Cisco UCS Integrated Infrastructure for Analytics with the MapR Converged Data Platform

Bring Flexibility, Scalability, and Performance for Big Data and Analytics.

Highlights

Proven enterprise-ready converged data platform
- The solution is the fifth generation of the Cisco UCS® Integrated Infrastructure platform, deployed across major industry verticals such as agriculture, education, entertainment, finance, healthcare, industrial, insurance, manufacturing, public sector, service provider, and utilities.
- The solution has been validated through industry-standard benchmarks.

Designed, tested, and validated for faster time to value
- Cisco® Validated Designs 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, providing business acceleration, a true on-demand infrastructure, and a system that grows gracefully and incrementally.

Designed to scale from small to very large deployments 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
- Deploy Cisco UCS Director Express for Big Data quickly and easily for big data infrastructure with one-click provisioning, installation, and configuration. It provides a holistic interface for end-to-end system management and detailed and precise visibility and control over every part of an enterprise data platform, thus making cluster management simple and straightforward.
Cisco and MapR Deliver an Enterprise-Ready Converged Data Platform

With ever-growing amounts of data being generated and stored longer, it is becoming increasingly important for organizations to be able to transform this large volume of data into actionable information. Enterprises are moving from traditional transactional data processing to high-speed, real-time streaming systems. Data generated by sensors, Internet of Things (IoT) devices, social media, online transactions, and other such sources is the main contributor to this change.

Data lakes today have evolved into ingest points for new data sources and are storing increasing amounts of data. Meeting the emerging big data requirements requires a scalable and integrated infrastructure. Today’s data lakes also are enabling new exploratory analytics that were not possible with traditional data warehouses.

Cisco and MapR together provide a comprehensive platform to address the fast-growing computing and data-intensive real-time and batch workloads in big data applications. This solution enables organizations to easily deploy the Apache Hadoop ecosystem of technologies with the simplified deployment model of the Cisco Unified Computing System™ (Cisco UCS®) and to can expand the solution on demand to support powerful big data analytics.

Cisco UCS Integrated Infrastructure for Big Data and Analytics

Organizations today must be sure 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 compute, network and storage resources 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. 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)

Cisco UCS M5 rack servers are dual-socket, 2-rack-unit (2RU) servers offering industry-leading performance and expandability for a wide range of storage and I/O-intensive infrastructure workloads, for big data and analytics. M5 servers use the new Intel® Xeon® Processor Scalable Family, with up to 28 cores per socket. They support 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. They have 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 Modular RAID Controller. Additionally, they have 2 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 1 shows Cisco UCS Integrated Infrastructure for Big Data and Analytics with Cisco UCS C240 M5 Rack Servers.
As one of the leaders in Hadoop technology, MapR, with the MapR Converged Data Platform, enables enterprise-class big data solutions that organizations can develop quickly and administer with ease. With significant investment in critical technologies, MapR offers one of the industry’s most comprehensive Hadoop platforms, fully optimized for performance and scalability.

The MapR distribution delivers more than a dozen tested and validated Hadoop software modules over a fortified data platform, offering exceptional ease of use, reliability, and performance for big data solutions, as shown in Figure 2.

Main features of the MapR Converged Data Platform include the following:

- **Performance**: Use of the MapR file system, designed for high performance and throughput
- **Scalability**: Up to a trillion files, with no restrictions on the number of nodes in a cluster
- **Standards-based APIs and tools**: Standard Hadoop APIs, Open Database Connectivity (ODBC), Java Database Connectivity (JDBC), Lightweight Directory Access Protocol (LDAP), Linux Pluggable Authentication Modules (PAM), and more
- **MapR Direct Access Network File System (NFS)**: Random read and write operations and real-time data flows and transparent support for existing non-Java applications
- **Manageability**: Advanced management console, rolling upgrades, and representational state transfer (REST) API support
- **Integrated security**: Kerberos and non-Kerberos options with wire-level encryption
- **Advanced multitenancy**: Volumes, data placement control, job placement control, queues, and more
- **Consistent snapshots**: Full data protection with point-in-time recovery
- **High-availability**: Ubiquitous high-availability with elimination of NameNode architecture and with YARN high-availability and NFS high-availability
- **Disaster recovery**: Cross-site replication with mirroring
- **MapR-DB**: Integrated enterprise-class NoSQL database
- **MapR Streams**: Event streaming on a global scale
Reference Architecture

The reference architectures for this 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. This next-generation infrastructure can be used to power extremely fast data access for the modern applications with its fast computing and memory and flexible storage options.

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>UCS-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></td>
</tr>
<tr>
<td>CPU</td>
<td>2 Intel Xeon Processor Scalable Family 6132 CPUs (2 x 14 cores and 2.6 GHz)</td>
<td>2 Intel Xeon Processor Scalable Family 4110 CPUs (2 x 8 cores and 2.1 GHz)</td>
<td>2 Intel Xeon processor E5-2680 v4 CPUs (2 x 14 cores and 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>
</tbody>
</table>
### 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. 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 big data solutions, see [https://www.cisco.com/go/bigdata](https://www.cisco.com/go/bigdata).
- For more information about Cisco UCS Integrated Infrastructure for Big Data and Analytics, see [https://blogs.cisco.com/datacenter/cpav5](https://blogs.cisco.com/datacenter/cpav5).
- For more information about Cisco’s big data validated designs, see [https://www.cisco.com/go/bigdata_design](https://www.cisco.com/go/bigdata_design).
- For more information about Cisco UCS, see [https://www.cisco.com/go/ucs](https://www.cisco.com/go/ucs).