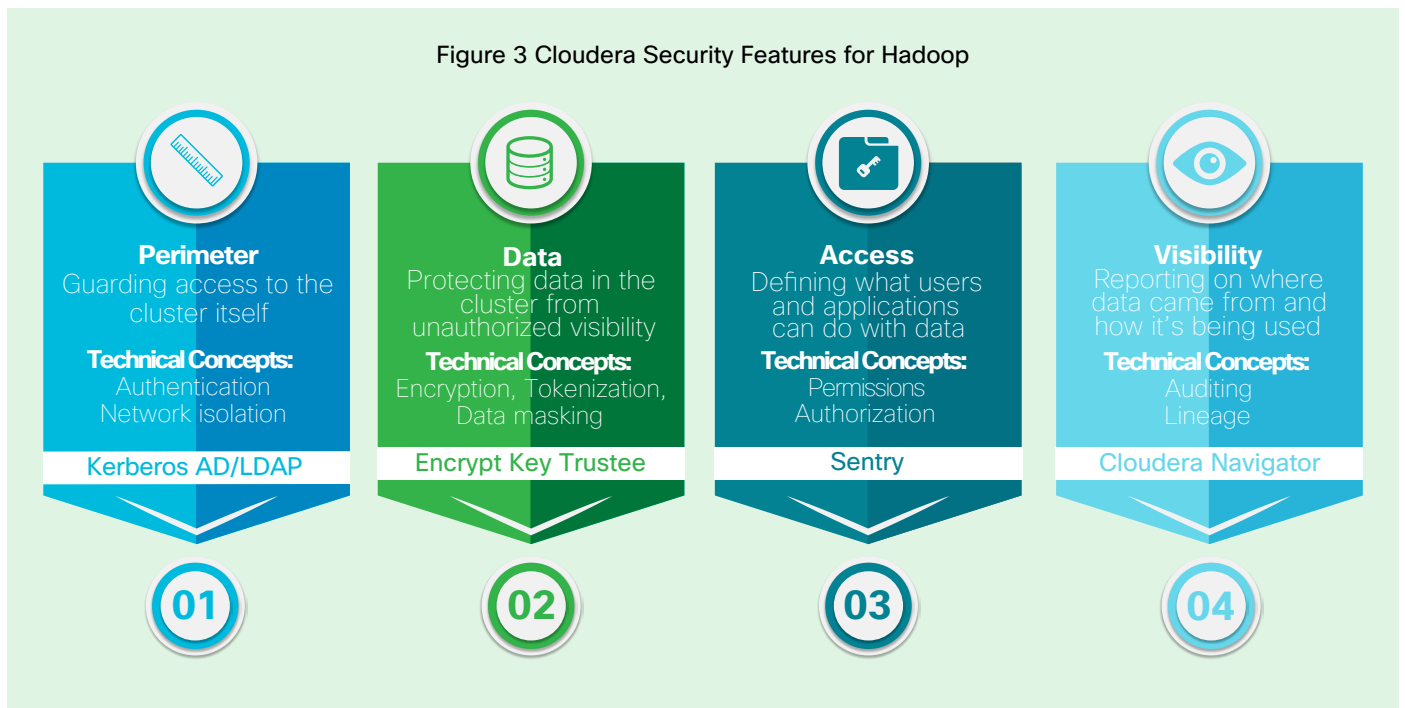
Cloudera Navigator is the only native end-to-end governance solution for Apache Hadoop-based systems. Through a single user interface, it provides visibility for administrators, data managers, data scientists, and analysts to secure, govern, and explore the large amounts of diverse data that land in Hadoop. Cloudera Navigator is part of Cloudera Enterprise’s comprehensive data security and governance offering and is a key part to meeting compliance and regulatory requirements. Features include:

- Guard **the perimeter**: Cloudera Manager makes it easy to secure

your cluster using industry-standard Kerberos, open Lightweight Directory Access Protocol (LDAP) and Microsoft Active Directory (AD), and Security Assertion Markup Language (SAML).

- Protect **data at rest and data in motion**: Cloudera can protect both data at rest and data in motion through encryption and powerful key management capabilities that are integrated into Cloudera Navigator.
- **Control access with Sentry**: Sentry integrates with the Hadoop platform and allows storage of sensitive data while meeting regulatory requirements.
- Gain visibility with **Cloudera Navigator**: Cloudera Navigator is the only native end-to-end governance solution for Hadoop. It provides full visibility into the source and use of the

Figure 3 Cloudera Security Features for Hadoop



data to verify authenticity and easily comply with regulatory requirements.

### Scaling the Architecture with Cisco Application Centric Infrastructure

The Cisco® Application Centric Infrastructure (Cisco ACI™) fabric consists of discrete components that

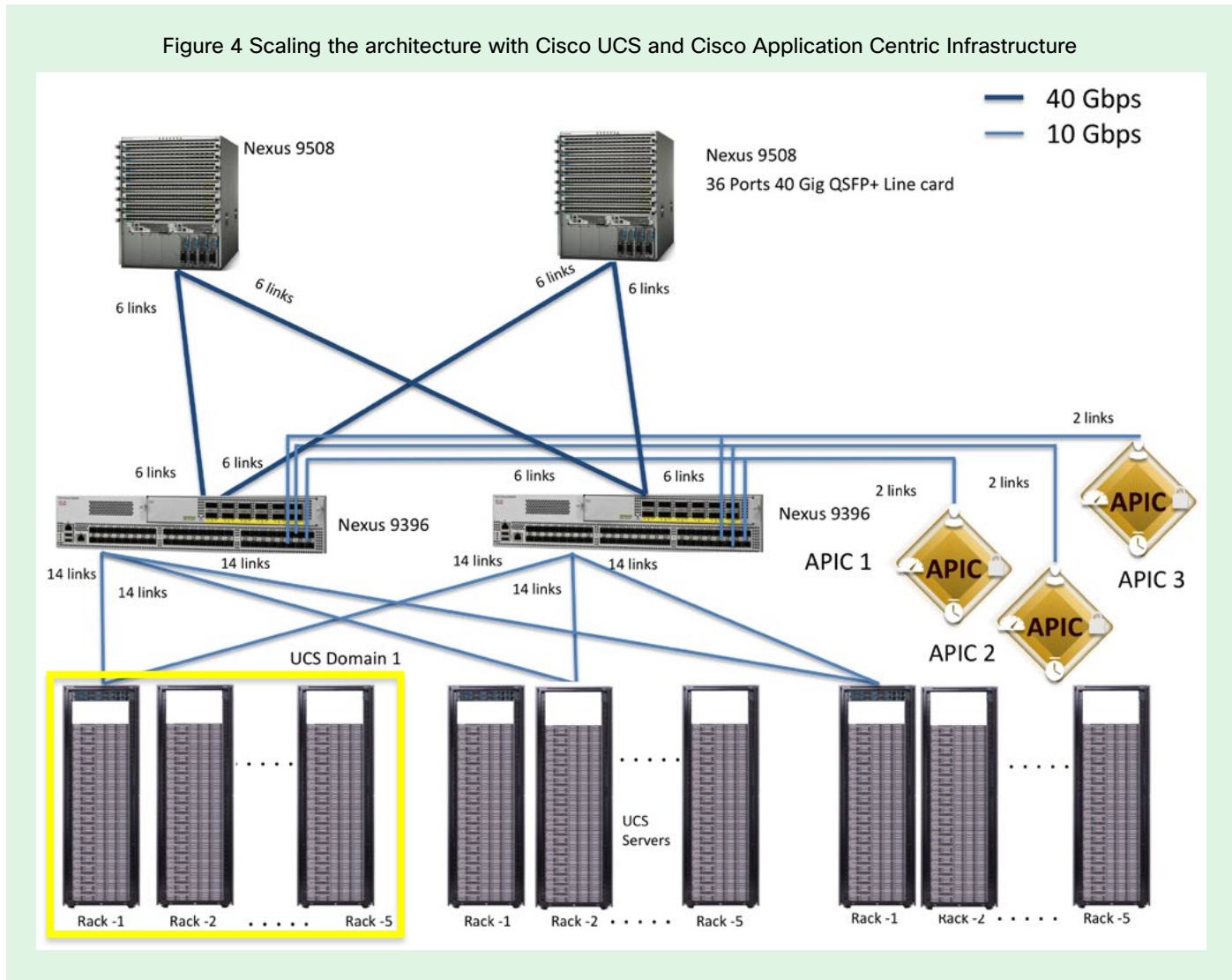
operate as routers and switches but is provisioned and monitored as a single entity. It operates like a single switch and router that provides advanced traffic optimization, security, and telemetry functions, stitching together virtual and physical workloads.

The fabric consists of three major components: the Cisco Application Policy Infrastructure Controller (APIC), spine switches, and leaf switches.

These three components handle both the application of network policy and the delivery of packets.

The system architecture consists of three domains (three pairs of fabric interconnects) connecting to Cisco ACI: two Cisco Nexus 9508 Switches acting as a spine, two Cisco Nexus 9396 (9396PX) Switches as the leaf switches, and three APIC devices as the APIC appliance (Figure 4).

Figure 4 Scaling the architecture with Cisco UCS and Cisco Application Centric Infrastructure



The Cisco UCS C240 M4 servers are directly connected to the fabric interconnect, which connects to the two Cisco Nexus 9396 leaf switches. This mode allows organizations to deploy Cisco UCS Manager capabilities in the fabric interconnects to provision the servers. Up to five racks can be connected to a pair of interconnects, forming a single domain. In addition, up to three domains can be connected to a pair of Cisco Nexus 9396 leaf switches. This topology can scale up to 5760 servers for a fully populated pair of Cisco Nexus 9508 Switches with all eight line cards populated and managed by UCS Central. The Cisco ACI infrastructure supports up to 12 spine switches, allowing tens of thousands of servers.

## Reference Architecture

The current version of the Cisco UCS Integrated Infrastructure for Big Data offers the configurations summarized in Table 1. The configuration used depends on the computing and storage requirements of Hadoop.

## For More Information

For more information about Cisco UCS big data solutions, please visit [http://www.cisco.com/go/bigdata\\_design](http://www.cisco.com/go/bigdata_design).

For more information about Cisco UCS Integrated Infrastructure for Big Data, please visit <http://blogs.cisco.com/datacenter/cpav3/>.



**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](http://www.cisco.com/go/offices).

Cisco and the Cisco Logo are trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and other countries. A listing of Cisco's trademarks can be found at [www.cisco.com/go/trademarks](http://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. (1005R)

Table 1: Cisco UCS Integrated Infrastructure for Big Data with Cloudera

 <h3>Performance Optimized</h3>	 <h3>Capacity Optimized</h3>
<p><b>Connectivity:</b></p> <ul style="list-style-type: none"> <li>• 2 Cisco UCS 6296UP 96 Port Fabric Interconnects</li> </ul>	<p><b>Connectivity:</b></p> <ul style="list-style-type: none"> <li>• 2 Cisco UCS 6296UP 96-Port Fabric Interconnects</li> </ul>
<p><b>Scaling:</b></p> <ul style="list-style-type: none"> <li>• Up to 80 servers per domain</li> <li>• Up to 160 servers per domain with Cisco Nexus 2232PP 10GE Fabric Extender</li> </ul>	<p><b>Scaling:</b></p> <ul style="list-style-type: none"> <li>• Up to 80 servers per domain</li> <li>• Up to 160 servers per domain with Cisco Nexus 2232PP 10GE Fabric Extender</li> </ul>
<p><b>16 Cisco UCS C240 M4 Rack Servers (SFF), each with:</b></p> <ul style="list-style-type: none"> <li>• 2 Intel Xeon processor E5-2680 v3 CPUs</li> <li>• 256 GB of memory</li> <li>• Cisco 12-Gbps SAS Modular RAID Controller with 2 GB of flash-based write cache (FBWC)</li> <li>• 24 x 1.2-TB 10,000-rpm SFF SAS drives (460 TB total)</li> <li>• 2 x 120-GB 6-Gbps 2.5-inch Enterprise Value SATA SSDs for boot</li> <li>• Cisco UCS VIC 1227 (with 2 x 10 Gigabit Ethernet SFP+ ports)</li> </ul>	<p><b>16 Cisco UCS C240 M4 Rack Servers (LFF), each with:</b></p> <ul style="list-style-type: none"> <li>• 2 Intel Xeon processor E5-2620 v3 CPUs</li> <li>• 128 GB of memory</li> <li>• Cisco 12-Gbps SAS Modular RAID Controller with 2 GB of FBWC</li> <li>• 12 x 4-TB 7200-rpm LFF SAS drives (768 TB total)</li> <li>• 2 x 120-GB 6-Gbps 2.5-inch Enterprise Value SATA SSDs for boot</li> <li>• Cisco UCS VIC 1227 (with 2 x 10 Gigabit Ethernet SFP+ ports)</li> </ul>

Scale to tens of thousands of servers with Cisco ACI

For more information about the Cisco Validated Design for the solution with Cisco ACI, please visit [http://www.cisco.com/c/dam/en/us/td/docs/unified\\_computing/ucs/UCS\\_CVDs/Cisco\\_UCS\\_Integrated\\_Infrastructure\\_for\\_Big\\_Data\\_with\\_ACI\\_and\\_Cloudera.pdf](http://www.cisco.com/c/dam/en/us/td/docs/unified_computing/ucs/UCS_CVDs/Cisco_UCS_Integrated_Infrastructure_for_Big_Data_with_ACI_and_Cloudera.pdf).

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