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

Deploy a highly scalable and secure data management platform.

Highlights

Real-time performance with linear scalability

Cisco UCS® Integrated Infrastructure for Big Data and Analytics, combined with DataStax Enterprise (DSE), built on Apache Cassandra™, providing operations management, developer tooling, advanced security, graph, analytics, and search. It scales from zero to hundreds of terabytes and billions of records without the need for downtime and handles the most challenging mixed workloads with ease.

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.

Cisco UCS uses fabric-centric architecture designed for business acceleration, providing true on-demand infrastructure with a system that grows gracefully and incrementally.

Ease of deployment

Cisco UCS Manager simplifies infrastructure deployment with an automated, policy-based mechanism that helps reduce configuration errors and system downtime.

Integrated real-time analytics and external Apache Hadoop support

Run real-time in-memory and batch analytics with DSE Analytics through Apache Spark™. Transparently integrate operational data in Apache Cassandra with historical information in Apache Hadoop™ distributions without the need for resource-intensive data integration.

Elastic scalability

The solution is designed to grow to maximum size without complexity or the need for downtime, allowing you to easily add capacity online to accommodate more customers and more data whenever you need.

Operational Simplicity and Integrated Analytics for Real-Time Online Applications

The massive growth of digital businesses is guiding the next generation of data architectures. To optimize business operations with real-time decisions and improve the overall customer experience, business owners need to be able to analyze data and gain actionable insights.

Big data technologies are becoming integral components of enterprise application ecosystems, moving from evaluation to adoption by enterprises of all sizes. As these important technologies help enterprises transform large volumes of data into actionable information, many organizations are seeking the most effective and reliable data management infrastructure that delivers the performance, agility, elasticity, and innovation needed to build high-performance, highly available online applications.

Cisco and DataStax have worked together to design and deliver an enterprise-class data management system that delivers continuous availability, outstanding performance, and linear scalability, plus full enterprise-class support. The result is a customer-centered solution that can be rapidly deployed, scaled on demand, and secured.

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 computing, 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 purposes.

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 the Cisco Unified Computing System™ (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 and C220 M5 Rack Servers

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, such as big data and analytics. These servers use the new Intel® Xeon® Processor Scalable Family with up to 28 cores per socket. They support up to 24 double-data-rate 4 (DDR4) dual in-line memory modules (DIMMs) for improved performance and lower power consumption. The DIMM slots are 3D XPoint ready, supporting next-generation nonvolatile memory technology. Depending on the server type, Cisco UCS rack servers offer a range of storage options. The Cisco UCS C240 M5 Rack Server supports up to 26 small-form-factor (SFF) 2.5-inch drives (with support for up to 10 Non-Volatile Memory Express [NVMe] Peripheral Component Interconnect Express [PCIe] solid-state drives [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. The Cisco UCS C220 M5 Rack Server supports up to 10 SFF 2.5-inch drives (with 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 (VIC) 1387.

DataStax Enterprise: Powered by the Best Distribution of Apache Cassandra

Today's digital economy demands your applications to be ready for anything – enormous growth, mixed workloads, or even catastrophic failure. Built on the best distribution of Apache Cassandra, DataStax Enterprise is the always-on data platform designed to allow you to effortlessly build and scale your apps, integrating graph, search, analytics, administration, developer tooling, and monitoring into a single unified platform. (Figure 1).

Figure 1. DataStax Enterprise Comprehensive Database Platform

The main DSE features include:

- Integrated streaming and real-time, ad hoc and batch analytics
- Transparent integration with solutions from external Apache Hadoop vendors
- In-memory database option for fast performance for transactional and analytical workloads
- Enterprise search with result highlighting, faceted search, robust document handling, and geospatial support
- Automatic management services and visual management for monitoring key administration and performance indicators
- Three hundred percent increase in read performance in DSE and Apache Spark for analytical queries versus standard open source alternatives, improving Spark interactive and impromptu use cases
- Simplified and tighter integration of mixed workloads, including expanded capabilities of the DataStax Enterprise File System (DSEFS) - distributed, scalable, fault-tolerant file system, with no single point of failure and full compatibility with Apache Hadoop Distributed File System (HDFS), for operational database support

- DataStax Enterprise File System (DSEFS) distributed, scalable, fault-tolerant file system, with no single point of failure and full compatibility with Apache Hadoop Distributed File System (HDFS), for operational database support

Table 1 summarizes the new features and major enhancements in the recent releases of DSE.

Feature	Description
DSE Analytics	Improvements include support for Apache Spark 2.0, including SparkR. Enhancements include operational improvements, performance improvements, structured streaming, DSE GraphFrames, Spark SQL, and geospatial types.
DSEFS	DataStax Enterprise File System provides a distributed file system for storing very large data sets, such as Spark Streaming data and analytics processing. DSEFS replaces the Apache Cassandra File System (CFS).
DSE Graph	Improvements include new fuzzy search, advanced configuration of search indexing, better data-loading performance, improved geospatial querying, Kerberos support for DSE Graph Loader, capability to customize graph visualizations, and stability improvements.
DSE Security	New security features include detailed access control that allows permissions on table rows and search indexes, DSE Proxy management for web services, and Java Management Extensions (JMX) authentication integrated with DSE Unified Authentication (Lightweight Directory Access Protocol [LDAP] or internal).
DSE Search	Enhancements include faster and simpler DSE Search feature using Apache Solr 6.0, new index management Apace Cassandra Query Language (CQL) shell (CQLSH) commands to simplify operations and development processes, improved query and indexing performance, enhanced error handling, and improved hardening and stability.
DSE In-Memory	Enhancements include MemoryOnlyStrategy using compression.

Table 1. DataStax Enterprise Release 5.0 main features and enhancements

DSE OpsCenter

DSE OpsCenter is the visual management and monitoring solution for DSE. OpsCenter includes integrated functions for real-time monitoring, tuning, provisioning, backup, and security management. It provides everything you need to intelligently manage and monitor your data management cluster deployment at scale. OpsCenter installs transparently, and you can easily use it as the single point of control for even the most complex workloads from any browser (Figure 2).

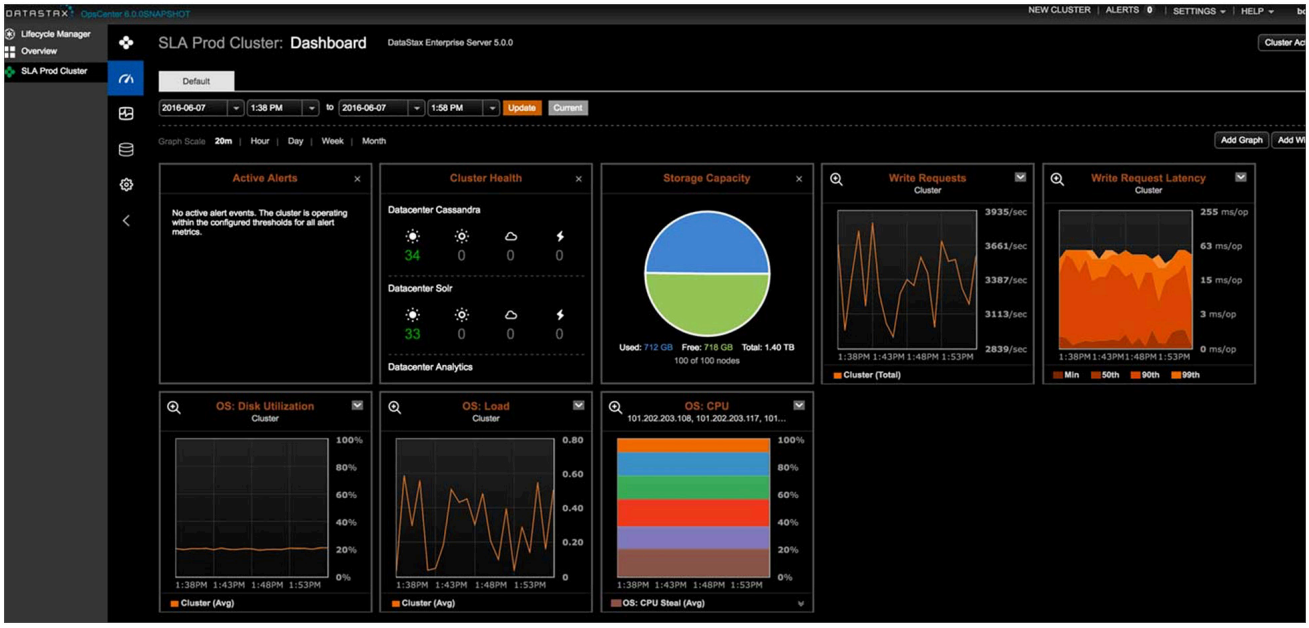


Figure 2. DSE OpsCenter Dashboard for Managing and Monitoring **for** DataStax Enterprise

Common use cases

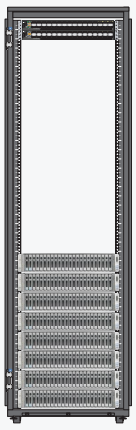
Because of its progressive architecture and built-in capabilities, DSE excels in the following application use cases:

- Inventory Management
- Logistics
- Supply Chain
- Security (ie, threat detection such as denial of service attacks, etc.)
- Identity Management (including Authentication, Authorization)
- E-commerce (shopping cart experience, etc.)
- Asset Monitoring
- Customer Loyalty
- Customer 360 (C360)
- Personalization/Recommendation
- Consumer Fraud Detection

Reference Architecture

Cisco UCS Integrated Infrastructure for Big Data and Analytics for DataStax includes eight or more C240 M5 (or C220 M5) Rack Servers, each with dual Intel® Xeon® Processor Scalable Family 6132 CPUs (2 x 14 cores and 2.6 GHz), 384 GB of RAM, dual 40-Gbps network connectivity, and 8 (or 16) solid-state disks (SSDs). The servers are connected to Cisco UCS 6332 Fabric Interconnects.

Figure 3 shows the reference architecture for DSE.



Network connectivity:

- Cisco UCS 6332 Fabric Interconnect (40 Gbps)

Starter

Servers: 8 Cisco UCS C220 M5 Rack Servers, each with:

- Dual Intel Xeon Processor Scalable Family 6132 CPUs (2 x 14 cores and 2.6 GHz)
- 12 x 16 GB 2666 MHz (192 GB) of DDR4 RAM
- 8 x 960-GB (or 1.6-TB) SSDs
- Cisco UCS VIC 1387 with 2 x 40-Gbps port connectivity
- Cisco® 12-Gbps SAS Modular RAID Controller with 2-GB flash-based write cache (FBWC)
- M.2 with 2 x 480-GB SSDs

Advanced

Servers: 8 Cisco UCS C240 M5 Rack Servers, each with:

- Dual Intel Xeon Processor Scalable Family 6132 CPUs (2 x 14 cores and 2.6 GHz)
- 12 x 16 GB 2666 MHz (192 GB) of DDR4 RAM
- 16 x 960-GB (or 1.6-TB) SSDs
- Cisco UCS VIC 1387 with 2 x 40-Gbps port connectivity
- Cisco 12-Gbps SAS Modular RAID Controller with 4-GB FBWC
- M.2 with 2 x 480-GB SSDs

Figure 3. Cisco UCS Reference Architecture for DataStax Enterprise

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 total cost of ownership (TCO) and scale from small to very large deployments as applications demand.

The combination of DSE, and Cisco UCS provides the power of an always-on distributed database with a dependable deployment model that can be implemented rapidly using Cisco UCS servers.

For more information

- To find out more about Cisco UCS big data solutions, see https://www.cisco.com/go/bigdata.
- For more information about DataStax Enterprise, see www.datastax.com .

C22-739254-00 07/17