VersaStack for IBM Cloud Object Storage on Cisco UCS C240 for Concentrated Dispersal Mode Design Guide

Published: October 2018

About the Cisco Validated Design Program

The Cisco Validated Design (CVD) program consists of systems and solutions designed, tested, and documented to facilitate faster, more reliable, and more predictable customer deployments. For more information, see:

http://www.cisco.com/go/designzone.

ALL DESIGNS, SPECIFICATIONS, STATEMENTS, INFORMATION, AND RECOMMENDATIONS (COLLECTIVELY, "DESIGNS") IN THIS MANUAL ARE PRESENTED "AS IS," WITH ALL FAULTS. CISCO AND ITS SUPPLIERS DISCLAIM ALL WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OR ARISING FROM A COURSE OF DEALING, USAGE, OR TRADE PRACTICE. IN NO EVENT SHALL CISCO OR ITS SUPPLIERS BE LIABLE FOR ANY INDIRECT, SPECIAL, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS OR LOSS OR DAMAGE TO DATA ARISING OUT OF THE USE OR INABILITY TO USE THE DESIGNS, EVEN IF CISCO OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

THE DESIGNS ARE SUBJECT TO CHANGE WITHOUT NOTICE. USERS ARE SOLELY RESPONSIBLE FOR THEIR APPLICATION OF THE DESIGNS. THE DESIGNS DO NOT CONSTITUTE THE TECHNICAL OR OTHER PROFESSIONAL ADVICE OF CISCO, ITS SUPPLIERS OR PARTNERS. USERS SHOULD CONSULT THEIR OWN TECHNICAL ADVISORS BEFORE IMPLEMENTING THE DESIGNS. RESULTS MAY VARY DEPENDING ON FACTORS NOT TESTED BY CISCO.

CCDE, CCENT, Cisco Eos, Cisco Lumin, Cisco Nexus, Cisco StadiumVision, Cisco TelePresence, Cisco WebEx, the Cisco logo, DCE, and Welcome to the Human Network are trademarks; Changing the Way We Work, Live, Play, and Learn and Cisco Store are service marks; and Access Registrar, Aironet, AsyncOS, Bringing the Meeting To You, Catalyst, CCDA, CCDP, CCIE, CCIP, CCNA, CCNP, CCSP, CCVP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unified Computing System (Cisco UCS), Cisco UCS B-Series Blade Servers, Cisco UCS C-Series Rack Servers, Cisco UCS S-Series Storage Servers, Cisco UCS Manager, Cisco UCS Management Software, Cisco Unified Fabric, Cisco Application Centric Infrastructure, Cisco Nexus 9000 Series, Cisco Nexus 7000 Series. Cisco Prime Data Center Network Manager, Cisco NX-OS Software, Cisco MDS Series, Cisco Unity, Collaboration Without Limitation, EtherFast, EtherSwitch, Event Center, Fast Step, Follow Me Browsing, FormShare, GigaDrive, HomeLink, Internet Quotient, IOS, iPhone, iQuick Study, LightStream, Linksys, MediaTone, MeetingPlace, MeetingPlace Chime Sound, MGX, Networkers, Networking Academy, Network Registrar, PCNow, PIX, PowerPanels, ProConnect, ScriptShare, SenderBase, SMARTnet, Spectrum Expert, StackWise, The Fastest Way to Increase Your Internet Quotient, TransPath, WebEx, and the WebEx logo are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries.

All other trademarks mentioned in this document or website are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0809R)

Table of Contents

Executive Summary

Cisco Validated Designs consist of systems and solutions that are designed, tested, and documented to facilitate and improve customer deployments. These designs incorporate a wide range of technologies and products into a portfolio of solutions that have been developed to address the business needs of our customers.

The purpose of this document is to describe the design of IBM Cloud Object Storage (COS) on latest generation of Cisco UCS C240 Rack Servers. This validated design provides the framework of designing IBM COS software on Cisco UCS C240 Rack Servers. The Cisco Unified Computing System provides the storage, network, and storage access components for IBM COS, deployed as a single cohesive system.

Cisco Validated design describes how the Cisco Unified Computing System can be used in conjunction with IBM COS 3.13.6 or later releases in a Concentrated Dispersal Mode. With the continuous evolution of Software Defined Storage (SDS), there has been increased demand to have small IBM COS solutions validated on Cisco UCS servers. The Cisco UCS C240 Rack Server, originally designed for the data center, together with IBM COS is optimized for such object storage solutions, making it an excellent fit for unstructured data workloads such as active archive and backup. The Cisco UCS C240 Rack Server delivers a complete infrastructure with exceptional scalability for computing and storage resources together with 25/40 Gigabit Ethernet networking.

Cisco and IBM are collaborating to offer customers a scalable object storage solution for unstructured data that is integrated with IBM COS. With the power of the Cisco UCS management framework, the solution is cost effective to deploy and manage and will enable the next-generation cloud deployments that drive business agility, lower operational costs and avoid vendor lock-in.

Solution Overview

Introduction

Traditional storage systems are limited in their ability to easily and cost-effectively scale to support large amounts of unstructured data. With about 80 percent of data being unstructured, new approaches using x86 servers are proving to be more cost effective, providing storage that can be expanded as easily as your data grows. Software Defined Storage is a scalable and cost-effective approach for handling large amounts of data.

But more and more there are requirements to store unstructured data even in smaller quantities as object storage. The advantage of identifying the data by metadata and not taking over management of the location is very attractive even for smaller quantities. As a result, new technologies need to be developed to provide similar levels of availability and reliability as large scale-out object storage solutions.

IBM's COS is a storage platform that is ideal for holding large amounts of colder production data, such as backups and archives, and very large individual files, such as video files, image files, and genomic data and can also include support of warm or even hot data, by increasing CPU performance and/or memory capacity. COS is highly reliable, durable, and resilient object storage for that is designed for scale and security.

With the Concentrated Dispersal mode (CD mode) feature, IBM has developed a way to map the COS solution to small environments without sacrificing availability and reliability. CD mode enables customers to initially deploy IBM COS at smaller capacity sizes, starting as low as 72 terabytes, and grow incrementally adding interfaces and capacity as needed. Often customers want to start their initial investment into object storage at a small scale based on one use case. The new CD mode feature provides a cost-efficient method to begin with a small deployment (72TB) and scale to petabytes and beyond. Customers can take advantage of the benefits of IBM COS in terms of data availability and system reliability along with management capabilities, allowing them to consider extending the infrastructure to other use cases. Customers will have the flexibility to deploy low-capacity configurations in one or two-site configurations as well as with multiple sites depending on their requirements.

Together with Cisco UCS, IBM COS can deliver a fully enterprise-ready solution that can manage different workloads and still remain flexible. The Cisco UCS C240 Rack Server is an excellent platform to use with the main types of object workloads, such as active archive or backup workloads but not only. It is best suited for sequential access, as opposed to random, to unstructured data, and to whatever the data size. It is designed for applications, not direct end-users.

Audience

The intended audience for this document includes, but is not limited to, sales engineers, field consultants, professional services, IT managers, partner engineering, and customers who want to deploy an IBM COS on the Cisco Unified Computing System (Cisco UCS) using Cisco UCS C240 Rack Servers.

Purpose of this Document

This document describes the architecture and design of an IBM COS solution using three Cisco UCS C240 Rack Servers and two Cisco UCS 6332 Fabric Interconnect managed by Cisco UCS Manager. Provided are

the steps required to design an IBM COS in CD mode on Cisco UCS. It shows the simplicity of installing and configuring Cisco UCS rack server and illustrates the need of a well-conceived network architecture for low-latency, high-bandwidth.

Solution Summary

This Cisco Validated Design (CVD) is a simple and linearly scalable architecture that provides Software Defined Storage for object on IBM COS 3.13.16 and Cisco UCS C240 rack server. This CVD describes in detail the design of IBM COS 3.13.16 on Cisco UCS C240 rack server. The solution includes the following features:

- Infrastructure for scale-out storage.

- Design of an IBM COS solution in CD mode together with Cisco UCS C240 rack server.

- Simplified infrastructure management with Cisco UCS Manager (UCSM).

The configuration uses the following architecture for the deployment:

- 3 x Cisco UCS C240 Rack Servers

- 2 x Cisco UCS 6332 Fabric Interconnect

- 1 x Cisco UCS Manager

- 2 x Cisco Nexus C9332PQ Switches

The solution has various options to scale capacity. The tested configuration uses an IDA (Information Dispersal Algorithm, s. page 13) of 18/9/11 with 3 Slicestors. A base capacity summary for the tested solution is listed in Table 1 .

Table 1   Usable Capacity Options for Tested Cisco Validated Design

HDD Type	Number of Disks	Concentrated Dispersal Mode
4 TB 7200-rpm LFF SAS drives	36	72 TB
6 TB 7200-rpm LFF SAS drives	36	108 TB
8 TB 7200-rpm LFF SAS drives	36	144 TB
10 TB 7200-rpm LFF SAS drives*	36	180 TB
12 TB 7200-rpm LFF SAS drives	36	216 TB

* Tested configuration

Technology Overview

Cisco Unified Computing System

The Cisco Unified Computing System™ (Cisco UCS) is a state-of-the-art data center platform that unites computing, network, storage access, and virtualization into a single cohesive system.

The main components of Cisco Unified Computing System are:

- Computing - The system is based on an entirely new class of computing system that incorporates rack-mount and blade servers based on Intel® Xeon® Scalable processors. The Cisco UCS servers offer the patented Cisco Extended Memory Technology to support applications with large datasets and allow more virtual machines (VM) per server.

- Network - The system is integrated onto a low-latency, lossless, 10/25/40-Gbps unified network fabric. This network foundation consolidates LANs, SANs, and high-performance computing networks which are separate networks today. The unified fabric lowers costs by reducing the number of network adapters, switches, and cables, and by decreasing the power and cooling requirements.

- Virtualization - The system unleashes the full potential of virtualization by enhancing the scalability, performance, and operational control of virtual environments. Cisco security, policy enforcement, and diagnostic features are now extended into virtualized environments to better support changing business and IT requirements.

- Storage access - The system provides consolidated access to both SAN storage and Network Attached Storage (NAS) over the unified fabric. By unifying the storage access, the Cisco Unified Computing System can access storage over Ethernet (NFS or iSCSI), Fibre Channel, and Fibre Channel over Ethernet (FCoE). This provides customers with choice for storage access and investment protection. In addition, the server administrators can pre-assign storage-access policies for system connectivity to storage resources, simplifying storage connectivity, and management for increased productivity.

The Cisco Unified Computing System is designed to deliver:

- A reduced Total Cost of Ownership (TCO) and increased business agility.

- Increased IT staff productivity through just-in-time provisioning and mobility support.

- A cohesive, integrated system, which unifies the technology in the data center.

- Industry standards supported by a partner ecosystem of industry leaders.

Cisco UCS C240 Rack Server

The Cisco UCS C240 Rack Server is a 2-socket, 2-Rack-Unit (2RU) rack server offering industry-leading performance and expandability. It supports a wide range of storage and I/O-intensive infrastructure workloads, from big data and analytics to collaboration. Cisco UCS C-Series Rack Servers can be deployed **as standalone servers or as part of a Cisco UCS managed environment to take advantage of Cisco's**

standards-based unified computing innovations that help reduce **customers' TCO** and increase their business agility.

Figure 1 Cisco UCS C240 Rack Server

In response to ever-increasing computing and data-intensive real-time workloads, the enterprise-class Cisco UCS C240 server extends the capabilities of the Cisco UCS portfolio in a 2RU form factor. It incorporates the Intel® Xeon® Scalable processors, supporting up to 20 percent more cores per socket, twice the memory capacity, and five times more Non-Volatile Memory Express (NVMe) PCI Express (PCIe) Solid-State Disks (SSDs) compared to the previous generation of servers. These improvements deliver significant performance and efficiency gains that will improve your application performance. The Cisco UCS C240 M5 delivers outstanding levels of storage expandability with exceptional performance, comprised of the following:

- Latest Intel Xeon Scalable CPUs with up to 28 cores per socket

- Up to 24 DDR4 DIMMs for improved performance

- Up to 26 hot-swappable Small-Form-Factor (SFF) 2.5-inch drives, including 2 rear hot-swappable SFF drives (up to 10 support NVMe PCIe SSDs on the NVMe-optimized chassis version), or 12 Large-Form-Factor (LFF) 3.5-inch drives plus 2 rear hot-swappable SFF drives

- Support for 12-Gbps SAS modular RAID controller in a dedicated slot, leaving the remaining PCIe Generation 3.0 slots available for other expansion cards

- Modular LAN-On-Motherboard (mLOM) slot that can be used to install a Cisco UCS Virtual Interface Card (VIC) without consuming a PCIe slot, supporting dual 10-, 25- or 40-Gbps network connectivity

- Dual embedded Intel x550 10GBASE-T LAN-On-Motherboard (LOM) ports

- Modular M.2 or Secure Digital (SD) cards that can be used for boot

The Cisco UCS C240 rack server is well suited for a wide range of enterprise workloads, including:

- Object Storage

- Big Data and analytics

- Collaboration

- Small and medium-sized business databases

- Virtualization and consolidation

- Storage servers

- High-performance appliances

Cisco UCS C240 rack servers can be deployed as standalone servers or in a Cisco UCS managed environment. When used in combination with Cisco UCS Manager, the Cisco UCS C240 brings the power and automation of unified computing to enterprise applications, including Cisco® SingleConnect technology, drastically reducing switching and cabling requirements.

Cisco UCS Manager uses service profiles, templates, and policy-based management to enable rapid deployment and help ensure deployment consistency. If also enables end-to-end server visibility, management, and control in both virtualized and bare-metal environments.

The Cisco Integrated Management Controller (IMC) delivers comprehensive out-of-band server management with support for many industry standards, including:

- Redfish Version 1.01 (v1.01)

- Intelligent Platform Management Interface (IPMI) v2.0

- Simple Network Management Protocol (SNMP) v2 and v3

- Syslog

- Simple Mail Transfer Protocol (SMTP)

- Key Management Interoperability Protocol (KMIP)

- HTML5 GUI

- HTML5 virtual Keyboard, Video, and Mouse (vKVM)

- Command-Line Interface (CLI)

- XML API

Management Software Development Kits (SDKs) and DevOps integrations exist for Python, Microsoft PowerShell, Ansible, Puppet, Chef, and more. For more information about integrations, see Cisco DevNet (https://developer.cisco.com/site/ucs-dev-center/).

The Cisco UCS C240 is Cisco Intersight™ ready. Cisco Intersight is a new cloud-based management platform that uses analytics to deliver proactive automation and support. By combining intelligence with automated actions, you can reduce costs dramatically and resolve issues more quickly.

Cisco UCS Virtual Interface Card 1387

The Cisco UCS Virtual Interface Card 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. The mLOM slot can be used to install a Cisco VIC without consuming a PCIe slot providing greater I/O expandability.

Figure 2 Cisco UCS Virtual Interface Card 1387

The Cisco UCS VIC 1387 provides high network performance and low latency for the most demanding applications, including:

- Big data, high-performance computing (HPC), and high-performance trading (HPT)

- Large-scale virtual machine deployments

- High-bandwidth storage targets and archives

The card is designed for the M5 generation of Cisco UCS C-Series Rack Servers and Cisco UCS S3260 dense storage servers. **It includes Cisco's next**-generation converged network adapter technology and offers a comprehensive feature set, so you gain investment protection for future feature software releases.

The card can present more than 256 PCIe standards-compliant interfaces to its host. These can be dynamically configured as either network interface cards (NICs) or host bus adapters (HBAs).

This engine provides support for advanced data center requirements, including stateless network offloads for:

- Network Virtualization Using Generic Routing Encapsulation (NVGRE)

- Virtual extensible LAN (VXLAN)

- Remote direct memory access (RDMA)

The engine also offers support for performance optimization applications such as:

- Server Message Block (SMB) Direct

- Virtual Machine Queue (VMQ)

- Data Plane Development Kit (DPDK)

- Cisco NetFlow

Cisco UCS 6300 Series Fabric Interconnect

The Cisco UCS 6300 Series Fabric Interconnects are a core part of Cisco UCS, providing both network connectivity and management capabilities for the system (Figure 3). The Cisco UCS 6300 Series offers line-rate, low-latency, lossless 10 and 40 Gigabit Ethernet, Fibre Channel over Ethernet (FCoE), and Fibre Channel functions.

Figure 3 Cisco UCS 6300 Series Fabric Interconnect

The Cisco UCS 6300 Series provides the management and communication backbone for the Cisco UCS B-Series Blade Servers, 5100 Series Blade Server Chassis, Cisco UCS C-Series Rack Servers, and Cisco UCS S-Series Storage Dense Server managed by Cisco UCS. All servers attached to the fabric interconnects become part of a single, highly available management domain. In addition, by supporting unified fabric, the Cisco UCS 6300 Series provides both LAN and SAN connectivity for all servers within its domain.

From a networking perspective, the Cisco UCS 6300 Series uses a cut-through architecture, supporting deterministic, low-latency, line-rate 10 and 40 Gigabit Ethernet ports, switching capacity of 2.56 terabits per second (Tbps), and 320 Gbps of bandwidth per chassis, independent of packet size and enabled services. The product family supports Cisco® low-latency, lossless 10 and 40 Gigabit Ethernet unified network fabric capabilities, which increase the reliability, efficiency, and scalability of Ethernet networks. The fabric interconnect supports multiple traffic classes over a lossless Ethernet fabric from the server through the fabric interconnect. Significant TCO savings can be achieved with an FCoE optimized server design in which network interface cards (NICs), host bus adapters (HBAs), cables, and switches can be consolidated.

The Cisco UCS 6332 32-Port Fabric Interconnect is a 1-rack-unit (1RU) Gigabit Ethernet, and FCoE switch offering up to 2.56 Tbps throughput and up to 32 ports. The switch has 32 fixed 40-Gbps Ethernet and FCoE ports.

Both the Cisco UCS 6332UP 32-Port Fabric Interconnect and the Cisco UCS 6332 16-UP 40-Port Fabric Interconnect have ports that can be configured for the breakout feature that supports connectivity between 40 Gigabit Ethernet ports and 10 Gigabit Ethernet ports. This feature provides backward compatibility to existing hardware that supports 10 Gigabit Ethernet. A 40 Gigabit Ethernet port can be used as four 10 Gigabit Ethernet ports. Using a 40 Gigabit Ethernet SFP, these ports on a Cisco UCS 6300 Series Fabric Interconnect can connect to another fabric interconnect that has four 10 Gigabit Ethernet SFPs. The breakout feature can be configured on ports 1 to 12 and ports 15 to 26 on the Cisco UCS 6332UP fabric interconnect. Ports 17 to 34 on the Cisco UCS 6332 16-UP fabric interconnect support the breakout feature.

Cisco Nexus 9332PQ Switch

The Cisco Nexus® 9000 Series Switches (Figure 4) include both modular and fixed-port switches that are designed to overcome these challenges with a flexible, agile, low-cost, application-centric infrastructure.

Figure 4 Cisco Nexus 9332PQ Switch

The Cisco Nexus 9300 platform consists of fixed-port switches designed for top-of-rack (ToR) and middle-of-row (MoR) deployment in data centers that support enterprise applications, service provider hosting, and cloud computing environments. They are Layer 2 and 3 nonblocking 10 and 40 Gigabit Ethernet switches with up to 2.56 terabits per second (Tbps) of internal bandwidth.

The Cisco Nexus 9332PQ Switch is a 1-rack-unit (1RU) switch that supports 2.56 Tbps of bandwidth and over 720 million packets per second (mpps) across thirty-two 40-Gbps Enhanced QSFP+ ports.

All the Cisco Nexus 9300 platform switches use dual- core 2.5-GHz x86 CPUs with 64-GB solid-state disk (SSD) drives and 16 GB of memory for enhanced network performance.

With the Cisco Nexus 9000 Series, organizations can quickly and easily upgrade existing data centers to carry 40 Gigabit Ethernet to the aggregation layer or to the spine (in a leaf-and-spine configuration) through advanced and cost-effective optics that enable the use of existing 10 Gigabit Ethernet fiber (a pair of multimode fiber strands).

Cisco provides two modes of operation for the Cisco Nexus 9000 Series. Organizations can use Cisco® NX-OS Software to deploy the Cisco Nexus 9000 Series in standard Cisco Nexus switch environments. Organizations also can use a hardware infrastructure that is ready to support Cisco Application Centric Infrastructure (Cisco ACI™) to take full advantage of an automated, policy-based, systems management approach.

Cisco UCS Manager

Cisco UCS® Manager (Figure 5) provides unified, embedded management of all software and hardware components of the Cisco Unified Computing System™ (Cisco UCS) across multiple chassis, rack servers, and thousands of virtual machines. It supports all Cisco UCS product models, including Cisco UCS B-Series Blade Servers, C-Series Rack Servers, and S- and M-Series composable infrastructure and Cisco UCS Mini, as well as the associated storage resources and networks. Cisco UCS Manager is embedded on a pair of Cisco UCS 6300 or 6200 Series Fabric Interconnects using a clustered, active-standby configuration for high availability. The manager participates in server provisioning, device discovery, inventory, configuration, diagnostics, monitoring, fault detection, auditing, and statistics collection.

Figure 5 Cisco UCS Manager
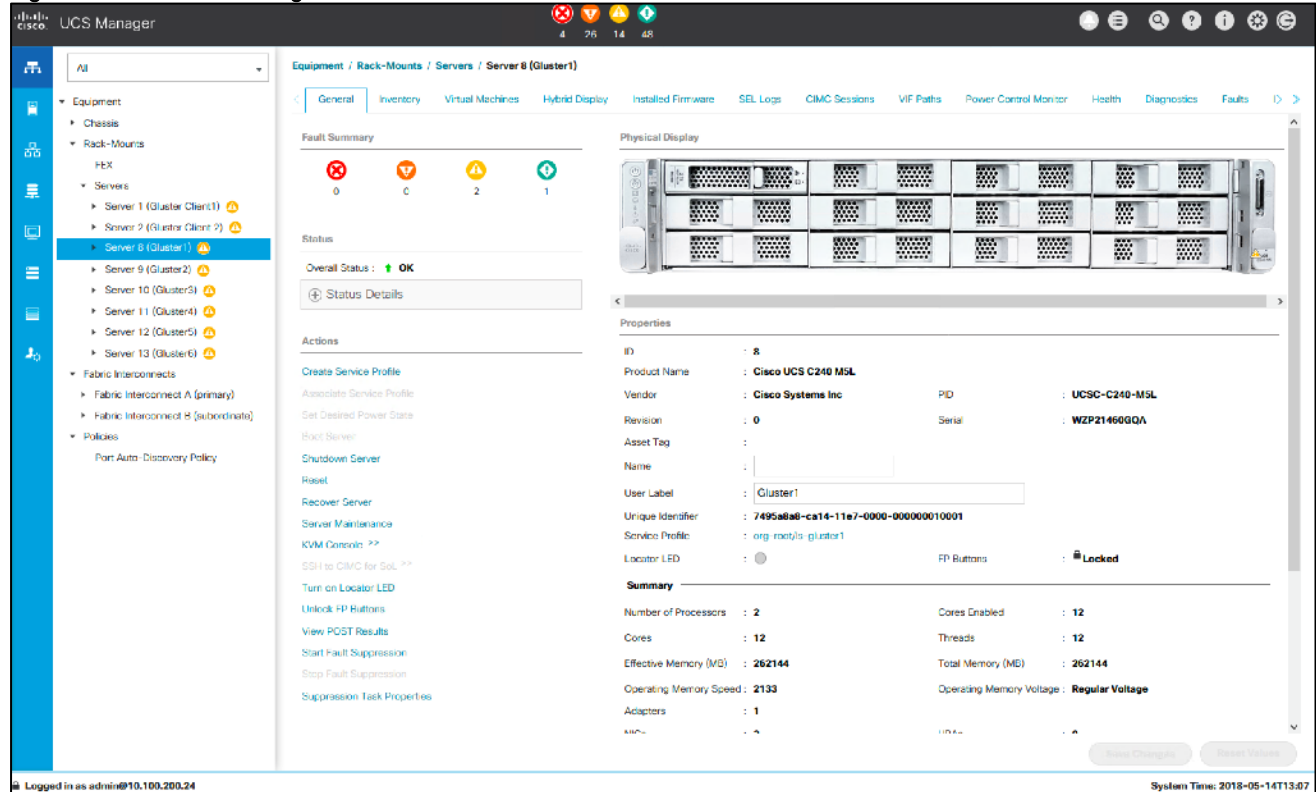
An instance of Cisco UCS Manager with all Cisco UCS components managed by it forms a Cisco UCS domain, which can include up to 160 servers. In addition to provisioning Cisco UCS resources, this infrastructure management software provides a model-based foundation for streamlining the day-to-day processes of updating, monitoring, and managing computing resources, local storage, storage connections, and network connections. By enabling better automation of processes, Cisco UCS Manager allows IT organizations to achieve greater agility and scale in their infrastructure operations while reducing complexity and risk. The manager provides flexible role- and policy-based management using service profiles and templates.

Cisco UCS Manager manages Cisco UCS systems through an intuitive HTML 5 or Java user interface and a command-line interface (CLI). It can register with Cisco UCS Central Software in a multi-domain Cisco UCS environment, enabling centralized management of distributed systems scaling to thousands of servers. Cisco UCS Manager can be integrated with Cisco UCS Director to facilitate orchestration and to provide support for converged infrastructure and Infrastructure as a Service (IaaS).

The Cisco UCS XML API provides comprehensive access to all Cisco UCS Manager functions. The API provides Cisco UCS system visibility to higher-level systems management tools from independent software vendors (ISVs) such as VMware, Microsoft, and Splunk as well as tools from BMC, CA, HP, IBM, and others. ISVs and in-house developers can use the XML API to enhance the value of the Cisco UCS platform according to their unique requirements. Cisco UCS PowerTool for Cisco UCS Manager and the Python Software Development Kit (SDK) help automate and manage configurations within Cisco UCS Manager.

IBM Cloud Object Storage

The IBM COS System platform is ideal whenever enterprises need to securely store large volumes of unstructured data with high availability and where latency is not a primary consideration.

With the unprecedented growth in new digital information, use cases have emerged that enable organizations to store and distribute limitless data. A distributed and decentralized storage architecture along with an Object Storage interface enables enterprises to deliver data to their users across the globe as never before. The use cases covered in the Cisco Validated Design include:

- Content Repository

- Storage-as-a-service

- Enterprise Collaboration

- Backup

- Archive

The IBM COS System software platform uses an approach for cost-effectively storing large volumes of unstructured data while still ensuring security, availability, and reliability.

The IBM COS System storage technology uses Information Dispersal Algorithms (IDA) to separate data into **unrecognizable "slices" that are distributed via network connections to storage nodes locally or** across the world. The collection of dispersed storage appliances creates what is called a dispersed storage network. With dispersed storage technology, transmission and storage of data are inherently private and secure. No complete copy of the data exists in any single storage node. Only a subset of nodes needs to be available to fully retrieve the data on the network.

IDA technology transforms data into slices by using equations such that a subset of the slices can be used to re-create the original data. These slices, which are like packets but are for data storage, are then stored across multiple storage appliances (or storage nodes). Slices are created by using a combination of erasure coding, encryption, and sophisticated dispersal algorithms.

Dispersed storage systems are well-suited for storing unstructured data like digital media of all types, documents that are produced by desktop productivity applications, and server log files, which are typically larger files. Dispersal is not optimized for transaction-oriented primary storage for databases and similar high IOP workloads because of the extra processing associated with slicing and dispersing.

At a basic level, the IBM COS System platform uses three steps for slicing, dispersing, and retrieving data (Figure 6):

1. Data is virtualized, transformed, sliced, and dispersed by using IDAs. In the first figure, the data is separated into 12 slices. So the "width" (n) of the system is 12.

2. Slices are distributed to some combination of separate disks, storage nodes, and geographic locations. In this example, the slices are distributed to three different sites.

3. The data is retrieved from a subset of slices. In this example, the number of slices that are needed to retrieve the data is 7. So the "threshold" (k) of the system is 7.

Given a width of 12 and a threshold of 7, you can refer to this example as a "7 of 12" (k of n) configuration.

The configuration of a system is determined by the level of reliability needed. In a "7 of 12" configuration," five slices can be lost or unavailable and the data can still be retrieved because the threshold of seven slices is met. With a "5 of 8" configuration, only three slices can be lost, so the level of reliability is lower. Conversely, with a "20 of 32" configuration, 12 slices can be lost, so the level of reliability is higher.

Figure 6 How Dispersed Storage Works
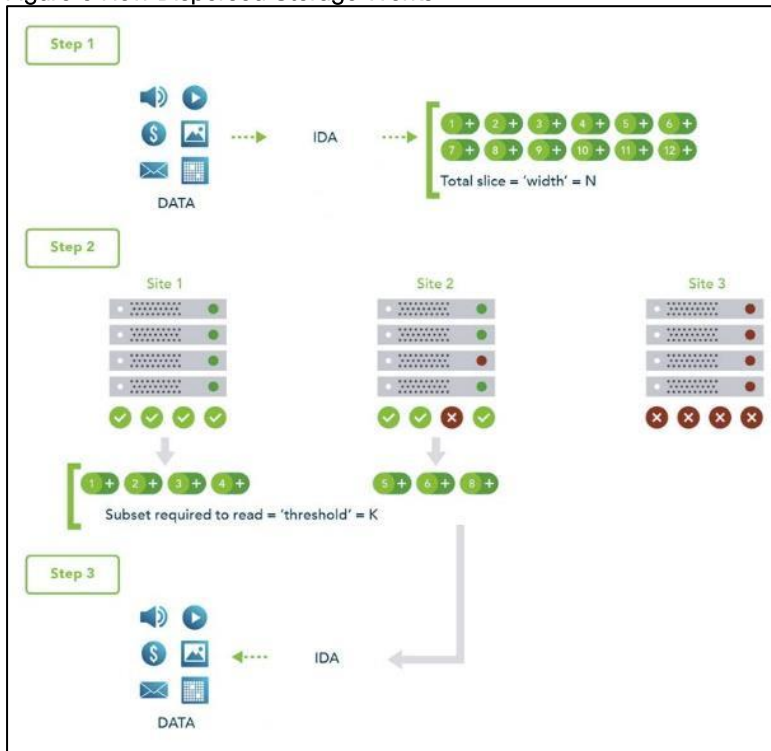
Concentrated Dispersal Mode

With the Concentrated Dispersal feature used in this CVD, each Slicestor device may be responsible for multiple slices of a given object stored in the system. Previously, each Slicestor device could only hold one slice of a given object. Prior to this the IDA Width set a lower bound on the number of Slicestor devices necessary in the deployment. For example, when using an IDA Width of 18, one would have to deploy at least 18 Slicestor devices. With the Concentrated Dispersal feature, each Slicestor device might hold 6 slices of each object, and so only 3 Slicestor devices would be necessary. By continuing to use a large IDA Width, high reliability and storage efficiency can be attained, but now with a much smaller hardware footprint.

Deploying a system requires the selection of an IDA Width, Threshold, and Write Threshold that provides an acceptable level of storage efficiency, reliability, and availability. Achieving good reliability requires a sufficient difference between Threshold and Write Threshold. A high write availability, similarly requires a sufficient difference between Width and Write Threshold. Finally, for the system to provide good storage efficiency requires the ratio between IDA Width and Threshold to be as small as possible while meeting the previous reliability and availability goals.

Previous design considerations used one pillar per Slicestor and had a few impacts:

- Initial deployments were expensive which precludes some customers with small data storage requirements from using dispersed storage.

- It set a high lower-bound on the amount of usable storage capacity the system provides and expands by. This made it difficult to support sub-PB systems efficiently.

- It forces some customers with small numbers of Slicestor devices to compromise with either less reliable, or less efficient IDA configurations.

Concentrated Dispersal enables systems on the order of a few hundred terabytes, while retaining the superb reliability and storage efficiency properties that previously could only be achieved in much larger systems comprised of dozens of devices.

When the number of Slicestor appliances is an integer divisor of the IDA Width all 18 pillars still exist, but each Slicestor appliance is now responsible for more than a single pillar's worth of data. Slicestor appliances maintain an ordered namespace assignment across all the disks, which together with the safety margin, prevents the case where one disk holds more than one slice of a given object. This can only happen if a sufficient number (half or more) of drives fail within a Slicestor appliance without replacement.

Figure 7 Concentrated Dispersal Stripe-Pillar Relationship for IDA 18 used in this Design CVD
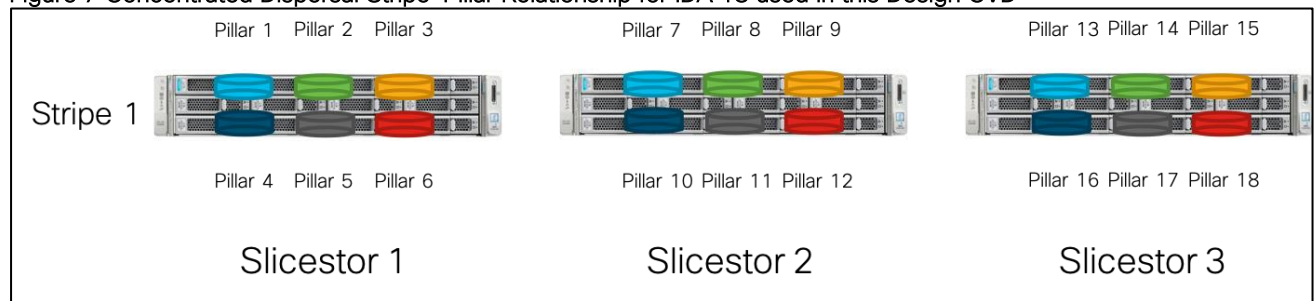
Use Cases

Concentrated Dispersal is ideally suited for the case where the system size is relatively small and is expected to remain small for the near term. There is no setting to enable or disable this feature. Instead it is activated whenever a Device Set is created which has 3 - 7 Slicestor devices. This feature should be used when the required usable capacity is on the order of 20 hard drives worth of usable storage capacity. Common reasons for deploying this quantity of Slicestor devices include:

- When the planned path of expanding the system needs to remain small (several Slicestor devices at a time)

- When seeking to minimize cost with an entry-level system

Reasons to reconsider using the Concentrated Dispersal feature:

- When it is necessary to use a feature that Concentrated Dispersal is not compatible with like Container mode or partially populated Slicestors

- When non-standard IDA Configurations must be explicitly configured like less pillars than drives

- When maintenance operations cannot readily replace failed hard drives

- When high performance is a primary concern

- When achieving the highest possible availability is a paramount concern

Cloud Object Storage Components

You can use the IBM COS System platform to create storage systems that have three software components: the IBM COS Manager software, IBM COS Accesser software and IBM COS Slicestor software.

The software components can be deployed on a wide range of compatible industry-standard hardware platforms, as virtual machines, and in the case of the IBM COS Accesser software, as a software application that is running on a Linux operating system. Physical and virtual deployment can be combined in a single system by using virtual machine deployment for the IBM COS Manager and IBM COS Accesser software and physical servers for the IBM COS Slicestor software as an example.

Each of the three software components serves a specific function:

- The IBM COS Manager software is responsible for monitoring the health and performance of the system, configuring the system and provisioning storage, managing faults, and other administrative and operational functions.

- The IBM COS Accesser software is responsible for encrypting/encoding data on ingest and decoding/decrypting it when read and managing the dispersal of slices of data resulting from this process across a set of IBM COS Slicestor nodes.

- The IBM COS Slicestor software is responsible for the storage of slices.

The underlying storage pool of a dispersed or concentrated dispersed storage system can be shared and is jointly accessible by multiple access protocols.

Object-based Access Methods

The Simple Object interface is accessed with a HTTP/REST API. Simple PUT, GET, DELETE, and LIST commands allow applications to access digital content, and the resulting object ID is stored directly within the application. With the IBM COS Accesser application, no IBM COS Accesser appliance is needed since the application server can talk directly to IBM COS Slicestor storage nodes.

REST API Access to Storage

Figure 8 REST API Storage Interfaces
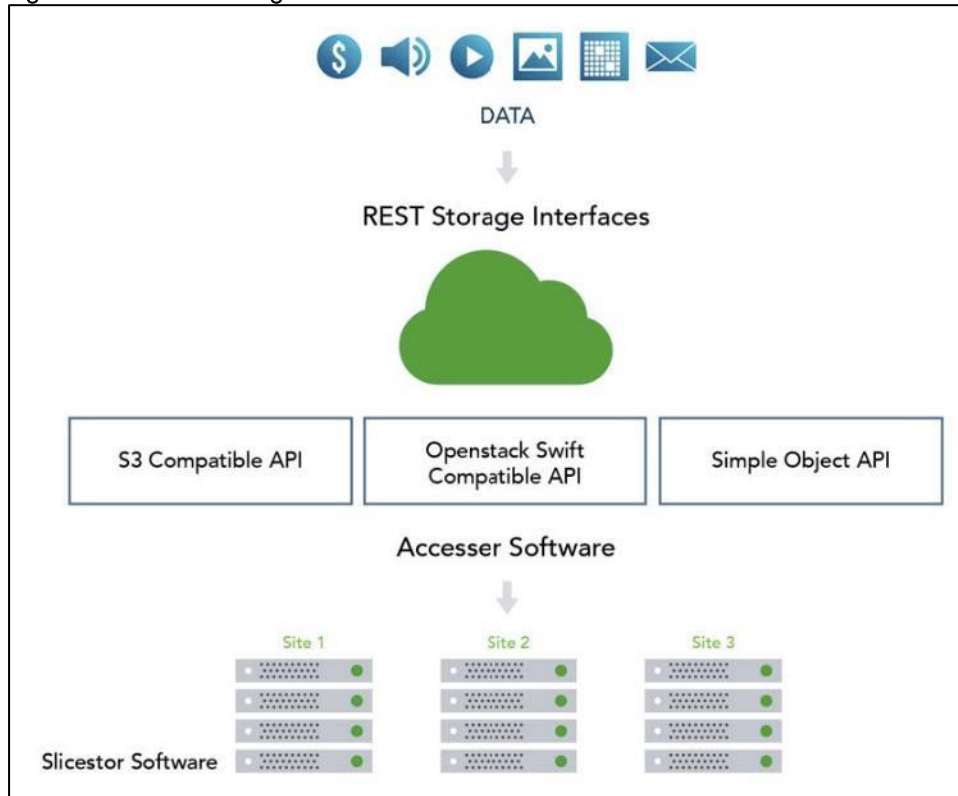
REST is a style of software architecture for distributed hypermedia information retrieval systems such as the World Wide Web. REST-style architectures consist of clients and servers. Clients initiate requests to servers. Servers process requests and return associated responses. Requests and responses are built around the transfer of various representations of the resources.

The REST API works in way that is similar to retrieving a Universal Resource Locator (URL). But instead of requesting a webpage, the application is referencing an object.

REST API access to storage offers several advantages:

- Tolerates internet latency

- Provides for "programmable" storage

- Provides efficient global access to large amounts of data

Embedded Accesser

This CVD uses an Embedded Accesser Appliance function. The Embedded Accesser Appliance feature provides Accesser Appliance functions on the IBM COS Slicestor Appliance. This feature provides customers an opportunity to save on capital expenses by using one physical appliance for both Accesser and Slicestor appliance functions. However, before you deploy this feature, careful consideration needs to be given to the Slicestor hardware and the workload presented to the servers and the load balancing between the available Slicestor appliances.
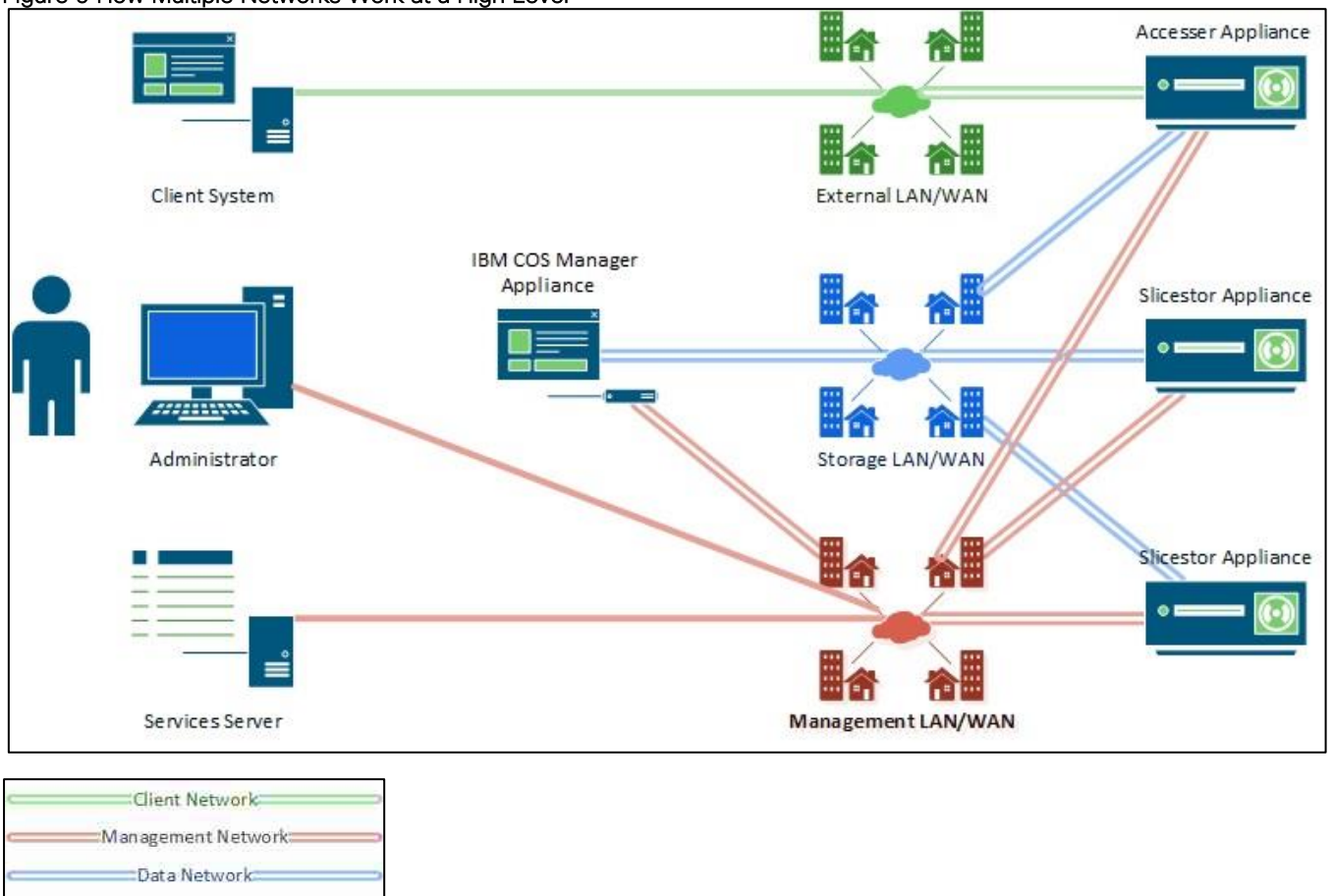
Network

Network administrators can configure the first four layers of the OSI model on a system to use separate network traffic between storage data, management information, and client data.

An IBM COS System that uses certified devices, can dedicate network interfaces (NICs) to three distinct networks to transfer:

- Data within the system

- Management information to management systems

- Data to a client application

These networks are referred to as channels.

Figure 9 How Multiple Networks Work at a High Level

In separating data into channels, the system provides better security, more flexible management options and minimizes network congestion for high-performance applications.

IBM Hardware Requirements

Hardware with a minimum of 10 GbE interconnect and a RAM capacity of 256 GB is recommended for a full-scale deployment of an IBM COS System with Embedded Accesser Appliance functions.

For Concentrated Dispersal Mode the following requirements are important:

- Slicestor requirements

  – Slicestor appliances in a Concentrated Dispersal system must have at least 12 Drives

  – Slicestor appliances in a Concentrated Dispersal system must have at least 32 GB of memory

  – Slicestor appliances in a Concentrated Dispersal system using embedded Accessers must have at least 128 GB of memory

- Device set requirements

  – Device sets must have between 3 and 6 Slicestor devices

  – System expansion requires the new set to equal the Concentrated Dispersal set size or an integer multiple of the full IDA width

- Device evacuation requirements

  – Device evacuation requires destination Slicestor appliances to have an equal or greater number of drives

- Mirroring requirements

  – Vault mirrors must use the same vault optimization on both sides of the mirror

  – Vault mirrors must use Concentrated Dispersal Device Set sizes (3 - 6 devices) on both sides of the mirror

Solution Design

Solution Overview

This Cisco Validated Design provides a comprehensive, end-to-end guide for deploying IBM COS in Concentrated Dispersal Mode with Embedded Accesser on Cisco UCS C240 within infrastructure made possible by Cisco UCS Manager and the Cisco UCS 6332 Fabric Interconnects.

One of the key design goals of this scale out architecture was to deploy all elements on 40GbE networking end to end within a single Cisco UCS domain and start small with CD mode for IBM COS. Both IBM COS components – Embedded Accesser, and Slicestor – utilize the robust throughput and low latency only provided by the Cisco UCS 6332 Fabric Interconnect. Additionally, both components take advantage of the flexibility provided by the stateless nature of Cisco UCS service profiles and service profile templates.

This design uses the Cisco Nexus 9000 series data center switches in NX-OS standalone mode but provides investment protection to migrate to ACI or higher network bandwidths (1/10/25/40/50/100Gbps) while enabling innovative analytics and visibility using Tetration and automation that support in-box and off-box Python scripting and Open NX-OS that support dev-ops tools (Chef, Puppet, Ansible).

The key design for IBM COS with CD mode on Cisco UCS C240 is shown in Figure 10.

Figure 10   Topology of IBM COS with CD Mode on Cisco UCS C240 M5L
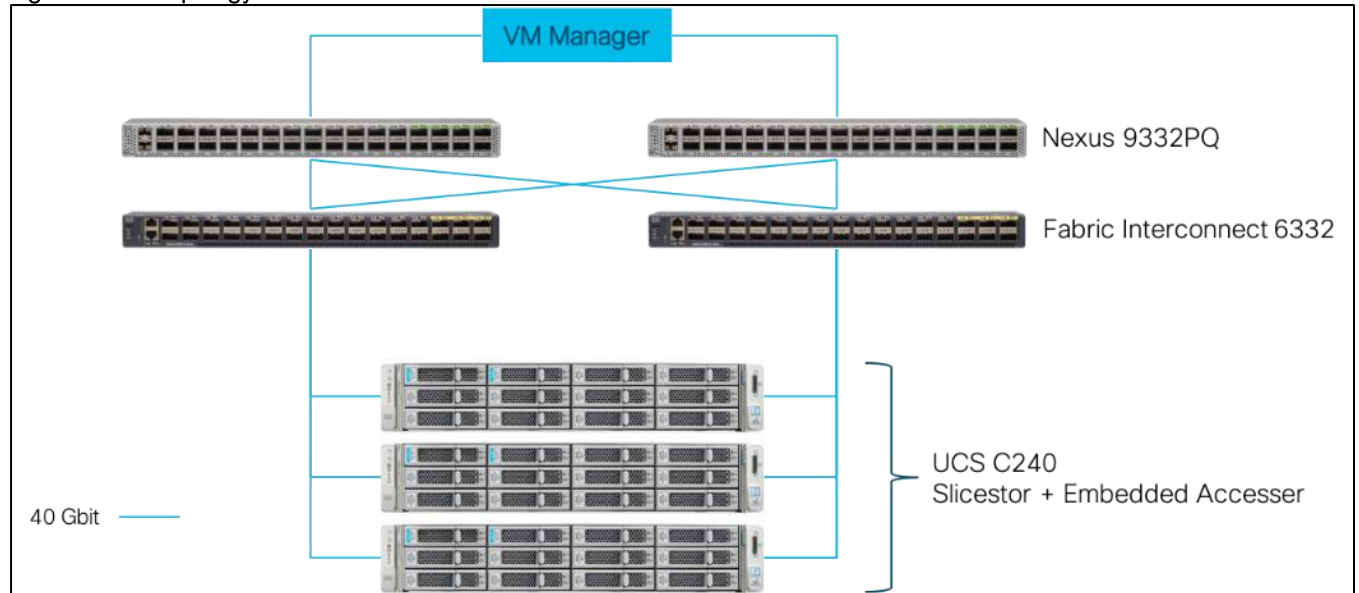
- Manager instance deployed as virtual machine OVA

- Embedded Accesser deployed on Slicestor

- Slicestor deployed on Cisco UCS C240

- Cisco UCS C240 connected to UCS 6332 Fabric Interconnect with 40Gbps line speed

- Cisco UCS 6332 Fabric Interconnect connected to Nexus 9332PQ with 40Gbps line speed

IBM Cloud Object Configuration for Cisco Validated Design

The current Design Guide uses the following configuration for IBM COS:

- Virtual Manager using an OVA

- Embedded Accesser

- Concentrated Dispersal Mode with IDA 18/9/11

Details for the specific CD Mode are listed in Table 2 .

Table 2   Configuration for CD Mode used in this Design Guide

Sites	IDA	Store Count	Disks per Store	Usable Capacity	Expansion Factor	Availability	Reliability
1	18/9/11	3	36	180 TB	2.00	5 nines	15+ nines

General Hardware Requirements

Table 3   List of Components

Component	Model	Quantity	Comments
IBM Slicestor/Accesser	Cisco UCS C240	3	Per server node: • 2 x Intel Skylake 4xxx/5xxx • 128-256 GB Memory • 1 x VIC 1380 • 12 Gbit SAS RAID Controller • Disks ○ 2 x SSD/HDD RAID 1 – Boot ○ 12 x NL-SAS HDD JBOD – Data
IBM Manager	Virtual Machine OVA	1	• 4 vCPU • 16 GB Memory • 128 GB Disk • 1 x Network
Cisco UCS Fabric Interconnects	Cisco UCS 6332 Fabric Interconnects	2	
Switches	Cisco Nexus 9332PQ	2	

Compute Layer Design

Each Cisco UCS C240 rackmount server is equipped with a Cisco UCS Virtual Interface Card (VIC) supporting dual 40-Gbps fabric connectivity. The Cisco UCS VICs eliminate the need for separate physical interface cards on each server for data and management connectivity. For this solution with IBM COS

ClevOS the VIC is configured with two virtual NICs, one on each physical VIC interface. IBM COS is configured to leverage these two vNICs to provide operational active-backup redundancy in software.

Cisco UCS Server Connectivity to Unified Fabric

Cisco UCS servers are typically deployed with a single VIC card for unified network and storage access. The Cisco VIC connects into a redundant unified fabric provided by a pair of Cisco UCS Fabric Interconnects. Fabric Interconnects are an integral part of the Cisco Unified Computing System, providing unified management and connectivity to all attached blades, chassis and rack servers. Fabric Interconnects provide a lossless and deterministic FCoE fabric. For the servers connected to it, the Fabric Interconnects provide LAN, SAN and management connectivity to the rest of the network.

Validated Compute Design

The connectivity of the solution is based on 40 Gbps. All components are connected together via 40 Gbips QSFP cables except the virtual Manager node, which uses a 10 Gbit connectivity. Between both Cisco Nexus 9332PQ switches are 2 x 40 Gbit cabling. Each Cisco UCS 6332 Fabric Interconnect is connected via 2 x 40 Gbps to each Cisco UCS 9332PQ switch. And each Cisco UCS C240 M5L rack server is connected with a single 40 Gbit cable to each Fabric Interconnect.

The exact cabling for the IBM COS solution is illustrated in following picture. It shows also the vNIC configuration for the Data, Client and Management channel.

The virtual IBM COS Management node is connected to both Nexus N9k-A and N9k-B and has access to the Slicestors/Accessers.

Figure 11   IBM COS Cabling Diagram
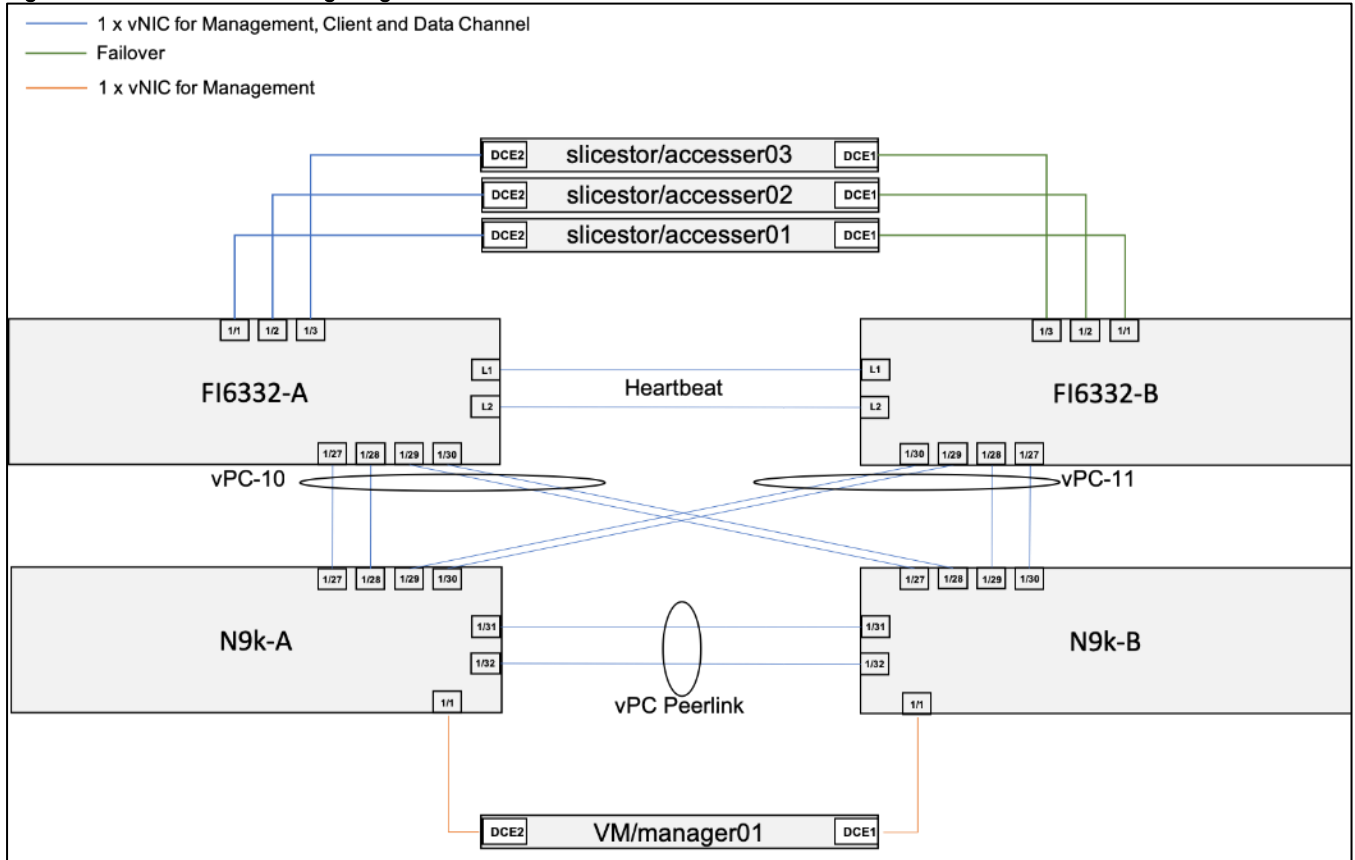
For a better reading and overview, the exact physical connectivity between the Cisco UCS 6332 Fabric Interconnects and the Cisco UCS C-Class server is listed in Table 4 .

Table 4   Physical Connectivity between FI 6332 and Cisco UCS C240 M5L

Port	Role	FI6332-A	FI6332-B
Eth1/1	Server	slicestor/accesser01, DCE2	slicestor/accesser01, DCE1
Eth1/2	Server	slicestor/accesser02, DCE2	slicestor/accesser02, DCE1
Eth1/3	Server	slicestor/accesser03, DCE2	slicestor/accesser03, DCE1
Eth1/27	Network	N9k-A, Eth1/27	N9k-B, Eth1/30
Eth1/28	Network	N9k-A, Eth1/28	N9k-B, Eth1/29
Eth1/29	Network	N9k-B, Eth1/27	N9k-A, Eth1/30
Eth1/30	Network	N9k-B, Eth1/28	N9k-B, Eth1/29

High Availability

The Cisco and IBM solution was designed for maximum availability of the complete infrastructure (compute, network, storage) with no single points of failure.

Compute

- Cisco UCS system provides redundancy at the component and link level and end-to-end path redundancy to the LAN network.

- Cisco UCS C240 rack server is highly redundant with redundant power supplies and fans.

- Each server is deployed using vNICs that provide redundant connectivity to the unified fabric. NIC failover is enabled between Cisco UCS Fabric Interconnects using Cisco UCS Manager. This is done for all Slicestor with Embedded Accesser node vNICs.

Network

- Link aggregation using port channels and virtual port channels can be used throughout the design for higher bandwidth and availability, if the optional Cisco UCS Nexus 9332 is deployed. Between each Cisco UCS 6332 Fabric Interconnect and both Cisco Nexus 9332PQ is one virtual Port Channel (vPC) configured. vPCs allow links that are physically connected to two different Cisco Nexus 9000 switches to appear to the Fabric Interconnect as coming from a single device and as part of a single port channel.

- Each Slicestor with Embedded Accesser is configured in mode 1 active-backup bonding mode at the ClevOS software layer.

Figure 12 illustrates the logical configuration of the network for the IBM COS solution with Embedded Accesser and CD mode.

Figure 12   Logical View of the Network Configuration used in the CVD
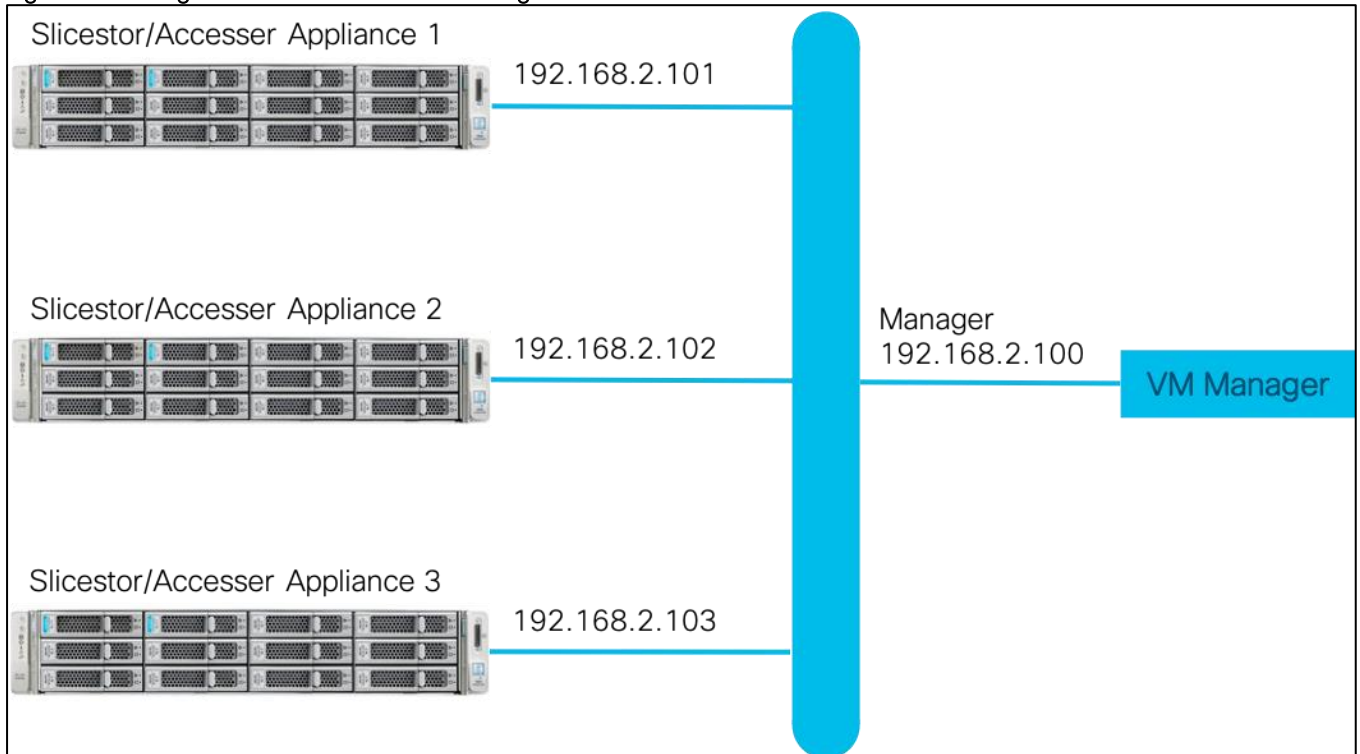
QoS and Jumbo Frames

Cisco UCS, Cisco Nexus, and IBM COS nodes in this solution provide QoS policies and features for handling congestion and traffic spikes. The network-based QoS capabilities in these components can alleviate and provide the priority that the different traffic types require.

This design also recommends end-to-end jumbo frames with an MTU of 9000 Bytes across the LAN and Unified Fabric links. Jumbo frames increase the throughput between devices by enabling larger sized frames to be sent and received on the wire while reducing the CPU resources necessary to process them. Jumbo

frames were enabled during validation on the LAN network links in the Cisco Nexus switching layer and on the Unified Fabric links.

Software Distributions and Versions

The required software distribution versions are listed below in Table 5 .

Table 5   Software Versions

Layer	Component	Version or Release
Cisco UCS C240	Adapter	4.3(1b)
	BIOS	C240M5.4.0.1c
	Board Controller	38.0
	CIMC Controller	4.0(1a)
	Storage Controller SAS 2	50.1.0-1456
Network 6332 Fabric Interconnect	UCS Manager	4.0(1a)
	Kernel	5.0(3)N2(4.01a)
	System	5.0(3)N2(4.01a)
Network Nexus 9332PQ	BIOS	07.59
	NXOS	7.0(3)I5(1)
Software	IBM COS	3.3.16

Summary

Object storage is an increasingly popular form of distributing data in a scale-out system. The entry size sinks to more and more smaller units. IBM with IBM COS is leading the pack with the new Concentrated Dispersal Mode technology when it comes to storing data as an object with high availability and reliability on small entry-level solutions.

The entry-level solution in this design guide provides customers and partners with everything necessary to store object data easily and securely. Cisco's leading technology of centralized management and advanced networking technology helps to easily deploy, manage and operate the IBM COS solution with Embedded Accessor in Concentrated Dispersal Mode.

Cisco and IBM are a perfect combination when it comes to reliably implementing new solutions in the area of object storage and showing customers new ways.

About the Authors

Oliver Walsdorf, Technical Marketing Engineer for Software Defined Storage, Computer Systems Product Group, Cisco Systems, Inc.

Oliver has more than 20 years of storage experience, working in different roles at different storage vendors, and is now an expert for software-defined storage at Cisco. For the past four years Oliver was focused on developing storage solutions at Cisco. He now works on IBM COS, develops Co-Solutions with IBM for the overall storage market and published several Cisco documents. With his focus on SDS he drives the overall attention in the market for new technologies. In his leisure time, Oliver enjoys hiking with his dog and motorcycling.

Acknowledgements

For their support and contribution to the design, validation, and creation of this Cisco Validated Design, we would like to acknowledge the significant contribution and expertise that resulted in developing this document:

- **Chris O'Brien**, Cisco Systems, Inc.
- Jawwad Memon, Cisco Systems, Inc.
- Ulrich Kleidon, Cisco Systems, Inc.