



Cisco UCS Integrated Infrastructure for Big Data with Optional Tiered Storage Extension



Tiered Storage Solution for Big Data with Cost Effective Performance.

Highlights



Comprehensive integrated infrastructure

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



Scalability for Big Data workloads

- Cisco UCS Integrated Infrastructure for Big Data offers linear scalability and simplification of essential operations for single-rack and multi-rack deployments without adding complex layers of switching infrastructure



Faster and easier big data infrastructure deployment Cisco UCS

- Director Express for Big Data offers one-click provisioning, installation, and configuration.



Tiered storage support with Hortonworks Data Platform (HDP) 2.2:

- With HDP 2.2, the Hadoop Distributed File System (HDFS) provides the capability to use heterogeneous storage media within the HDFS cluster to enable both an archival storage tier (Cisco UCS C3160 Rack Server) and a hot data tier (Cisco UCS C240 M4 Rack Server).



Enterprise SQL on HDP 2.2

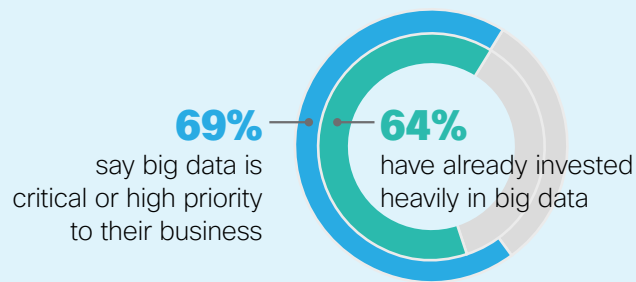
- With HDP 2.2, Apache Hive embeds Apache Tez so that it can translate complex SQL statements into highly optimized, purpose-built data processing graphs for performance, throughput, and scalability, making Hive interactive

Is your enterprise using big data? And are you challenged by the need to store it all? If so, you're not alone.

Big data is here. According to IDC¹, the market for big data and related analytics technologies will reach US\$125 billion worldwide in 2015.

Chief among the big data technologies being adopted by enterprises is Hadoop, an Apache open-source project, Java-based programming framework that processes large data sets in a distributed environment. Hadoop was developed because the size and complexity of the analytics required by enterprises in the era of big data is overpowering the computing capabilities of traditional databases.

Figure 1: Big Data Has Big Traction with CIOs



Source: CDC Sixth Annual Global CIO Survey 2014-2015

Almost 70 percent of enterprises around the world say that these technologies are either critical or high priorities for their businesses, with 64 percent saying that they have already invested heavily in them (Figure 1).²

Hadoop can handle large volumes of structured and unstructured data much more efficiently than traditional enterprise data warehouses. Cost savings for customers is significant because Hadoop is open source and can scale linearly as data volume grows.

¹ Six Predictions for the \$125 Billion Big Data and Analytics Market for 2015, by Gil Press, Forbes, December 11, 2014, <http://www.forbes.com/sites/gilpress/2014/12/11/6-predictions-for-the-125-billion-big-data-analytics-market-in-2015/>

² CDC Sixth Annual Global CIO Survey 2014-2015



Enterprises today facing Big Data challenges

One of the major challenges of big data is the management of fast-growing data and associated costs. To contain costs, many businesses previously had to make difficult choices about which data was most valuable given their business priorities. Frequently, large amounts of raw data were deleted, because it was too expensive to keep. Subsequently, when business priorities changed, the necessary data was not readily available to users. This hampered the ability of businesses to respond to market shifts and react with agility to the competitive environment.

Storage tiering with Cisco UCS and Hadoop 2.0

New data is more frequently accessed than older data. Therefore, over time the frequency of read operations on a given dataset naturally decreases, with data becoming less frequently accessed as it ages. New data is deemed “hot” and old data is “cold or archival”. An in-between type of data is “warm” data. As enterprises collect increasing volumes of data of all three types, they have a growing need to cost-effectively store it all.

This is where tiered storage can help.

With Hortonworks Data Platform (HDP) 2.2, Hadoop provides the capability to use heterogeneous storage media to support two types of storage:

- Hot Data Tier delivers a storage tier that consists of C240 M4 Servers to store datasets that require high speed storage access.
- Cold or Archival data tier uses Cisco UCS® C3160 Rack Servers, designed to offer industry leading storage density.

HDP 2.2 offers data placement policies (storage policies) to manage tiers of hot, warm, and cold data, with the placement of data tied to its temperature (Figure 3).

Figure 2: Investing in Hadoop

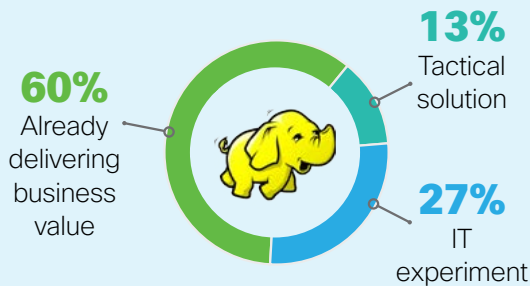
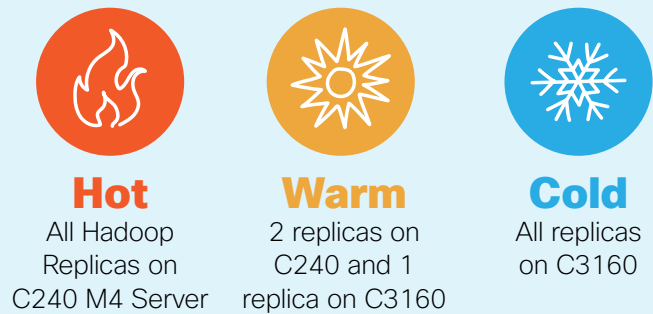


Figure 3: Tired Storage Options with Cisco UCS Integrated Infrastructure for Big Data and Hadoop 2.0



The Cisco Unified Computing System™ (Cisco UCS) Solution for Hortonworks

The Cisco UCS solution for Hortonworks is based on **Cisco UCS Integrated Infrastructure for Big Data**, a highly scalable architecture that includes computing, storage, connectivity, and unified management capabilities. It is designed to meet a variety of scale-out application demands with transparent data integration and management integration capabilities built using the following components shown in Figure 4.

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

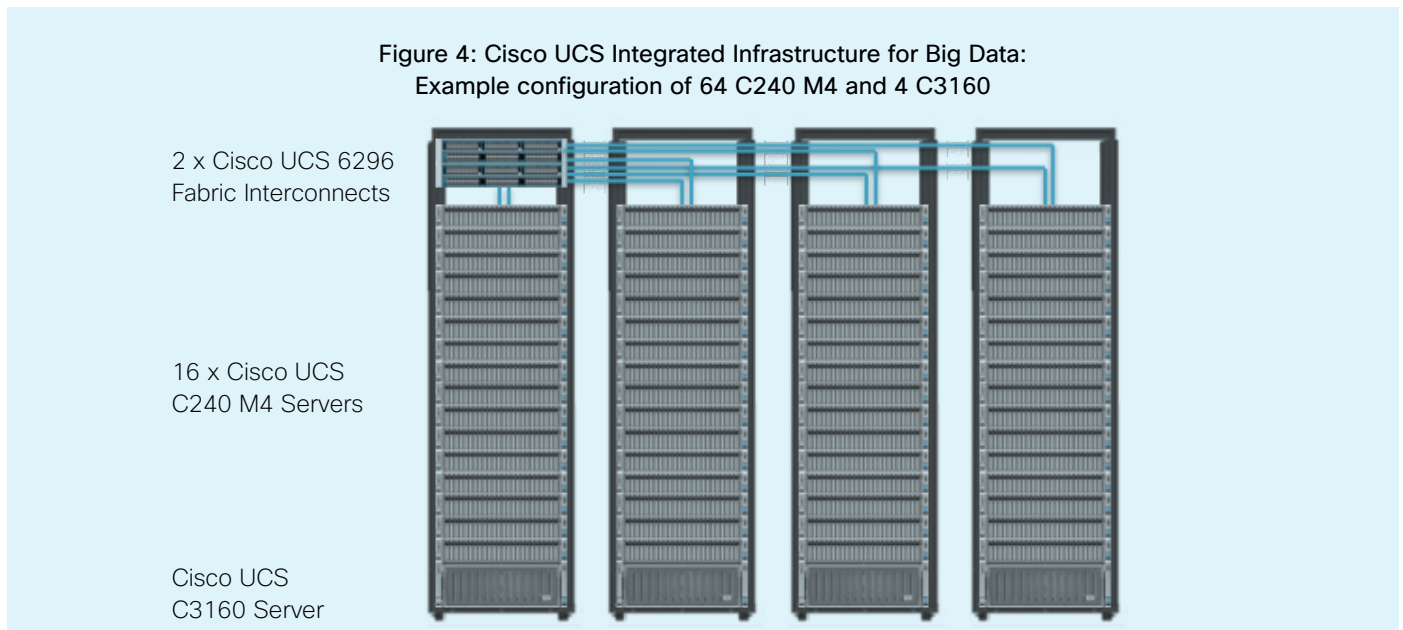
by Cisco UCS Manager. Deployed in redundant pairs, Cisco fabric interconnects offer the full active-active redundancy, performance, and exceptional scalability needed to support the large number of nodes that are typical in clusters serving big data applications. Cisco UCS Manager enables rapid and consistent 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 UCS C240 M4 Rack Server

The Cisco UCS C240 M4 is designed for performance and expandability over a wide range of storage-intensive workloads. The servers are based on the Intel® Xeon® processor E5-2600 v3

series and 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 support 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 2x1 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.

Figure 4: Cisco UCS Integrated Infrastructure for Big Data:
Example configuration of 64 C240 M4 and 4 C3160



Cisco UCS Integrated Infrastructure for Big Data with Optional Tiered Storage Extension
 Tiered Storage Solution for Big Data with Cost Effective Performance.

Cisco UCS C3160 Rack Server

The Cisco UCS C3160 Rack Server, as archive storage, is an advanced, modular rack server with extremely high storage density (Figure 5). Based on the Intel Xeon processor E5-2600 v2 series, the Cisco UCS C3160 offers up to 360 TB of local storage in a compact 4-rack-unit (4RU) form factor. It works extremely well for the archival storage tier in an HDP 2.2 environment. These nodes can be used to store less frequently accessed “cold” data much more cost effectively.

Because all of its hard-disk drives are individually hot-swappable, and with its built-in enterprise-class redundant array of independent disks (RAID) redundancy, the Cisco UCS

C3160 helps you achieve the highest levels of data availability. It’s well suited for snapshots, active archiving, compliance, media storage, and distributed file systems for scenarios in which high storage capacity is important.

Hortonworks Data Platform

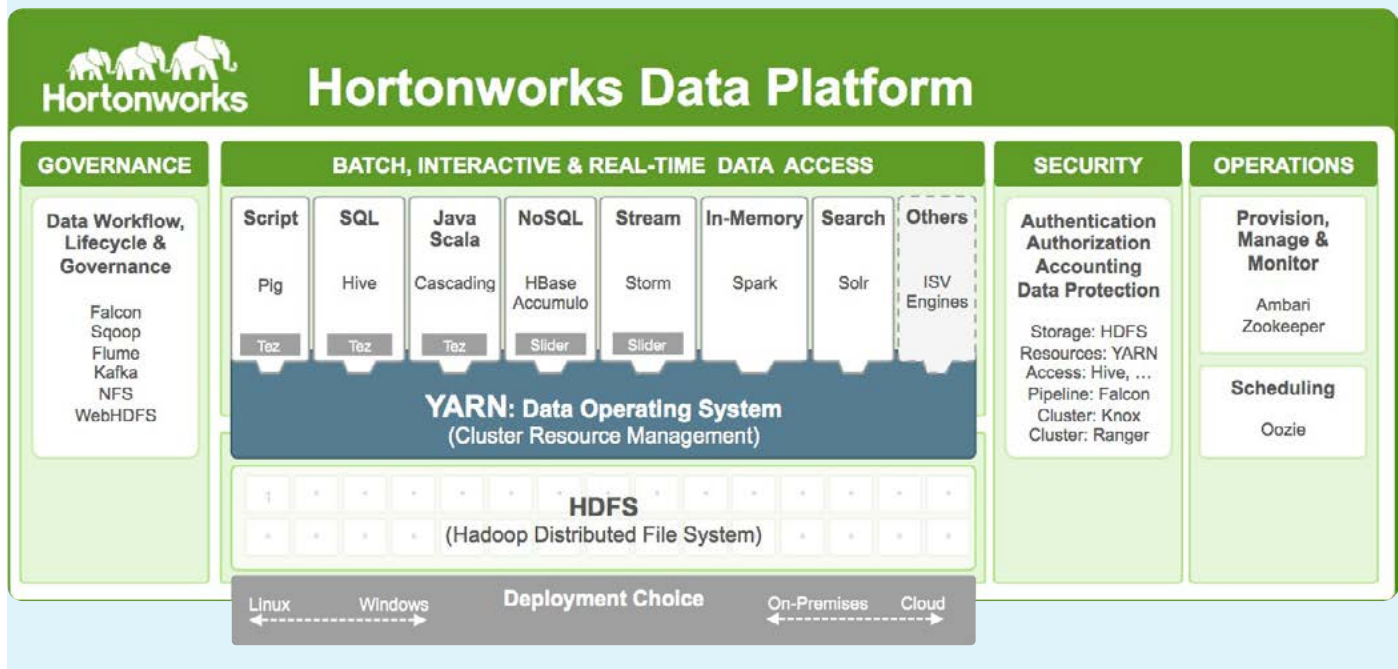
Hortonworks Data Platform enables Enterprise Hadoop: the full suite of essential Hadoop capabilities that are required by the enterprise and that serve as the functional definition of any data platform technology. This comprehensive set of capabilities is aligned with data management, data access, data governance and integration, security, and operations(Figure 6).

Together, Hortonworks Data Platform and the Cisco UCS Integrated Infrastructure for Big Data provide an industry-leading platform for Hadoop-based applications to take advantage of today’s best-in-class tiered storage solutions.

Figure 5: Cisco UCS C3160 Rack Server



Figure 6: Hortonworks Data Platform



Reference Architecture

The current version of Cisco UCS Integrated Infrastructure for Big Data offers several configurations according to the computing and storage requirements (Figure 6). It works transparently with the HDP.

For More Information

For more information about Cisco UCS Big Data solutions, please visit http://www.cisco.com/go/bigdata_design

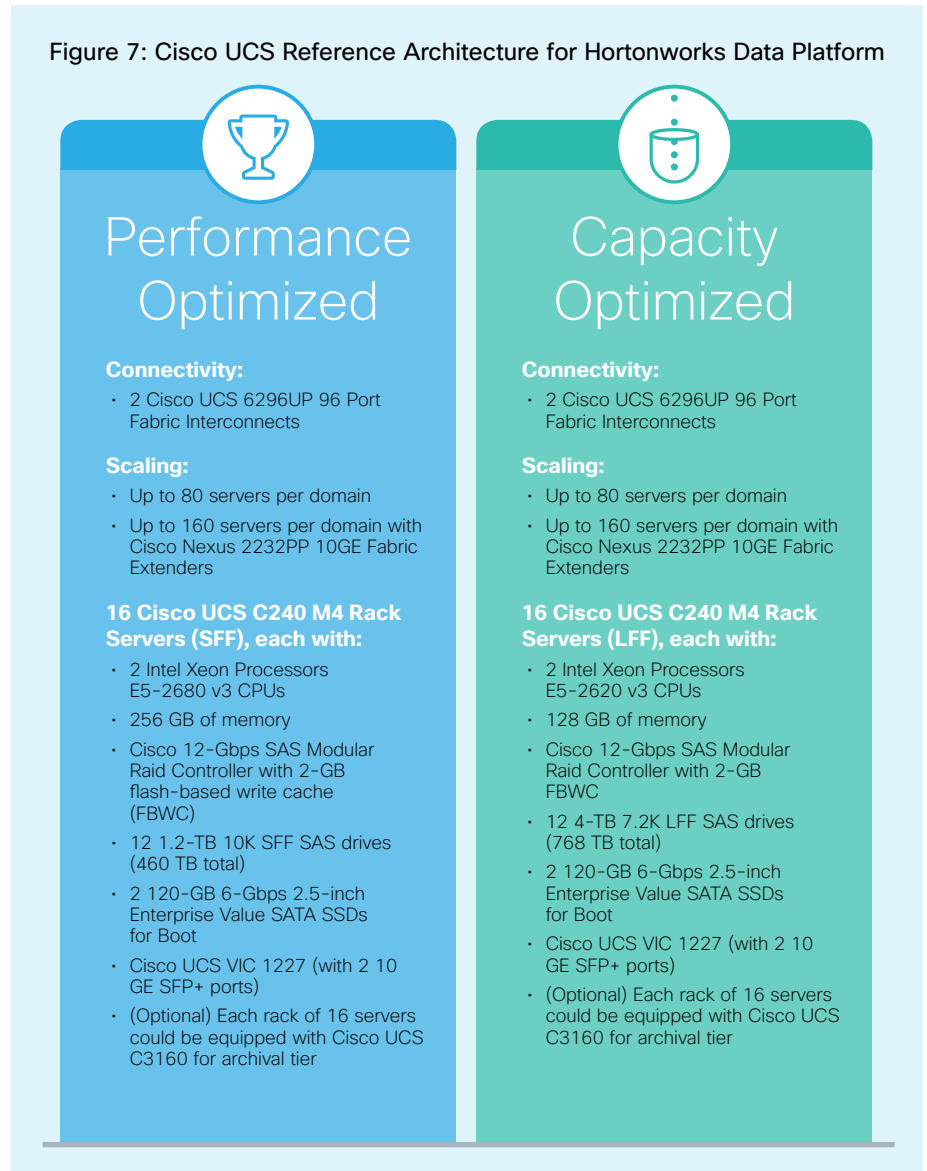
For more information about Cisco UCS Integrated Infrastructure for Big Data, please visit <http://blogs.cisco.com/datacenter/cpav3/>

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

For more information on the Cisco Validated Design (CVD) 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_Bigdata_with_Hortonworks.pdf

For more information about Hortonworks, please visit <http://hortonworks.com/>

Figure 7: Cisco UCS Reference Architecture for Hortonworks Data Platform



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.