Cloudera Hadoop Infrastructure with Cisco UCS Servers and Cisco Nexus Switches Powered by Intel Xeon Processors

Solution Brief
November 2011

Highlights

Tested and Certified Building Blocks to Reduce Cost and Risk

- Cisco offers infrastructure building blocks that help organizations deploy Cloudera's Apache Hadoop distribution quickly, while scaling configurations rapidly and predictably as demand dictates.

Powerful and Cost-Effective Cisco UCS C-Series Rack-Mount Servers

- A choice of Cisco Unified Computing System™ (Cisco UCS™) servers with Intel® Xeon® processors delivers a best-in-class, next-generation platform that can be deployed quickly and scaled to meet data processing needs.

A Choice of Cisco Nexus Switches

- A choice of popular Cisco Nexus® switches can be selected based on workload needs and help ensure Hadoop infrastructure growth without arbitrary restrictions.

Supported Cloudera Enterprise for Comprehensive Cluster Management

- By combining expert support with a software suite that delivers deep visibility into and across Hadoop clusters, Cloudera Enterprise gives Hadoop operators an efficient and precise way to provision and manage cluster resources.

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

As competitive pressure in enterprises has increased, so too has the amount of data and the need for timely analytics, in turn Increasing the need for cost-effective, flexible, and scalable infrastructure to host Apache Hadoop deployments. Cisco and Cloudera have partnered to deliver tested and certified infrastructure solutions and ongoing support that helps take the time and risk out of deploying Hadoop infrastructure, providing:

- A comprehensive, enterprise-class platform for Hadoop applications, powered by Cloudera's Distribution including Apache Hadoop (CDH), tested by Cisco and certified by through the Cloudera Certified Technology program to streamline deployment and reduce risk
- Proven data center architecture featuring Cisco UCS™ C-Series Rack-Mount Servers powered by Intel® Xeon® processors
- High-bandwidth, low-latency Cisco Nexus® switching to meet the most demanding performance and scalability requirements at lower cost

Together these powerful and tightly integrated components yield building blocks that can be used to rapidly deploy and easily scale the CDH infrastructure.

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 does not fit easily into traditional relational models, and it requires a new type of platform to adequately capture and extract value from it. Apache Hadoop 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, the Apache Hadoop framework is rapidly emerging as an essential enterprise platform. Consumer, commercial, and financial industries are all finding applications for data analytics, particularly as they are faced with new challenges in today's web and social-oriented content and interaction models.

Together, Cisco and Cloudera are well positioned to help organizations exploit the valuable business insights in their unstructured data. Cloudera is an active contributor to the Hadoop project and provides an enterprise-ready, commercial distribution of Hadoop. 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 integrated Hadoop infrastructure.

Cisco Solution Overview

The combination of Cisco Nexus high-performance interconnects with Cisco UCS C-Series servers is part of the foundation for the Cisco® Data Center Business Advantage solution. The solution addresses the needs of high-performance Hadoop solutions as well as the business and operational requirements of evolving data centers. This approach offers:

- Infrastructure simplicity and a building-block approach to reduce total cost of ownership (TCO)
- Enhanced business resilience with greater operational continuity based on the modular Cisco NX-OS Software operating system
- Capability to use existing operational models and administrative domains for easy deployment

Cisco's modular solution offers customers the choice of two compute modules for powering the Hadoop infrastructure.

Cisco UCS C-Series Rack-Mount Servers

Cisco UCS C-Series Rack-Mount Servers deliver world-record-setting performance in an industry-standard form factor to reduce TCO, increase agility, and increase customer choice. Cisco UCS C-Series servers address a range of workloads through a balance of processing, memory, I/O, and internal storage resources. Cisco has tested and certified two popular server configurations for CDH to provide enterprise-class processing power for a range of workloads.

Figure 1. Cisco UCS C200 M2 High-Density Rack-Mount Server (Top) and Cisco UCS C210 M2 General-Purpose Rack-Mount Server (Bottom)

For dense and compute-intensive applications, the Cisco UCS C200 M2 High-Density Rack-Mount Server (Figure 1) is a 2-socket, 1-rack-unit (1RU) server that offers:

- Up to two Intel Xeon 5500 or 5600 series processors
- Up to 192 GB of industry-standard double-data-rate 3 (DDR3) main memory
- Up to eight 2.5-inch or four 3.5-inch internal SAS or SATA disk drives for up to 8 terabytes (TB) of local storage

For applications that require a balance of processor and I/O bandwidth with additional local storage capacity, the Cisco UCS C210 M2 General-Purpose Rack-Mount Server is a 2-socket, 2RU server. The server balances performance, density, and efficiency

for storage-intensive workloads, and it supports:

- Up to two Intel Xeon 5500 or 5600 Series processors
- Up to 192 GB of industry-standard double data rate (DDR3) main memory
- Up to 16 2.5-inch internal SAS or SATA disk drives for up to 16 terabytes of total storage

Cisco Nexus 5500 Platform with Cisco Nexus 2200 Series Fabric Extenders

To meet business requirements, enterprise Big Data solutions must deliver low TCO and ease of management without sacrificing performance.

The Cisco Nexus 5500 switching platform combined with the Cisco Nexus 2200 Series Fabric Extenders (Figure 2) offers a state-of-the art alternative to traditional architectures. The solution delivers the industry's lowest TCO along with ease of management and high performance.

The Cisco Nexus 2248TP-E Fabric Extender is optimized for specialized data center workloads, with a large buffer space that is designed to sustain the bursty traffic often imposed by Big Data workloads.

The combination of the Cisco Nexus 5500 platform and the Cisco Nexus 2200 Series offers:

- Low-latency, line-rate, cut-through switching architecture

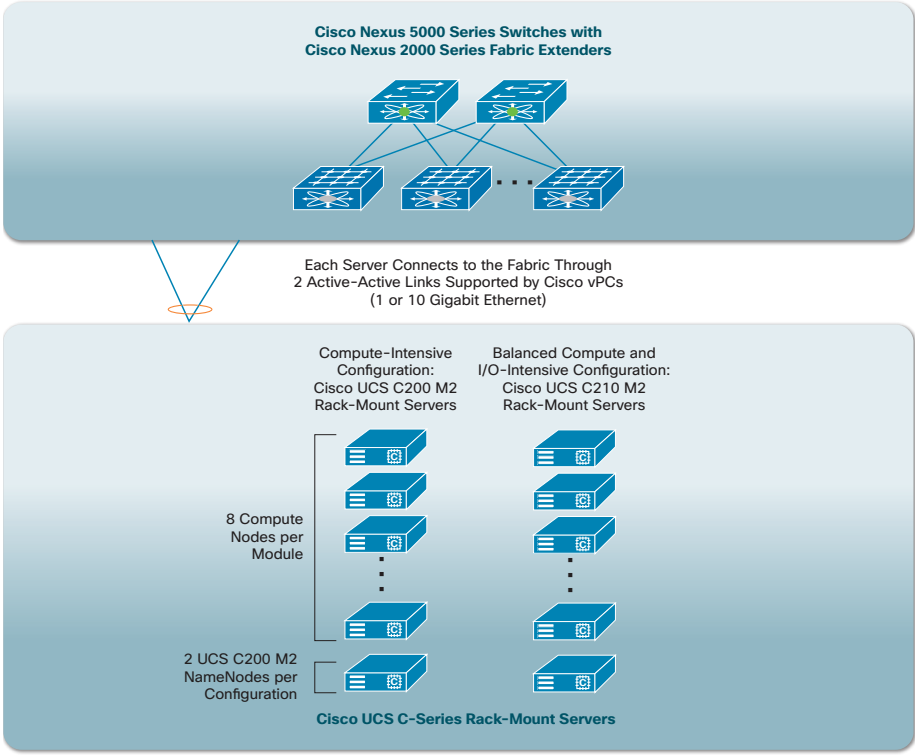

Figure 2. Cisco Building Blocks Tested and Certified with Cloudera Supply a Choice of Network Fabrics as Well as a Choice of Server Building Blocks, Allowing Organizations to Deploy Solutions Quickly and Scale Them as Required.

- A virtual modular system in which the switching infrastructure is logically centralized yet physically distributed, with the fabric extenders acting as remote line cards for the parent switch
- The benefits of both fixed and modular systems
- High scalability with up to 768 servers with 10 Gigabit Ethernet connectivity or 1152 servers with Gigabit Ethernet connectivity in a single management domain
- Architecture flexibility that supports both Gigabit Ethernet and 10 Gigabit Ethernet connectivity

Cisco and Cloudera: Better Together

Tested and certified infrastructure from Cisco and Cloudera delivers an enterprise-class Hadoop platform in a modular, easy-to-deploy package with ongoing support options to enable data-focused enterprises to run

Apache Hadoop environments cost effectively and with repeatable success.

As depicted in Figure 2, the solution provides two Hadoop NameNodes with each configuration for redundancy. Organizations can deploy either Cisco UCS C200 M2 or Cisco UCS C210 M2 servers to meet performance and capacity needs that are satisfied with anywhere from a few servers to multiple racks. The result is a tested and certified Cloudera Hadoop infrastructure designed to meet to meet your performance, capacity, and scalability requirements.

About Cloudera's Distribution Including Apache Hadoop

Cloudera's distribution delivers a streamlined path for putting Apache Hadoop to work solving business problems in production. Well suited for enterprises seeking a stable, tested Hadoop solution, CDH is the bridge between the insights of organizations using Hadoop in production and the continuous stream of innovations from the Apache community.

About Cloudera

Cloudera is the leading provider of Apache Hadoop-based software and services and works with customers in financial services, web, telecommunications, government, and other industries. The company's products Cloudera Enterprise and CDH help organizations profit from all their information. CDH is one of the most comprehensive Apache Hadoop-based platforms in the industry. Cloudera Enterprise offers a cost-effective way to perform large-scale data storage and analysis and includes the tools, platform, and support necessary to use Apache Hadoop in a production environment.

For More Information

- For complete details on Cisco UCS C-Series Rack-Mount Servers, please visit: http://www.cisco.com/en/US/products/ps10493/index.html
- For complete details about the Cisco Nexus 5000 Series, please visit : http://www.cisco.com/en/US/products/ps9670/index.html
- For complete details on the Nexus Fabric Extender architecture please visit: http://www.cisco.com/en/US/products/ps10110/index.html
- For more information about Cloudera and for free downloads of CDH, please visit : http://www.cloudera.com

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**.

LE-33201-01 11/11