Cisco UCS Integrated Infrastructure for Big Data and Splunk Enterprise for an Advanced Big Data Analytics Platform

Cisco UCS and Splunk Solution Brief

Highlights

Comprehensive Integrated infrastructure

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

Horizontal Scalability to Support Splunk Enterprise Clusters

- Cisco UCS Integrated Infrastructure for Big Data offers linear scalability along with essential operation simplification for single-rack and multiple-rack deployments without adding complex layers of switching infrastructure.

Ease of Deployment and Consistent Server Management

- Cisco UCS Manager automates deployment and scaling, reduces risk of configuration errors, and parallelizes new server installations. It provides a single tool for managing your infrastructure through the GUI, command-line interface (CLI), or XML API, offering you full control of your system.

Frozen Data Storage Support

- The Cisco UCS C3160 Rack Server offers exceptional high-density storage capacity in a small form factor, allowing you to store frozen data that is readily accessible to Splunk Enterprise clusters.

Support for Advanced Splunk Enterprise Features

- Cisco UCS C240 M4 Rack Servers serve as indexers to form the Splunk indexer cluster, providing industry-leading scalability and reliability for mission-critical data storage.
- Cisco UCS C220 M4 Rack Servers serve as search heads to form the Splunk search head cluster, providing a highly available analytics interface for the end user.

Machine data is big data and contains critical insights about IT infrastructure and applications.

Machine-generated data is one of the fastest growing and most complex types of big data. It's also one of the most valuable because it contains a definitive record of all the activity and behavior of user transactions, customer activity, sensor readings, machine behavior, security threats, fraudulent activity, and more.

Modern enterprises generate terabytes (TB) to hundreds of terabytes of machine data every day. This large amount of data poses huge challenges for the organization's IT operations:

- Where does the data come from?
- What are the format and structure of the data?
- What is the volume of the data and at what rate is it generated?
- How is the data collected?
- How is the data stored and its growth managed?
- What data is critical, and what data can the organization do without?
- How can critical but old data be saved in a cost-effective way, yet kept in close quarters?

Splunk Enterprise provides a fast, easy, and secure way to analyze the massive streams of machine data generated by IT systems and technical infrastructure and turn them into invaluable business insights. Splunk software collects, indexes, and harnesses live data generated from almost any source, format, or location, whether data is physical, virtual, or in the cloud. It can collect data from packaged and custom applications, application servers, web servers, databases, networks, virtual machines, hypervisors, and operating systems—without requiring custom parsers, adapters, or a back-end database.

This operational intelligence provides a real-time understanding of what's happening across the IT systems and technology infrastructure so that the business can make informed decisions.

Cisco UCS Integrated Infrastructure for Big Data with Splunk Enterprise Delivers an Advanced Analytics Solution

Cisco UCS® Integrated Infrastructure for Big Data offers a balance of computing power, I/O bandwidth, and storage capacity to meet the needs of Splunk Enterprise. It is built on a modular and unified framework that can rapidly scale from small to very large as needs change. It can grow efficiently along with the organization's desire to process machine data.

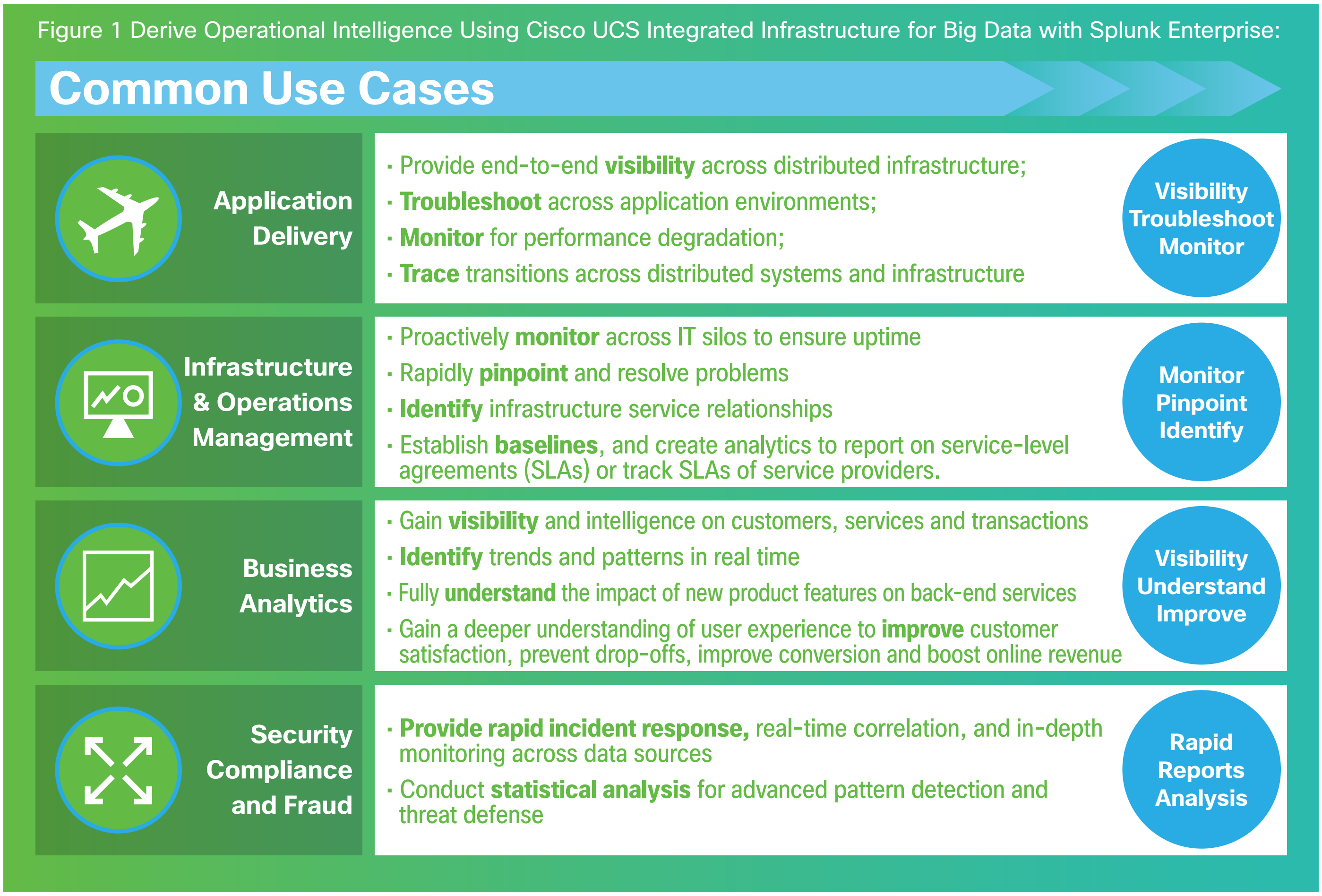
Cisco UCS Manager enables rapid and consistent server configuration using Cisco UCS service profiles, advanced monitoring, and automation of ongoing system maintenance activities across the entire cluster as a single operation. The service profiles help enable consistent and rapid deployment, delivering out-of-the-box performance and scale-out capabilities.

Manage Splunk's Hot, Warm, Cold, and Frozen Index Data with Cisco UCS Integrated Infrastructure for Big Data

This prevalidated solution incorporates advanced features such as indexer clusters, search head clusters, deployment servers, and archival capability to meet the needs of any organization deploying Splunk Enterprise at scale. The solution is built on the Distributed Deployment with High Capacity reference architecture with the addition of a server with Network File System (NFS) capabilities for archiving frozen data.

High-performing storage and I/O-optimized Cisco UCS C240 M4 Rack Servers configured as a Splunk indexer cluster manage the hot, warm, and cold index data generated by Splunk software. New to this solution is the addition of the modular high-density Cisco UCS C3160 Rack Server: an archival node configured as an NFS server for storing Splunk's frozen index data.

Figure 1 Derive Operational Intelligence Using Cisco UCS Integrated Infrastructure for Big Data with Splunk Enterprise:



Common Use Cases

Application Delivery

- Provide end-to-end **visibility** across distributed infrastructure;
- **Troubleshoot** across application environments;
- **Monitor** for performance degradation;
- **Trace** transitions across distributed systems and infrastructure

Visibility Troubleshoot Monitor

Infrastructure & Operations Management

- Proactively **monitor** across IT silos to ensure uptime
- Rapidly **pinpoint** and resolve problems
- **Identify** infrastructure service relationships
- Establish **baselines**, and create analytics to report on service-level agreements (SLAs) or track SLAs of service providers.

Monitor Pinpoint Identify

Business Analytics

- Gain **visibility** and intelligence on customers, services and transactions
- **Identify** trends and patterns in real time
- Fully **understand** the impact of new product features on back-end services
- Gain a deeper understanding of user experience to **improve** customer satisfaction, prevent drop-offs, improve conversion and boost online revenue

Visibility Understand Improve

Security Compliance and Fraud

- **Provide rapid incident response,** real-time correlation, and in-depth monitoring across data sources
- Conduct **statistical analysis** for advanced pattern detection and threat defense

Rapid Reports Analysis

- **Increase application and infrastructure performance** by providing real-time, end-to-end, single-pane visibility across applications and physical, virtual, and cloud infrastructure.
- **Improve IT productivity with comprehensive IT analytics and management functions.** Perform proactive infrastructure monitoring, application availability monitoring, incident management, problem management, and capacity management.
- **Detect, investigate, and view** network, server, storage, virtualization, and cloud infrastructure problems and correlate them with application- or user-related problems.
- **Collect and analyze** real-time metrics about resource use and performance from disparate systems and connected devices (the Internet of Things [IoT]) to meet SLAs.

Deploying and Scaling Splunk's Distributed Search Architecture Made Easy With Cisco UCS

The Cisco UCS Integrated Infrastructure for Big Data for Splunk Enterprise allows the enterprise to scale its Splunk Enterprise deployment to meet its changing needs. Splunk software deployments scale horizontally; so does the Cisco Unified Computing System™ (Cisco UCS). Scalability is achieved by following simple steps: Rack and stack the servers, and with a few clicks in UCS-Manager, auto-discover, provision and scale the infrastructure. Figures 2 and 3 present application and hardware views of the clustered distributed search deployment architecture.

Figure 2 Clustered Distributed Search Deployment Architecture on Cisco UCS: Application View

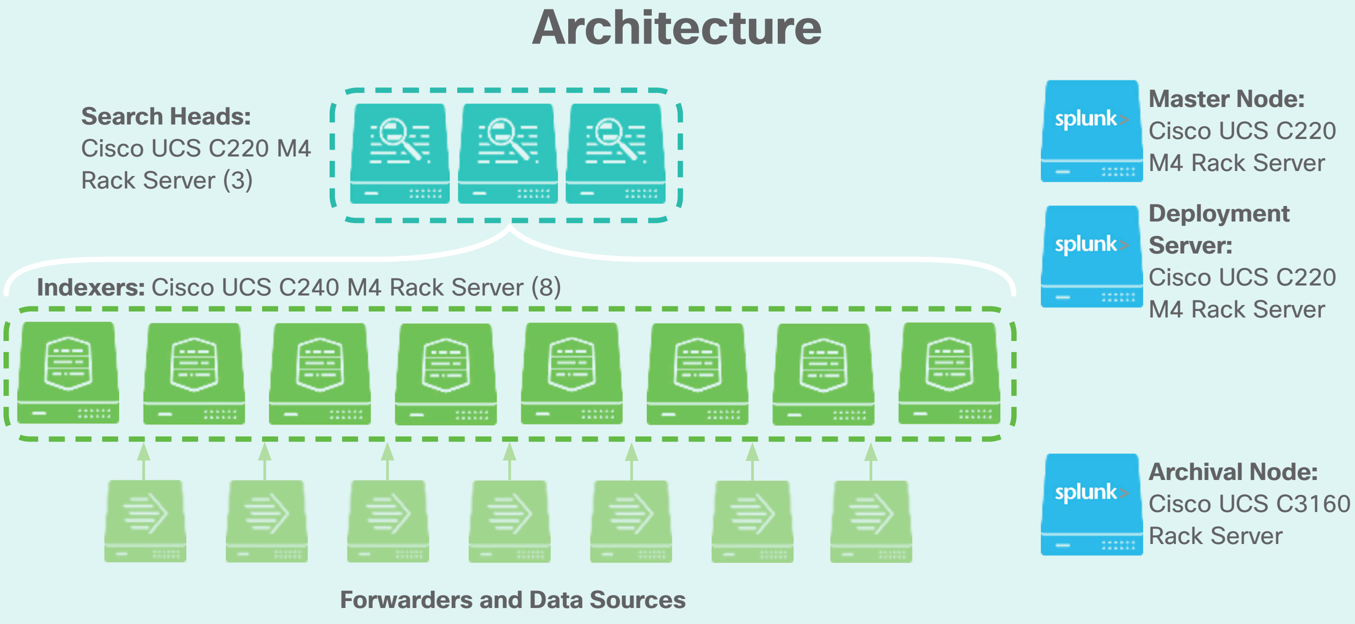
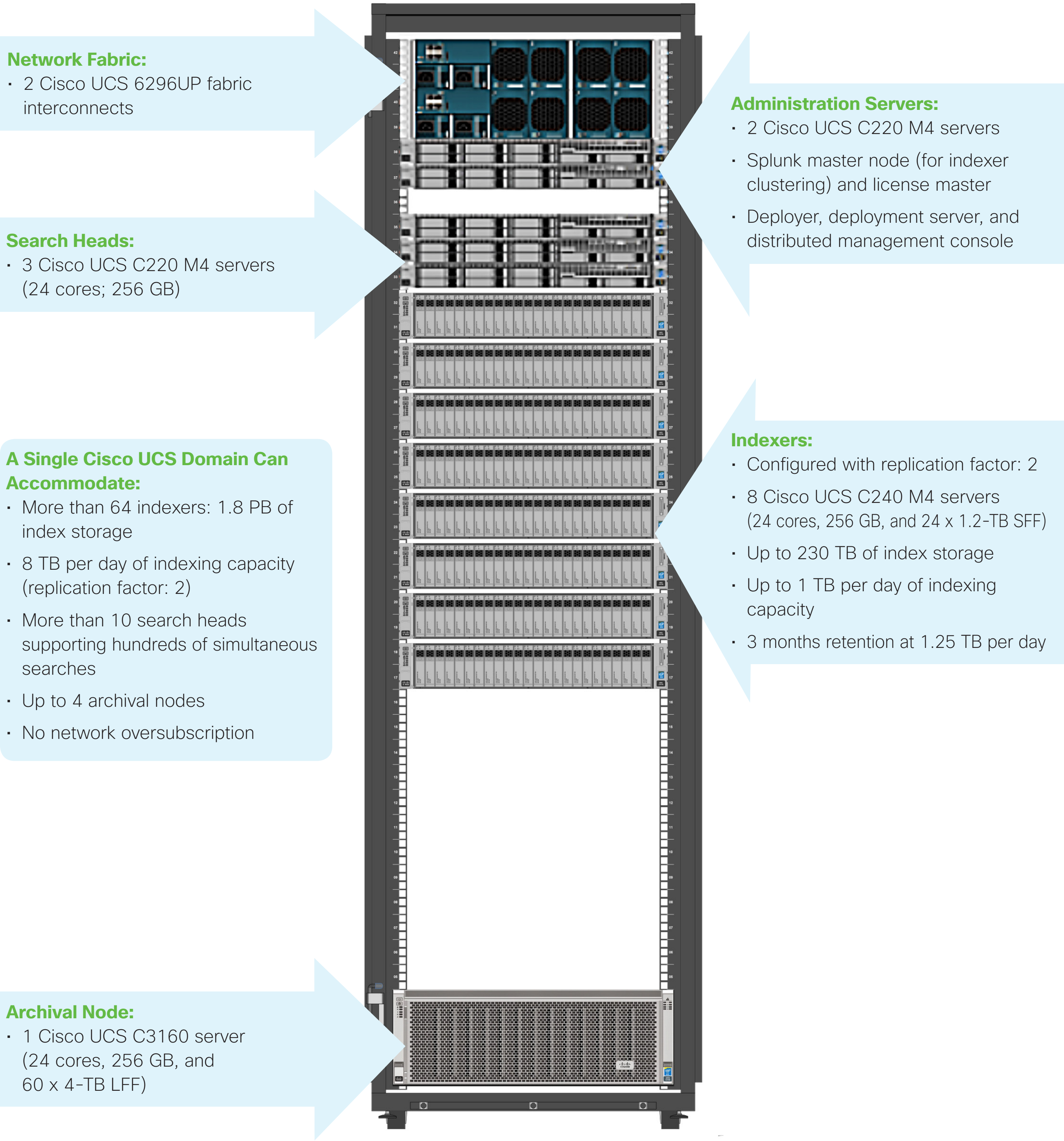

Architecture
Search Heads: Cisco UCS C220 M4 Rack Server (3)
Indexers: Cisco UCS C240 M4 Rack Server (8)
Forwarders and Data Sources
splunk>
Master Node: Cisco UCS C220 M4 Rack Server
splunk>
Deployment Server: Cisco UCS C220 M4 Rack Server
splunk>
Archival Node: Cisco UCS C3160 Rack Server

Figure 3 Clustered Distributed Search Deployment Architecture on Cisco UCS: Hardware View



Network Fabric:

- 2 Cisco UCS 6296UP fabric interconnects

Administration Servers:

- 2 Cisco UCS C220 M4 servers
- Splunk master node (for indexer clustering) and license master
- Deployer, deployment server, and distributed management console

Search Heads:

- 3 Cisco UCS C220 M4 servers (24 cores; 256 GB)

Indexers:

- Configured with replication factor: 2
- 8 Cisco UCS C240 M4 servers (24 cores, 256 GB, and 24 x 1.2-TB SFF)
- Up to 230 TB of index storage
- Up to 1 TB per day of indexing capacity
- 3 months retention at 1.25 TB per day

A Single Cisco UCS Domain Can Accommodate:

- More than 64 indexers: 1.8 PB of index storage
- 8 TB per day of indexing capacity (replication factor: 2)
- More than 10 search heads supporting hundreds of simultaneous searches
- Up to 4 archival nodes
- No network oversubscription

Archival Node:

- 1 Cisco UCS C3160 server (24 cores, 256 GB, and 60 x 4-TB LFF)

Table 1 presents a scenario in which an enterprise deploys Splunk Enterprise to monitor the IT infrastructure to show how Cisco UCS Integrated Infrastructure for Big Data easily scales with Splunk software.

Table 1 Splunk Enterprise Indexer Cluster and Cisco UCS Scale Out Together

Time	Customer Requirement	Cisco UCS Hardware Configuration Details
Day 1	· **Initial** indexing requirement: Index 500 GB per day. · Retain the data for 3 months.	· 2 Cisco UCS 6296UP Fabric Interconnects · 2 administrative nodes (Cisco UCS C220 M4 servers), · 3 search heads (Cisco UCS C220 M4 servers), · 4 indexers (Cisco UCS C240 M4 servers)
6 months later	· **Change** indexing requirement: Increase indexing to 1 TB per day. · Retain the data for 3 months.	· Add 3 indexers (Cisco UCS C240 M4 servers)
9 months later	· **New** requirement: Store frozen data for 6 months.	· Add 1 archival node (Cisco UCS C3160 server)

Table 2 provides high-level sizing and scaling guidelines for Splunk Enterprise deployment using Cisco UCS C240 M4 servers as indexers for hot, warm, and cold data, and Cisco UCS C3160 servers as the archival nodes for storing frozen data.

Daily Indexing (TB per day)	Number of Indexers (Cisco UCS C240 M4 servers with 24 1.2TB HDDs to manage hot, warm, and cold data providing about 4 months retention)	Number of Archival Nodes (Cisco UCS C3160 Servers with 60 4TB HDDs to manage frozen data providing over 1 year retention)
1TB	4	1
2TB	8	2
4TB	16	4
8TB	32	8

NOTE: The retention period of data is determined by available storage without considering any index replication. Additional servers will need to be added to this configuration for increasing storage and for incorporating index replication by means of indexer clustering.

Deploying Splunk in clustered distributed search mode on a Cisco UCS domain powered by a pair of fabric interconnects with 64 Cisco UCS C240 M4 servers as indexers and 4 Cisco UCS C3160 servers as archival nodes can provide up to 1.8 petabytes (PB) of hot, warm, and cold index data storage and 1.6 PB of frozen index data storage. In such a configuration, Splunk software could index up to 2.5 TB per day with a replication factor of 2 and a search factor of 2, with retention of hot, warm, and cold data for 12 months and retention of frozen data for up to an additional 10 months.

The Cisco UCS reference architectures for Splunk Enterprise support the massive scalability that Splunk deployments demand. The configuration described in this document can be extended to support up to 80 servers with a pair of 96-port Cisco UCS fabric interconnects.

Multiple Cisco UCS domains, with up to thousands of servers, can be supported using Cisco Nexus® 9000 or 7000 Series Switches.

Together, Splunk Enterprise and Cisco UCS Integrated Infrastructure for Big Data provide a comprehensive solution for delivering operational intelligence to monitor, manage, scale, and operate the IT infrastructure that controls the nervous system of today's enterprises and help mine valuable business insights from ever-growing machine data.

For More Information

- For more information about Cisco UCS Big Data solutions, please visit **http://www.cisco.com/go/bigdata_design**.
- For more information about Cisco® Validated Designs for the solution, 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_Splunk.pdf**.
- For more information about Cisco UCS Integrated Infrastructure for Big Data, please visit **http://blogs.cisco.com/datacenter/cpav3/**.
- For more information about Cisco UCS Integrated Infrastructure for Splunk Enterprise, please visit **http://blogs.cisco.com/datacenter/splunk/**.
- For more information about Splunk Enterprise, please visit **http://www.splunk.com/**.
- For more information about the Cisco UCS SmartPlay program, please visit **http://www.cisco.com/go/smartplay**.

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.