



Cisco Data Intelligence Platform (CDIP)

Modernizing your data lake to the evolving landscape

In today's environment, voluminous amount of data, reaching exabytes in scale, ends up being stored in a data ecosystem. Enterprises are constantly evaluating new sets of data management for processing, transforming, and analyzing these large amounts of data, leading to newer data pipelines that are evolving beyond the standard data lake.

The rapid advancement of Artificial Intelligence and Machine Learning (AI/ML) introduces a new set of challenges to the business and the IT organization's data strategy when implementing high-performance, scalable, and agile-fashion cloud-scale architecture.

The next generation of distributed systems for big data analytics needs to address data silos between different tiers such as data lakes, data warehouses, AI/compute, and object storage. It is imperative to develop an infrastructure that sustains a healthy data pipeline between storage devices and computing devices (CPU, GPU, FPGA). Also, reducing network bandwidth and achieving overall low latency for parallel data processing is critical for supporting an organization's data-driven goals.

Cisco® Data Intelligence Platform (CDIP) is a cloud-scale architecture that brings together big data, AI/compute farms, and storage tiers to work together as a single entity while also being able to scale independently to address the IT issues in the modern data center.

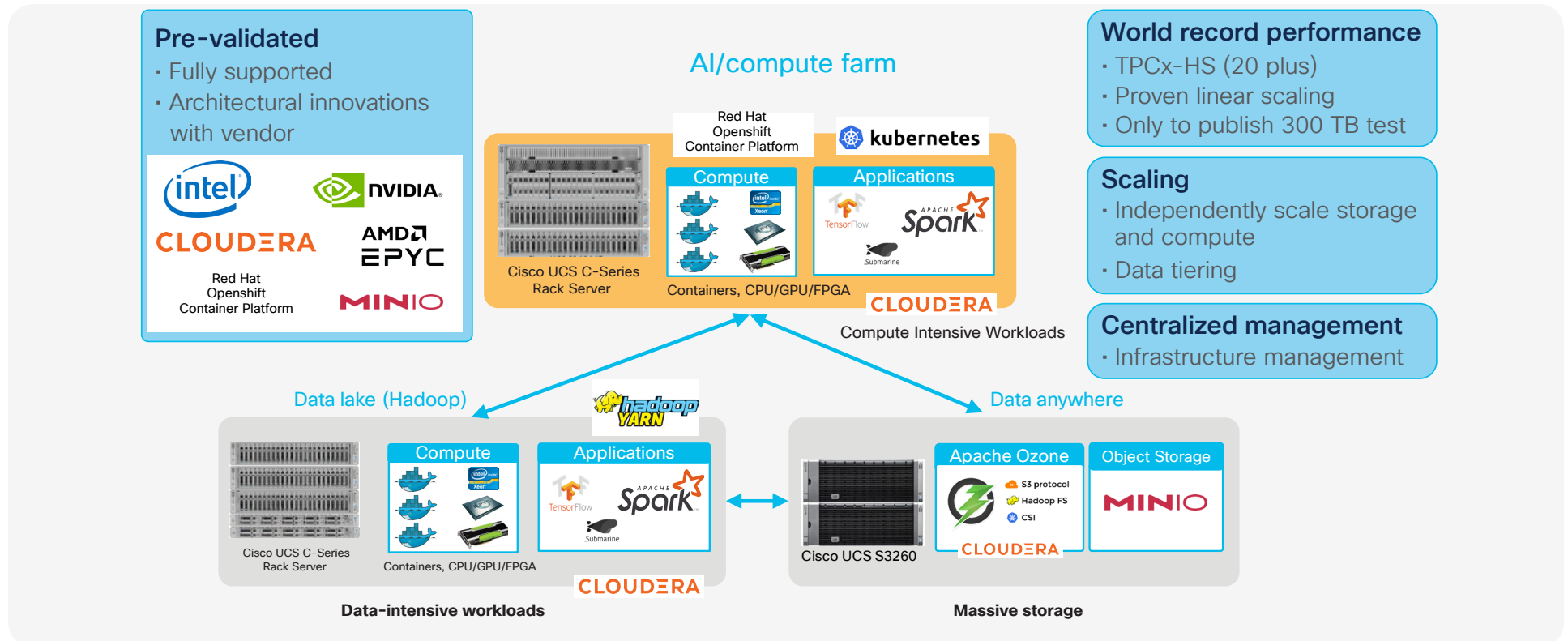
Cisco Data Intelligence Platform

The Cisco Data Intelligence Platform (CDIP) supports today’s evolving architecture for big data analytics. CDIP combines a fully scalable infrastructure with centralized management and a fully supported software stack (in partnership with industry leaders) to each of three independently scalable components of the architecture: the data lake, AI/ML technologies, and object stores.

Hadoop 3.0 introduced Docker support along with GPU isolation and scheduling. This opened up a plethora of opportunities for modern applications, such as micro-services and distributed applications running on thousands of containers to execute AI/ML algorithms on petabytes of data with ease and in a speedy fashion. CIDP is fully capable of addressing those application needs managed by either YARN or Kubernetes.

As the journey continues in the Hadoop ecosystem, more staggering and impressive frameworks and technologies, such as Apache submarine and Spark 3.0, are available to further complement it. Along with these new technologies, CDIP offers an extremely adaptable architecture and it evolves as underlying technologies, the platform, and frameworks changes, providing total investment protection. Figure 1 shows an overview of the CDIP architecture.

Figure 1. Cisco Data Intelligence Platform Architecture



CDIP delivers world-record performance

Cisco recently published five Hadoop performance world records on Cisco Data Intelligence Platform using TPCx-HS, an industry standard and leading benchmark for Hadoop throughput with both MapReduce and Spark.

By running different scale-factor workloads with Spark ([1 TB](#), [3 TB](#), and [10 TB](#)) and with MapReduce ([1 TB](#), [3 TB](#), and [10 TB](#)), we can demonstrate that Cisco Data Intelligence Platform is not only a leader in performance; it also scales linearly.

CDIP Highlights

Intelligent multi-domain management with Cisco Intersight

Cisco Intersight® enables IT to operationalize heterogeneous infrastructure and the application platform at scale to seamlessly function as a single cohesive unit through single-plane-of-glass management.

Powered by the latest generation in CPU from Intel and AMD

The latest generation of processors from Intel (Cascade Lake Refresh) and AMD (EPYC 7002 series) provides the foundation for powerful data center platforms with an evolutionary leap in agility and scalability.

Elimination of infrastructure silos with CDIP

CDIP is a highly modular platform that brings big data, AI compute farms, and object storage to work together as a single entity, while each component can scale independently to address the IT issues in the modern data center.

Disaggregated architecture

CDIP is a disaggregated architecture that brings together a more integrated and scalable solution for big data analytics and AI. It is specifically designed to improve resource utilization, elasticity, heterogeneity, and failure handling. It can also consume continuously evolving AI/ML frameworks and landscapes.

Pre-validated and fully supported

[Cisco Validated Design](#) (CVD) facilitates faster, more reliable, and more predictable customer deployments by providing configuration and integration of all components into a fully working optimized design. CVDs also provide scalability and performance recommendations.

ESG paper on the value of CDIP for our customers

The paper by ESG [Optimizing Analytics Workloads with the Cisco Data Intelligence Platform](#) goes into details on various aspects of considerations when designing your data lakes to cater to the evolving data pipelines supporting data engineers, data scientists, and data analysts and why CDIP is a platform of choice.

Fully supported and pre-validated architectural innovations with partners

Pre-tested and pre-validated through industry-standard benchmarks, tighter integration, and performance optimization with industry-leading Independent Software Vendor (ISV) partners in each of these areas: big data, AI, and object storage, Cisco Data Intelligence Platform offers best of the breed end-to-end validated architectures reduces integration and deployment risk by eliminating guesswork.

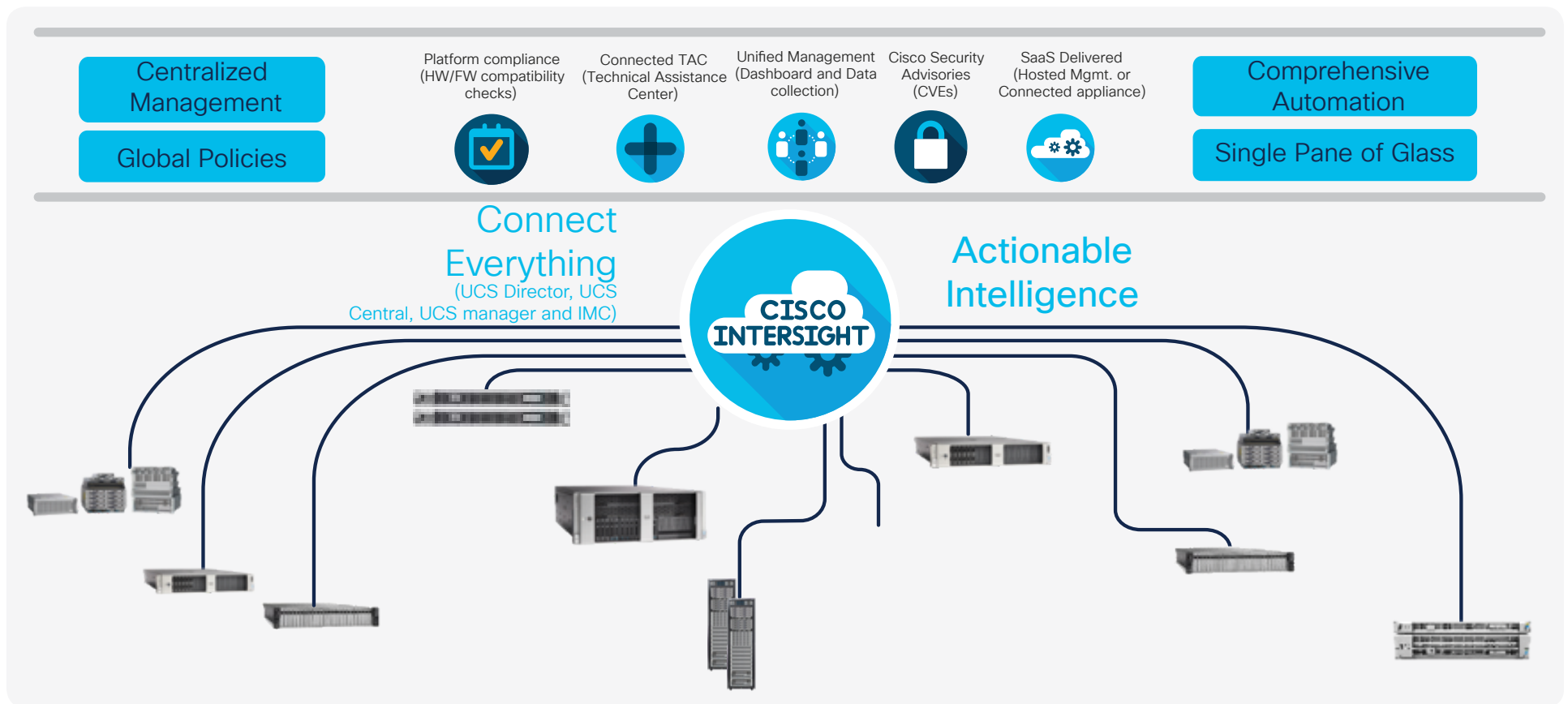
For more information, see: https://www.cisco.com/go/bigdata_design

Centralized management through Cisco Intersight

Cisco Intersight is a Software-as-a-Service (SaaS) infrastructure management tool that provides single-pane-of-glass management of CDIP infrastructure in the data center. Cisco Intersight scales easily, and frequent updates are implemented without impact.

Cisco Intersight (Figure 2) includes Cisco Technical Assistance Center (TAC) support, security advisories, and other capabilities for CDIP customers. CDIP with Cloudera Data Platform (CDP) can be fully deployed with Cisco Intersight.

Figure 2. Features of Cisco Intersight and how it fits in the infrastructure



Reference architecture

Cisco Data Intelligence Platform provides data lifecycle management with full integration and a Cisco Validated Design (CVD) for big data analytics. Our reference architectures are carefully designed, optimized, and tested with the leading big data and analytics software distributions to achieve a balance of performance and capacity to address specific application requirements. You can deploy these configurations as is or use them as templates for building custom configurations. You can scale your solution as your workloads demand, including expansion to thousands of servers through the use of Cisco Nexus® 9000 Series Switches. The configurations vary in disk capacity, bandwidth, price, and performance characteristics. Base configurations for each solution are listed in Tables 1, 2, 3, and 4. Figure 3 showcases various Cisco UCS® servers used in big data architectures.

Figure 3. Cisco UCS Integrated Infrastructure for Big Data and Analytics – Modernize Hadoop

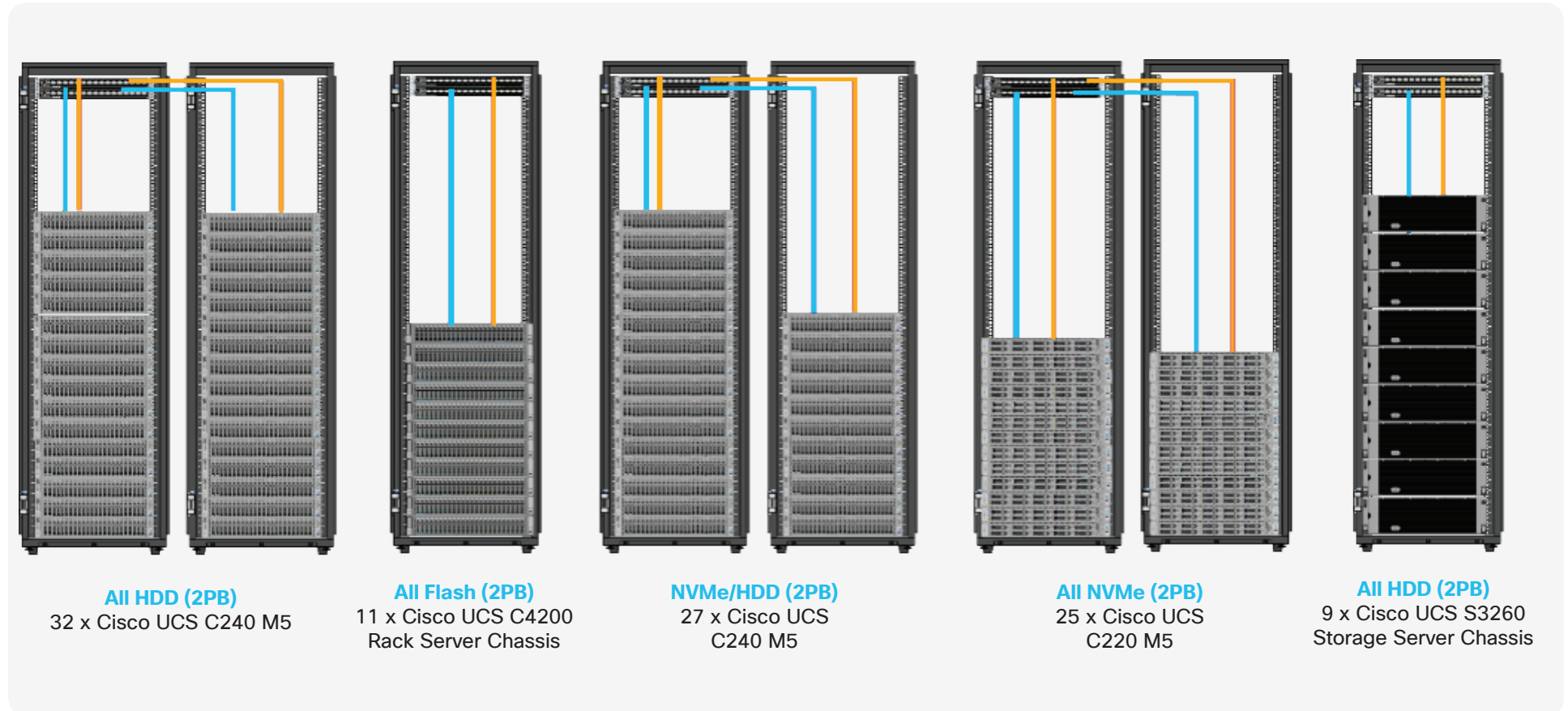


Table 1. Cisco UCS Integrated Infrastructure for Big Data and Analytics configuration options for data lakes

	NVMe performance	Flash performance	Performance	Capacity	High capacity
Servers	16 x Cisco UCS C220 M5SN Rack Servers with Small- Form-Factor (SFF) drives (UCSC-C220-M5SN)	8 x Cisco UCS C4200 Series Rack Servers with 4 x Cisco UCS C125 M5 Rack Server Nodes	16 x Cisco UCS C240 M5 Rack Servers with Small-Form-Factor (SFF) drives	16 x Cisco UCS C240 M5 Rack Servers with Large-Form-Factor (LFF) drives	8 x Cisco UCS S3260 Storage Servers with two S3260 M5 server nodes
CPU	2 x 2 nd Gen Intel® Xeon® Scalable Processor 6230R (2 x 26 cores, at 2.1 GHz)	1 x AMD 7352 Processor (24 cores, at 2.3 GHz)	2 x 2 nd Gen Intel® Xeon® Scalable Processor 5218R (2 x 20 cores, at 2.1 GHz)	2 x 2 nd Gen Intel Xeon Scalable Processor 4210R (2 x 10 cores, at 2.4 GHz)	2 x 2 nd Gen Intel Xeon Scalable Processor 6230R (2 x 26 cores, 2.1 GHz)
Memory	12 x 32-GB DDR4 (384 GB)	16 x 32-GB DDR4 (512 GB)	12 x 32-GB DDR4 (384 GB)	12 x 32-GB DDR4 (384 GB)	12 x 32-GB 2666 MHz (384 GB)
Boot	Cisco Boot-Optimized M.2 RAID Controller with 2 x 240-GB SSDs	M.2 with 2 x 240-GB SATA SSDs	Cisco Boot-Optimized M.2 RAID Controller with 2 x 240-GB SSDs	Cisco Boot-Optimized M.2 RAID Controller with 2 x 240-GB SSDs	2 x 240-GB SATA Boot SSDs
Storage	10 x 8-TB 2.5-in U.2 Intel P4510 NVMe high-performance value endurance	6 x 7.6-TB enterprise-value SATA SSDs	26 x 2.4-TB 10K RPM SFF SAS HDDs or 12 x 1.6-TB enterprise-value SATA SSDs	12 x 8-TB 7.2K RPM LFF SAS HDDs	14 x 8-TB 7.2K RPM LFF SAS HDDs
Virtual Interface Card (VIC)	25 Gigabit Ethernet (Cisco UCS VIC 1457) or 40/100 Gigabit Ethernet (Cisco UCS VIC 1497)	25 Gigabit Ethernet (Cisco UCS VIC 1455)	25 Gigabit Ethernet (Cisco UCS VIC 1457) or 40/100 Gigabit Ethernet (Cisco UCS VIC 1497)	25 Gigabit Ethernet (Cisco UCS VIC 1457) or 40/100 Gigabit Ethernet (Cisco UCS VIC 1497)	25 Gigabit Ethernet (Cisco UCS VIC 1455) or 40/100 Gigabit Ethernet (Cisco UCS VIC 1495)
Storage controller	NVMe switch included in the optimized server	Cisco 12-Gbps SAS 9460-8i RAID controller with 2-GB FBWC	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)	Cisco 12-Gbps SAS modular RAID controller with 2-GB FBWC or Cisco 12-Gbps modular SAS Host Bus Adapter (HBA)	Cisco UCS S3260 dual RAID controller
Network connectivity	Cisco UCS 6332 Fabric Interconnect or Cisco UCS 6454/64108 Fabric Interconnect	Cisco UCS 6454/64108 Fabric Interconnect	Cisco UCS 6332 Fabric Interconnect or Cisco UCS 6454/64108 Fabric Interconnect	Cisco UCS 6332 Fabric Interconnect or Cisco UCS 6454/64108 Fabric Interconnect	Cisco UCS 6332 Fabric Interconnect or Cisco UCS 6454/64108 Fabric Interconnect
GPU (optional)	Up to 2 x NVIDIA Tesla T4 with 16 GB of memory each		Up to 2 x NVIDIA Tesla V100 with 32 GB of memory each or up to 6 x NVIDIA Tesla T4 with 16 GB of memory each	2 x NVIDIA Tesla V100 with 32 GB memory of each or up to 6 x NVIDIA Tesla T4 with 16 GB of memory each	

Note: The Cisco UCS C240 M5 SFF Hybrid with NVMe reference architecture is described in the UCS C240 M5 section.

Table 2. Cisco UCS Integrated Infrastructure for Big Data and Analytics configuration options for high-density CPU cores and GPU nodes

	Select stack	Elite stack	Premier stack
Servers	8 x Cisco UCS C240 M5 Rack Servers 4 x Cisco UCS C480 M5 Rack Servers	8 x Cisco UCS C240 M5 Rack Servers 4 x Cisco UCS C480 ML M5 Rack Servers	8 x Cisco UCS C4200 Rack Server Chassis, each with 4 x Cisco UCS C125 M5 Rack Server Nodes
CPU	2 x 2 nd Gen Intel Xeon Scalable Processor 6230R (2 x 26 cores at 2.1 GHz)	2 x 2 nd Gen Intel Xeon Scalable Processor 6230R (2 x 26 cores at 2.1 GHz)	2 x AMD 7552 processor (2 x 48 cores at 2.2 GHz)
Memory	12 x 32-GB DDR4 (384 GB)	12 x 32-GB DDR4 (384 GB)	16 x 32-GB DDR4 (512 GB)
Boot	M.2 with 2 x 960-GB SSDs	M.2 with 2 x 960-GB SSDs	M.2 with 2 x 240-GB SATA SSDs
Storage	24 x 1.8-TB 10K rpm SFF SAS HDDs or 12 x 1.6-TB enterprise-value SATA SSDs	24 x 1.8-TB 10K rpm SFF SAS HDDs or 12 x 1.6-TB enterprise-value SATA SSDs	6 x 3.8-TB enterprise-value SATA SSDs
Virtual Interface Card (VIC)	25 Gigabit Ethernet (Cisco UCS VIC 1457) or 40/100 Gigabit Ethernet (Cisco UCS VIC 1497)	25 Gigabit Ethernet (Cisco UCS VIC 1457) or 40/100 Gigabit Ethernet (Cisco UCS VIC 1497)	25 Gigabit Ethernet (Cisco UCS VIC 1455)
Storage controller	Cisco 12-Gbps SAS modular RAID controller with 4-GB FBWC or Cisco 12-Gbps modular SAS HBA	Cisco 12-Gbps SAS modular RAID controller with 4-GB FBWC or Cisco 12-Gbps modular SAS HBA	Cisco 12-Gbps SAS 9460-8i RAID controller with 2-GB FBWC
Network connectivity	Cisco UCS 6332 Fabric Interconnect or Cisco UCS 6454/64108 Fabric Interconnect	Cisco UCS 6332 Fabric Interconnect or Cisco UCS 6454/64108 Fabric Interconnect	Cisco UCS 6454/64108 Fabric Interconnect
GPU	For C240 M5: 2 x NVIDIA TESLA V100 with 32 GB of memory each or up to 6 x NVIDIA T4 For C480 M5: 4 x NVIDIA TESLA v100 with 32 GB of memory each or 4 x NVIDIA T4	For C240 M5: 2 x NVIDIA TESLA V100 with 32 GB of memory each or up to 6 x NVIDIA T4 For C480 M5 ML: 8 x NVIDIA TESLA V100 with 32 GB of memory each and with NVLink	

Table 3. Cisco UCS Integrated Infrastructure for Big Data and Analytics configuration options for object storage

Servers	Cisco UCS S3260 with single node	Cisco UCS S3260 with dual node
CPU	2 x 2 nd Gen Intel Xeon Scalable Processor 6230R (2 x 26 cores, 2.1 GHz)	2 x 2 nd Gen Intel Xeon Scalable Processor 6230R (2 x 26 cores, 2.1 GHz)
Memory	12 x 32GB 2666 MHz (384 GB)	6 x 32GB 2666 MHz (192 GB) per node
Boot	2 x 1.6TB SATA Boot SSDs	4 x 1.6TB SATA Boot SSDs
Storage	UCS S3260, 4 rows of drives – 56 x 14 TB per node	UCS S3260, 2 rows of drives – 28 x 14 TB per node
Virtual Interface Card (VIC)	25 Gigabit Ethernet (Cisco UCS VIC 1455) or 40/100 Gigabit Ethernet (Cisco UCS VIC 1495)	25 Gigabit Ethernet (Cisco UCS VIC 1455) or 40/100 Gigabit Ethernet (Cisco UCS VIC 1495)
Storage controller	Cisco UCS S3260 dual RAID controller	Cisco UCS S3260 dual RAID controller
Network connectivity	Cisco UCS 6332 Fabric Interconnect or Cisco UCS 6454/64108 Fabric Interconnect	Cisco UCS 6332 Fabric Interconnect or Cisco UCS 6454/64108 Fabric Interconnect
Cache	2x UCS S3260 M5 SIOC 2TB NVMe	1x UCS S3260 M5 SIOC 2TB NVMe/node

Future-proofing Advanced analytics deployment with CDIP

As enterprises embark on the journey of digital transformation, an integrated, extensible infrastructure implementation purpose-built to keep pace with the constant challenges of technological advancement for each workload can reduce bottlenecks, improve performance, decrease bandwidth constraints, and minimize business disruption.

Cisco UCS C240 M5 Rack Server

The Cisco UCS C240 M5 Rack Server is a dual-socket, 2-rack-unit (2RU) server with the latest second-generation Intel Xeon and Intel Xeon Scalable processors. It features 24 DIMM slots for DDR4 DIMMs, one dedicated internal slot for a 12-Gbps SAS storage controller card, and up to 26 internal SFF drives or up to 12 front-facing internal LFF drives. The C240 M5 server offers industry-leading performance and expandability for a wide range of storage and I/O-intensive infrastructure workloads, such as big data, analytics, and collaboration. In addition, the server has two modular M.2 cards that can be configured for boot. A modular LAN-on-motherboard (mLOM) slot supports dual 40/100 Gigabit Ethernet network connectivity with the Cisco UCS VIC 1497 or quad 25 Gigabit Ethernet connectivity with the Cisco VIC 1457.

A 6-PCIe 3.0 slot supports various PCIe card options based on the workload requirement. The Cisco UCS C240 M5 supports up to six NVIDIA Tesla T4 or up to two NVIDIA Tesla V100 GPU cards. Hadoop 3.0 and Spark 3.0 add native-support GPU to accelerate AI/ML, ETL, and other workloads.

Figure 4 shows the available ports and slots on the C240 M5. Figure 5 shows the rack server's reference architecture.

Figure 4. Cisco UCS C240 M5 Rack Server

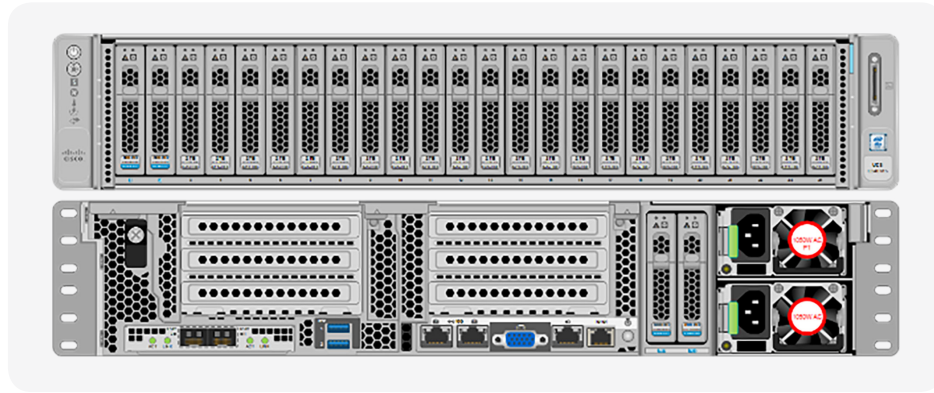
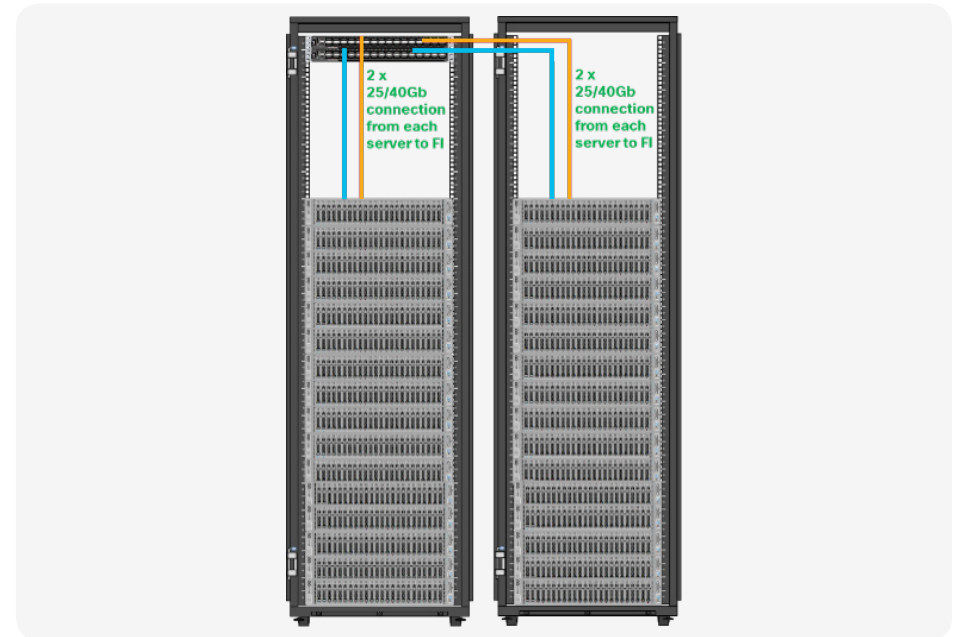


Table 4. Cisco UCS C240 M5 storage configuration

Configuration	Storage media	Per node capacity
All HDD	26 x 2.4-TB 12G SAS 10K RPM SFF HDD	62.4 TB
Hybrid - HDD + NVMe	24 x 2.4-TB 12G SAS 10K RPM SFF HDD 2 x 8TB high-performance value endurance NVMe	73.6 TB
All Flash	16 x 3.8-TB enterprise-value SATA SSD Or 12 x 7.6-TB enterprise-value SATA SS	60.8 TB or 91.2 TB
NVMe	10 x 8-TB high-performance value endurance NVMe	80 TB

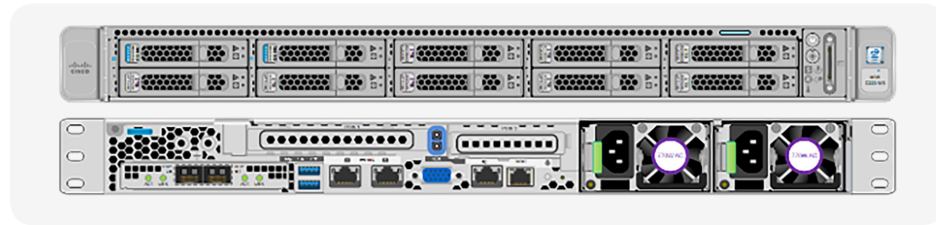
Figure 5. Cisco UCS C240 M5 reference architecture



Cisco UCS C220 M5 Rack Server

The Cisco UCS C220 M5 Rack Server is a dual-socket, 1-rack-unit (1RU) server with the latest second-generation Intel Xeon and Intel Xeon Scalable processors. It features 24 DIMM slots for DDR4 DIMMs, one dedicated internal slot for a 12-Gbps SAS storage controller card, and up to 10 internal SFF drives or up to 4 front-facing internal LFF drives. The C220 M5 offers industry-leading performance and expandability for a wide range of storage and I/O-intensive infrastructure workloads, such as big data, analytics, and collaboration. In addition, the server has two modular M.2 cards that can be configured for boot. A modular LAN-on-motherboard (mLOM) slot supports dual 40/100 Gigabit Ethernet network connectivity with the Cisco UCS VIC 1497 or quad 25 Gigabit Ethernet connectivity with the Cisco VIC 1457. Figure 6 shows the available ports and slots on the C220 M5. Figure 7 shows the rack server's reference architecture.

Figure 6. Cisco UCS C220 M5 Rack Server



The Cisco UCS C220 M5 all-NVMe—with a system that can support up to 10 SFF 8-TB NVMe PCIe SSDs—provides low latency, high performance, cost-effective storage with less power, cooling, and a data center footprint.

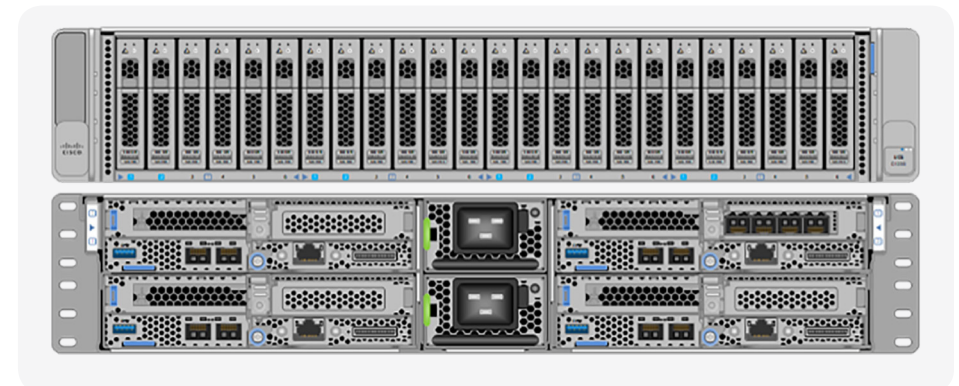
Figure 7. Cisco UCS C220 M5 reference architecture



Cisco UCS C4200 Series Rack Server Chassis with C125 M5 Server Node

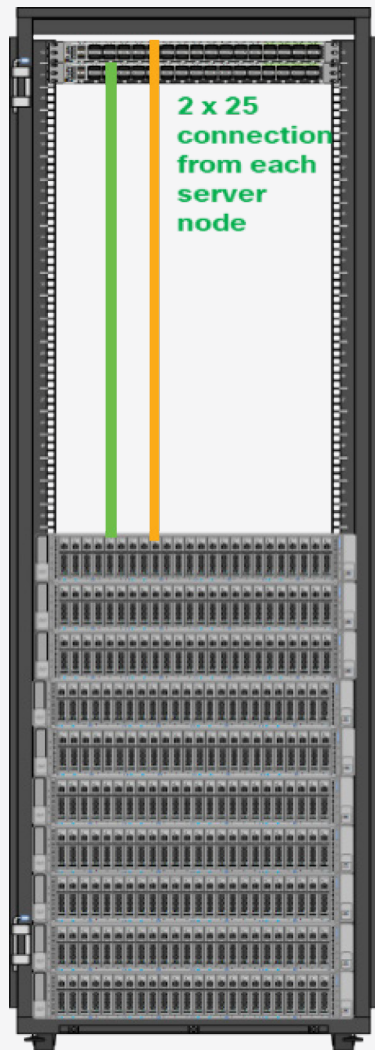
The Cisco UCS C4200 Series Rack Server Chassis is Cisco's densest computing solution, with up to four Cisco UCS C125 M5 Rack Server Nodes in 2RU of rack space. This density makes it well suited for the network edge or in the data center for scale-out applications. The C125 M5 node contains AMD EPYC processors, up to 2 TB of memory, and up to six SAS/SATA drives or four plus two Non-Volatile Memory Express (NVMe) drives. Additional Secure Digital (SD) or M.2 storage modules can be used as boot devices or additional storage. Fourth-generation VICs and an OCP 2.0 mezzanine slot offer exceptional levels of performance, flexibility, and I/O throughput to run your applications. Figure 8 shows the nodes on the C4200 chassis with C125 M5 server nodes. Figure 9 shows the reference architecture.

Figure 8. Cisco UCS C4200 Rack Server Chassis with four C125 M5 Server Node



The Cisco UCS C4200 Series Rack Server Chassis with Cisco UCS C125 M5 Rack Server nodes is modular rack server, 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.

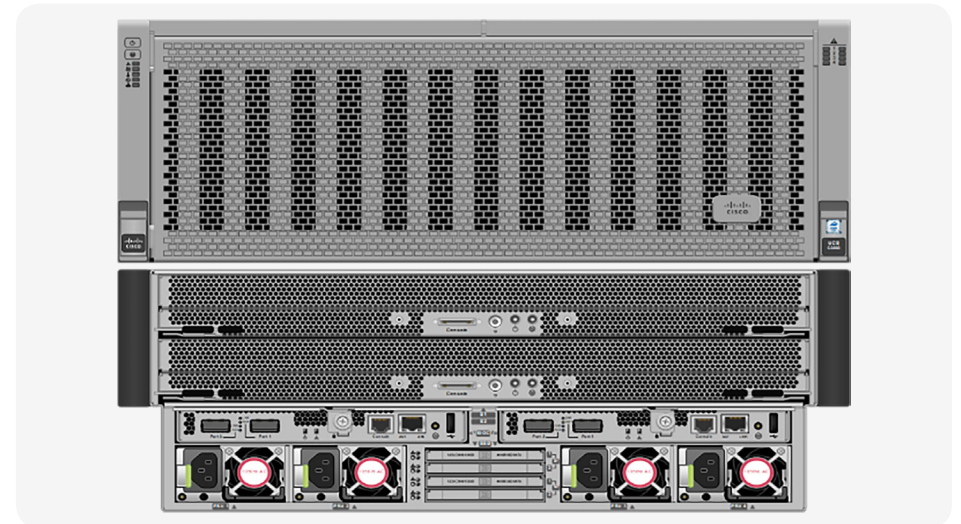
Figure 9. Cisco UCS C4200 Rack Server Chassis with C125 M5 Server Node reference architecture



Cisco UCS S3260 Storage Server

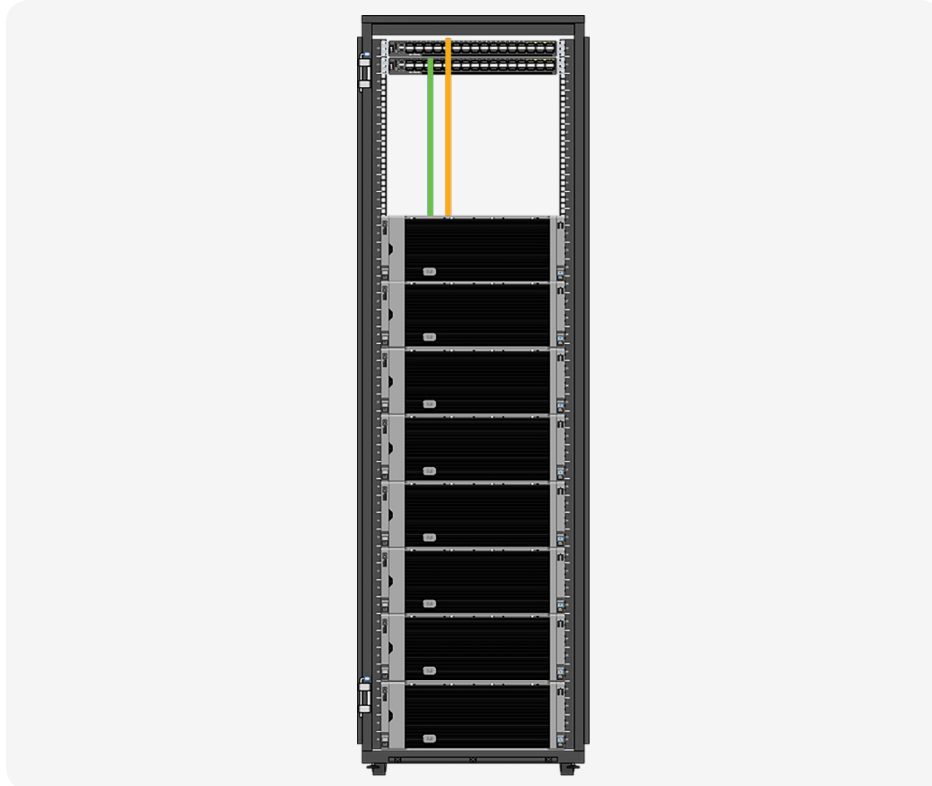
The Cisco UCS S3260 Storage Server is a modular storage server with dual server nodes. It is optimized for large data sets used in scenarios such as big data, cloud, object storage, video surveillance, and content delivery environments. The S3260 server helps achieve the highest levels of data availability and performance. With a dual-node capability that is based on the latest second-generation Intel Xeon and Intel Xeon Scalable processors, it offers up to 840 TB of local storage in a compact 4RU form factor. Network connectivity is provided with dual-port 40-Gbps nodes in each server. Figure 10 is an image of the back of the S3260 server. Figure 11 shows its reference architecture.

Figure 10. Cisco UCS S3260 Storage Server



The Cisco UCS S3260 Storage Server with S3260 M5 server node provides a high-capacity and lower cost of storage per TB. Software-Defined Storage (SDS) with the Cisco UCS S3260 brings together the simplicity and agility of the cloud and the cost benefits of industry-standard servers. It offers an excellent S3-compatible object-storage platform that is highly scalable and optimized for capacity and I/O performance. Data can be offloaded from high-performance HDFS containing hot data onto the object store, or second-tier storage configured on the S3260 storage server.

Figure 11. Cisco UCS S3260 Storage Server with S3260 M5 server node reference architecture



Cisco UCS 6300 and 6400 Series Fabric Interconnects

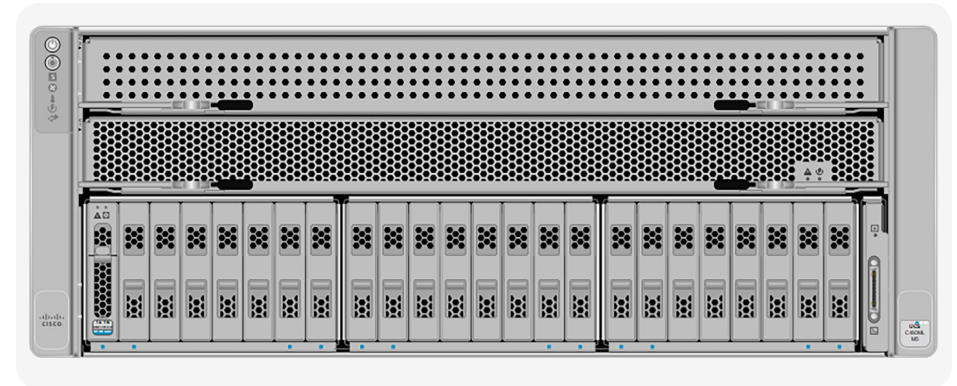
Cisco UCS 6300 and 6400 Series Fabric Interconnects are a core part of the Cisco UCS portfolio, providing both network connectivity and management capabilities for the UCS system. The Cisco UCS 6300 Series offers line-rate, low-latency, lossless 40-Gigabit Ethernet, Fibre Channel over Ethernet (FCoE), and Fibre Channel functions, as well as unified ports capable of either Ethernet or Fibre Channel operation. The Cisco UCS 6400 series Fabric Interconnect offers line-rate, low-latency, lossless 10, 25, 40, and 100 Gigabit Ethernet, FCoE, and Fibre Channel functions.

The Cisco UCS 6300 and 6400 Series Fabric Interconnects provide the management and communication backbone for the Cisco UCS B-Series Blade Servers and C-Series Rack Servers. All servers attached to the Cisco UCS fabric interconnects become part of a single, highly available management domain.

Cisco UCS C480 ML M5 Rack Server

The Cisco UCS C480 ML M5 Rack Server (Figure 12), is a purpose-built 4RU server for deep-learning environments. It supports the latest second-generation Intel Xeon and Intel Xeon Scalable processors and 8 NVIDIA Tesla V100 32-GB Tensor Core GPUs with NVLink interconnects. It supports up to 3 TB of DDR4 memory in 24 slots, up to 24 SFF hot-swappable SAS/SATA SSDs and HDDs, up to 6 PCIe NVMe disk drives, and up to 2 internal M.2 drives.

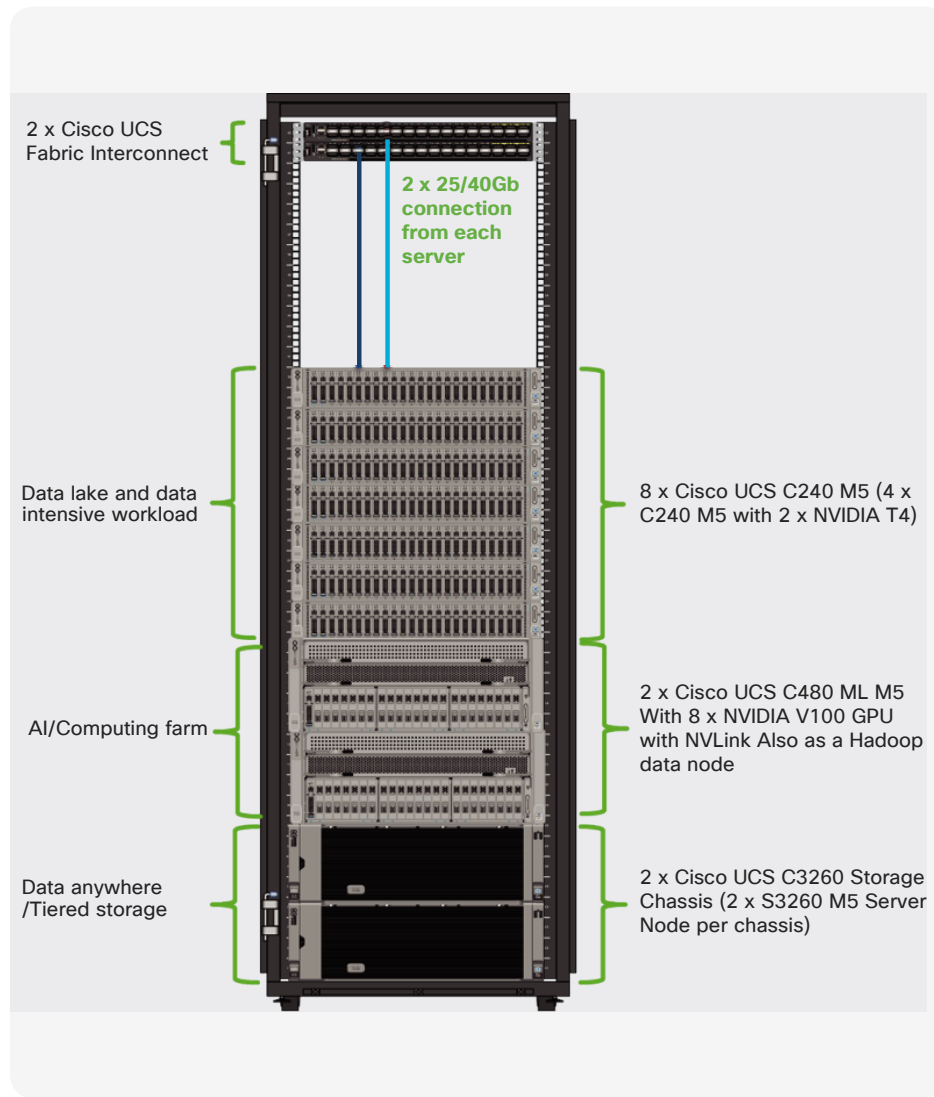
Figure 12. Cisco UCS C480 ML M5 purpose-built deep-learning server



Scaling the architecture

The Cisco Data Intelligence Platform can be deployed in a single rack and can be scaled to thousands of nodes. Figure 13 shows the reference architecture for a single rack. This design can be scaled to support existing or future workload demand independently.

Figure 13. Cisco Data Intelligence Platform deployed in a single rack

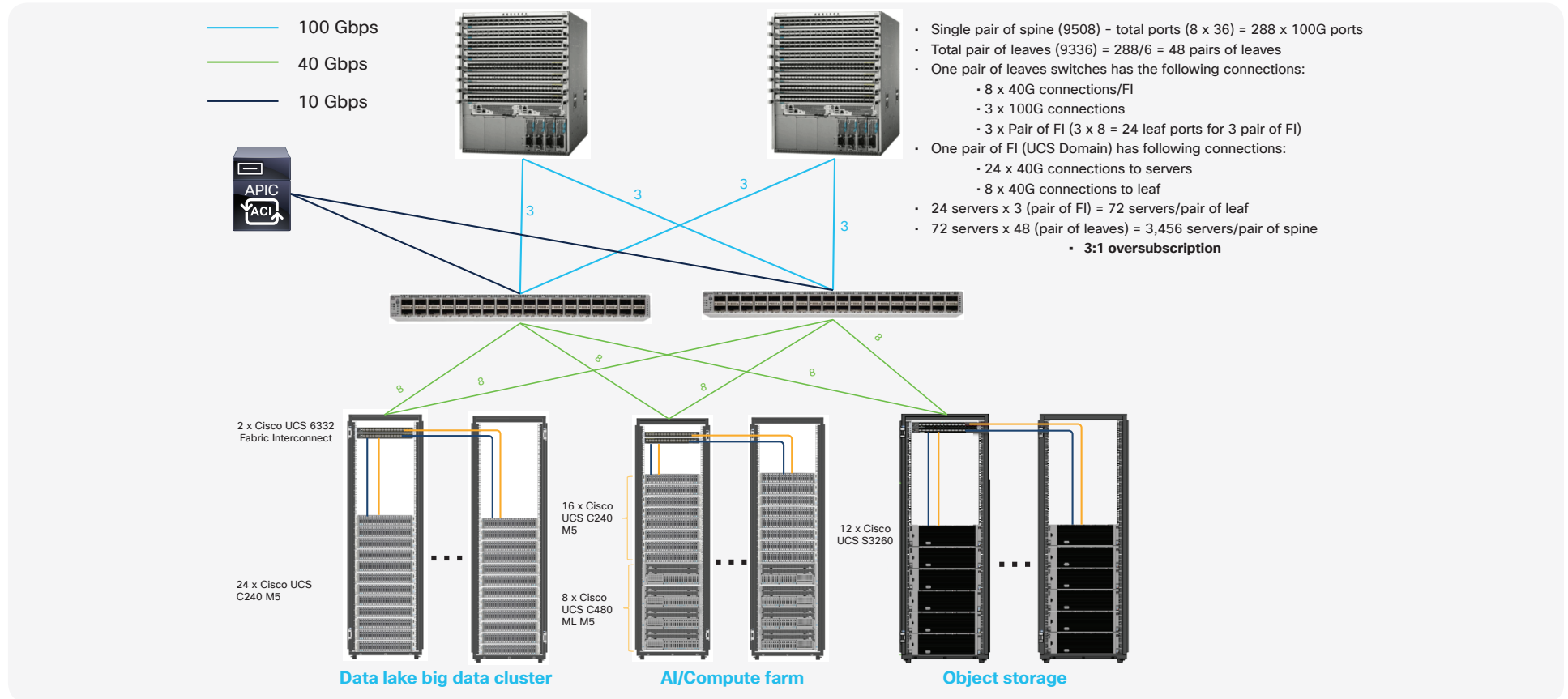


Scaled architecture with 3:1 oversubscription with Cisco UCS Fabric Interconnects and Cisco ACI

The architecture discussed here and shown in Figure 14 supports 3:1 network oversubscription from every node to every other node across a multidomain cluster (nodes in a single domain within a pair of Cisco fabric interconnects are locally switched and not oversubscribed). The Cisco Nexus 9508 Switch with the Cisco N9K-X9736C-FX line card can support up to 36 x 100-Gbps ports each and 8 N9K-X9736C-FX line cards.

From the viewpoint of the data lake, 24 Cisco UCS C240 M5 Rack Servers are connected to a pair of Cisco UCS 6332 Fabric Interconnects (with 32 x 40-Gbps throughput). From each fabric interconnect, 8 x 40-Gbps links connect to a pair of Cisco Nexus 9336 Switches. Three pairs of fabric interconnects can connect to a single pair of Cisco Nexus 9336 Switches (8 x 2 40-Gbps links). Each of these Cisco Nexus 9336 Switches connects to a pair of Cisco Nexus 9508 switches with 6 x 100-Gbps uplinks (connecting to a Cisco N9K-X9736C-FX line card). Figure 14 provides a layout of the infrastructure.

Figure 14. Scaled architecture with 3:1 oversubscription with Cisco fabric interconnects and Cisco ACI



Scaled architecture with 2:1 oversubscription with Cisco ACI

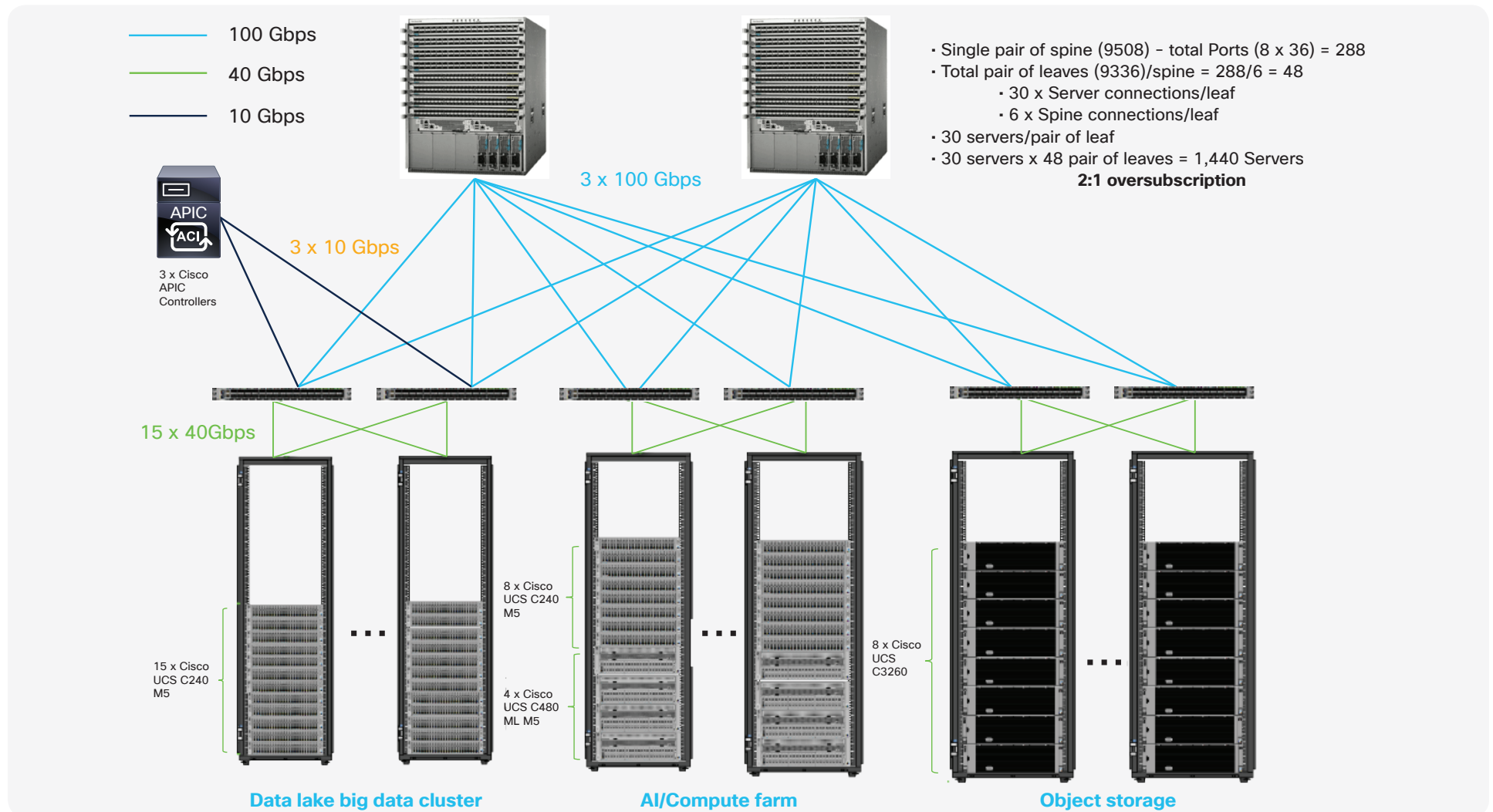
In the scenario discussed here and shown in Figure 15, the Cisco Nexus 9508 Switch with the Cisco N9K-X9736C-FX line card can support up to 36 x 100-Gbps ports each and 8 N9K-X9736C-FX line cards.

For the 2:1 oversubscription, 30 Cisco UCS C240 M5 Rack Servers are connected to a pair of Cisco Nexus 9336 Switches and each Cisco Nexus 9336 connects to a pair of Cisco Nexus 9508 Switches with three uplinks each. A pair of Cisco Nexus 9336 Switches can support 30 servers and connect to a spine with 6 x 100-Gbps links on each spine. This single pod (a pair of Cisco Nexus 9336 Switches connecting to 30 Cisco UCS C240 M5 servers and 6 uplinks to each spine) can be repeated 48 times (288/6) for a given Cisco Nexus 9508 Switch and can support up to 1440 servers.

To reduce the oversubscription ratio (to get 1:1 network subscription from any node to any node), you can use just 15 servers under a pair of Cisco Nexus 9336 Switches and then move to Cisco Nexus 9516 Switches (the number of leaf nodes would double).

To scale beyond this number, multiple spines can be aggregated. Figure 15 provides a layout of the scenario.

Figure 15. Scaled architecture with 2:1 oversubscription with Cisco ACI



For more information

- To find out more about Cisco UCS big data solutions, visit <https://www.cisco.com/go/bigdata>
- To find out more about Cisco UCS big data validated designs, visit <https://www.cisco.com/c/en/us/solutions/design-zone/data-center-design-guides/data-center-big-data.html>
- To find out more about the Cisco Data Intelligence Platform, visit <https://www.cisco.com/c/dam/en/us/products/servers-unified-computing/ucs-c-series-rack-servers/solution-overview-c22-742432.pdf>
- To find out more about Cisco UCS AI/ML solutions, visit <https://www.cisco.com/c/en/us/solutions/data-center/artificial-intelligence-machine-learning/index.html>
- To find out more about Cisco ACI solutions, visit <https://www.cisco.com/c/en/us/solutions/data-center-virtualization/application-centric-infrastructure/index.html>
- To find out more about Cisco validated solutions based on software-defined storage (SDS), visit <https://www.cisco.com/c/en/us/solutions/data-center-virtualization/software-defined-storage-solutions/index.html>

Conclusion

Evolving workloads need a highly flexible platform to cater to various requirements, whether data-intensive (data lake) or compute-intensive (AI/ML/DL) or just storage-dense (object store). An infrastructure to enable this evolving architecture—one that is able to scale to thousands of nodes—requires strong attention to operational efficiency.

To bring in seamless operation of the application at this scale, one needs:

- An infrastructure automation with centralized management
- Deep telemetry and simplified granular troubleshooting capabilities
- Multi-tenancy for application workloads, including containers and micro-services, with the right level of security and SLA for each workload

Cisco UCS with Cisco Intersight and Cisco ACI can enable this next-generation cloud-scale architecture, deployed and managed with ease.