Cisco UCS Integrated Infrastructure for Big Data and Analytics with Cloudera Enterprise

Bring faster performance and scalability for big data analytics.

Highlights

Proven platform for the Enterprise Data Hub
- The solution is the fifth generation of the Cisco UCS® Integrated Infrastructure platform, deployed across major industries, such as agriculture, education, entertainment, finance, healthcare, industrial, insurance, manufacturing, public sector, service provider, and utilities.
- The infrastructure has been validated with industry-standard benchmarks.

Designed, tested, and validated for faster time to value
- Cisco® Validated Designs (CVD) facilitate faster, more reliable, and more predictable customer deployments and provide design, scalability, and performance recommendations.

Built on the Cisco Unified Computing System
- The Cisco Unified Computing System™ (Cisco UCS) M5 platform offers complete integration of computing, networking, and storage resources with unified management and provides high performance, expandable storage, and scalability for big data systems.
- The solution uses a fabric-centric architecture designed for business acceleration, providing a true on-demand infrastructure and 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.
- Used in combination with Cloudera Manager, a holistic interface that provides end-to-end system management and detailed and precise visibility and control over every part of an enterprise data hub, the solution makes cluster management simple and straightforward.
Cisco and Cloudera Deliver World-Class Solutions for Powering the Enterprise Data Hub

Business users are continually envisioning new and innovative ways to use data for operational reporting and advanced analytics. And data is being generated at an unprecedented scale, and more data is being collected faster and stored longer than ever before. Traditional transactional data is being supplemented with data from high-speed, real-time streaming systems and then stored for longer periods of time to derive insights and for archival and regulatory purposes.

A major challenge of big data systems is managing the rapidly growing amount of data and the corresponding increasing costs. 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 to provide sentiment and exploratory analytics, trigger any alerts, etc. Organizations need the right technology and infrastructure to deploy workflows and processes that can help optimize the use of this data.

A data hub can be an effective data management solution for both advanced analytics experts and business users. From processing data science and engineering workloads, to powering an operational database, to running large-scale analytics, a data hub allows users to handle everything from a single place.

A data hub allows users to analyze large volumes and varieties of data on demand. It can be built from a multitude of products and provides users with extensive, scalable customization to address a variety of use cases.

Figure 1 shows the typical architecture for a data hub powered by Cloudera Enterprise Data Hub. It shows the flow of data from various data sources to Kafka nodes and then to Apache Spark or to Hadoop Distributed File System (HDFS), NoSQL databases, SQL databases, Apache Solr, and other systems for additional processing.

Figure 1. Typical Architecture with Cloudera Enterprise Data Hub
Cisco UCS Integrated Infrastructure for Big Data and Analytics

Organizations today must help 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 managing computing, network, and storage resources, allowing organizations to successfully address the business needs of IT innovation and acceleration. Cisco UCS Integrated Infrastructure for Big Data and Analytics provides an end-to-end architecture for processing high volumes of structured and unstructured data for both real-time processing 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. 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 C240 M5 Rack Servers

The Cisco UCS C240 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 new Intel Xeon Processor Scalable Family, with up to 28 cores per socket. It supports up to 24 DDR4 DIMMs for improved performance and lower power consumption.

The DIMM slots are also 3D XPoint ready, supporting next-generation nonvolatile memory technology. It has a range of storage options, with up to 24 small-form-factor (SFF) 2.5-inch drives, with support for up to 10 Non-Volatile Memory Express (NVMe) PCIe solid-state disks (SSDs) on the NVMe-optimized chassis version, or 12 large-form-factor (LFF) 3.5-inch drives plus 2 rear hot-swappable SFF drives with a Cisco 12-Gbps SAS Module RAID Controller. Additionally, it has two modular M.2 cards that can be used for bootup. A modular LAN-on-motherboard (mLOM) slot supports dual 40-Gbps network connectivity with the Cisco UCS Virtual Interface Card (VIC) 1387.

Figure 2. Cisco UCS Integrated Infrastructure for Big Data and Analytics with Cloudera Enterprise
Cloudera Enterprise

Cisco and Cloudera provide organizations with an enterprise-ready data management platform as well as management integration with the enterprise application ecosystem. They transparently combine to provide a uniquely capable industry-leading architectural platform for Hadoop-based applications (Figure 3).

Cloudera Enterprise provides a scalable, flexible, integrated platform that makes it easy to manage rapidly increasing volumes and varieties of data in your enterprise. Industry-leading Cloudera products and solutions enable businesses to deploy and manage Apache Hadoop and related projects and manipulate and analyze data and keep that data secure and protected.

Cloudera Enterprise enables more insights for more users from within a single platform. With the most powerful open-source tools and the only active data optimization designed for Hadoop, you can move from big data to results faster.

Main features include:

- In-memory data processing: The most experience with Apache Spark
- Fast analytic SQL: The lowest latency and best concurrency for business intelligence with Apache Impala
- Native search: Complete user accessibility built in to the platform with Apache Solr
- Updatable analytic storage: The only Hadoop storage for fast analytics for fast-changing data with Apache Kudu

Figure 3. Cloudera Enterprise and Cisco UCS Integrated Infrastructure
Reference Architecture

The reference architectures for the solution are optimally designed and tested to achieve a balance between performance and capacity. These configurations can be deployed as is or used 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. With its extremely fast computing and memory and flexible storage options, this next-generation infrastructure can be used to power fast data access to the large storage resources required for modern applications.

Table 1 lists configuration options for the solution.

Table 1. Cisco UCS Integrated Infrastructure for Big Data and Analytics Configuration Options

<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 Rack Servers with SFF drives</td>
<td>16 x Cisco UCS C240 M5 Rack Servers with LFF drives</td>
<td>8 x Cisco UCS S3260 Storage Server, each server node with:</td>
</tr>
<tr>
<td>CPU</td>
<td>2 x Intel Xeon Processor Scalable Family 6132 (2 x 14 cores, 2.6 GHz)</td>
<td>2 x Intel Xeon Processor Scalable Family 4110 (2 x 8 cores, 2.1 GHz)</td>
<td>2 x 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 2400MHz (256 GB)</td>
</tr>
<tr>
<td>Boot</td>
<td>M.2 with 2 x 480-GB SSDs</td>
<td>M.2 with 2 x 480-GB SSDs</td>
<td>2 x 480-GB Enterprise Value Boot SSDs</td>
</tr>
<tr>
<td>Storage</td>
<td>26 x 1.8 TB 10K rpm SFF SAS HDDs or 12 x 1.6 TB Enterprise Value SATA SSDs.</td>
<td>12 x 8 TB 7.2K rpm LFF SAS HDDs + 2 SFF rear hot-swappable 1.6 TB Enterprise Value SATA SSDs</td>
<td>24 x 6 TB 7.2K rpm LFF SAS HDDs</td>
</tr>
<tr>
<td>VIC</td>
<td>40 Gigabit Ethernet (Cisco UCS VIC 1387)</td>
<td>40 Gigabit Ethernet (Cisco UCS VIC 1387)</td>
<td>40 Gigabit Ethernet (Cisco UCS 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 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 two Intel Xeon Processor Scalable Family 6132 CPUs, 384 GB of memory, Cisco 12-Gbps SAS RAID Controller with a 4-GB cache, and Cisco UCS VIC 1387 (two 40 Gigabit Ethernet Quad Small Form-Factor Pluggable [QSFP] interfaces).
Conclusion

The fifth generation of Cisco UCS Integrated Infrastructure for Big Data and Analytics builds on the previous generation of platforms 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 deployments as applications demand. With Cisco Application Centric Infrastructure (Cisco ACI™), it can scale to thousands of nodes. This integrated infrastructure delivers an optimal combination of high availability, performance, and flexibility while protecting your long-term investments.

Reference

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