cloudera

SOLUTION BRIEF

Cloudera Enterprise with Cisco Unified Computing System

Today's enterprise must store and analyze massive amounts of unstructured data to uncover crucial insights that can lead to competitive advantage. Cisco and Cloudera offer high-performance infrastructure for big data analytics that is cost effective, flexible, and scalable.

While the competitive pressure on enterprises has increased, the amount of data being ingested and managed has exploded and is accelerating quickly. At the same time, the need for timely and more accurate analytics has also increased. As a result, the need for a cost-effective, flexible, and scalable infrastructure to store and process data has never been greater. Cisco and Cloudera have partnered to deliver tested and certified Hadoop infrastructure solutions and ongoing support that help take the time and risk out of deploying Hadoop. These solutions provide a comprehensive, enterprise-class platform for Hadoop applications powered by Cloudera Enterprise, which includes CDH (Cloudera's Distribution Including Apache Hadoop). They've been tested by Cisco and certified by the Cloudera Certified Technology program to streamline deployment and reduce risk.

The Rise of Big Data and Apache Hadoop

The volume, variety, and velocity of unstructured data coming from a profusion of Internet-connected devices is unprecedented. "Big Data" refers to data that just doesn't fit easily into traditional relational models because it's often a mix of structured and unstructured data, it comes in too fast, and it's too expensive to store in a way that's accessible. The ability to leverage big data requires a new type of data management platform that can adequately capture and extract value from all of it. Apache Hadoop is the open source framework that lets organizations mine the insights of new and emerging types of information, a capability that simply did not exist before.

Moving beyond its roots in Web 2.0 technology, Apache Hadoop is rapidly emerging as an essential enterprise platform. Consumer and commercial industries are all finding applications for big data analytics, particularly as they are faced with new challenges in today's web and social-oriented content and interaction models. But taking advantage of Hadoop is not simple – it's a complex distributed system comprised of a dozen different open source projects.

Together, Cisco and Cloudera are well positioned to help organizations exploit the valuable business insights in all their data, regardless of whether it's structured, semi structured or unstructured. Cloudera is the leading provider of enterprise-grade Hadoop infrastructure software and services, and the leading contributor to the Apache Hadoop project overall. Cloudera provides an enterprise-ready Hadoop-based solution known as Cloudera Enterprise, which includes their market leading open source Hadoop distribution (CDH), their comprehensive management system (Cloudera Manager), and technical support. Cisco has been the leader in networking for decades, providing proven solutions that meet critical business requirements. Cisco UCS C-Series Rack-Mount Servers based on Intel® Xeon® processors complete these offerings, delivering an integrated Hadoop infrastructure.

Highlights

- Best-of-breed, integrated big data infrastructure combines enterprise-grade Hadoop software & services with networking solutions
- Streamlined big data platform deployment: tested by Cisco, certified by Cloudera
- Configurable for high performance or high capacity
- Simplified management with Cisco UCS Manager & Cloudera Manager

Solution Components

Cisco's Big Data Common Platform Architecture:

- Cisco UCS 6200 Series Fabric Interconnects
- Cisco UCS 2200 Series Fabric Extenders
- Cisco UCS C240 M3 Rack-Mount Servers
- Cisco UCS Virtual Interface Cards
- Cisco UCS Manager

Cloudera Enterprise:

- Cloudera Distribution Including Apache Hadoop
- Cloudera Manager
- Cloudera Technical Support

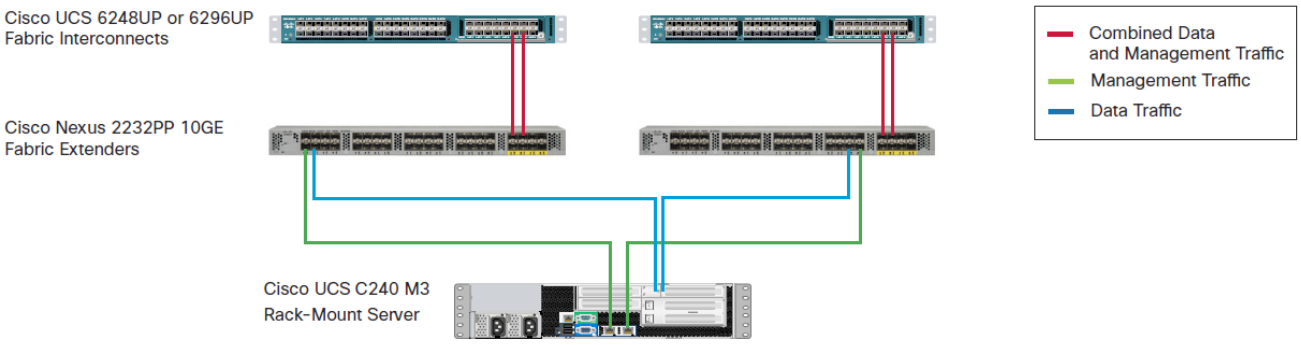
Cisco Solution Overview

The Cloudera Hadoop Reference Configuration is based on Cisco's Big Data Common Platform Architecture (CPA), a highly scalable architecture designed to meet a variety of scale-out application demands with seamless data integration and management integration capabilities built using the following components:

- **Cisco UCS 6200 Series Fabric Interconnects:** The Cisco UCS 6200 Series Fabric Interconnects are a core part of Cisco UCS, providing both network connectivity and management capabilities across Cisco UCS 5100 Series Blade Server Chassis as well as Cisco UCS C-Series Rack-Mount Servers. Typically deployed in redundant pairs, the fabric Interconnects offer line-rate, low-latency, lossless 10 Gigabit Ethernet connectivity and unified management with Cisco UCS Manager in a highly available management domain.
- **Cisco UCS 2200 Series Fabric Extenders:** Cisco UCS 2200 Series Fabric Extenders behave as remote line cards for a parent switch and provide a highly scalable and extremely cost-effective unified server-access platform.
- **Cisco UCS C240 M3 Rack-Mount Servers:** Cisco UCS C240 M3 Rack-Mount Servers are 2-socket servers based on Intel Xeon E-2600 series processors and up to 768 GB of main memory. Up to 24 Small-Form Factor (high performance option) or 12 Large-Form Factor (high capacity option) disk drives are supported, along with 4 Gigabit Ethernet LAN-on-motherboard (LOM) ports.
- **Cisco UCS Virtual Interface Cards (VICs):** Unique to Cisco, Cisco UCS Virtual Interface Cards incorporate next-generation converged network adapter (CNA) technology from Cisco, and offer dual 10-Gbps ports designed for use with Cisco UCS C-Series Rack-Mount Servers. Optimized for virtualized networking, these cards deliver high performance and bandwidth utilization and support up to 256 virtual devices.
- **Cisco UCS Manager:** Cisco UCS Manager resides within the Cisco UCS 6200 Series Fabric Interconnects. It makes the system self-aware and self-integrating, managing all of the system components as a single logical entity. Cisco UCS Manager can be accessed through an intuitive graphical user interface (GUI), a command-line interface (CLI), or an XML application programming interface (API). Cisco UCS Manager uses service profiles to define the personality, configuration, and connectivity of all resources within Cisco UCS, radically simplifying provisioning of resources so that the process takes minutes instead of days. This simplification allows IT departments to shift their focus from constant maintenance to strategic business initiatives.

The single-rack configuration consists of two fully redundant Cisco UCS 6248UP 48-Port Fabric Interconnects (for up to five racks) or Cisco UCS 6296UP 96-port Fabric Interconnects (up to 10 racks) along with two Cisco Nexus® 2232PP 10GE Fabric Extenders and 16 Cisco UCS C240 M3 Rack-Mount Servers. Each server in the configuration connects to the unified fabric through two active-active 10-Gbps links using a Cisco UCS Virtual Interface Card (VIC, for data traffic) and Cisco Integrated Management Controller (IMC, for management traffic). See Figure 1. Multi-rack configurations include two Cisco Nexus 2232PP fabric extenders and 16 Cisco UCS C240 M3 Rack-Mount Servers for every additional rack.

Figure 1: UCS Fabric Architecture



High-Performance Configuration[1]

- 16 x Cisco UCS C240 M3 Rack-Mount Servers
- 2 x Intel® Xeon® E5-2665 processors
- 256 GB Memory
- 1 x Cisco UCS VIC 1225
- 1 x LSI MegaRAID SAS 9226CV-8i Card
- 24 x 1-TB SATA 7200 RPM SFF Disk Drive

[1] Available as Cisco Global SmartPlay Solution bundle UCS-EZ-BD-HP

High-Capacity Configuration[2]

- 16 x Cisco UCS C240 M3 Rack-Mount Servers
- 2 x Intel® Xeon® E5-2640 processors
- 128 GB Memory
- 1 x Cisco UCS VIC 1225
- 1 x LSI MegaRAID SAS 9226CV-8i Card
- 12 x 3-TB SAS 7200 RPM LFF Disk Drive

[2] Available as Cisco Global SmartPlay Solution bundle UCS-EZ-BD-HC

2 x Cisco UCS 6200 Fabric Interconnects
2 x Cisco Nexus® 2232PP 10GE Fabric Extenders

Figure 2: High Performance and High Capacity Reference Configuration

The performance and capacity characteristics of high performance and high capacity configurations are shown in table 2 and table 3.

Table 2. High Performance Reference Configurations

	Component	Single Rack	Multi-Rack
Network Fabric	Fabric interconnects	2	2 per cluster
	Fabric extenders	2	2 per rack
Computing	Servers	16	16 per rack
	Computer processor cores	256	256 per rack
	Memory	4 TB (up to 12 TB supported)	4 TB (up to 12 TB supported)
	Unformatted storage capacity	384 TB (using 24x1 TB SFF SATA disk drives)	384 TB per rack

Table 3. High Capacity Reference Configurations

	Component	Single Rack	Multi-Rack
Network Fabric	Fabric interconnects	2	2 per cluster
	Fabric extenders	2	2 per rack
Computing	Servers	16	16 per rack
	Computer processor cores	192 (up to 256 supported per rack)	192 (up to 256 supported per rack)
	Memory	2 TB (up to 12 TB supported)	2 TB (up to 12 TB supported)
	Unformatted storage capacity	576 TB (using 24x3 TB LFF SAS disk drives)	576 TB per rack

High Performance and Exceptional Scalability

Cisco UCS unified fabric architecture provides fully redundant, highly scalable lossless 10-Gbps unified fabric connectivity for big data traffic. Powered by the latest Intel Xeon processor, the joint Cisco Cloudera solution delivers best-in-class performance and internal storage capacity. The Cisco Cloudera Reference Configurations can easily scale to support a large number of nodes when required by business demands. The advanced management capabilities of Cisco UCS radically simplify this process with a single point of management that spans all nodes in the cluster.

Simplified Management

Big data analytics implementations tend to involve very large numbers of servers. In traditional environments, it can be challenging to manage these large numbers of servers effectively. Cisco UCS Manager delivers unified, model-based management that applies personality and configures server, network, and storage connectivity resources, making it as easy to deploy hundreds of servers as it is to deploy a single server. Additionally, Cisco UCS Manager can perform system maintenance activities such as firmware updates across the entire cluster as a single operation.

Coexistence with Enterprise Applications

In building big data solutions that involve Hadoop and/or NoSQL, organizations need ways to transfer data transparently between their enterprise applications and big data platforms. This solution can connect to other Cisco UCS deployments running enterprise applications across the same management plane, thereby radically simplifying data center management and connectivity. Cloudera Enterprise provides a comprehensive platform for designing and managing solutions that cross the boundaries of traditional and big data platforms. By providing easy-to-use tools and familiar design concepts for both traditional and big data platforms, Cloudera empowers organizations to leverage existing IT skillsets to build big data solutions.

Rapid Deployment

Deployment of large numbers of servers can take time. Systems need to be racked, networked, configured, and provisioned before they can be put into use. Cisco UCS Manager uses a model-based approach to provision servers by applying a desired configuration to physical infrastructure quickly, accurately, and automatically. The ability to create consistent configurations improves business agility and eliminates a major source of errors that can cause downtime. Cloudera Enterprise's tightly integrated platform demystifies the challenges of building end-to-end solutions that take you from data acquisition and processing to rich analytics solutions.

Enterprise Service and Support

Enterprises want know that the vendors providing a solution have the expertise to help them quickly proceed through the design, deployment, and testing of strategic big data initiatives. Businesses also need to have confidence that if a critical system fails, they will be able to get timely and competent support. The Cisco Reference Configurations bring together world-class service and support from Cisco and Cloudera.

Cloudera, Inc. 220 Portage Avenue, Palo Alto, CA 94306 USA | 1-888-789-1488 or 1-650-362-0488 | **cloudera.com**

cloudera