High Performance Hadoop Solution with Cisco UCS powered by AMD, Samsung and Hortonworks

Solution highlights

Proven Cisco Unified Computing System™ (Cisco UCS®) architecture foundation
Cisco UCS Integrated Infrastructure for Big Data and Analytics is a proven platform for enterprise analytics applications. The Cisco UCS platform offers complete integration of computing, networking, and storage resources with unified management, providing easy, linear scalability of the architecture.

Linear scalability: from small to very large
Using the Cisco Application Centric Infrastructure (Cisco ACI™) platform, a cluster can be scaled to thousands of nodes. The Cisco ACI platform implements an application-aware, policy-based approach that treats the network as a single entity rather than a collection of switches.

Ease of deployment with UCS Manager and Intersight
Cisco UCS Manager simplifies infrastructure deployment with an automated, policy-based mechanism that helps reduce configuration errors and system downtime. It offers proven, high-performance linear scalability and easy scaling of the architecture with single and multiple-rack deployments. Apart from UCS Manager, the Cisco Intersight™ platform provides intelligent cloud-powered infrastructure management for the Cisco Unified Computing System™ (Cisco UCS®) and Cisco HyperFlex™ platforms.

Modular design
The Cisco UCS C4200 Series Rack Server Chassis with Cisco UCS C125 M5 Rack Server nodes powered by AMD EPYC processors is optimized for use in environments requiring dense computing form factors and high core densities, such as scale-out, computing-intensive, general service provider, and bare-metal applications.

Platform for enterprise data lake with Hortonworks Data Platform (HDP)
The Hortonworks product suite addresses the complete needs of data in motion and data at rest. Hortonworks offers a scalable, fault-tolerant, and cost-efficient platform. It provides consistent data management, governance, and security capabilities while delivering robust analytics from real-time customer applications that accelerate decision making and innovation.

Enterprise storage lifecycle management with Samsung solid-state disks (SSDs)
Samsung SSDs deliver optimized performance and value for big data and analytics applications, and reliable enterprise storage solutions with low power consumption. High-density computing architecture in combination with Samsung SSDs helps reduce total cost of ownership (TCO) by reducing the number of servers needed to support existing and future applications.

## Cisco and Hortonworks solution for big data analytics

Data is an organization's future and its most valuable asset. The accumulation of data—from multiple sources, such as sensors, Internet of Things (IoT) devices, social networking, and online transactions— is growing at an unprecedented rate. These sources are generating data that needs to be captured, monitored, and rapidly processed regardless of origin to enable organizations to make informed decisions based on this data in a timely manner.

A huge challenge for enterprises is managing all this data, with its increasing volume and variety and a growth rate that was rarely seen in the past. Big data and analytics architectures are making this data increasingly available for processing. With both data in motion and data at rest now readily available, customers are discovering that they can extract real business value from insights gained through data analytics and data science.

Data lakes enhance and amplify existing IT investments and provide new ways to create business value. They also provide a cost-effective and technologically feasible way to meet today's day big data challenge. They enable the storage and analysis of large volumes and wide varieties 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.

The Cisco Unified Computing System™ (Cisco UCS®) and Hortonworks together create a solution that helps enterprises transform their businesses by unlocking the full potential of big data. Cisco UCS Integrated Infrastructure for Big Data and Analytics with Hortonworks Data Platform (HDP) powers the next-generation architecture for big data systems, encompassing myriad use cases, including IoT, fraud analytics, and precision medicine through genome sequencing (Figure 1).

Figure 1.   Enterprise data lake with HDP
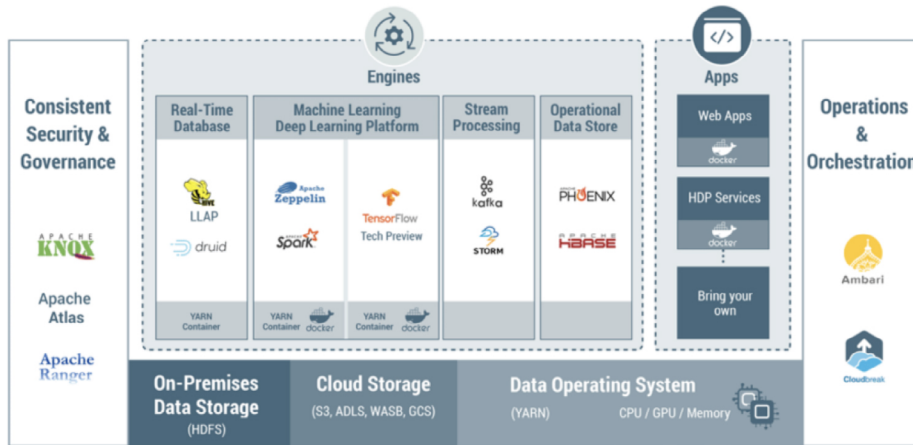
Cisco UCS Integrated Infrastructure for Big Data and Analytics

Cisco UCS Integrated Infrastructure for Big Data and Analytics provides an end-to-end architecture for processing high volumes of real-time and historical data, both structured and unstructured. At the same time, it quickly and efficiently delivers out-of-the-box performance while allowing organizations to start small and then scale to very large deployments as the needs of the business and big data and analytics requirements grow.

Organizations today must be sure that the underlying physical infrastructure can be deployed, scaled, and managed in a way that is agile enough to adapt 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.

Reference architecture

The reference architecture for Cisco UCS C4200 Series Rack Service Chassis with Cisco UCSs C125 M5 Rack Server nodes powered by AMD EPYC processors and Hortonworks' big data distribution is optimally designed and tested to help ensure a balance between performance and capacity. It can scale out to meet big data and analytics requirements. It can expand to thousands of servers with Cisco Nexus® 9000 Series Switches using the Cisco® Application Policy Infrastructure Controller (APIC) with a leaf-and-spine design using the Cisco Application Centric Infrastructure (Cisco ACI™) platform. This next-generation infrastructure can be deployed to meet a wide variety of computing, storage, and connectivity options.

Figure 2.   Solution Reference Architecture
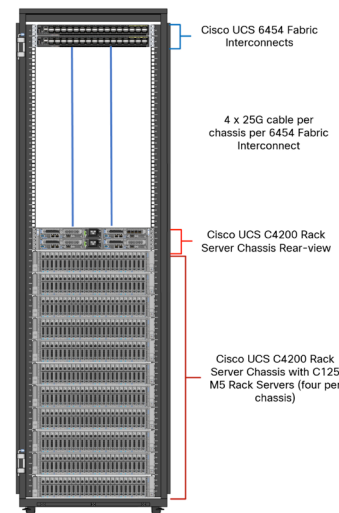
Components	High Performance
Servers	12 x Cisco UCS C4200 Rack Server
	Each with:4 x Cisco UCS C125 M5 Rack Servers
CPU	2 x AMD EPYC 7401 Processors (2 x 24 cores, 2.0 GHz)
Memory	8 x 32 GB DDR4 (256GB)
Boot	M.2 with 2 x 240 GB SATA SSDs
Storage	6 x 3.8 TB Enterprise Value SATA SSD
VIC	25 Gigabit Ethernet (Cisco UCS VIC 1455)
Storage Controller	Cisco 12-Gbps SAS 9460-8i RAID Controller 2GB cache (FBWC)
Network Connectivity	2 x Cisco UCS 6454 Fabric Interconnect

Cisco UCS 6454 Fabric Interconnects

The Cisco UCS 6454 Fabric Interconnect is a core part of Cisco UCS, providing both network connectivity and management capabilities for the system. The Cisco UCS 6454 offers line-rate, low-latency, lossless 10, 25, 40, and 100 Gigabit Ethernet, Fibre Channel over Ethernet (FCoE), and Fibre Channel functions.

The Cisco UCS 6454 provides the management and communication backbone for the Cisco UCS B-Series Blade Servers, Cisco UCS 5108 B-Series Server Chassis, Cisco UCS Managed C-Series Rack Servers, and Cisco UCS S-Series Storage Servers. All servers attached to the Cisco UCS 6454 Fabric Interconnect become part of a single, highly available management domain. In addition, by supporting a unified fabric, the Cisco UCS 6454 provides both the LAN and SAN connectivity for all servers within its domain.

Cisco UCS C4200 Series Rack Server Chassis

The Cisco UCS C4200 Series Rack Server Chassis is a modular, density-optimized rack-server chassis that supports:

- Up to four Cisco UCS C125 M5 Rack Server nodes and up to 256 cores per chassis with AMD EPYC processors: It is excellent for environments requiring dense computing form factors and high core densities, such as scale-out, computing-intensive, general service provider, and bare-metal applications.

- 24 small-form-factor (SFF) drives: The drive bays are allocated so that each rack server node has access to six SAS and SATA HDD or solid-state disks (SSDs) or up to four SSDs and two Non-Volatile Memory Express (NVMe) drives.

Cisco UCS C125 M5 Rack Server

Cisco UCS C125 M5 Rack Server node has the greatest number of cores commercially available in a multinode system. It offers 2 sockets per node and from 8 to 32 cores per processor, with the support of the AMD EPYC 7000 series processors, 16 DIMM slots for 2666-MHz DDR4 DIMMs, and capacity points of up to 128 GB per slot for a total of 2 TB per socket. It offers up to 2 half-height and half-length PCI Express (PCIe) 3.0 slots and an optional M.2

SSD module. The C125 supports either SAS RAID through a PCIe 12-Gbps SAS storage controller card or SATA directly from the AMD EPYC processor. The node also includes a dedicated internal LAN mezzanine slot based on the Open Compute Project (OCP) 2.0 standard, supporting networking speeds of up to 100 Gbps. Additionally, a fourth-generation Cisco PCIe virtual interface card (VIC) can be added in the x16 PCIe 3.0 slot.

AMD EPYC 7000 series processor

Designed from the foundation for a new generation of solutions, AMD EPYC server processors implement a philosophy of choice without restriction. Choose the number of cores that meet your needs without sacrificing key features such as memory and I/O.

Each EPYC processor can have from 8 to 32 cores with access to an exceptional amount of I/O and memory regardless of the number of cores in use. The processors include 128 PCIe Generation 3 lanes and support for up to 2 TB of high-speed memory per socket.

The innovative AMD EPYC architecture provides outstanding performance. I/O-intensive workloads can use the plentiful I/O bandwidth with the right number of cores, helping organizations avoid overpaying for unneeded power. And computing-intensive workloads can make use of fully loaded core counts, dual sockets, and plenty of memory.

Samsung enterprise SSDs

Samsung enterprise SSDs offer a simple yet versatile and comprehensive selection of enterprise data storage and caching options suitable for nearly any application. They are engineered to provide high throughput and a consistent rate of I/O operations per second (IOPS). They offer large capacities of up to 3.84 TB, making them well suited for enterprise storage solutions for businesses seeking to enhance performance and cost effectiveness by upgrading their current servers or workstations from hard-disk drives (HDDs). These newer SSDs dramatically reduce latency and improve the IOPS rate.

Hortonworks Data Platform

HDP is built on Apache Hadoop and Spark, an open-source framework for distributed storage and processing of large, multisource data sets. It helps enterprises create a secure enterprise data lake from structured and unstructured data and delivers the analytics needed to help enterprises innovate quickly and gain valuable business insights.

HDP supports agile application deployment, machine-learning and deep-learning workloads, real-time data warehousing, and security and governance. It is a critical component of a modern data architecture for data at rest.

The Hadoop Distributed File System (HDFS) and YARN are the key components of HDP. HDFS provides 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 includes a versatile range of processing engines that enable interaction with the same data in multiple ways at the same time. Applications therefore can access the data using the best tool for the job: batch processing, interactive analysis, or low-latency access with NoSQL. Emerging use cases for data science, search, and streaming are also supported with Spark, Storm, and Kafka.

HDP extends data access and management with 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 at multiple layers. Critical features for authentication, authorization, accountability, and data protection are in place to help secure HDP across these important components.

Conclusion

Cisco UCS Integrated Infrastructure for Big Data and Analytics with Cisco UCS C4200 Series Rack Server Chassis and Cisco UCS C125 M5 Rack Servers powered by AMD Zen architecture–based EPYC processors, SAMSUNG SSDs, and Hortonworks' big data distribution help deliver a highly computing-intensive and scalable solution while aiming to solve a diverse set of real-world business problems. The proposed solution enables batch through real-time interactive processing, providing an enterprise-ready platform to support modern data applications.

For more information

- To find out more about Cisco UCS big data solutions, see http://www.cisco.com/go/bigdata.
- To find out more about AMD EPYC processors, see https://www.amd.com/en/products/epyc-server.
- To find out more about Hortonworks Data Platform, see https://hortonworks.com/products/data-platforms/hdp/.
- For more information about Samsung's line of SSDs, visit: https://www.samsung.com/enterprisessd