

Cisco HyperFlex Systems

Hyperconverged multicloud platform

You need infrastructure that can adapt to match the speed of your business— whether deploying enterprise applications or containers in multicloud environments. Cisco HyperFlex™ systems with Intel® Xeon® Scalable processors deliver hyperconvergence with the power and simplicity for any application, on any cloud, and at any scale. Engineered on the Cisco Unified Computing System™ (Cisco UCS®), Cisco HyperFlex systems deliver the agility, scalability, and pay-as-you-grow economics of the cloud with the benefits of on-premises infrastructure.

- **Ready for any application:** The latest release of Cisco HyperFlex software supports virtualized and containerized deployment with multiple hypervisors, and has been tested and validated for numerous enterprise applications.
- **Ready for any cloud:** The platform includes tools for application performance monitoring, application placement, and cloud mobility so that you can design how to deploy your applications and place them wherever your business needs dictate.
- **Ready for any scale:** This release features greater scalability with larger clusters, and logical availability zones to mitigate the impact of node and component failures. Stretch clusters span geographies for continuous deployment, and cloud-based deployment allows you to ship equipment anywhere and install from the cloud.

Cisco HyperFlex™
Systems with Intel®
Xeon® Scalable
processors



Contents

The challenge: a changing landscape	3
The solution: Cisco HyperFlex Systems	3
Complete end-to-end solution	4
Ready for your enterprise applications	5
Engineered on Cisco UCS technology	5
Your choice of management tools	6
Powered by next-generation data technology	7
Agile, efficient, adaptable, and scalable solution	7
Solution architecture	8
Cluster composition with Cisco Hyperflex HX-Series nodes .	9
Complete hyperconvergence with Cisco networking	9
Integration with Cisco Application Centric Infrastructure	10
Cisco HyperFlex HX Data Platform	10
Cisco HyperFlex HX Data Platform controller	11
Data distribution	12
Logical availability zones	13
Data read and write operations	13
Enterprise-class storage features	14
Engineered on Cisco UCS technology	15
Cisco HyperFlex HX-Series Nodes	16
Hyperconvergence at the network edge	16
Scaling with Cisco UCS servers	16
Connect to external shared storage	17
Powered by Intel Xeon scalable processors	17
Cisco UCS management	18
Conclusion	18
Part of our broad data center strategy	18
Cisco HyperFlex services	19

The challenge: a changing landscape

The digital business transformation is powered by applications. Today you must support an even wider set of application deployment models:

- **Traditional enterprise applications** are maintaining some monolithic components but are transitioning to a more scalable, cloud-like model for components such as presentation layers.
- **Big data and analytic applications** are now commonplace, and they need scale-out architectures with large amounts of high-performance storage.
- **Cloud-native architectures** based on microservices demand both virtualized and container-based deployment. Your developers are creating these applications so that they can run across multiple clouds with automatic deployment and scaling.

Reducing complexity is the only way to support both existing and new applications, ensure service delivery, maintain control over data, and attain necessary performance. This need is fueling demand for hyperconverged solutions because it is one of the best ways to deliver the wide range of support that IT organizations must provide.

But you can't solve these challenges with just any hyperconverged infrastructure. You need solutions that are:

- **Ready for any application.** You need support for virtualized and containerized software, and support for multiple hypervisors. Validation for running your existing enterprise applications is a must.
- **Ready for any cloud.** Today's developers are already creating microservices that deploy literally anywhere. You need tools that help you deploy all of your applications in the cloud that makes most business sense.
- **Ready for any scale.** You need to host the range of applications above, with scale beyond what most hyperconverged products support. But you need scale combined with the intelligence to support continuous operations even with multiple node or component failures.

The solution: Cisco HyperFlex Systems

To help you meet the multicloud challenge, we introduce Cisco HyperFlex Systems™, an adaptive system to power any application with the simplicity of hyperconvergence. Our solution delivers flexibility to support multiple hypervisors, containerized environments, and multicloud services to efficiently and economically deploy, monitor, and manage hybrid applications.

Powered by Intel® Xeon® Scalable processors, this is a new generation of more flexible, more scalable, enterprise-class hyperconverged solutions. We deliver a complete solution based on a next-generation data platform—one that smoothly integrates into the data center you have today. In contrast to first-generation products, our solution includes an integrated network fabric

and powerful data optimization capabilities that unlock the full potential of hyperconvergence for a wider range of workloads and use cases.

Our solution is faster to deploy, simpler to manage, easier to scale, and ready to provide a unified pool of resources to power your business applications. You harness these resources with your choice of centralized management tools: Cisco Intersight™ cloud-based management as a service; Cisco HyperFlex Connect; Microsoft System Center Virtual Machine Manager, Hyper-V Manager, or a VMware vSphere plug-in. Cisco HyperFlex Systems integrate into the data center you have today without creating an island. You can deploy our solution wherever you need it, from central data center environments to remote locations and edge-computing environments (Figure 1).

Complete end-to-end solution

We offer the first hyperconverged platform that is designed as an end-to-end software-defined infrastructure that eliminates the compromises found in first-generation products. We designed Cisco HyperFlex Systems to support a broader range of applications and workloads in the data center, private and hybrid clouds, remote locations, and edge-computing environments. This new generation extends the ease of hyperconverged system deployment,

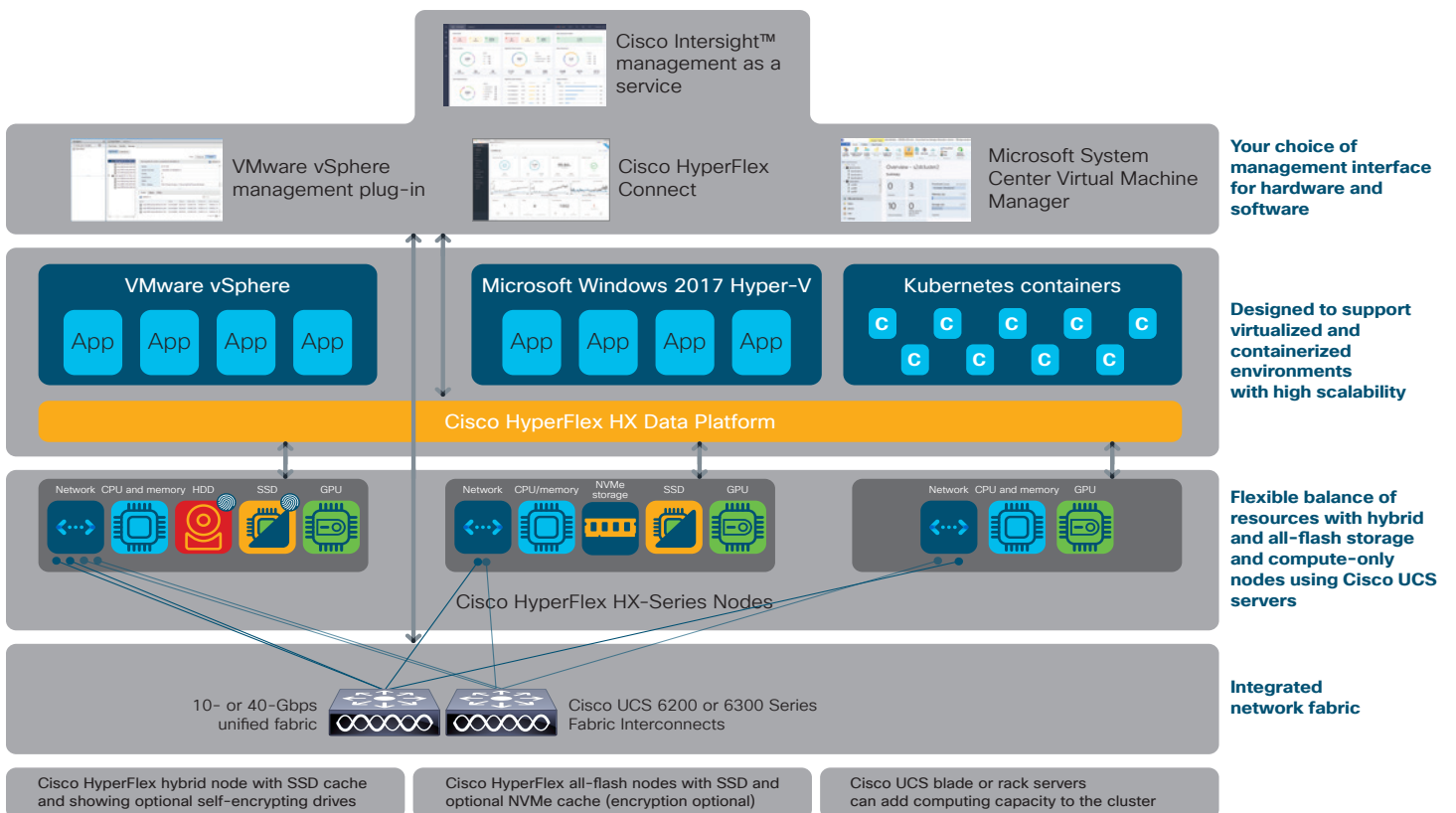


Figure 1 Cisco HyperFlex Systems support virtualized and containerized applications and let you choose the exact combination of resources to power your enterprise applications

What's new?

Here are the enhancements we have incorporated into our 3rd-generation product:

- **Multihypervisor support.** We now support Microsoft Windows Server 2016 Hyper-V including failover clustering and backup software integration.
- **Containerized application support.** A prepackaged Kubernetes platform enables containers-as-a-service for multicloud deployment. A FlexVolume driver enables persistent data.
- **Enterprise application validation.** Our engineering team has developed Cisco® Validated Designs that demonstrate our readiness for Oracle Database, Microsoft SQL Server, Microsoft Exchange Server, and both Citrix and VMware virtual desktop environments.
- **Flexible multicloud services.** These features, discussed in our multicloud solution brief, support deployment to any cloud with application performance monitoring, application placement, and cloud mobility services. Cisco Intersight integration. Configure, deploy, manage, and monitor global operations from a single HTML 5 interface.

management, and support beyond your central data center to multicloud environments and to the network edge.

Cisco HyperFlex Systems combine:

- **Software-defined computing** in the form of Cisco Unified Computing System™ (Cisco UCS®) servers with Intel Xeon Scalable processors
- **Software-defined storage** with the powerful Cisco HyperFlex HX Data Platform software
- **Software-defined networking** with Cisco UCS fabric that integrates smoothly with Cisco® Application Centric Infrastructure (Cisco ACI™)

Together, these elements comprise an adaptive infrastructure that lets you integrate easily with your existing infrastructure. The result is a cluster that comes up and configures itself in an hour or less and that scales resources independently to closely match your application resource needs.

Ready for your enterprise applications

The power and performance of Cisco HyperFlex M5 nodes brings lower latency and readiness for a wider range of applications than ever before. We have invested in developing Cisco Validated Designs that have tested and validated many traditional enterprise applications in the Cisco HyperFlex environment to help you speed deployment and reduce risk. And with flexible configurations, you can deploy Cisco HyperFlex Systems in your enterprise data center, to create private clouds, or for edge computing.

Now you can use hyperconvergence for enterprise databases such as Oracle database and Microsoft SQL Server. The system's high storage capacity makes it ready for big data applications such as Splunk, and for in-memory databases such as SAP HANA. The capability to add graphics processor acceleration makes Cisco HyperFlex Systems ready for high-performance computing (HPC) environments, and it helps provide a smooth user experience in virtual desktop infrastructure (VDI) environments. Container support allows you to deploy containers in your own private cloud and multicloud environments so that you can write once and deploy anywhere. Of course our solution is ready to serve all of your VMware vSphere and Microsoft Windows Server 2016 Hyper-V virtualized environments and Internet infrastructure services such as Microsoft Exchange.

Engineered on Cisco UCS technology

Cisco UCS provides a single point of connectivity and hardware management that integrates Cisco HyperFlex nodes into a single unified cluster. The system is built to be self-aware and self-integrating so that when the system's fabric interconnects notice a new component attached, that component is automatically incorporated into the cluster. Rather than requiring you to configure each element in the system manually through a variety of element managers, Cisco UCS was built so that every aspect

What's new, continued

- **Increased scale and resiliency.** To support any application, the platform has been expanded to support up to 64 nodes in a cluster with added resiliency in the fully-automated logical availability zones. Individual nodes can be configured with higher capacity through large-form-factor drives.
- **Stretch clusters.** Active-active clusters stretched across geographic distance with integrated failover.
- **Cloud-based deployment.** Cisco Intersight™ management as a service supports deploying Cisco HyperFlex systems from the cloud, enabling you to install your hardware in any datacenter and install and configure remotely.

of server personality, configuration, and connectivity can be set through software.

Your choice of management tools

A high level of automation is controlled through the Cisco UCS management API by higher-level tools that allow you to configure a cluster in minutes, with no risk of configuration creep or noncompliant settings. You can choose the approach that best suit your needs.

- **Cisco HyperFlex Connect.** This intuitive, data center-hosted HTML 5-based management tool is device independent, giving you access to all cluster features through any device, anywhere. All of the cluster data platform features (such as logical availability zones and native replication) can be managed through this interface.
- **Cisco Intersight management as a service.** This cloud-based interface gives you instant access to all of your clusters regardless of where they are deployed. High-level resource inventory and status is provided by Cisco Intersight dashboards. A recommendation engine helps you proactively respond to impending issues such as the need to scale capacity. Install Edge deployments remotely simply by claiming your nodes in the interface and the platform will configure a cluster remotely in minutes. Drill down into data platform operations and you have exactly the same control as Cisco HyperFlex Connect provides—except that you do not have to host management software and you always have the most up-to-date versions.
- **Hypervisor management integration.** Cisco HyperFlex Systems management integrates with both Microsoft Windows 2016 Hyper-V and VMware vCenter. This level of management integration enables you to manage the lifecycle of virtual machines—including storage management through data platform—without leaving the interface that your administrators are probably already familiar with. These management tools include Microsoft System Center Virtual Machine Manager (SCVMM), Microsoft Hyper-V Manager, and a VMware vCenter plug-in.
- **Container and multicloud deployment.** Our integration of Kubernetes container management and orchestration allows you to deploy Docker-based containerized applications in your local private cloud or the Google public cloud. With click-of-the-mouse simplicity, you can write your applications once and deploy anywhere.

Other tools make use of the management API to support your application lifecycle. Cisco UCS Director can implement end-to-end application lifecycle orchestration and automation. Cisco ONE™ Enterprise Cloud Suite can turn your cluster into an on-premises cloud service with smooth migration of workloads to and from public cloud facilities.

You can independently scale the computing power of your cluster by configuring Cisco UCS servers as computing-only nodes. This feature lets you adjust the balance of CPU and storage capacity, optimizing your total cost of ownership (TCO). Incremental scalability allows you to start small and

More agile

Cisco HyperFlex Systems are more agile because they perform, scale, and interoperate better:

- **Deployment is fast and easy.** Your cluster ships with the hypervisor and data platform preinstalled and ready to launch through the installation wizard.
- **Integrated networking brings high performance.** Your cluster is interconnected with low, consistent latency, and with 10- and 40-Gbps network bandwidth.
- **Scaling is fast and simple.** The system automatically discovers new hardware when it is installed. Then adding it to the cluster takes only a few mouse clicks. New logical availability zones allow clusters to grow to up to 64 nodes while minimizing the impact of multiple node failures.
- **Interoperability is straightforward.** Management capabilities enable you to install and operate your Cisco HyperFlex System in the data center you have today, with high-level management tools that support operations across your hyperconverged and your traditional infrastructure.

scale as your needs grow. You can accelerate virtual desktop environments with graphics processing units (GPUs) that integrate directly into nodes, speeding graphics-intensive workloads for a smooth user experience.

Powered by next-generation data technology

The HX Data Platform combines the cluster's storage devices into a single distributed, multitier, object-based data store. It makes this data available through the file system protocols and mechanisms needed by the higher-level hypervisors and container platforms.

Performance scales linearly as you scale the cluster because all components contribute processing and storage capacity to the cluster. The data platform optimizes storage tiers for an excellent balance between price and performance. For example, hybrid nodes use solid-state disk (SSD) drives for caching and hard-disk drives (HDDs) for capacity; all-flash nodes use fast SSD drives or Nonvolatile Memory Express (NVMe) storage for caching and SSDs for capacity.

A full set of enterprise-class data management features are built in: for example, snapshots, thin provisioning, synchronous and asynchronous replication, data encryption, integration with third-party backup tools, and instant and space-efficient cloning. Our networking secures your data through its lifecycle, with security and compliance controls for protection when you distribute, migrate, and replicate data across storage environments. New security features help you comply with industry and governmental standards. The platform delivers high availability through parallel data distribution and replication, accelerated by the low latency and high bandwidth of the Cisco unified fabric. Data is continuously optimized with real-time, always-on deduplication, compression, and optional encryption, helping reduce your storage costs without affecting performance. Dynamic data placement in server memory, caching, and capacity tiers increases application performance and reduces performance bottlenecks.

Agile, efficient, adaptable, and scalable solution

Bringing benefits to your IT organization and to your business, Cisco HyperFlex Systems overcome the limitations of first-generation products, delivering real solutions that are agile, efficient, and adaptable. Our solution is well suited for hosting environments such as virtual desktops, server virtualization deployments, and test and development environments.

- **More agile.** Cisco HyperFlex Systems are more agile because they perform, scale, and interoperate better (see "More agile" on page 7)
- **More efficient.** Our solution was designed from the beginning with a purpose-built, highly efficient data platform that combines the cluster's scale-out storage resources into a single, distributed, multitier, object-based data store. Features you expect of enterprise storage systems are built into Cisco HyperFlex Systems (see "More efficient" on page 9).

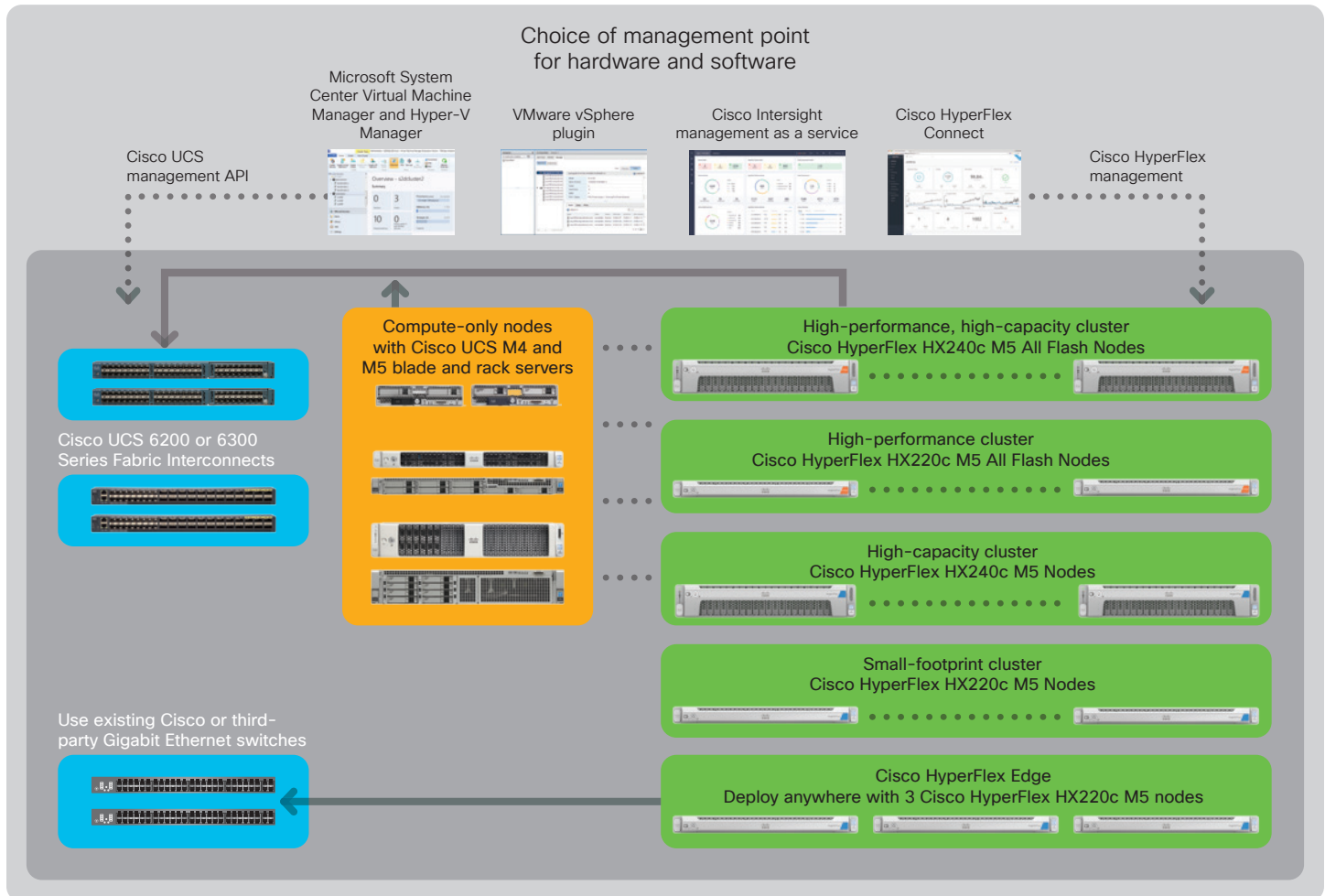


Figure 2 Cisco HyperFlex Systems product family

- **More adaptable.** Your business needs and your workloads are constantly changing. Your infrastructure needs to quickly adapt to support your workloads and your business (see “More adaptable” on page 10).
- **More scalable.** Clusters can scale to up to 64 nodes, with protection from multiple node and component failures (see “More scalable” on page 11).

Solution architecture

Cisco HyperFlex Systems combine Cisco UCS networking and computing technology, powerful Intel Xeon Scalable processors, and the HX Data Platform to deliver a complete, preintegrated solution. After you run the configuration wizard, your cluster is ready to work for you whether you need to support virtualized or containerized applications. You get a uniform pool of computing, networking, and storage resources that is designed to power your applications.

More efficient

Our solution is more efficient because of the following:

- **Buy only the storage you need.** Continuous inline data deduplication and compression, fast, space-efficient clones, and thin provisioning contribute to lowering the cost of your storage.
- **Data protection is built in.** You can use native HX Data Platform capabilities or use the same data protection solutions you use in your data center because Cisco HyperFlex Systems interoperate with leading backup tools.
- **Your data is secure.** You don't need to take extra steps to secure your data at rest when using self-encrypting drives. Cisco HyperFlex Systems help you save time and ensure compliance.
- **Save on storage administration.** You don't need to install a complex storage network or worry about logical unit numbers (LUNs). If you already have enterprise shared storage, you can connect it directly to your cluster to run virtual machines from it, migrate the storage to your cluster, or use the storage for backup operations.

Logically, a cluster of three or more Cisco HyperFlex HX-Series nodes that provide computing and storage resources with a full-featured data platform. The data platform is implemented using a data platform controller that runs on each node. This approach creates a uniform pool of resources that supports applications across the cluster. When you need to provision computing or storage capacity, that capacity is drawn from the entire pool.

Cluster composition with Cisco Hyperflex HX-Series nodes

Physically, a cluster is composed from a set of three or more HX-Series nodes. These are integrated to form a single system by a pair of Cisco UCS 6200 or 6300 Series Fabric Interconnects with 10- or 40-Gbps connectivity (Figure 2). Up to an equal number of Cisco UCS servers can act as computing-only nodes, participating in the data platform but contributing only computing power (and possibly graphics acceleration) to the cluster.

Cisco HyperFlex Edge solutions are deployed as fixed sets of three nodes that use existing Gigabit Ethernet switches, offering the utmost in flexibility for deployment in remote and branch-office environments (see “Designed for the network edge” on page 16).

A cluster can reside in a single location or it can be stretched across short geographic distances. An active-active “stretched” cluster synchronously replicates data between the two sites and has a very short recovery time objective (RTO) and zero data loss. Across longer geographic distances, native replication can synchronize data from a primary to a secondary site to support more traditional disaster recovery strategies.

Complete hyperconvergence with Cisco networking

Networking in most hyperconverged environments is an afterthought. We give you complete hyperconvergence with networking as an integral and essential part of the system. Using Cisco UCS fabric interconnects, you have a single point of connectivity and management that incorporates HX-Series nodes and Cisco UCS servers: a feature that no other hyperconverged vendor can offer. After you deploy a cluster, you can scale it to its maximum size without needing to redesign the network. The solution is designed for easy, smooth scaling. Hyperconverged systems need massive amounts of east-west traffic bandwidth and low latency, and we deliver both with your choice of fabric interconnects supporting either 10- or 40-Gbps networking.

Networking is important in hyperconverged systems because the data platform performance depends on it. With Cisco UCS fabric interconnects, you get high-bandwidth, low-latency unified fabric connectivity that carries all production IP traffic, hyperconvergence-layer traffic, and management traffic over a single set of cables. Every connection in the cluster is treated as its own microsegment, with the same level of security as if it were supported with a separate physical link, making the integrated network more secure than when commodity approaches are used.

More adaptable

Your business needs and your workloads are constantly changing. Your infrastructure needs to quickly adapt to support your workloads and your business.

- **Easy resource expansion and contraction.** You can scale resources up and out without having to adjust your software or networking capabilities.
- **Multi-site support.** You can stretch a cluster across two sites for immediate recovery from disasters. Or you can distribute clusters geographically and replicate data between primary and backup sites.
- **Nondisruptive scaling.** Your infrastructure can easily scale out without the need to take down your cluster to add a node.
- **Ready for Microsoft Windows 10 migration.** With support for powerful graphics accelerators in our nodes, Cisco HyperFlex systems can help you support Windows 10 in a virtual desktop environment with a superior user experience.
- **Pay as you grow.** You can grow in small increments that won't break your budget, and you can independently scale your computing and capacity resources so they adapt to fit your specific application needs.

The system is designed so that all traffic, even from different blade server chassis, reaches any other node in the cluster with only a single network hop. This single-hop architecture accelerates east-west traffic, enhancing cluster performance. Our latency is deterministic, so you get consistent network performance for the data platform, and you don't have to worry about network constraints on workload placement.

Integration with Cisco Application Centric Infrastructure

As your environment grows and begins to span your enterprise, you can use Cisco ACI to implement a software-defined network. Cisco ACI provides automated, policy-based network deployment that secures your applications within secure, isolated containers. The network can attach directly to virtual machines and physical servers with increased security, real-time monitoring and telemetry, and automated performance optimization. You get consistency at scale because you can use Cisco ACI to interconnect your entire data center network, integrating your hyperconverged cluster into your existing infrastructure.

Cisco HyperFlex HX Data Platform

The HX Data Platform is a purpose-built, high-performance, scale-out file system with a wide array of enterprise-class data management services. The data platform's innovations redefine distributed storage technology, giving you complete hyperconvergence with enterprise storage features:

- **Enterprise-class data management features** required for complete lifecycle management and enhanced data protection in distributed storage environments are provided. Features include deduplication, compression, thin provisioning, fast space-efficient clones, snapshots, native replication, and integration with backup solutions from leading vendors.
- **Continuous data optimization** with inline data deduplication and compression increases resource utilization with more headroom for data scaling.
- **Securely encrypted storage** optionally encrypts both the caching and persistent layers of the data platform. Integrated with enterprise key management software or with passphrase-protected keys, encryption of data at rest helps you comply with industry and government regulations. The platform itself is hardened to Federal Information Processing Standard (FIPS) 140-1, and the encrypted drives with key management comply with the FIPS 140-2 standard.
- **Dynamic data placement** optimizes performance and resilience by enabling all cluster resources to participate in I/O responsiveness. Hybrid nodes use a combination of SSD drives for caching and HDDs for the capacity layer; All-flash nodes use SSD drives or NVME storage for the caching layer and SSDs for the capacity layer.

More scalable

You need a hyperconverged solution that is up to the scale of today's enterprise and big data applications that need large numbers of nodes to support your business.

- **More nodes.** We have doubled the maximum scale to 64 nodes, with a maximum of 32 Cisco HyperFlex nodes and 32 Cisco UCS compute-only nodes.
- **More resiliency.** We have implemented logical availability zones that help you to scale without compromising availability.
- **More capacity.** You can choose nodes with large-form-factor disk drives for even greater capacity. This allows your cluster to scale to higher capacity for storage-intensive applications.

- **Clusterwide parallel data distribution** implements clusterwide data replication for high availability and performance, accelerated by the low latency and high bandwidth of the Cisco UCS network fabric.
- **Stretch clusters** span geographic distance with continuous operation in the event of a data center failure with no data loss.
- **Logical availability zones** increase scalability and availability because they can automatically protect your data against multiple component and node failures.
- **Native replication** creates remote copies of individual virtual machines for disaster recovery purposes. With all nodes in both the local and remote clusters participating in the replication activity, the overhead imposed on each cluster node is minimal.
- **Linear and incremental scaling** provides faster, easier scalability compared to enterprise shared-storage systems, in which controller resources become a bottleneck and necessitate a complete upgrade of the storage system. Instead, whenever you add an increment of storage in Cisco HyperFlex Systems, you also increment the data platform processing capacity.
- **API-based data platform architecture** provides data virtualization flexibility to support existing and new cloud-native data types. An API for data protection allows enterprise backup solutions to create snapshot-based backups of virtual machines.
- **A simplified approach** eliminates the need to configure LUNs or to require a storage administrator to configure SANs. Storage and data services are managed completely through hypervisor tools such as VMware vCenter.

Cisco HyperFlex HX Data Platform controller

An HX Data Platform controller resides on each node and implements a distributed file system (Figure 3). The controller runs in user space within a virtual machine and intercepts and handles all I/O from guest virtual machines. Dedicated CPU cores and memory allow the controller to deliver consistent performance without affecting performance of the other virtual machines in the cluster. When nodes are configured with self-encrypting drives, the controller negotiates with Cisco UCS Manager to receive the encryption keys that enable the drives to encrypt and decrypt data that flows to and from the various storage layers.

The data platform has modules to support the specific hypervisor or container platform in use. The controller accesses all of the node's disk storage through hypervisor bypass mechanisms for excellent performance. It uses the node's memory and SSD drives or NVMe storage as part of a distributed caching layer, and it uses the node's HDDs for distributed storage. The data platform controller interfaces with the hypervisor in two ways:

- **IOvisor:** The data platform controller intercepts all I/O requests and routes requests to the nodes responsible for storing or

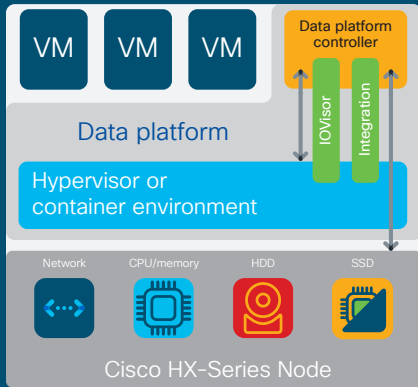


Figure 3 Cisco HyperFlex Data Platform Controller plugs into the hypervisor in each node

retrieving the blocks. The IOvisor makes the existence of the hyperconvergence layer transparent to the hypervisor.

- Advanced feature integration.** A module uses the hypervisor APIs to support advanced storage system operations such as snapshots and cloning. These are accessed through the hypervisor so that the hyperconvergence layer appears just as enterprise shared storage does. The controller accelerates operations through manipulation of metadata rather than actual data copying, providing rapid response, and thus rapid deployment of new application environments.

Data distribution

The HX Data Platform controller handles all read and write requests for volumes that the hypervisor accesses and thus intermediates all I/O from the virtual machines. Recognizing the importance of data distribution, the HX Data Platform is designed to exploit low network latencies and parallelism, in contrast to other approaches that emphasize data affinity.

With data distribution, the data platform stripes data evenly across all nodes, with the number of data replicas determined by the policies you set (Figure 4). This approach helps prevent both network and storage hot spots and makes I/O performance the same regardless of virtual machine location. This feature gives you more flexibility in workload placement and contrasts with other architectures in which a locality approach does not fully utilize all available networking and I/O resources.

- Data write operations:** For write operations, data is written to the local SSD or NVMe cache, and the replicas are written to remote caches in parallel before the write operation is acknowledged. Writes are later flushed to the capacity layer HDDs (for hybrid nodes) or SSD drives (for all-flash nodes) in a synchronous fashion.
- Data read operations:** For read operations in all-flash nodes, local and remote data is read directly from the SSD drive in the distributed capacity layer. For read operations in hybrid configurations, data

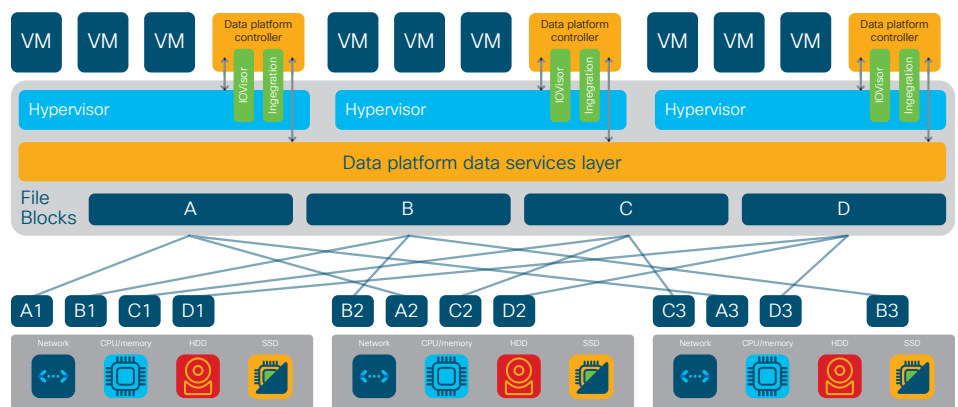


Figure 4 Data blocks are replicated across the cluster

that is local usually is read directly from the cache. This process allows the platform to use all SSD drives for read operations, reducing bottlenecks and delivering excellent performance.

In addition, when migrating a virtual machine to a new location, the data platform does not require data movement because any virtual machine can read its data from any location. Thus, moving virtual machines has no performance impact or cost.

Logical availability zones

As the maximum cluster size increases, the risk of a two-node failure shutting down the cluster increases. To mitigate this risk, logical availability zones partitions the cluster into several logical zones. The HX data platform then ensures that there is no more than one copy of data in every zone. This feature can be enabled and automatically configured with the click of a mouse.

When logical availability zones are configured, multiple node failures in a single zone don't cause the entire cluster to fail because other zones contain data replicas.

Data read and write operations

The data platform implements a log-structured file system that uses a caching layer to accelerate read requests and write responses, and it implements a capacity layer for persistent storage.

Incoming data is striped across the number of nodes that you define to meet your data availability requirements. The log-structured file system assembles blocks to be written to a configurable cache until the buffer is full or workload conditions dictate that it be destaged to a spinning disk. When existing data is (logically) overwritten, the log-structured approach simply appends a new block and updates the metadata, requiring little use of the server's Intel Xeon Scalable processors. When data is destaged, the write operation consists of a single seek operation with a large amount of data written. This approach improves performance significantly compared to the traditional read-modify-write model, which is characterized by numerous seek operations and small amounts of data written at a time.

When data is destaged to disk in each node, the data is deduplicated and compressed. This process occurs after the write operation is acknowledged, so there is no performance penalty for these operations. A small deduplication block size helps increase the deduplication rate. Compression further reduces the data footprint. Data is then moved to the capacity tier as cache segments become free (Figure 5).

Read operations in hybrid nodes cache data on the SSD drives and in main memory for high performance. In all-flash nodes they read directly from SSD drives. Having the most frequently used data stored in the caching layer helps make Cisco HyperFlex Systems with Intel Xeon Scalable processors

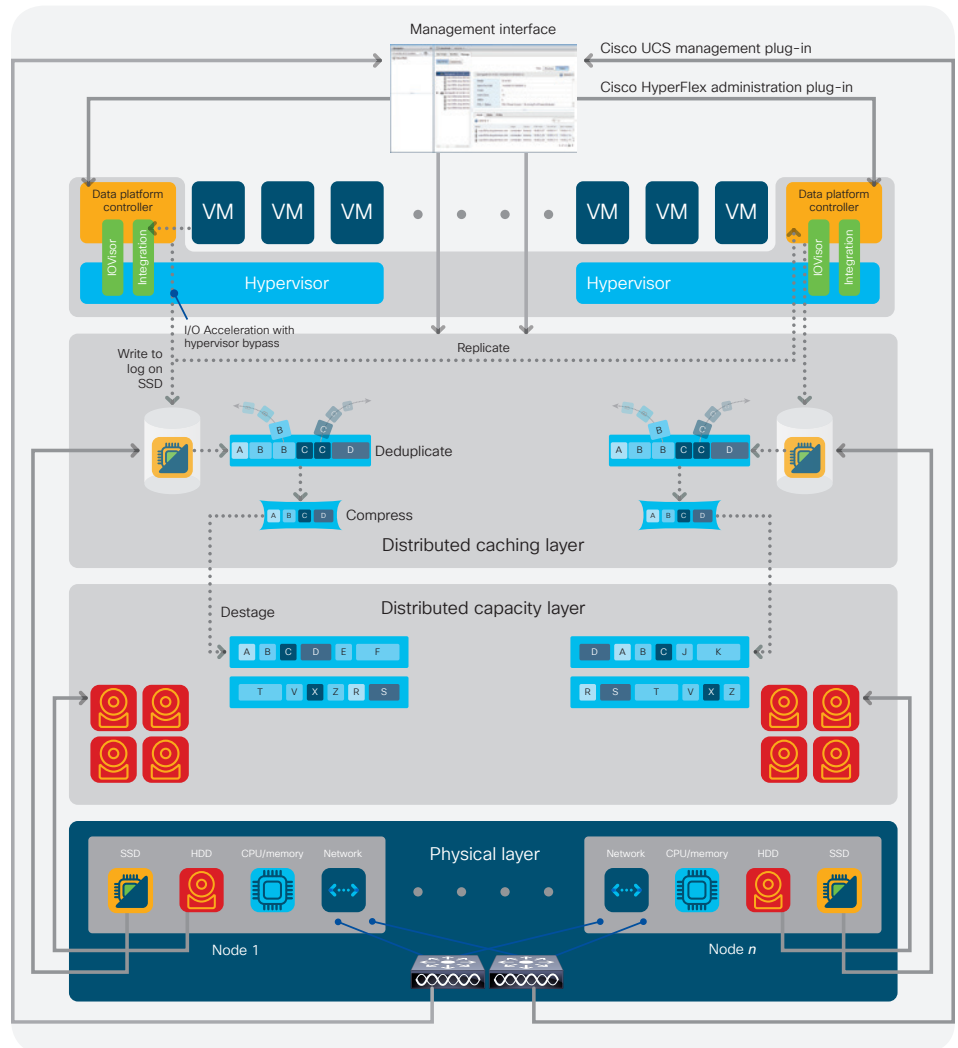


Figure 5 Data write flow through the Cisco HyperFlex HX Data Platform

perform well for virtualized applications. When virtual machines modify data, the original block is likely read from the cache, so there is often no need to read and then expand the data on a spinning disk. The data platform decouples the caching tier from the capacity tier and allows independent scaling of I/O performance and storage capacity.

Enterprise-class storage features

The data platform has all the features that you would expect of an enterprise shared storage system, eliminating the need to configure and maintain complex Fibre Channel storage networks and devices. The platform simplifies operations and helps ensure data availability. Enterprise-class storage features include the following:

- **Replication** stripes and replicates data across the cluster so that data availability is not affected if single or multiple components fail (depending on the replication factor configured).
- **Native replication** transfers cluster data to local or remote clusters for backup or disaster-recovery purposes. Data transfer is based on a many-to-many relationship between the local and remote cluster nodes: each node participates in the data transfer, minimizing the performance impact.
- **Synchronous replication**, or stretch clusters, places copies of data in two locations at the same time, enabling cluster failover with a zero RPO and very short recovery time objective (RTO).
- **Deduplication** is always on, helping reduce storage requirements in virtualization clusters in which multiple operating system instances in client virtual machines result in large amounts of replicated data.
- **Compression** further reduces storage requirements, reducing costs, and the log-structured file system is designed to store variable-sized blocks, reducing internal fragmentation.
- **Encryption** protects your data at rest with self-encrypting drives combined with enterprise key management software.
- **Thin provisioning** allows large data volumes to be created without requiring storage to support them until the need arises, simplifying data volume growth and making storage a “pay as you grow” proposition.
- **Fast, space-efficient clones** rapidly replicate storage volumes so that virtual machines can be replicated simply through metadata operations, with actual data copied only for write operations.
- **Data protection API** enables enterprise backup tools to access data volumes for consistent, per-virtual-machine backup operations.

Engineered on Cisco UCS technology

Cisco UCS, the foundation for Cisco HyperFlex Systems, is built with a single point of management and connectivity for the entire system. The system is designed as a single virtual blade server chassis that can span multiple chassis and racks of blade and rack server-based nodes. Cisco thus is in the unique position of being able to deliver a hyperconverged solution that can incorporate blade and rack systems in its architecture, offering greater flexibility than any other solution. Because Cisco develops the servers upon which Cisco HyperFlex nodes are based, you can count on having the latest processor technology available to speed your performance.

You can optimize your system with the amount of computing and storage capacity that you need by changing the ratio of CPU-intensive Cisco UCS blade and rack servers to storage-intensive Cisco HyperFlex capacity nodes. You can further optimize environments requiring graphics acceleration such as virtual desktop infrastructure by configuring GPUs into your nodes, speeding graphics rendering and helping to deliver a superior user experience.

Designed for the network edge

Our edge-optimized solution offers a lower-cost, small-footprint option, with automated management and the full feature set that you expect from Cisco HyperFlex Systems.

- Simplified three-node cluster
- Tuned to use existing 1-Gbps network
- Same easy deployment and management as your data center Cisco HyperFlex systems

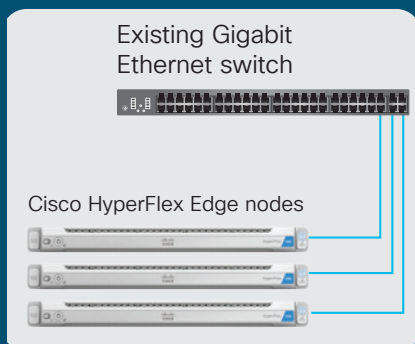


Figure 6 Cisco HyperFlex Edge is designed for remote- and branch-office environments

Cisco HyperFlex HX-Series Nodes

A cluster requires a minimum of three homogeneous nodes (with disk storage). Data is replicated across at least two of these nodes, and a third node is required for continuous operation in the event of a single-node failure. Each node is equipped with at least one high-performance SSD drive for data caching and rapid acknowledgment of write requests.

- **Hybrid nodes** use SSD drives for caching and HDDs for the capacity layer.
- **All-flash nodes** use SSD drives or NVMe storage for caching, and SSD drives for the capacity layer.

Hybrid and all-flash options are available in three types of nodes:

- **Cisco HyperFlex HX220c M5 Hybrid and All-Flash Nodes:** these balance up to 8 drives of disk capacity in a 2-socket, 1-rack-unit (1RU) package ideal for small-footprint environments.
- **Cisco HyperFlex HX240c M5 Hybrid and All-Flash Nodes:** these balance high disk capacity (up to 23 drives) in a 2-socket, 2RU package ideal for storage-intensive applications.
- **Cisco HyperFlex HX220c M5 Edge Nodes:** these nodes are designed to work in simplified three-node clusters using existing Gigabit Ethernet networks. Cisco HyperFlex Edge systems offer the same easy deployment and management as all Cisco HyperFlex systems.

Hyperconvergence at the network edge

While Cisco HyperFlex clusters support the needs of enterprise applications in data centers, Cisco HyperFlex Edge is designed for remote and branch locations including retail, finance, education, healthcare, transportation, manufacturing. Cisco HyperFlex Edge brings the robust feature set of Cisco HyperFlex systems to edge environments with a low-cost, easy-to-deploy, yet centrally managed solution (Figure 6).

Scaling with Cisco UCS servers

If you need more computing power than a cluster of HX-Series nodes provides, you can add Cisco UCS servers to your cluster to increase the ratio of computing power to storage. These servers become computing-only nodes that participate in the data platform layer but with no local cache or storage. You need to have a minimum of three HX-Series nodes in your cluster, and you can add Cisco UCS servers up to the point where you have one Cisco UCS server for each HX-Series node in the cluster. Supported servers include the Cisco UCS B200 M4 and M5 Blade Server and the Cisco UCS C220 and C240 M4 and M5 Rack Servers (Figure 7).

Scalability is enhanced with logical availability zones which minimize the impact of multiple-node failures in large clusters. If this feature is enabled, the hyperconvergence layer can automatically configure zones for foolproof operation.

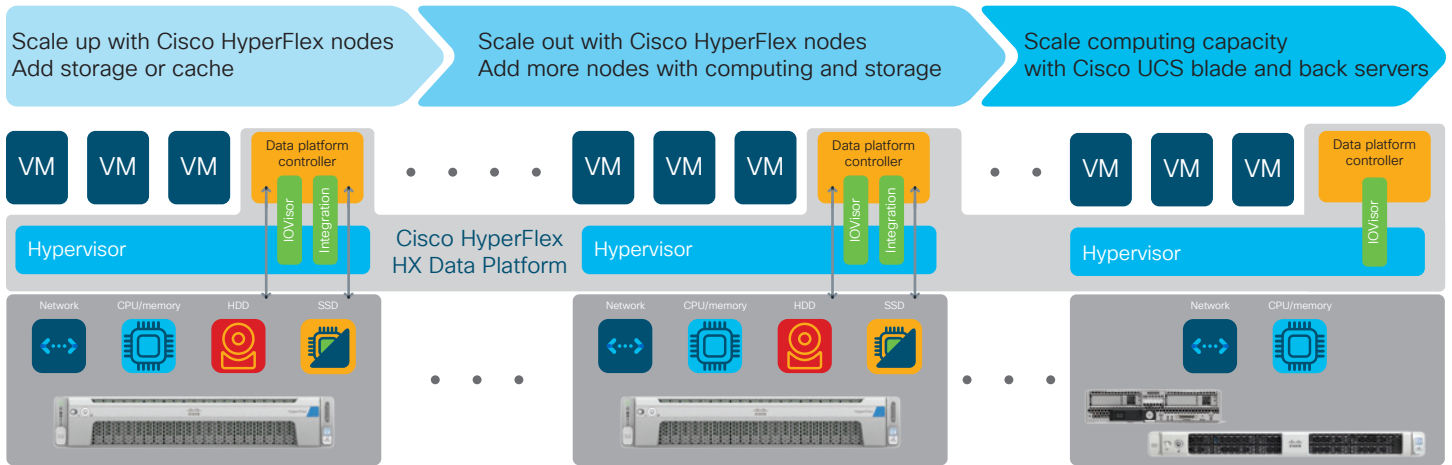


Figure 7 Scale-up and scale-out options for Cisco HyperFlex clusters

Connect to external shared storage

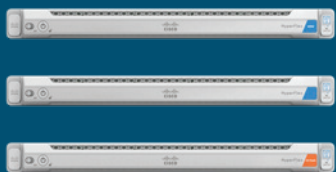
Many organizations use external shared storage in existing environments. Your Cisco HyperFlex Systems can connect to that shared storage directly through the fabric interconnects. You can connect through 8-Gbps Fibre Channel or up to 40-Gbps Small Computer System Interface over IP (iSCSI). With this capability you can:

- **Boot and run virtual machines** stored on the shared storage system
- **Migrate virtual machines** to your more scalable hyperconverged cluster
- **Use shared storage** for backing up your existing environment

Fibre Channel storage can be connected directly to each hypervisor with separate Fibre Channel interfaces that are configured through software on the Cisco UCS virtual interface cards (VICs) in each node. These interfaces allow the cluster to be configured to follow the hypervisor vendor's recommended best practices for traffic separation by creating separate network interfaces for each type of traffic, configured through software.

Powered by Intel Xeon scalable processors

Each Cisco HyperFlex M5 node is powered by two Intel Xeon scalable processors. These processors deliver significantly improved performance and can serve a much wider arrange of application needs than prior servers. The family delivers highly robust capabilities with outstanding performance, security, and agility. The CPUs provide top-of-the-line memory channel performance and include three Intel UltraPath Interconnect (UPI) links across the sockets for improved scalability and intercore data flow. Our world-record-setting benchmarks show improvements of up to 86 percent over the prior generation of servers. Each HX-Series node includes a range of processor choices so that you can best serve your application requirements.



Cisco HyperFlex HX220c M5 and All Flash and Edge Nodes



Cisco HyperFlex HX240c M5 and All Flash Nodes

Cisco Intersight management as a service

This [cloud-based interface](#) gives you instant access to all of your clusters—including installation capabilities—regardless of where they are located. This means that you can ship your nodes to the facility of your choosing and install them remotely. High-level resource inventory and status is provided by the Intersight dashboard, with drill down into specific clusters and nodes provided by a hosted version of Cisco HyperFlex Connect.

Cisco UCS management

Bringing you complete infrastructure automation, Cisco UCS management detects any component plugged into the system, making it self-aware and self-integrating. With the system itself able to adapt to changes in hardware configuration, you need only a few mouse clicks to incorporate new servers into a cluster. With Cisco UCS service profiles, every aspect of a node's identity, configuration, and connectivity is set through software, increasing efficiency and security and reducing deployment time.

Cisco HyperFlex Systems integrate easily into existing environments and operational processes. Cisco UCS management's API enables integration into higher-level management tools from Cisco and more than a dozen independent software vendors (ISVs). Tools include cloud-based Cisco Intersight management as a service, Cisco HyperFlex Connect interface, Microsoft SCVMM, Hyper-V Manager, and the VMware vSphere plug-in.

Tools that also use the Cisco UCS Management APIs include monitoring and analysis tools such as VMware vRealize Operations Manager and vCenter, other deployment and configuration tools, and service orchestration tools such as VMware vRealize Orchestrator. Cisco UCS management is also integrated into Cisco UCS Performance Manager for monitoring and analysis. When Cisco HyperFlex Systems are managed with Cisco UCS Director or Cisco ONE Enterprise Cloud Suite, they can be managed as infrastructure as a service (IaaS) or even as part of a hybrid cloud along with other Cisco and third-party hardware.

Conclusion

With Cisco HyperFlex Systems and Intel Xeon Scalable processors, we deliver a complete, next-generation hyperconverged solution that you can use for a wider range of applications and in any locations including your enterprise data center, remote locations, private cloud, and multicloud environments.

We unlock the full potential of hyperconvergence so that you can use a common platform to support more of your applications and use cases, including virtualized and containerized applications, virtual desktops, big data environments, enterprise applications, databases, and Internet infrastructure applications, along with test and development environments. Cisco HyperFlex Systems deliver the operational requirements for agility, scalability, and pay-as-you-grow economics of the cloud—but with the benefits of on-premises infrastructure.

Part of our broad data center strategy

As part of our overall data center vision, integration with Cisco ACI Enterprise Cloud Suite will put you on the road to a hybrid cloud environment (Figure 8). Cisco UCS Director enables you to use your Cisco HyperFlex Systems to

For More Information

- [Cisco HyperFlex Systems](#)
- [Cisco HyperFlex Services](#)

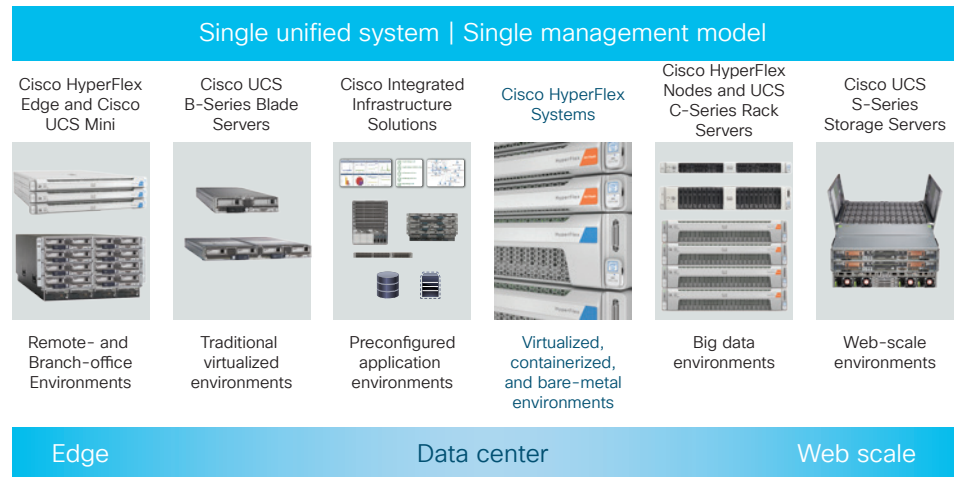


Figure 8 A single management model supports the entire Cisco product portfolio deliver IaaS and create a hybrid cloud. Your customers and clients can use the integrated service catalog to order customized application infrastructure. And when you need to augment capacity to handle periodic peaks, Enterprise Cloud Suite will help you use a policy-based approach for engaging public cloud services. When you choose Cisco HyperFlex Systems and Intel Xeon Scalable processors, you take your organization beyond a point-product solution, putting your business on a path to a more agile, adaptable, and efficient future.

Cisco HyperFlex services

Cisco Hyperflex services offers expertise that can help you rapidly migrate to Cisco Hyperflex systems and gain faster time to value. Cisco Hyperflex services can help you identify workloads, applications, and storage to migrate as well as verify I/O profiles and availability requirements of your applications.