Cisco UCS Integrated Infrastructure for Big Data and Analytics with Hortonworks Data Platform

Bringing Performance and Scalability for Big Data and Analytics.

Highlights

Proven Platform for Enterprise Data Lake
- Fifth generation of the platform deployed across major industries such as agriculture, education, entertainment, finance, healthcare, industrial, insurance, manufacturing, public-sector, service provider, and utilities
- Demonstrated through industry-standard benchmarks

Designed, tested, and validated for Faster Time to Value
Cisco Validated Designs (CVD) is designed to facilitate faster, more reliable, and more predictable customer deployments; it provides design, scalability, and performance recommendations.

Built on Cisco UCS foundations
The Cisco Unified Computing System™ (Cisco UCS®) M5 platform offers complete integration of computing, networking, and storage resources with unified management while providing high performance, expandable storage, and scalability for big data systems.

Being a fabric-centric architecture designed for business acceleration, it provides a true on-demand infrastructure, with a system that grows gracefully and incrementally.

Designed to Scale from Small to Very Large as applications demand
With Cisco® Application Centric Infrastructure (Cisco ACI), you can easily scale a cluster to thousands of nodes. Cisco ACI implements an application-aware, policy-based approach that treats the network as a single entity rather than a collection of switches.

Automated deployment and configuration
Enable one-click provisioning, installation, and configuration of big data infrastructure using Cisco UCS Director Express for Big Data.

Powering enterprise data lake
Hortonworks Data Platform (HDP) and Hortonworks DataFlow (HDF) together address the complete needs of data in motion and data at rest. It powers real-time customer applications and delivers robust analytics that accelerate decision making and innovation.
Cisco and Hortonworks Deliver Next-Generation Data-Driven Solutions for Enterprise

Every business is now a data business. Data is any organization's future and its most valuable asset. Sensors, Internet of Things (IoT) devices, social networking, and online transactions are all generating data that needs to be captured, monitored, and rapidly processed to make data-based decisions instantly. The biggest challenge for enterprises is to manage and use data with increased volume, variety, and a velocity that was rarely seen in the past. As this data becomes increasingly available for processing of both data in motion and data at rest, customers are discovering that they could extract insights through new applications to better serve their business needs.

Data Lake can enhance and amplify existing investments and create new forms of business value. It also provides a cost-effective and technologically feasible way to meet these data challenges and allow you to store and analyze large volumes and variety of data both for real time and batch processing. Building a next-generation data lake architecture requires simplified and centralized management, high performance, and a linearly scaling infrastructure and software platform.

Cisco and Hortonworks have partnered to build architectures that help enterprises transform their business by unlocking the full potential of big data. Cisco UCS Integrated Infrastructure for Big Data and Analytics with Hortonworks Data Platform and Hortonworks Dataflow powers the next-generation architecture for big data systems, spanning a myriad of use cases including IoT, fraud analytics, and precision medicine through genome sequencing.

Figure 1 shows the Enterprise Data Lake providing actionable intelligence, content enrichment, and rapid changes to real-time data flows. This platform provides analytics on data at rest using HDP and real-time analytics using HDF.

![Figure 1. Enterprise Data Lake with HDP and HDF](image-url)
Cisco UCS Integrated Infrastructure for Big Data and Analytics

Organizations today must ensure that the underlying physical infrastructure can be deployed, scaled, and managed in a way that is agile enough to change as workloads and business requirements change. Cisco UCS Integrated Infrastructure for Big Data and Analytics has redefined the potential of the data center with its revolutionary approach to manage compute, network, and storage to successfully address the business needs of IT innovation and acceleration. This solution provides an end-to-end architecture for processing high volumes of structured and unstructured data, for both real-time and archival.

Cisco UCS 6300 Series Fabric Interconnects

Cisco UCS 6300 Series Fabric Interconnects provide high-bandwidth, low-latency connectivity for servers, with Cisco UCS Manager providing integrated, unified management for all connected devices. The Cisco UCS 6300 Series Fabric Interconnects are a core part of Cisco UCS, providing low-latency, lossless 40 Gigabit Ethernet, Fibre Channel over Ethernet (FCoE), and Fibre Channel functions.

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 and automates 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 Rack Servers (C240 M5 and C220 M5)

The Cisco UCS M5 Rack Server is a dual-socket, 2-Rack-Unit (2RU) server offering industry-leading performance and expandability for a wide range of storage and I/O-intensive infrastructure workloads for big data and analytics. This server uses the latest Intel Xeon Processor Scalable Family with up to 28 cores per socket.

It supports up to 24 Double-Data-Rate 4 (DDR4) Dual In-line Memory Modules (DIMMs) for improved performance and lower power consumption. These DIMM slots are 3D XPoint ready, supporting next-generation nonvolatile memory technology. Based on the server type, Cisco UCS Rack servers have a range of storage options. Cisco UCS C240 M5 supports up to 24 Small Form-Factor Pluggable (SFF) 2.5-inch drives (support for up to 10 NonVolatile Memory express [NVMe] PCIe Solid-State Drives [SSDs] on the NVMe-optimized chassis version) or 12 Large-Form-Factor (LFF) 3.5-inch drives + 2 rear hot-swappable SFF drives with a 12-Gbps SAS Module RAID controller. Cisco UCS C220 M5 supports up to 10 Small Form-Factor Pluggable (SFF) 2.5-inch drives (support for up to 10 NVMe PCIe SSDs on the NVMe-optimized chassis version) Additionally, all the servers have two modular M.2 cards that you can use for boot. A modular LAN-on-motherboard (mLOM) slot supports dual 40 Gigabit Ethernet network connectivity with the Cisco UCS Virtual Interface Card 1387 (VIC 1387).

Figure 2. Cisco UCS Integrated Infrastructure for Big Data and Analytics with Hortonworks Data Platform
Hortonworks Data Platform
Hortonworks Data Platform (HDP), built on Apache Hadoop and Spark, enables the creation of a secure enterprise data lake and delivers the analytics you need to innovate quickly and power real-time business insights.

HDP includes a range of processing engines that enable simultaneous interaction with the same data in multiple ways. These data processing engines include interactive Structured Query Language (SQL), real-time streaming, data science, and batch processing to take advantage of data stored in a single platform, thereby unlocking an entirely new approach to analytics.

HDFS and YARN are the cornerstone components of HDP. HDFS provides the scalable, fault-tolerant, cost-efficient storage for the data. YARN is a cluster resource management system that provides a centralized architecture for the processing of multiple workloads simultaneously.

HDP extends data access and management through powerful tools for data governance and integration. It provides a reliable, repeatable, and simple framework for managing the flow of data in and out of Hadoop. Security is integrated into HDP in multiple layers. Critical features for authentication, authorization, accountability, and data protection are in place to help secure HDP across these important components.

Hortonworks DataFlow
Hortonworks DataFlow (HDF) provides an integrated solution with Apache NiFi, Kafka, and Storm. It makes streaming analytics faster and easier by enabling accelerated data collection, curation, analysis, and delivery in real time.

HDF offers the following main features:

- **Operational efficiency**: Fast, effective drag-and-drop interface for creating, managing, tuning, and troubleshooting data flows, enabling coding-free creation and adjustment of data flows in five minutes or less
- **Data collection**: Integrated collection from dynamic, disparate, and distributed sources of differing formats, schemas, protocols, speeds, and sizes, such as machines, geolocation devices, clickstreams, files, social media feeds, log files, and videos
- **Real-time decision making**: Real-time evaluation of perishable insights at the network edge to determine whether data is pertinent and to perform processing based on consequent decisions to send, drop, or locally store data as needed

Figure 3. Hortonworks Data Platform 2.6

Cisco UCS Integrated Infrastructure for Big Data and Analytics
Reference Architecture

The reference architectures are optimally designed and tested to ensure a balance between performance and capacity. These configurations can be deployed as-is or use as templates for building custom configurations. The solution can be customized based on workload demands, including expansion to thousands of servers through the use of Cisco Nexus® 9000 Series Switches. This next-generation infrastructure can be used to power extremely fast data access, when large storage is required for the modern applications with its blazingly fast compute and memory and flexible storage options.

Table 1 lists the performance and capacity options for the Cisco UCS Integrated Infrastructure for Big Data and Analytics.

<table>
<thead>
<tr>
<th>Bundle</th>
<th>Performance</th>
<th>Capacity</th>
<th>High Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Server SKU</td>
<td>UCS-SP-C240M5-A2</td>
<td>UCS-SPC240M5L-S1</td>
<td>UCSS-SP-S3260-BV</td>
</tr>
<tr>
<td>Servers</td>
<td>16 x Cisco UCS C240 M5 with SFF drives</td>
<td>16 x Cisco UCS C240 M5 with LFF drives</td>
<td>8 x Cisco UCS S3260 Storage Server, each server node with:</td>
</tr>
<tr>
<td>CPU</td>
<td>2 Intel Xeon Processor Scalable Family 6132 (2 x 14 cores, 2.6 GHz)</td>
<td>2 Intel Xeon Processor Scalable Family 4110 (2 x 8 cores, 2.1 GHz)</td>
<td>2 Intel Xeon processor E5-2680 v4 CPUs (2 x 14 cores, 2.4 GHz)</td>
</tr>
<tr>
<td>Memory</td>
<td>12 x 16 GB 2666 MHz (192 GB)</td>
<td>12 x 16 GB 2666 MHz (192 GB)</td>
<td>8 x 32 GB 2400 MHz (256 GB)</td>
</tr>
<tr>
<td>Boot</td>
<td>M.2 with 2 x 480-GB SSD</td>
<td>M.2 with 2 x 480-GB SSD</td>
<td>2 x 480-GB enterprise value boot SSD</td>
</tr>
<tr>
<td>Storage</td>
<td>26 x 1.8TB 10K rpm SFF SAS HDDs or 12 x 1.6 TB Enterprise Value SATA SSDs.</td>
<td>12 x 8-TB 7.2K LFF SAS HDDs + 2 SFF rear hot-swappable 1.6TB Enterprise Value SATA SSDs</td>
<td>24 x 6 TB 7.2K LFF SAS HDDs</td>
</tr>
<tr>
<td>VIC</td>
<td>40 Gigabit Ethernet (VIC 1387)</td>
<td>40 Gigabit Ethernet (VIC 1387)</td>
<td>40 Gigabit Ethernet (VIC 1387)</td>
</tr>
<tr>
<td>Storage Controller</td>
<td>Cisco 12-Gbps SAS modular RAID controller with 4-GB flash-based write cache (FBWC) or Cisco 12-Gbps modular SAS host bus adapter (HBA)</td>
<td>Cisco 12-Gbps SAS modular RAID controller with 2-GB flash-based write cache (FBWC) or Cisco 12-Gbps modular SAS HBA</td>
<td>Cisco 12-Gbps SAS modular RAID controller with 4-GB FBWC</td>
</tr>
<tr>
<td>Network Connectivity</td>
<td>Cisco UCS 6332 Fabric Interconnect</td>
<td>Cisco UCS 6332 Fabric Interconnect</td>
<td>Cisco UCS 6332 Fabric Interconnect</td>
</tr>
</tbody>
</table>

Note: For the management nodes use three Cisco UCS C240 M5 Rack Servers each with 2 Intel Xeon Processor Scalable Family 6132 Processors, 384 GB of memory, a 12-Gbps SAS RAID controller with a 4-GB cache, 10 x 1.8-TB 10,000-rpm SFF SAS drives, and Cisco UCS VIC 1387 (two 40 Gigabit Ethernet QSFP interfaces).

For HDF Deployments use 8 Cisco UCS C220M5 Rack Servers with 2 Intel Xeon Processor Scalable Family 6132 Processors, 384 GB of memory, a 12-Gbps SAS RAID controller with a 2-GB cache, 10 x 1.8-TB 10,000-rpm SFF SAS drives, and Cisco UCS VIC 1387 (two 40 Gigabit Ethernet QSFP interfaces).
Conclusion

The fifth generation of Cisco UCS Integrated Infrastructure for Big Data and Analytics is the next generation of platform with new processors, faster memory, and more storage options. It is designed, tested, and validated for enterprises to lower the cost of ownership, and to scale from small to very large as applications demand. With Cisco ACI, it can scale to thousands of nodes. The Cisco UCS delivers an optimal combination of high availability, performance, and flexibility while protecting your long-term investments.

Reference

- For more information about Cisco UCS, please visit https://www.cisco.com/go/ucs.
- For more information about Cisco UCS big data solutions, please visit https://www.cisco.com/go/bigdata.
- For more information about Cisco’s Big Data validated designs, visit https://www.cisco.com/go/bigdata_design.
- For more information about Cisco UCS Integrated Infrastructure for Big Data and Analytics, visit https://blogs.cisco.com/datacenter/cpav5.