

Easily Manage Your Structured and Unstructured Data for Analytics

Solution Brief
August 2016



With Software-Defined Storage Using IBM Spectrum Scale and Cisco UCS Integrated Infrastructure



Highlights

Deploy Integrated Infrastructure for Software-Defined Storage

- Cisco Unified Computing System™ (Cisco UCS®) and IBM Spectrum Scale make it easy to quickly deploy consistent software-defined storage.

Automate Storage Tiering and Data Lifecycle Management

- Storage policies allow data to be automatically tiered, compressed, and migrated to the right storage platform.

Keep Information Secure

- Native encryption capabilities help protect data from unauthorized access, and cryptographic erase features provide fast and secure file deletion.

Simplify Management

- A single namespace and a single point of management make it easy for you to manage and monitor large volumes of file and object data.

Reduce Cost

- The solution supports a range of disk storage options, and inexpensive solid-state disk (SSD) drives or flash-memory can be used in local caches to balance capacity, performance, and cost.

Reduce Risk

- This prevalidated solution from Cisco and IBM reduces integration and deployment risk by giving your IT staff proven configurations and guidelines for deployment.

Cisco and IBM provide the right foundation for software-defined storage so that you can easily store, access, and manage your structured and unstructured data for big data analytics.



Cisco in
collaboration
with IBM

With digital transformation changing the way your business works, your data center must do more than simply store digital records. You rely on your computing, storage, and networking resources to store, retrieve, and analyze data to help you achieve a competitive advantage. IBM Spectrum Scale running on Cisco UCS® Integrated Infrastructure for Big Data and Analytics delivers the performance and scalability you need to quickly store data to and retrieve data from your content repositories.

Cisco UCS Integrated Infrastructure for Big Data and Analytics with IBM Spectrum Scale

Cisco UCS Integrated Infrastructure for Big Data and Analytics with IBM Spectrum Scale integrates computing, network, storage, and management resources into a cohesive programmable infrastructure that can respond to the demands of users and workloads. The solution unifies Cisco UCS C240 M4 Rack Servers, Cisco Unified Computing System™ (Cisco UCS) fabric interconnects, and IBM Spectrum Scale storage with virtualization, analytics, and file-level and object-level access in a scale-out solution that is prevalidated to reduce integration and deployment risk (Figure 1).

Cisco UCS C240 M4 Rack Servers

Cisco UCS C240 M4 high-density rack servers support a range of computing, I/O, and storage-capacity demands in a compact design. The server uses dual Intel® Xeon® processor E5-2600 v4 series CPUs and supports up to 1.5 terabytes (TB) of main memory and a range of hard-disk drive (HDD) and solid-state disk (SSD) drive options. Twenty-four small-form-factor (SFF) disk drives are supported

in the performance-optimized option, and 12 large-form-factor (LFF) disk drives are supported in the capacity-optimized option.

Cisco UCS C240 M4 servers can be used with the Cisco UCS Virtual Interface Card (VIC) 1227 or 1387, depending on the fabric interconnect that is being used. The VIC 1227 is designed to optimize high-bandwidth and low-latency cluster connectivity. The VIC 1387 offers dual-port Enhanced Quad Small Form-Factor Pluggable (QSFP+) 40 Gigabit Ethernet and Fibre Channel over Ethernet (FCoE) in a modular LAN-on-motherboard (mLOM) form factor.

Cisco UCS 6200 and 6300 Series Fabric Interconnects

Cisco UCS 6200 Series Fabric Interconnects provide high-bandwidth, low-latency connectivity for servers, with integrated, unified management provided for all connected devices by Cisco UCS Manager. The latest version of this technology, Cisco UCS 6300 Series Fabric Interconnects expand these capabilities with support for 10 and 40 Gigabit Ethernet, FCoE, and Fibre Channel connectivity. Deployed in redundant pairs, Cisco fabric interconnects offer the full active-active redundancy, performance, and scalability needed to support the large number of nodes that are typical in clusters serving big data applications.

Cisco UCS Manager

Unified management simplifies your deployment and provisioning processes and provides the automation you need to be efficient. Using the role- and policy-based management capabilities of Cisco UCS Manager, your IT staff can provision servers in minutes rather than the days or weeks required in traditional environments. Ongoing maintenance activities are automated, and advanced monitoring allow the system to raise alarms and send notifications about the health of the solution.

IBM Spectrum Scale

IBM Spectrum Scale allows you to combine flash-memory, hard-disk, and tape storage into a high-performance, low-cost, unified system. This software-defined storage solution provides many advanced capabilities.

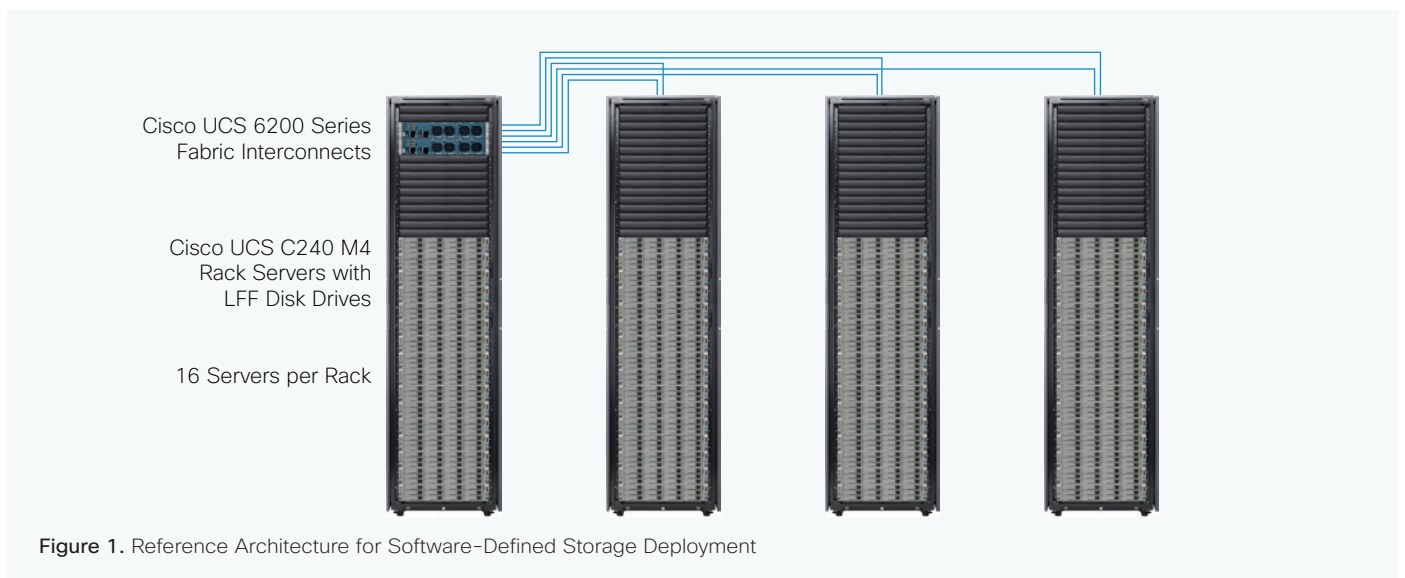


Figure 1. Reference Architecture for Software-Defined Storage Deployment

- **Automated storage tiering and data lifecycle management:** Storage policies allow data to be automatically tiered, compressed, and migrated to the right storage platform. You can group your storage devices (flash memory, SSD drives, and HDDs) based on your latency, performance, locality, or cost requirements.
- **Scalability:** You can independently scale storage capacity, performance, and protocols. With the flexibility to choose what to scale and when to scale it, you can start with a small configuration and expand to petabytes of capacity.
- **Global file sharing:** By using Active File Management (AFM) capabilities, you can deliver the right data to the right user at the right time regardless of location. This distributed disk-caching technology expands the namespace across geographic locations while delivering accelerated read and write performance to every user.
- **Performance:** Using inexpensive SSD drives or flash-memory located in your servers for your local cache helps accelerate I/O performance. CPUs spend less time waiting for data, and the load on your network and storage resources is significantly reduced, allowing other applications to benefit from available bandwidth.
- **Information security:** Native encryption capabilities can help you protect your data from unauthorized access, theft, loss, and inadvertent or improper deletion. When you want to make sure that your data is deleted, cryptographic erase features provide fast and secure file deletion.
- **Data reliability, availability, and integrity:** With no single point of failure, even in large-scale deployments, you can have confidence that the system will automatically recover so that your data remains available in the event of a node, storage, or other infrastructure failure.
- **Management simplicity:** With a single namespace and a single point of management, you can easily manage very large quantities of file and object data. Your administrators can monitor multiple installations from a single interface, improving visibility, control, and productivity.

Delivering Performance

IOzone was used to validate the performance characteristics of Cisco UCS Integrated Infrastructure for Big Data and Analytics with IBM Spectrum Scale. These tests used four Cisco UCS C240 M4 Rack Servers with 6-TB LFF disk drives. The four servers and eight client nodes were connected to two fabric interconnects.

Test Configuration

To avoid a single point of failure, systems and data were protected at several levels.

- **Server:** Three Cisco UCS C240 M4 servers ran the IBM Spectrum Scale software. The servers were configured into a cluster running in peer mode so that the failure of a server would not affect operation.
- **Disk:** The servers were configured with two data replicas to balance cost and reliability. These two copies of data help ensure that information is available in the event of a node failure. Furthermore, the disks in the servers used RAID 6 mechanisms to help ensure data integrity. By using striping, distributed parity with two independent parity blocks, and at least three drives in the configuration, data access was assured even if two drives failed.
- **Network:** Linux bonding (mode 6) was used to combine multiple network links from the servers to the fabric interconnects into a single interface for improved reliability. In mode 6, adaptive load balancing distributes network traffic to optimize the use of the underlying network links and increase throughput.

Tests with Different Block Sizes

The solution was tested using multiple transfer block sizes to determine the I/O throughput characteristics of the solution. To fully saturate the servers, IOzone cluster mode was used to start multiple IOzone test threads on multiple client nodes. Doing so allowed the aggregate throughput of the server to be gauged.

Four IOzone threads ran on each of the eight client nodes, resulting in a total of 32 IOzone threads. The sequential read and write test was performed multiple times using transfer block sizes ranging from 4 KB to 16 MB (Figure 2). These results show that the system performs consistently and can adapt to many application scenarios regardless of the application's I/O block size. Because the system used two data replicas, the total throughput measured is twice that shown in Figure 2.

Tests for Scalability

Several tests were run to highlight the scale-out performance of the solution, with the number of storage nodes increased in each test iteration. IOzone was used to perform a sequential read and write test on multiple nodes. This test was run multiple times over different numbers of server nodes. The solution scales linearly with the number of server nodes, validating that IBM Spectrum Scale can be used in clusters ranging from a few nodes to thousands of nodes to support your applications and growing volumes of data (Figure 3).

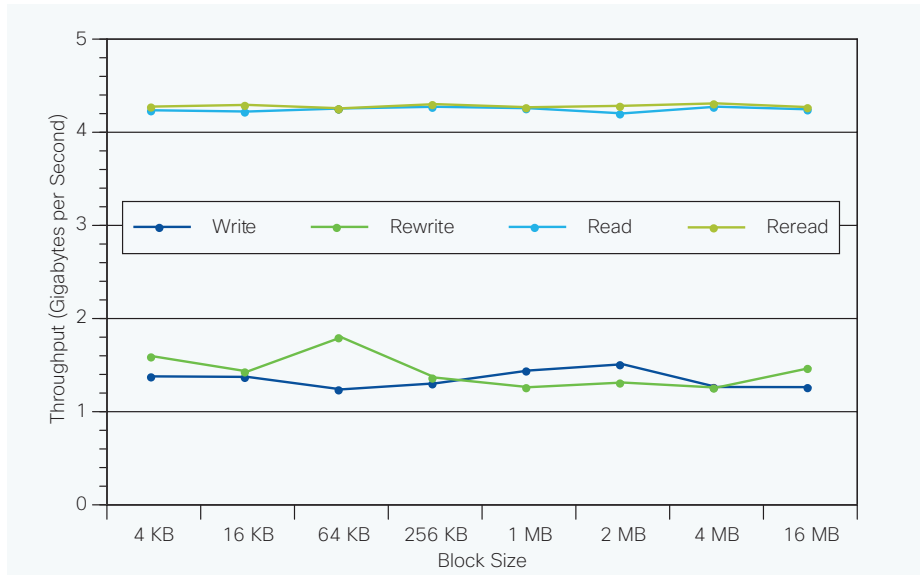


Figure 2. Test Results Using Different Block Sizes

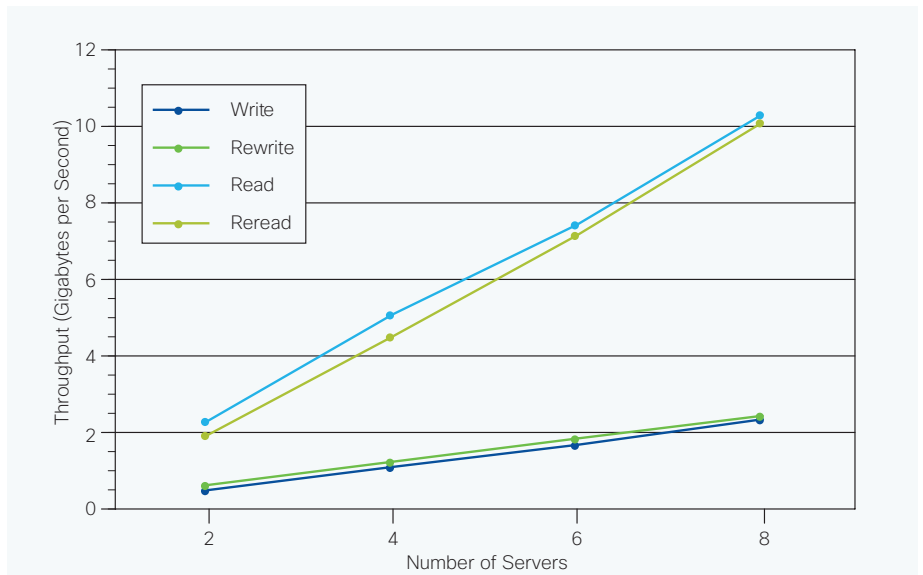


Figure 3. Scalability Test Results

Easy Ordering

With Cisco UCS Solution Accelerator Paks, you can quickly and easily deploy the software-defined storage you need in your enterprise without the expense entailed in designing and building your own custom solution (Table 1). You can scale the capacity of the solution by adding servers as needed, and the servers are integrated into the cluster in minutes.

Conclusion

If you need software-defined storage, consider programmable infrastructure from Cisco and IBM. This enterprise platform can help you support your data-intensive applications and content repositories. And you can simplify your data workflows, improve service levels, reduce costs, manage risk, and deliver business results today while positioning your data center for growth.

For More Information

- For more information about Cisco® big data solutions, visit <http://www.cisco.com/go/bigdata>.
- For more information about Cisco UCS Integrated Infrastructure for Big Data, visit <http://blogs.cisco.com/datacenter/cpav4>.
- For more information about IBM Spectrum Scale, visit <http://www.ibm.com/systems/storage/spectrum/scale>.

Table 1. Cisco UCS Solution Accelerator Paks for Software-Defined Storage

Solution Solution SKU Server SKU	Capacity Optimized Option 1 UCS-SL-CPA4-C1 UCS-SPBD-C240M4-C1	Capacity Optimized Option 2 UCS-SL-CPA4-C2 UCS-SPBD-C240M4-C2
Connectivity	2 Cisco UCS 6296UP 96-Port Fabric Interconnects	2 Cisco UCS 6296UP 96-Port Fabric Interconnects
Servers	16 Cisco UCS C240 M4 Rack Servers, each with: <ul style="list-style-type: none"> • 2 Intel Xeon processor E5-2620 v4 CPUs (8 cores; 256 cores for solution) • 128 GB of memory • 2 x 240-GB 6-Gbps SSD drives • 12 x 6-TB 7.2K LFF SAS drives • Total of 72 TB of storage and 2.5 GBps of I/O bandwidth • Cisco UCS VIC 1227 	16 Cisco UCS C240 M4 Rack Servers, each with: <ul style="list-style-type: none"> • 2 Intel Xeon processor E5-2620 v4 CPUs (8 cores; 256 cores for solution) • 256 GB of memory • 2 x 240-GB 6-Gbps SSD drives • 12 x 8-TB 7.2K LFF SAS drives • Total of 96 TB of storage and 2.3 GBps of I/O bandwidth • Cisco UCS VIC 1227
Storage controller	• Cisco 12-Gbps SAS Modular RAID Controller with 2-GB flash-based write cache (FBWC)	
Rack space	• 36RU	• 36RU
Scaling	<ul style="list-style-type: none"> • Up to 80 servers per domain with no oversubscription • Scalability to thousands of servers with Cisco Nexus® 7000 or 9000 Series Switches 	• Up to 80 servers per domain with no oversubscription
Key to abbreviations: 10,000-rpm (10K); 7200-rpm (7.2K) large form factor (LFF); rack units (RU); small form factor (SFF); terabyte (TB); and virtual interface card (VIC)		



Americas Headquarters
 Cisco Systems, Inc.
 San Jose, CA

Asia Pacific Headquarters
 Cisco Systems (USA) Pte. Ltd.
 Singapore

Europe Headquarters
 Cisco Systems International BV Amsterdam,
 The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.