Version 2.2

Overview

Information

Design Zone

CISCO

Library View

Sitemap View

Outline View

Alphabetical

Data Center Design Playbook
Big Data

November 2016 | Version 2.2

Executive Summary

Go to CVD sitemap

Visit DC Design Zone for the full and most current set of validated designs:
www.cisco.com/go/dcdesignzone

What are CVDs?

A Cisco Validated Design (CVD) is a specific bundle of products—Cisco products as well as products from our partners—designed to address the business needs of customers. The CVDs are created based on Cisco's observation of market trends and inside knowledge of future directions of Cisco and its partners. As a complete solution, each CVD consists of both hardware and software, and optimizes the capabilities of the Cisco components to maximize speed, performance, stability, and reliability of the CVD bundle.

Most importantly, each CVD has been extensively tested, validated, and documented. The CVDs were designed to facilitate faster, more reliable, and more predictable customer deployments and are guaranteed to work as specified.

The initial sections of this playbook are educational. You can browse through them to get up to speed on the market forces driving Cisco's United Computing System (UCS) value proposition and competitive differentiators, and to learn about the benefits of using CVD solutions.

What a CVD does:

- Provides a reference design for a system architecture (A reference design is a general design that applies to 60 to 80 percent of enterprise customers.)
- Documents a deployable system / architecture
- Provides design recommendations that compare different design options and trade-offs
- Includes the most common product scalability and performance recommendations
- Sets the "rules" for deploying a system with confidence

What a CVD does not do:

- Document a specific customer implementation
- Advocate a technology / system that is too early for deployment or one with major gaps
- Define how every Cisco product will perform in the system
- Provide product-specific performance limits (a.k.a. "drag strip" numbers)
- Act as a comprehensive troubleshooting guide

Version 2.2

Overview

Information

Design Zone

CISCO

Library View

Sitemap View

Outline View

Alphabetical

What's in it for customers?

Two words: *minimized risk.* There is always risk in any large-scale IT initiative, especially one that involves switching out hardware and networking equipment while implementing, migrating, or upgrading mission-critical applications.

There are two types of risk: *integration risk*, risk that products won't work together, and *performance risk*, risk that they won't perform as promised.

Using a CVD minimizes both these risks. CVDs are not simply an ad hoc group of products. Each CVD forms an integrated, tested, and documented solution. Like the reference architectures produced by other vendors, CVDs incorporate best-of-breed combinations of equipment and software to optimize the value of a configuration for a customer. But unlike reference architectures, CVDs are validated and supported. In most cases, multiple technical teams, from both Cisco and its partners, have put each CVD configuration through rigorous testing to ensure that all products work together to enable rapid deployment.

By implementing the solution presented in a CVD, you are guaranteed to have a successful deployment for the specific workload covered in the CVD. Of course, each specific workload may vary.

Cisco UCS Integrated Infrastructure Solutions

5-Year ROI

483%

5-Year Total Business Benefit per Organization

$13M

Break-Even Period

7 months

Key Performance Improvements Realized from Customers Who Deployed Cisco UCS Integrated Infrastructure

Lower IT Infrastructure Costs

46%

Reduction in Unplanned Downtime

89%

Reduction of Staff Time "Keeping the Lights on"

38%

Business Value Benefits – Average Annual Benefits per Organization over Five Years

Business Productivity	User Productivity	IT Staff Productivity	IT Infrastructure Cost Reduction
$1.44M	$1M	$1.37M	$0.15M

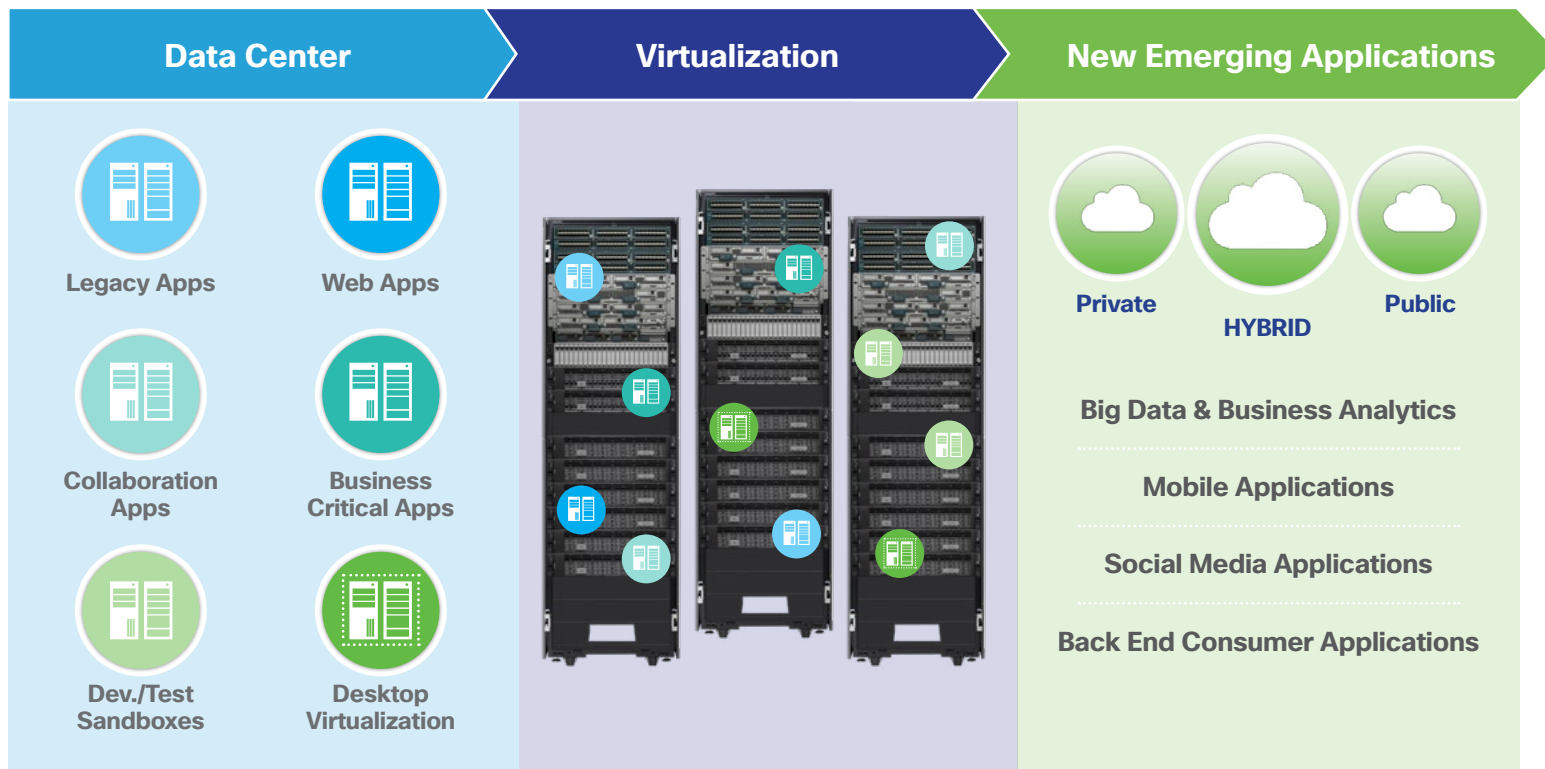Document # #US41066816

IDC
Analyze the Future

Building Momentum: UCS Value Proposition

In today's data center, a broad range of applications need to coexist: legacy, web, collaboration, and business-critical applications, as well as test / development sandboxes and desktop virtualization. All these applications are being virtualized, saving both operating expenses (OPEX) and capital expenditures (CAPEX). In addition, emerging applications need to be supported: Big Data and business analytics, and mobile, social media, and back-end consumer applications.

Data Center

- Legacy Apps
- Web Apps
- Collaboration Apps
- Business Critical Apps
- Dev./Test Sandboxes
- Desktop Virtualization

Virtualization

New Emerging Applications

Private — HYBRID — Public

Big Data & Business Analytics

Mobile Applications

Social Media Applications

Back End Consumer Applications

Version 2.2

Overview

Information

Design Zone

CISCO

Library View

Sitemap View

Outline View

Alphabetical

Benefits of UCS CVD Solutions

Here are just some of the reasons you should use CVDs:

Demonstrate best practices for deploying Cisco products

These tested solutions use all the best practices Cisco has internally identified for its products, as well the best practices recommended by Cisco partners. By following the instructions in a CVD design, customers can set performance expectations when they deploy their desired solution.

Provide much greater efficiency in deployment

Because CVDs provide everything–from designs to configuration instructions to bill of materials (BOMs)–enterprises require a much shorter time to deploy their solutions. Everything is clearly and precisely laid out; there are no surprises during the design phase. And because CVD solutions are guaranteed to work, Cisco provides 24/7 support for any issues that might arise.

Demonstrate Cisco's credibility and commitment to solutions

Cisco UCS CVDs are evidence that Cisco's recommended solutions are credible. They are proof of

Cisco's commitment to these solutions and provide a viable path for enterprises that depend on the Cisco technology roadmap.

Offer flexibility through scalability options

Most Cisco UCS CVDs offer scalability options that allow customers to scale solutions to meet their specific needs. Customers still receive Cisco's CVD guarantees for performance, even if they choose to adjust the scale of the solution.

Version 2.2

Overview

Information

Design Zone

CISCO

Library View

Sitemap View

Outline View

Alphabetical

Using UCS CVD Solutions to Meet Business Needs

CVDs are innovative, proven solutions to help address the business needs of customers.

For implementations of Cisco solutions as infrastructure, CVDs can help customers realize optimal performance and optimization of Cisco UCS with hardware and software partners.

For implementations of Oracle, SAP, or other enterprise applications, CVDs incorporate Cisco's expertise in both applications and infrastructure to provide detailed guidance for installing a particular application or suite at both the hardware and software levels.

By providing demonstrated solutions along with success stories from other customers, CVDs give enterprises that are Cisco customers—but haven't yet shifted their servers to UCS—a level of confidence in the technology.

Benefits provided by CVDs

Cisco UCS CVDs provide customers with the following benefits

- **Removes risks.** No one likes to be on the bleeding edge. CVDs show that the trail has been blazed. By providing tested solutions, CVDs remove the risk from using Cisco UCS in a broad range of deployment scenarios.
- **Proves Cisco is more than a networking vendor.** Although Cisco is predominantly a network provider, CVDs is proof that Cisco is also a provider of a viable server option, UCS.
- **Demonstrates possibilities.** Cisco has carefully calculated which products are put in each solution bundle based on its understanding of both customers' existing needs and the direction its technology partners are heading. Customers can use the CVDs as a jumping-off point to design solutions that meet their needs.
- **Bridges multiple partners.** If you have a customer account that has already standardized on a particular software or storage vendor, the CVDs show how well Cisco UCS works with partners like Oracle, Red Hat, NetApp, VMware, and others.

Version 2.2

Overview

Information

Design Zone

CISCO

Library View

Sitemap View

Outline View

Alphabetical

Potential shortcomings

For all the benefits that CVDs provide, customers should be aware of a few potential shortcomings:

- **CVDs may not always be completely up-to-date.** The Cisco CVD teams can only work so fast,and thus, there is always newer code available that has not been validated. It's important to understand the difference between what is supported and what has been validated.

- **CVDs probably won't precisely match a customer's needs.** The chances that a CVD will meet a customer's complete requirements are slim. Instead, CVDs can be used as a jumping-off point to discuss designs that do fully match the customer's needs.
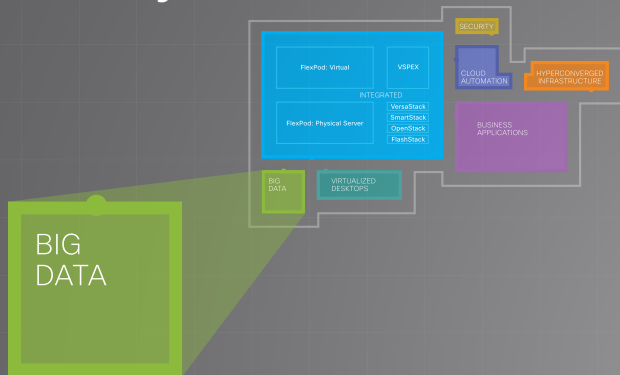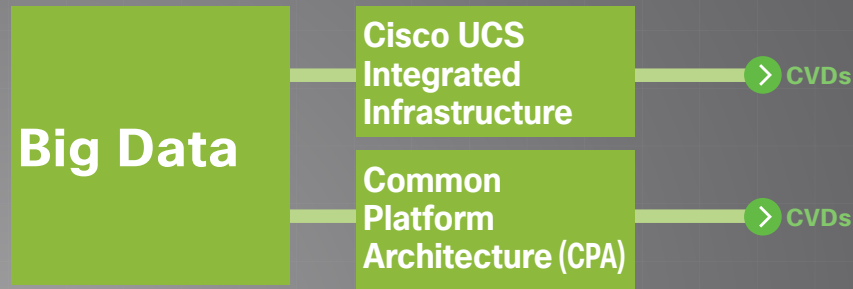
Let's Get Started

Welcome to the DC Design library. Click on a section below or the icons in the top header to see different tours of the CVDs.

Library View

BIG DATA

SECURITY
FlexPod: Virtual | VSPEX
CLOUD AUTOMATION | HYPERCONVERGED INFRASTRUCTURE
INTEGRATED
VersaStack
SmartStack
OpenStack
FlashStack
FlexPod: Physical Server
BUSINESS APPLICATIONS
BIG DATA | VIRTUALIZED DESKTOPS

Sitemap View

Big Data

Cisco UCS Integrated Infrastructure → CVDs

Common Platform Architecture (CPA) → CVDs

Outline View

BIG DATA

A. Cisco UCS Integrated Infrastructure
Cisco UCS Petabyte-Scale Solution for Splunk Enterprise - *New*
Cisco UCS S3260 Storage Server with Cloudera Enterprises - *New*
Cisco UCS S3260 Storage Server with Hortonworks Data Platform - *New*
Cisco UCS S3260 Storage Server with MapR Converged Data Platform - *New*
Cisco UCS Integrated Infrastructure for Big Data and Analytics with Cloudera for Real-time Analytics
Cisco UCS Integrated Infrastructure for Big Data with SAP HANA Vora
Cisco UCS Integrated Infrastructure for SAP HANA
Hadoop as a Service on BareMetal with UCS Director Express (UCSDE) for Big Data on Cisco UCS
   Integrated Infrastructure for Big Data and Cisco ACI
Cisco UCS Integrated Infrastructure with IBM BigInsights for Apach Hadoop
Cisco UCS Integrated Infrastructure with ACI with Cloudera
Cisco UCS Integrated Infrastructure with Hortonworks
Cisco UCS Integrated Infrastructure with MapR
Cisco UCS Integrated Infrastructure with Splunk Enterprise
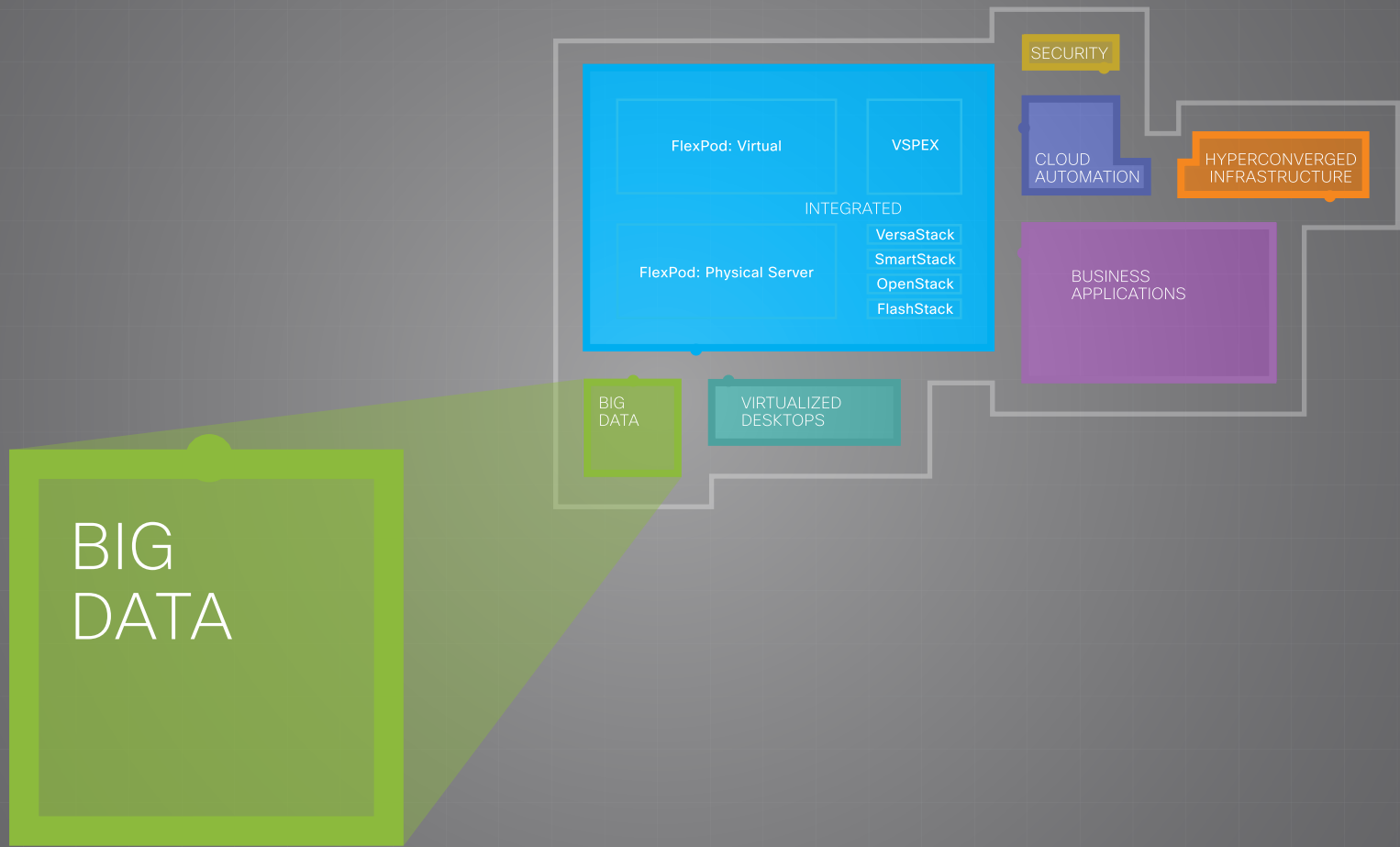
B. Common Platform Architecture (CPA)
Big Data 60 node Hadoop Cluster with EMC Isilon
Big Data Cisco ACI with Cloudera
Cisco UCS CPAv2 for Big Data with Cloudera
Cisco UCS CPAv2 for Big Data with Hortonworks
HaaS with Cisco UCS CPAv2 for Big Data and OpenStack
Cisco UCS CPAv2 for Big Data with Intel Distribution
Cisco UCS CPAv2 for Big Data with Pivotal HD and HAWQ

Alphabetical View

B
Big Data 60 node Hadoop Cluster with EMC Isilon
Big Data Cisco ACI with Cloudera

C
Cisco UCS CPAv2 for Big Data with Cloudera
Cisco UCS CPAv2 for Big Data with Hortonworks
Cisco UCS CPAv2 for Big Data with Intel Distribution
Cisco UCS CPAv2 for Big Data with Pivotal HD and HAWQ
Cisco UCS Integrated Infrastructure for Big Data and Analytics with Cloudera for Real-time Analytics
Cisco UCS Integrated Infrastructure for Big Data with IBM BigInsights for Apach Hadoop
Cisco UCS Integrated Infrastructure for Big Data with SAP HANA Vora
Cisco UCS Integrated Infrastructure for SAP HANA
Cisco UCS Integrated Infrastructure with ACI with Cloudera
Cisco UCS Integrated Infrastructure with Hortonworks
Cisco UCS Integrated Infrastructure with MapR
Cisco UCS Integrated Infrastructure with Splunk Enterprise
Cisco UCS Petabyte-Scale Solution for Splunk Enterprise - *New*
Cisco UCS S3260 Storage Server with Cloudera Enterprises - *New*
Cisco UCS S3260 Storage Server with Hortonworks Data Platform - *New*
Cisco UCS S3260 Storage Server with MapR Converged Data Platform - *New*

H
HaaS with Cisco UCS CPAv2 for Big Data and OpenStack
Hadoop as a Service on BareMetal with UCS Director Express (UCSDE) for Big Data on Cisco UCS
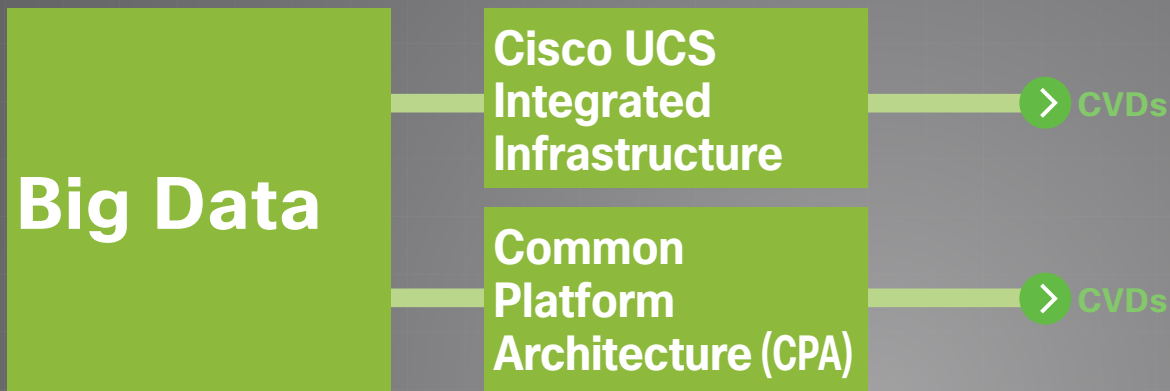   Integrated Infrastructure for Big Data and Cisco ACI

Visit DC Design Zone for the full and most current set of validated designs: **www.cisco.com/go/dcdesignzone**

Version 2.2

Overview

Information

Design Zone

CISCO

Library View

Sitemap View

Outline View

Alphabetical

Data Center Design Library

SECURITY

FlexPod: Virtual

VSPEX

INTEGRATED

VersaStack
SmartStack
OpenStack
FlashStack

FlexPod: Physical Server

CLOUD
AUTOMATION

HYPERCONVERGED
INFRASTRUCTURE

BUSINESS
APPLICATIONS

BIG
DATA

VIRTUALIZED
DESKTOPS

BIG
DATA

Version 2.2

Overview

Information

Design Zone

CISCO

Library View

Sitemap View

Outline View

Alphabetical

DC Design Sitemap

Big Data

Cisco UCS Integrated Infrastructure

> **CVDs**

Common Platform Architecture (CPA)

> **CVDs**

CVD Outline

BIG DATA

A. Cisco UCS Integrated Infrastructure

Cisco UCS Petabyte-Scale Solution for Splunk Enterprise - *New*

Cisco UCS S3260 Storage Server with Cloudera Enterprises - *New*

Cisco UCS S3260 Storage Server with Hortonworks Data Platform - *New*

Cisco UCS S3260 Storage Server with MapR Converged Data Platform - *New*

Cisco UCS Integrated Infrastructure for Big Data and Analytics with Cloudera for Real-time Analytics

Cisco UCS Integrated Infrastructure for Big Data with SAP HANA Vora

Cisco UCS Integrated Infrastructure for SAP HANA

Hadoop as a Service on BareMetal with UCS Director Express (UCSDE) for Big Data on Cisco UCS
Integrated Infrastructure for Big Data and Cisco ACI

Cisco UCS Integrated Infrastructure with IBM BigInsights for Apach Hadoop

Cisco UCS Integrated Infrastructure with ACI with Cloudera

Cisco UCS Integrated Infrastructure with Hortonworks

Cisco UCS Integrated Infrastructure with MapR

Cisco UCS Integrated Infrastructure with Splunk Enterprise

B. Common Platform Architecture (CPA)

Big Data 60 node Hadoop Cluster with EMC Isilon

Big Data Cisco ACI with Cloudera

Cisco UCS CPAv2 for Big Data with Cloudera

Cisco UCS CPAv2 for Big Data with Hortonworks

HaaS with Cisco UCS CPAv2 for Big Data and OpenStack

Cisco UCS CPAv2 for Big Data with Intel Distribution

Cisco UCS CPAv2 for Big Data with Pivotal HD and HAWQ

Version 2.2

Overview

Information

Design Zone

CISCO

Library View

Sitemap View

Outline View

Alphabetical

Alphabetical CVDs

B

Big Data 60 node Hadoop Cluster with EMC Isilon

Big Data Cisco ACI with Cloudera

C

Cisco UCS CPAv2 for Big Data with Cloudera

Cisco UCS CPAv2 for Big Data with Hortonworks

Cisco UCS CPAv2 for Big Data with Intel Distribution

Cisco UCS CPAv2 for Big Data with Pivotal HD and HAWQ

Cisco UCS Integrated Infrastructure for Big Data and Analytics with Cloudera for Real-time Analytics

Cisco UCS Integrated Infrastructure for Big Data with IBM BigInsights for Apach Hadoop

Cisco UCS Integrated Infrastructure for Big Data with SAP HANA Vora

Cisco UCS Integrated Infrastructure for SAP HANA

Cisco UCS Integrated Infrastructure with ACI with Cloudera

Cisco UCS Integrated Infrastructure with Hortonworks

Cisco UCS Integrated Infrastructure with MapR

Cisco UCS Integrated Infrastructure with Splunk Enterprise

Cisco UCS Petabyte-Scale Solution for Splunk Enterprise – *New*

Cisco UCS S3260 Storage Server with Cloudera Enterprises – *New*

Cisco UCS S3260 Storage Server with Hortonworks Data Platform – *New*

Cisco UCS S3260 Storage Server with MapR Converged Data Platform – *New*

H

HaaS with Cisco UCS CPAv2 for Big Data and OpenStack

Hadoop as a Service on BareMetal with UCS Director Express (UCSDE) for Big Data on Cisco UCS
Integrated Infrastructure for Big Data and Cisco ACI

Visit DC Design Zone for the full and most current set of validated designs: **www.cisco.com/go/dcdesignzone**

Version 2.2

Overview

Information

Design Zone

CISCO

Library View

Sitemap View

Outline View

A
Alphabetical

Big Data

Cisco UCS Integrated Infrastructure

Cisco UCS Petabyte-Scale Solution for Splunk Enterprise
- New

Cisco UCS S3260 Storage Server with Cloudera Enterprises
- New

Cisco UCS S3260 Storage Server with Hortonworks Data Platform
- New

Cisco UCS S3260 Storage Server with MapR Converged Data Platform
- New

Cisco UCS Integrated Infrastructure for Big Data and Analytics with Cloudera for Real-time Analytics

Cisco UCS Integrated Infrastructure for Big Data with SAP HANA Vora

Cisco UCS Integrated Infrastructure for SAP HANA

Hadoop as a Service on BareMetal with UCS Diretor Express (UCSDE) for Big Data on Cisco UCS Integrated Infrastructure for Big Data and Cisco ACI

Cisco UCS Integrated Infrastructure for Big Data with IBM BigInsights for Apach Hadoop

Cisco UCS Integrated Infrastructure with Application Centric Infrastructure (ACI) with Cloudera

Cisco UCS Integrated Infrastructure with Hortonworks

Cisco UCS Integrated Infrastructure with MapR

Cisco UCS Integrated Infrastructure with Splunk Enterprise

Version 2.2

Overview

Information

Design Zone

CISCO

Library View

Sitemap View

Outline View

Alphabetical

Big Data

Common Platform Architecture (CPA)

Big Data 60 node
Hadoop Cluster with
EMC Isilon

Big Data Cisco
Application Centric
Infrastructure (ACI)
with Cloudera

Cisco UCS CPAv2
for Big Data with
Cloudera

Cisco UCS CPAv2
for Big Data with
Hortonworks

Hadoop-as-a-Service
(HaaS) with Cisco
UCS CPA v2 for
Big Data and
OpenStack

Cisco UCS CPAv2
for Big Data with
Intel Distribution

Cisco UCS CPAv2
for Big Data with
Pivotal HD and
HAWQ

Version 2.2

Overview

Information

Design Zone

CISCO

Library View

Sitemap View

Outline View

Alphabetical

Big Data

CISCO UCS INTEGRATED INFRASTRUCTURE

- Cisco UCS Petabyte-Scale Solution for Splunk Enterprise *– New*
- Cisco UCS S3260 Storage Server with Cloudera Enterprises *– New*
- Cisco UCS S3260 Storage Server with Hortonworks Data Platform *– New*
- Cisco UCS S3260 Storage Server with MapR Converged Data Platform *– New*
- Cisco UCS Integrated Infrastructure for Big Data and Analytics with Cloudera for Real-time Analytics
- Cisco UCS Integrated Infrastructure for Big Data with SAP HANA Vora
- Cisco UCS Integrated Infrastructure for SAP HANA

- Hadoop as a Service on BareMetal with UCS Director Express (UCSDE) for Big Data on Cisco UCS Integrated Infrastructure for Big Data and Cisco ACI
- Cisco UCS Integrated Infrastructure for Big Data with IBM BigInsights for Apach Hadoop
- Cisco UCS Integrated Infrastructure for Big Data with Application Centric Infrastructure (ACI) and Cloudera
- Cisco UCS Integrated Infrastructure for Big Data with Hortonworks
- Cisco UCS Integrated Infrastructure for Big Data with MapR
- Cisco UCS Integrated Infrastructure for Big Data with Splunk Enterprise

The Big Data market is poised to explode. Industry analysts project that it will exceed $50 billion by 2016, with a whopping compound annual growth rate (CAGR) of 58 percent.[1] These estimates include hardware, software, and services revenue. With the growing awareness of the power of Big Data and the development of new analytics capabilities, organizations seek to use Big Data to gain competitive advances and boost operational efficiencies.[2]

Cisco UCS Petabyte-Scale Solution for Splunk Enterprise

TECHNICAL HIGHLIGHTS

- Peta-byte scale Splunk solution (ITOA, ES and ITSI) on S-series
- Tiered storage comes standard: SSDs for HOT+WARM data and HDDs for COLD & Frozen
- Scales up to 8.9PB in single UCS domain

SUMMARY

- Powered by Cisco UCS S-series Storage Servers, Splunk real-time collection and indexing of machine data, and distributed high capacity and performance architecture

BUSINESS CHALLENGES

- Scale horizontally while continuously delivering exceptional performance
- Highly scalable infrastructure ensuring rapid and predictable delivery of insights
- Optimize application and infrastructure performance

ARCHITECTURE

Network Fabric:
- Two Cisco UCS Fabric Interconnects 6332
- 32 40 Gigabit-Ethernet ports

Search Head Servers:
Three Cisco UCS C220 M4 Servers with:
- 28 cores, 256 GB RAM

Total Storage:
- Hot/Warm Tier: 88-TB
- Cold Tier: 800-TB

A single UCS domain can accomodate:
- 25 indexers
- 275 TB of Hot/Warm storage
- 2.5 PB of Cold storage

Admin Servers:
Two Cisco UCS C220 M4 Servers with:
- Splunk cluster master, deployer
- DMC, License Master, Deployment server

Indexers:
Four Cisco UCS S3260 Storage Servers with:
- Two processing node (per server) with
- 28 cores (total 56 cores)
- 16 SSDs for HOT/WARM, 40HDDs for COLD tiers
- 4 40 Gigabit Ethernet ports

Total Storage:
- Hot/Warm Tier:88-TB
- Cold Tier: 800-TB

Indexing Capacity:
- Up to 2.4 terabytes (TB) per day
- HOT/WARM: 30 days (1 month)
- COLD:300 days (10 months)

CVD: http://www.cisco.com/c/dam/en/us/td/docs/unified_computing/ucs/UCS_CVDs/Cisco_UCS_Petabyte_Scale_Solution_for_Splunk_Enterprise.pdf
Solution: http://www.cisco.com/c/dam/en/us/products/collateral/servers-unified-computing/le-44701-sb-splunk.pdf

Cisco UCS S3260 Storage Server with Cloudera Enterprise

TECHNICAL HIGHLIGHTS

Configuration Details

- 2 Cisco UCS 6332 Fabric Interconnects

8 Cisco UCS S3260 Storage Servers, each with 2 nodes with:

- 2 Intel Xeon processor
- E5-2680 v4 CPUs(14 cores on each CPU)
- 256 GB of memory
- Cisco 12-Gbps SAS Modular RAID Controller with 4-GB FBWC
- 24 x 4-TB 7200-rpm LFF SAS drives(1.54 petabytes [PB] total)
- 2 x 480-GB 6-Gbps 2.5-inch enterprise value SATA SSD drives for boot
- Cisco UCS VIC 1387 (with 2 x 40 Gigabit Ethernet QSFP ports)

** 3 Cisco UCS C240 M4, as master nodes

BUSINESS CHALLENGES

- Bringing flexibility and scalability to dense storage for Big Data
- Flexible modular architecture to handle both high performance and high capacity workloads

SUMMARY

- Flexible Big Data platform powering the enterprise data hub with Cloudera Enterprise

ARCHITECTURE

2 Cisco UCS 6332 Fabric Interconnects

3 Cisco UCS C240 Servers

8 Cisco UCS S3260 Storage Chassis

Cisco UCS S3260 Storage Server with Hortonworks Data Platform

TECHNICAL HIGHLIGHTS

Configuration Details

- 2 Cisco UCS 6332 Fabric Interconnects

8 Cisco UCS S3260 Storage Servers, each with 2 nodes with:

- 2 Intel Xeon processor
- E5-2680 v4 CPUs(14 cores on each CPU)
- 256 GB of memory
- Cisco 12-Gbps SAS Modular RAID Controller with 4-GB FBWC
- 24 x 4-TB 7200-rpm LFF SAS drives(1.54 petabytes [PB] total)
- 2 x 480-GB 6-Gbps 2.5-inch enterprise value SATA SSD drives for boot
- Cisco UCS VIC 1387 (with 2 x 40 Gigabit Ethernet QSFP ports)

** 3 Cisco UCS C240 M4, as master nodes

BUSINESS CHALLENGES

- Bringing flexibility and scalability to dense storage for Big Data
- Flexible modular architecture to handle both high performance and high capacity workloads

SUMMARY

- Enterprise data lake using Cisco UCS S3260 Storage Server with Hortonworks Data Platform

ARCHITECTURE

2 Cisco UCS 6332 Fabric Interconnects

3 Cisco UCS C240 Servers

8 Cisco UCS S3260 Storage Chassis
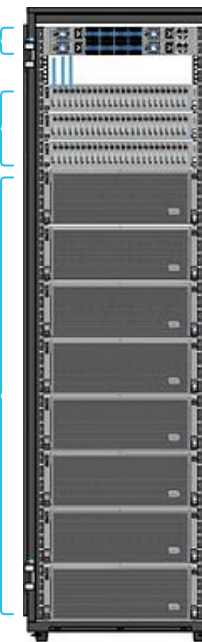
Cisco UCS S3260 Storage Server with MapR Converged Data Platform

TECHNICAL HIGHLIGHTS

Configuration Details

- 2 Cisco UCS 6332 Fabric Interconnects

8 Cisco UCS S3260 Storage Servers, each with 2 nodes with:

- 2 Intel Xeon processor
- E5-2680 v4 CPUs(14 cores on each CPU)
- 256 GB of memory
- Cisco 12-Gbps SAS Modular RAID Controller with 4-GB FBWC
- 24 x 4-TB 7200-rpm LFF SAS drives(1.54 petabytes [PB] total)
- 2 x 480-GB 6-Gbps 2.5-inch enterprise value SATA SSD drives for boot
- Cisco UCS VIC 1387 (with 2 x 40 Gigabit Ethernet QSFP ports)

BUSINESS CHALLENGES

- Bringing flexibility and scalability to dense storage for Big Data
- Flexible modular architecture to handle both high performance and high capacity workloads

SUMMARY

- Comprehensive integrated infrastructure for Big Data with MapR Converged Data Platform

ARCHITECTURE

2 Cisco UCS 6332 Fabric Interconnects

3 Cisco UCS C240 Servers

8 Cisco UCS S3260 Storage Chassis
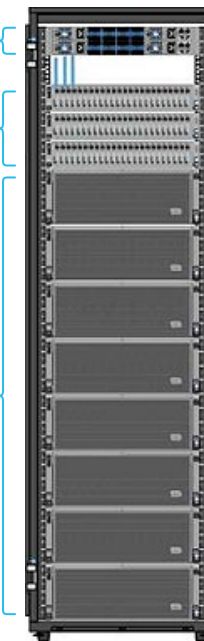
Cisco UCS Integrated Infrastructure for Big Data and Analytics with Cloudera for Real-time Analytics

TECHNICAL HIGHLIGHTS

- Base configuration of 64 nodes with SFF (1.8TB) drives. This also offers HA with 3 management nodes. This solution can be scaled further by adding data nodes, ideally in sets of 16 Cisco UCS C240 M4 servers.

- Up to 80 servers (5 racks) can be supported with no additional switching in a single Cisco UCS domain with no network over-subscription

- Offers scalable/sizable reference architectures for Spark batch processing, and Spark Streaming with Kafka, etc.

BUSINESS CHALLENGES

- Apache Spark brings batch and real-time data processing together to offer deeper and interactive insights to customers

- Spark Ecosystem offers a data-execution engine for all workloads with Spark Streaming, SparkSQL, and Spark Mlib

- Cloudera offers the fastest and easiest secure-data platform for Hadoop, along with Spark, and helps solves a magnitude of business challenges

SUMMARY

- Cisco and Cloudera provide companies with enterprise data management on a unified platform, with high performance, low cost, and help in utilizing data to drive better business insights

- This solution helps organizations exploit the valuable hidden potential of the data, regardless of whether it's structured, semi-structured or unstructured. Cloudera 5.7-is the leading version of enterprise-grade Hadoop infrastructure software and services along with a strong analytics stack

ARCHITECTURE

Cisco UCS with Cloudera and Apache Spark

2 Cisco UCS 6296UP
Fabric Interconnects

Kafka Nodes
8 Cisco UCS C240 M4

16 x 10 Gigabit Links

cloudera

Handoop Clusters
(NameNodes/Resource Manager/Data Nodes/Spark Executors)
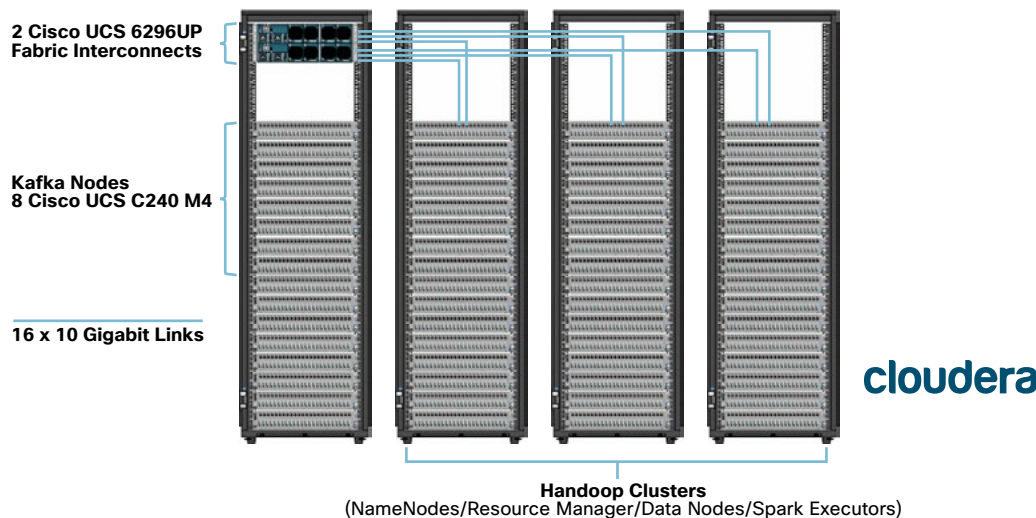
Cisco UCS Integrated Infrastructure for Big Data with SAP HANA Vora

TECHNICAL HIGHLIGHTS

- Cisco Application Centric Infrastructure as the backbone of the solution
- FlexPod datacenter solution for SAP HANA
- Cisco UCS integrated infrastructure for big data – Up to 80 servers per Cisco UCS domain without network over-subscription
- SAP HANA Vora

BUSINESS CHALLENGES

- Data from various sources are received and processed in their own silos
- Need correlation between various data sources: enterprise data, unstructured/semi-structured big data(customer, partner, geo-location, mobile devices, social media), and newer sources like IoT sensors and smart devices
- Big data lives in Hadoop, enterprise data lives in enterprise DB, need to bridge the gap
- Need an infrastructure that can scale-up and scale-out to break down this digital divide

SUMMARY

- Detailed procedure and documentation for building the solution with:
  - Cisco Application Centric Infrastructure (ACI)
  - Cisco UCS Infrastructure
  - SAP HANA Vora on Hortonworks Data Platform
- Works with any SAP HANA datacenter solution implementation that connects to ACI infrastructure

ARCHITECTURE
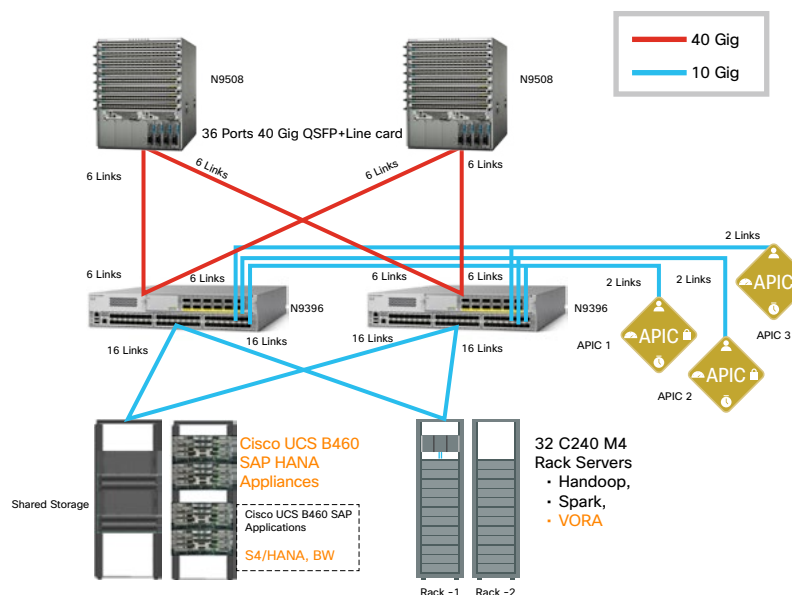
Cisco UCS Integrated Infrastructure for SAP HANA

TECHNICAL HIGHLIGHTS

- Solution is designed with next gen Cisco UCS fabric interconnect with 40Gb end-to-end network
- Persistent storage is configured on UCS C240 C-series servers with MapR Converged Data Platform
- MapR Data Plaform provides distributed, reliable, high performance, scalable, and full read/write NF S-based storage for SAP HANA
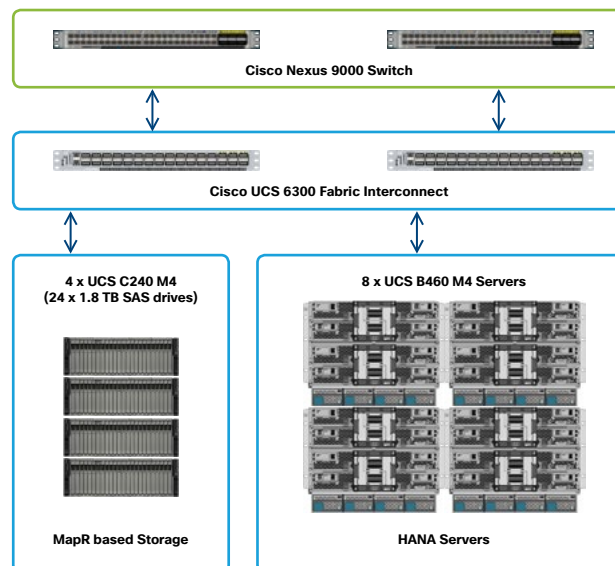
BUSINESS CHALLENGES

- Provide an end-to-end implementation of SAP HANA utilizing the capabilities of the unified infrastructure for compute, network and storage
- Cisco's integrated infrastructure with policy-driven programmable infrastructure enables customers to better fit SAP applications into their data center
- Provide a reliable, flexible and scalable reference design with Cisco only hardware

SUMMARY

- Rapid provisioning of SAP HANA using UCS Service Profile with 40Gb end-to-end network including Storage Network. Single management endpoint for compute and storage elements for SAP HANA
- Cisco UCS servers provides persistent storage with MapR Converged Data Platform, a modern NFS-mountable distributed file-system with enterprise grade storage
- UCS integrated infrastructure provides policy-based models to deliver operational simplicity, comprehensive and consistent performance to run SAP HANA

ARCHITECTURE

Cisco Nexus 9000 Switch

Cisco UCS 6300 Fabric Interconnect

4 x UCS C240 M4 (24 x 1.8 TB SAS drives)

8 x UCS B460 M4 Servers

MapR based Storage

HANA Servers

CVD: http://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/UCS_CVDs/ucsii_saphana_mapr.html

Hadoop as a Service on BareMetal with UCS Director Express (UCSDE) for Big Data on Cisco UCS Integrated Infrastructure for Big Data and Cisco ACI

TECHNICAL HIGHLIGHTS

- The base configuration consists of 1 UCSDE management node and 80 Hadoop nodes with SFF/LFF drives. This solution could be scaled further just by adding nodes ideally in sets of 16 Cisco UCS C240 M4 servers
- Up to 80 servers (5 racks) can be supported with no additional switching in a single Cisco UCS domain with no network over-subscription

BUSINESS CHALLENGES

- Data center's biggest challenge is the provisioning and managing of the large number of Hadoop nodes
- Consolidating multiple-Hadoop clusters in a single, centrally managed physical cluster to improve infrastructure utilization, and provide access controls and security isolation between tenants

SUMMARY

- Cisco UCSDE provides centralized visibility into the complete infrastructure and big data application to identify potential failures and latent threats before they affect application and business performance
- ACI provides centralized visibility to the entire network with deep telemetry and real-time network health status for each tenant
- Cisco UCS Integrated Infrastructure for Big Data with ACI, offers a linearly scalable architecture and simplification of essential operations for single-rack and multiple-rack deployments spanning thousands of servers

ARCHITECTURE

- Multi Tenancy
- Resource Fragmentation
- Single point of management for Network
- Scale-Out Penalty-Free Overlay Network
- Single point of Management for Hadoop

Cisco UCS Integrated Infrastructure for Big Data with IBM BigInsights for Apache Hadoop

TECHNICAL HIGHLIGHTS

- The base configuration consists of 3 management nodes and 16 Data nodes with LFF (6TB) drives. This solution could be scaled further just by adding data nodes ideally in sets of 16 Cisco UCS C240 M4 servers

- Up to 80 servers (5 racks) can be supported with no additional switching in a single Cisco UCS domain with no network over-subscription

BUSINESS CHALLENGES

- Biggest challenge for data scientists and analysts to reuse their experience and tools in Big Data landscape

- IBM BigInsights offers BigSQL a feature rich SQL engine on Hadoop that can deliver analytic capabilities on Hadoop

- IBM BigInsights introduces enterprise capabilities for Hadoop, including machine learning with Big R to help data scientists, analysts and administrators accelerate

SUMMARY

- Cisco and IBM provide enterprises with transparent, simplified data as well as management integration with an enterprise application ecosystem

- This solution helps organizations to exploit the valuable business insights in all their data, regardless of whether it's structured, semi structured or unstructured. Big Insights 3.1 is the leading version of enterprise-grade Hadoop infrastructure software and services along with a strong analytics stack and machine learning capabilities

ARCHITECTURE

2 X Cisco UCS 6296 Fabric Interconnect

3 X Cisco UCS C220 M4 Server

64 X Cisco UCS C240 M4 Servers

IBM
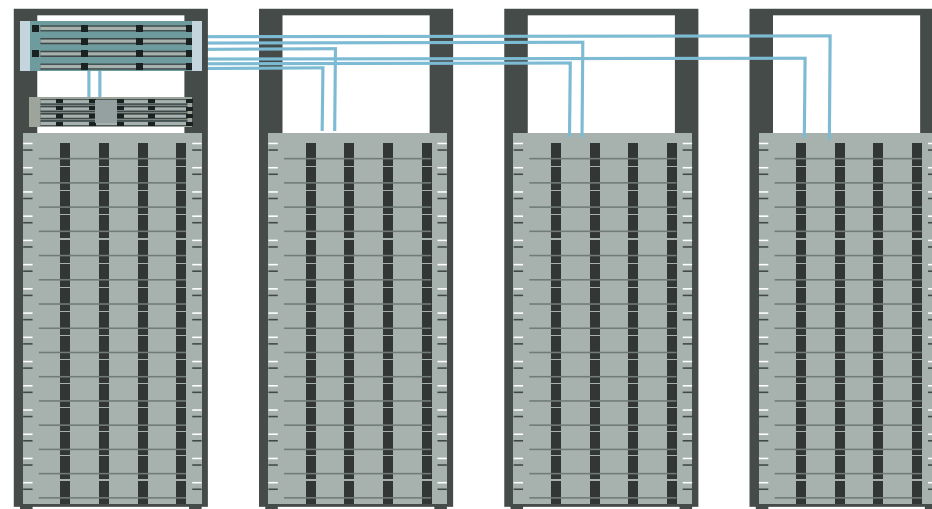
Cisco UCS Integrated Infrastructure and Application Centric Infrastructure (ACI) and Cloudera

TECHNICAL HIGHLIGHTS

- 160 UCS C240 M4 (SFF/LFF) servers with scale-out option up to 5760 servers in a single switching domain
- Two Cisco N9K-C9508
- 23 Cisco N9K-C9396PX
- Three APIC-L1
- Cloudera Hadoop 5.0

SUMMARY

- Scaling to thousands of servers with no over-subscription within a UCS domain, and 5.7.1 over-subscription between domains
- Radically simplifies, optimizes, and accelerates the entire application deployment lifecycle
- The Cisco ACI policy model is designed top down using a promise theory model to control a scalable architecture of defined network and service objects
- Centralized management for servers and network independently
- Best practices for Hadoop 2.0 services tuning

BUSINESS CHALLENGES

- Scalability, performance and support for multi-tenancy that enforces proper isolation and SLA's for workloads of different tenants
- Centralized network visibility with real time application health monitoring
- Simplified automation through application driven policy model
- Best practices for installing Cloudera Distribution for Apache Hadoop 5.0 tuned for performance and scalability

ARCHITECTURE
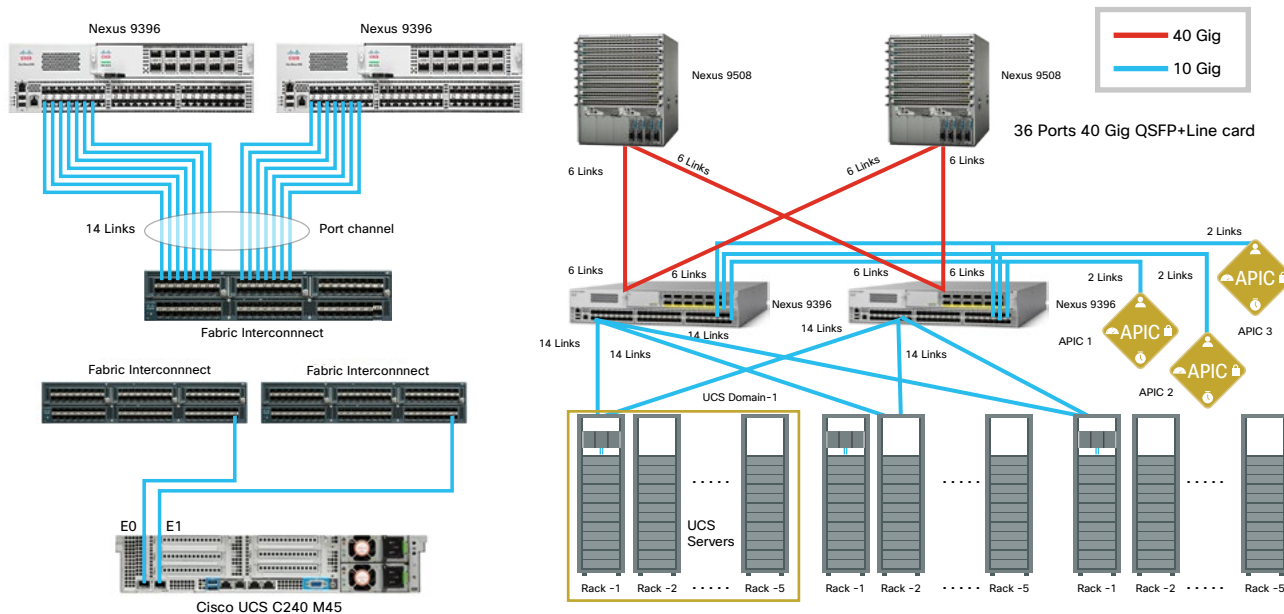
Cisco UCS Integrated Infrastructure with Hortonworks

TECHNICAL HIGHLIGHTS

- 64 UCS C240 M4 (SFF/LFF) with four C3160 servers with scale-out option up to five racks in a single switching domain
- Cisco UCS® Director Express for Big Data offers one-click provisioning, installation, and configuration
- HDP 2.2 includes HDFS tiered-storage support (with related storage policies) to manage tiers of hot, warm, and cold data. Solution focuses on moving cold/archival data to Cisco UCS C3160 through while maintaining hot and warm data in faster Cisco UCS C240 M4 series servers

BUSINESS CHALLENGES

- Data lifecycle management for Hadoop where usage and utility of data can be categorized into hot (recent data), warm (not-so-recent) and cold (archival data)
- Cost-effective solution for the full lifecycle of data, processing it as applications demand
- Enterprise SQL on Hadoop with cost optimizations and interactive queries

SUMMARY

- Committed to 100% open source distribution
- Cisco UCS Integrated Infrastructure for Big Data and Hortonworks Data Platform offer comprehensive set of capabilities for data management, data access, data governance and integration and operations
- Next version of Hadoop HDP 2.2 with enterprise SQL and tiered-storage support

ARCHITECTURE

2 X Cisco UCS 6296 Fabric Interconnect

16 X Cisco UCS C240 M4 Server

Cisco UCS C3160 Server

Hortonworks

CVD: http://www.cisco.com/c/dam/en/us/td/docs/unified_computing/ucs/UCS_CVDs/Cisco_UCS_Integrated_Infrastructure_for_Bigdata_with_Hortonworks.pdf

Cisco UCS Integrated Infrastructure with MapR

TECHNICAL HIGHLIGHTS

- Detailed instructions for installation and fine tuning MapR on a 64 node UCS C240 M4 (SFF/LFF) cluster
- Detailed configuration of multi-tenancy with:
  - Data and compute isolation
  - Enforced quotas
  - High availability
  - Associating tenants with volumes
  - Ability to manage SLAs with data placement and job placement control

BUSINESS CHALLENGES

- Multi-tenancy is the capability of a single instance of software to serve multiple tenants. A tenant is a group of users that have the same view of the system
- Hadoop is an enterprise data hub, and it demands multi-tenancy. Big Data platforms are increasingly expected to support multi-tenancy by default
- Multi-tenancy requires isolation of the distinct tenants: both the data in the data platform and the computing aspect

SUMMARY

- Multi-tenancy with MapR supports data and job placement control to isolate sensitive data and applications
- Ubiquitous high availability with no-NameNode architecture, YARN high availability, and NFS high availability
- Simplified management through Cisco UCSM and MapR Control System (MCS)
- Ease of deployment
- Scalability for Big Data workloads

ARCHITECTURE

2 X Cisco UCS 6296 Fabric Interconnect

16 X Cisco UCS C240 M4 Server
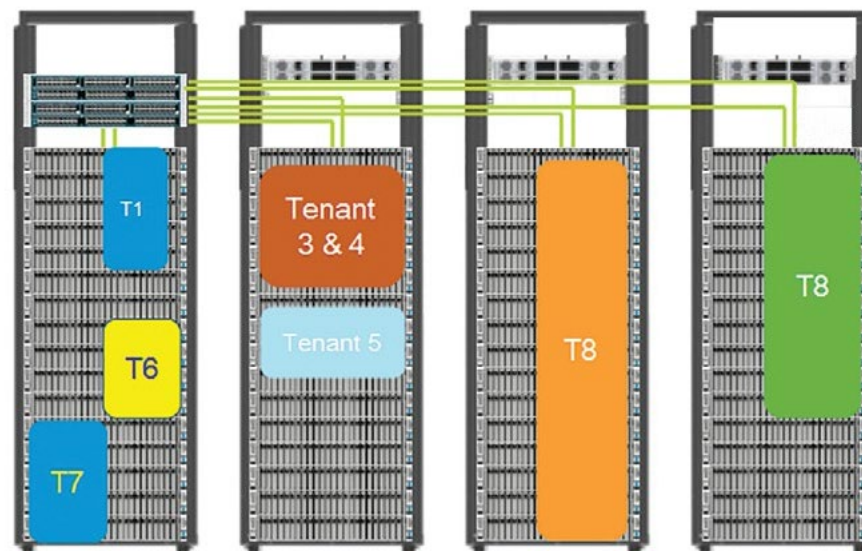
Cisco UCS Integrated Infrastructure with Splunk Enterprise

TECHNICAL HIGHLIGHTS

- 8 Cisco UCS C240 M4 rack servers
- 5 Cisco UCS C220 M4 rack servers
- 1 Cisco UCS C3160 rack server
- 2 Cisco UCS 6296UP fabric interconnects
- Splunk Enterprise 6.2.2 and greater
- Cisco UCS C220 M4 rack servers serve as search heads to form the Splunk search head cluster, providing a highly available analytics interface for the end user
- Cisco UCS C3160 rack server as NFS server to archive frozen data
- Cisco UCS C240 M4 rack servers serve as indexers to form the Splunk indexer cluster, providing industry-leading scalability and reliability mission critical data storage
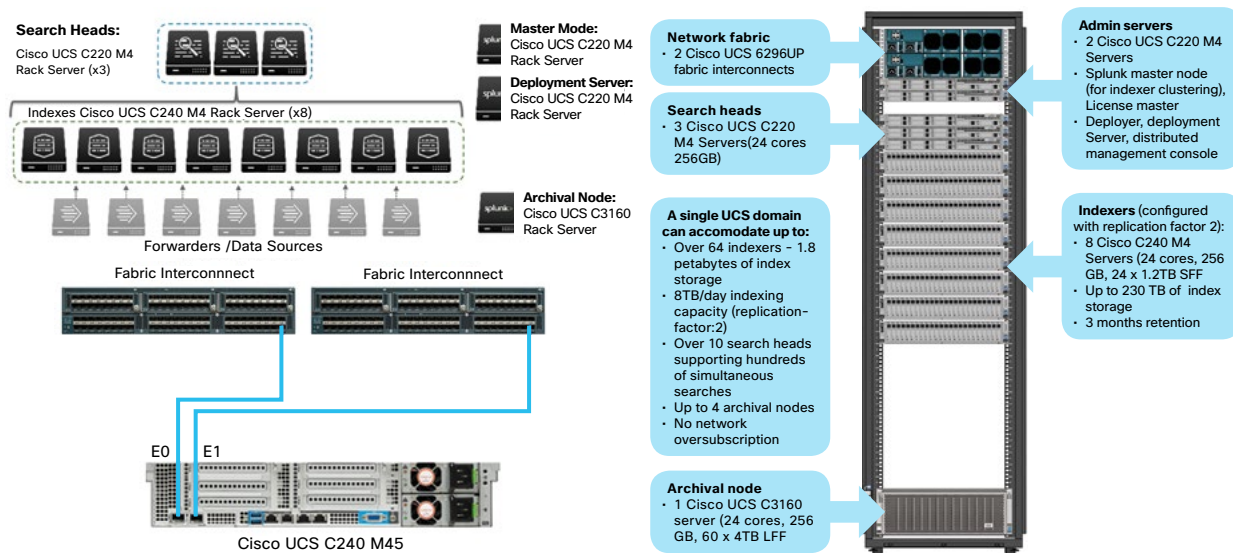
BUSINESS CHALLENGES

- Modern day datacenters generate machine data in the order of terabytes (TBs) to hundreds of TBs every day. The organization needs to collect, process, manage, derive business insights from this large amount of data
- Organizations need an integrated infrastructure for processing machine data that's efficient, reliable and secure
- Need to retain collected for longer term analysis, or need to put in an integrated storage medium for retrieval at a later point in time

SUMMARY

- Built on the Cisco UCS Integrated Infrastructure for Big Data for Splunk Enterprise distributed search with High Capacity reference architecture
- Cisco UCS Integrated Infrastructure offers horizontal scalability for Splunk Enterprise clusters
- Pre-validated configuration radically simplifies, and accelerates the entire Splunk distributed search deployment process
- Cisco UCS C3160 rack server offers unmatched high density storage capacity in small form factor to store frozen data that is readily accessible to Splunk Enterprise clusters
- Best practices for highly available distributed search deployment using Splunk Enterprise

ARCHITECTURE

Search Heads:
Cisco UCS C220 M4 Rack Server (x3)

Indexes Cisco UCS C240 M4 Rack Server (x8)

Forwarders /Data Sources

Fabric Interconnnect   Fabric Interconnnect

E0   E1

Cisco UCS C240 M45

Master Mode:
Cisco UCS C220 M4 Rack Server
Deployment Server:
Cisco UCS C220 M4 Rack Server

Archival Node:
Cisco UCS C3160 Rack Server

Network fabric
- 2 Cisco UCS 6296UP fabric interconnects

Search heads
- 3 Cisco UCS C220 M4 Servers(24 cores 256GB)

A single UCS domain can accomodate up to:
- Over 64 indexers - 1.8 petabytes of index storage
- 8TB/day indexing capacity (replication-factor:2)
- Over 10 search heads supporting hundreds of simultaneous searches
- Up to 4 archival nodes
- No network oversubscription

Archival node
- 1 Cisco UCS C3160 server (24 cores, 256 GB, 60 x 4TB LFF)

Admin servers
- 2 Cisco UCS C220 M4 Servers
- Splunk master node (for indexer clustering), License master
- Deployer, deployment Server, distributed management console

Indexers (configured with replication factor 2):
- 8 Cisco C240 M4 Servers (24 cores, 256 GB, 24 x 1.2TB SFF
- Up to 230 TB of index storage
- 3 months retention

Big Data

COMMON PLATFORM ARCHITECTURE (CPA)

- Big Data 60 node Hadoop Cluster with EMC Isilon
- Big Data Cisco Application Centric Infrastructure (ACI) with Cloudera
- Cisco UCS CPAv2 for Big Data with Cloudera
- Cisco UCS CPAv2 for Big Data with Hortonworks

- Hadoop-as-a-Service (HaaS) with Cisco UCS CPAv2 for Big Data and OpenStack
- Cisco UCS CPAv2 for Big Data with Intel Distribution
- Cisco UCS CPAv2 for Big Data with Pivotal HD and HAWQ

The Big Data market is poised to explode. Industry analysts project that it will exceed $50 billion by 2016, with a whopping compound annual growth rate (CAGR) of 58 percent.[1] These estimates include hardware, software, and services revenue. With the growing awareness of the power of Big Data and the development of new analytics capabilities, organizations seek to use Big Data to gain competitive advances and boost operational efficiencies.[2]

Big Data 60 node Hadoop Cluster with EMC Isilon

TECHNICAL HIGHLIGHTS

- Cisco UCS 5108 blade chassis with 2208XP fabric extender modules
- Cisco UCS B200 M3 blade servers
- Cisco UCS 6296UP fabric interconnects
- EMC Isilon S200 with OneFS 7.2.0.0
- VMware vSphere 5.5, vCenter Server 5.5 with Big Data extensions 2.1
- Cloudera Manager 5.2 with CDH 5.1.3

BUSINESS CHALLENGES

- Need to address scalability and deployment challenges in enterprise Hadoop deployments to rapidly increase or decrease the compute, storage and memory allocated to the Hadoop nodes
- Best practices for creating a virtualized enterprise grade Hadoop solution using Cisco UCS, EMC Isilon and Cloudera Distribution Apache Hadoop to meet the changing needs of enterprises, while minimizing downtime
- Ability to scale compute nodes and HDFS storage nodes independent of each other

SUMMARY

- End-to-end installation virtualized Hadoop solution with predictable scalability implementing Business Data Lake architecture
- Unified compute for MapReduce and Isilon multi protocol scale-out NAS platform for distributed NameNode and DataNode services
- Detailed instructions for installation and fine tuning for Big Data workloads

ARCHITECTURE
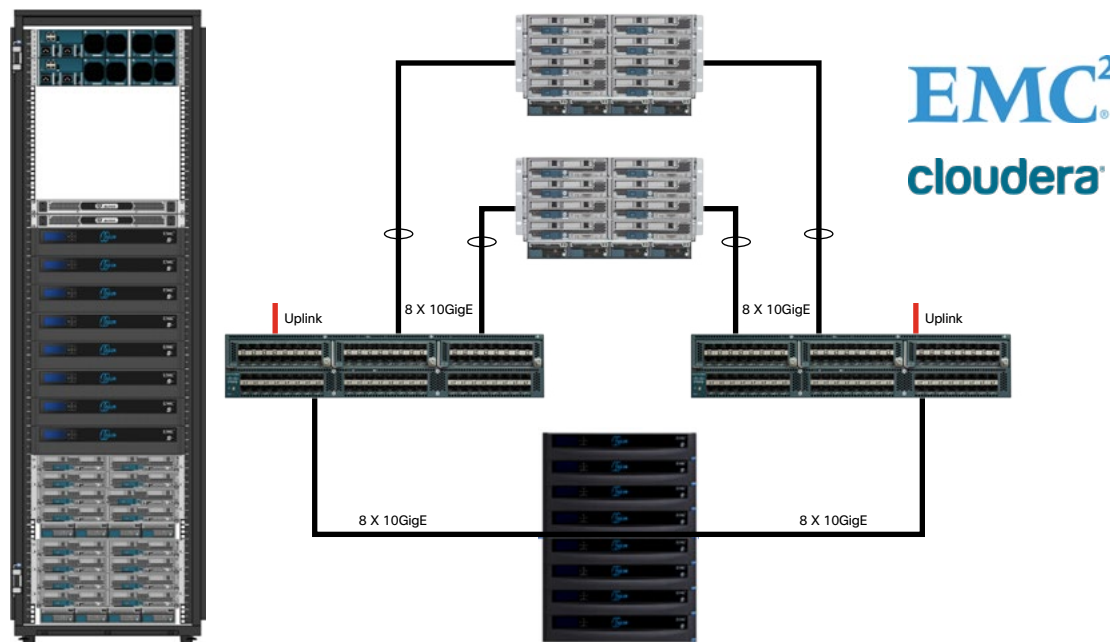
Big Data Cisco Application Centric Infrastructure (ACI) with Cloudera

TECHNICAL HIGHLIGHTS

- 42 UCS C240 M3 (SFF/LFF) servers with scale-out option up to 964 servers in a single switching domain
- Two Cisco NK-C9508
- 23 Cisco N9K-C9396PX
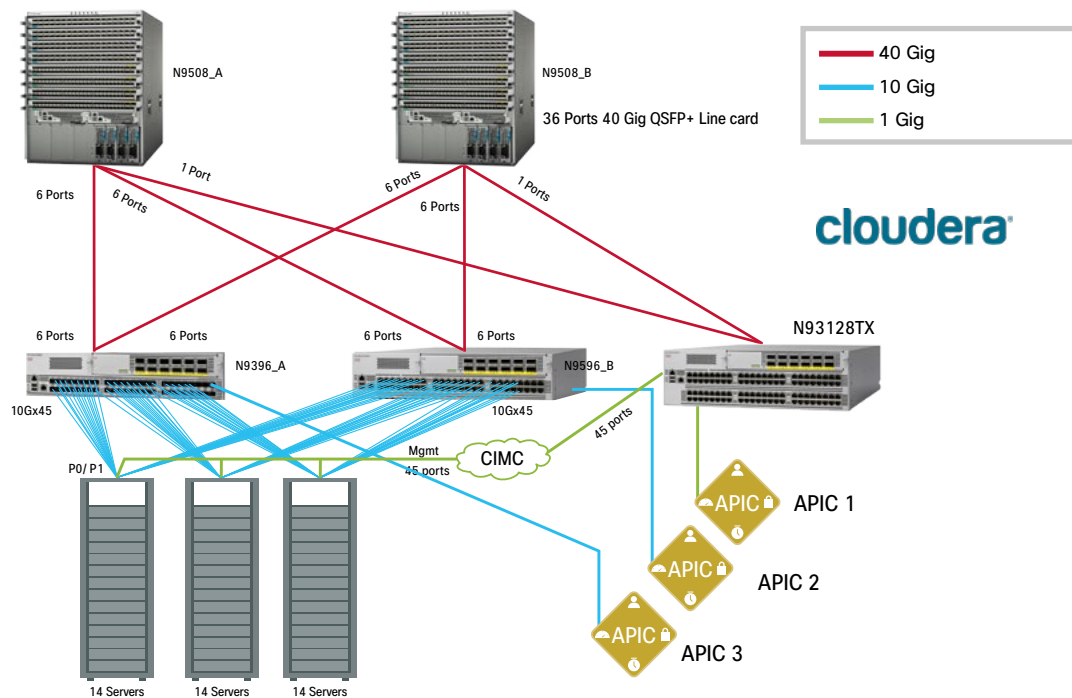- 12 Cisco N9K-C93128TX
- Three APIC-L1
- Cloudera Hadoop 5.0

BUSINESS CHALLENGES

- Scalable performance and multi-tenancy in hardware
- Centralized network visibility with real time application health monitoring
- Simplified automation through application driven policy model
- Best practices for installing Cloudera Distribution for Apache Hadoop 5.0 tuned for performance and scalability

SUMMARY

- Scaling to thousands of servers with no network over-subscription
- Application centric management
- Best practices for Hadoop 2.0 services tuning

ARCHITECTURE

Version 2.2

Overview

Information

Design Zone

CISCO

Library View

Sitemap View

Outline View

Alphabetical

Cisco UCS CPAv2 for Big Data with Cloudera

TECHNICAL HIGHLIGHTS

- 64 UCS C240 M3 (SFF/LFF) servers with scale out option up to 10 racks in a single switching domain
- Detailed instructions for installation and fine tuning for Big Data workloads
- Detailed Raid configuration and operating system tuning for Big Data application performance

BUSINESS CHALLENGES

- Highly scalable architecture designed to meet a variety of scale-out application (Big Data) demands with seamless data integration and management
- Support for multiple distributed computational frameworks on same Hadoop cluster
- Best practices for installing Cloudera Distribution for Apache Hadoop 5.0 tuned for performance and scalability

SUMMARY

- Next version of Hadoop Map/Reduce (Yarn)
- End-to-end installation for Cloudera Distribution of Apache Hadoop for performance and scalability
- Best practices for Hadoop 2.0 services tuning
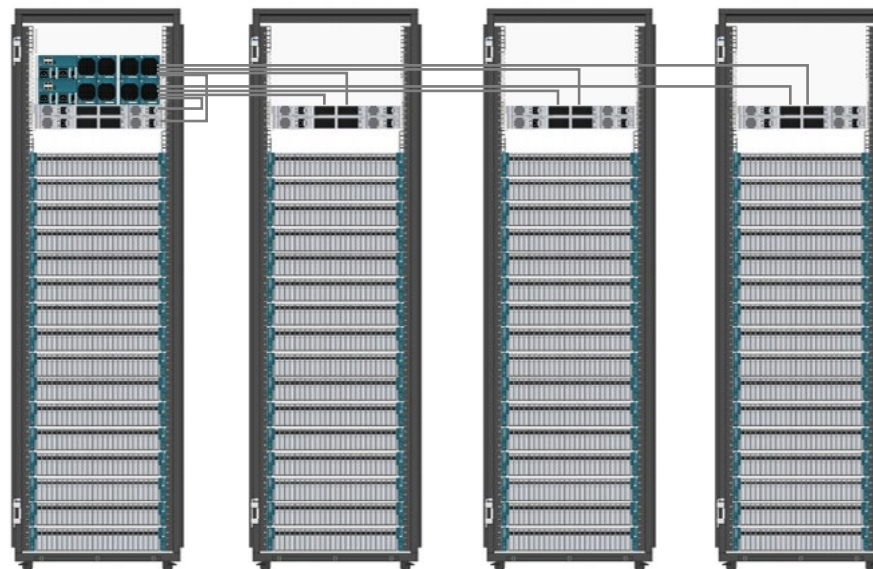
ARCHITECTURE

2x Cisco UCS Fabric Interconnect 6296UP

2x Cisco UCS Fabric Extender 2232PP

16X Cisco UCS C240M3 Server

8 x 10 Gigabit Links

cloudera

Cisco UCS CPAv2 for Big Data with Hortonworks

TECHNICAL HIGHLIGHTS

- 64 UCS C240 M3 (SFF/LFF) servers with scale-out option up to 10 racks in a single switching domain
- Detailed instructions for installation and fine tuning for Big Data workloads
- Detailed Raid configuration and operating system tuning for Big Data application performance

BUSINESS CHALLENGES

- Highly scalable architecture designed to meet a variety of scale-out application (Big Data) demands with seamless data integration and management
- Support for multiple distributed computational frameworks on same Hadoop cluster
- 100 percent open source distribution with enterprise support

SUMMARY

- Committed to 100% open source distribution
- End-to-end installation for Hortonworks HDP 2.0
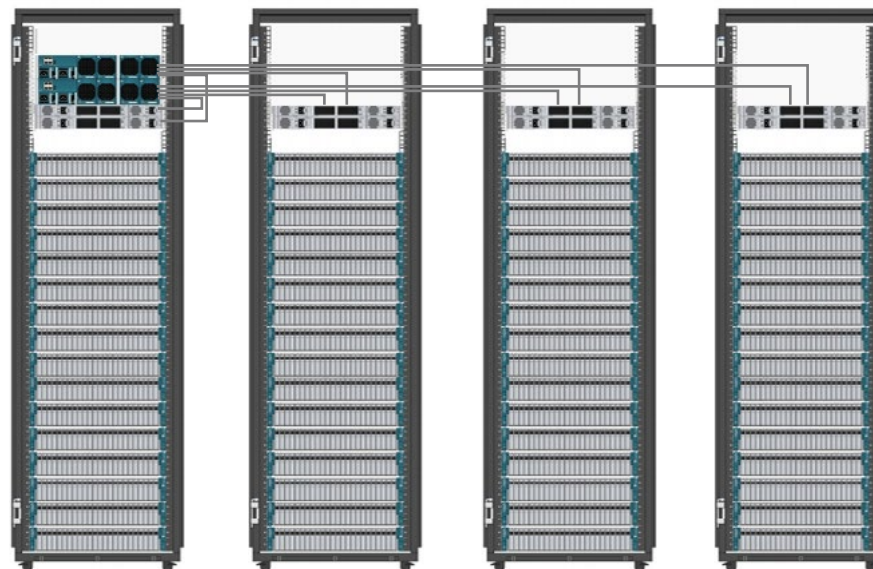- Next version of Hadoop Map/Reduce (Yarn)

ARCHITECTURE

2x Cisco UCS Fabric Interconnect 6296UP

2x Cisco UCS Fabric Extender 2232PP

16X Cisco UCS C240M3 Server

8 x 10 Gigabit Links

Hortonworks

Hadoop-as-a-Service (HaaS) with Cisco UCS CPAv2 for Big Data and OpenStack

TECHNICAL HIGHLIGHTS

- 64 UCS C240 M3 (SFF/LFF) servers with scale-out option up to 10 racks in a single switching domain
- Detailed instructions for installation and fine tuning for Big Data workloads
- Detailed Raid configuration and operating system tuning for Big Data application performance

BUSINESS CHALLENGES

- Highly scalable architecture offering Hadoop-as-a-Service
- Running Hadoop on OpenStack
- Enterprise-grade, hardened Apache Hadoop distribution with excellent support
- 100 percent open source distribution with enterprise support

SUMMARY

- Self service provisioning of Hadoop
- Tuning OpenStack for Hadoop
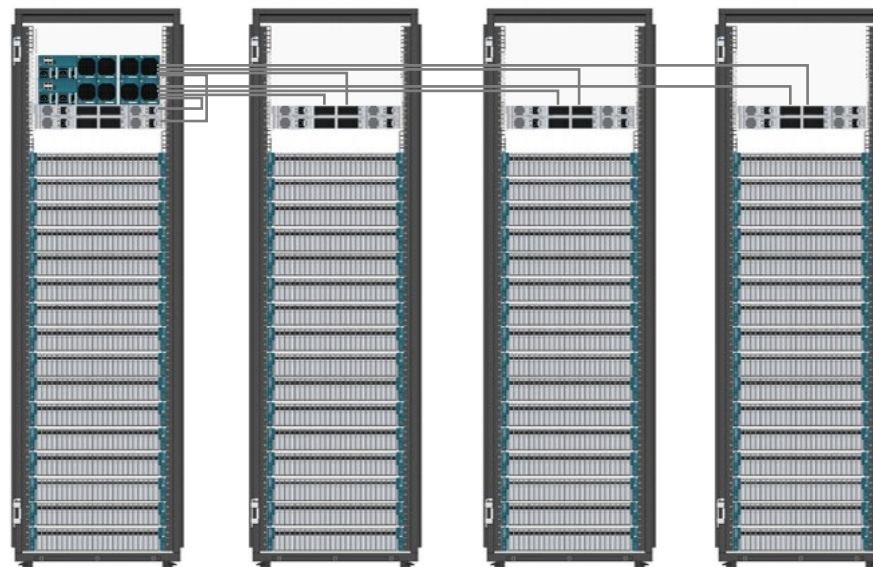- End-to-end installation for Hortonworks HDP 2.0 on VM's managed by OpenStack

ARCHITECTURE

2x Cisco UCS Fabric Interconnect 6296UP

2x Cisco UCS Fabric Extender 2232PP

16X Cisco UCS C240M3 Server

8 x 10 Gigabit Links

Hortonworks

Cisco UCS CPAv2 for Big Data with Intel Distribution

TECHNICAL HIGHLIGHTS

- 64 UCS C240 M3 (SFF/LFF) servers with scale-out option for up to 10 racks in a single switching domain
- Detailed role-based access control and other security setting
- Detailed RAID configuration and operating system tuning for Big Data application performance

BUSINESS CHALLENGES

- Highly scalable architecture designed to meet a variety of scale-out application (Big Data) demands with seamless data integration and management
- Enterprise-grade, hardened Apache Hadoop distribution with excellent support
- Hardware-assisted encryption using AES-NI technology beneficial for HIPPA and payment card industry (PCI) data security

SUMMARY

- Focused on hardware-assisted encryption, security, fine-grained access control (ACL), role-based access control, and Kerberos
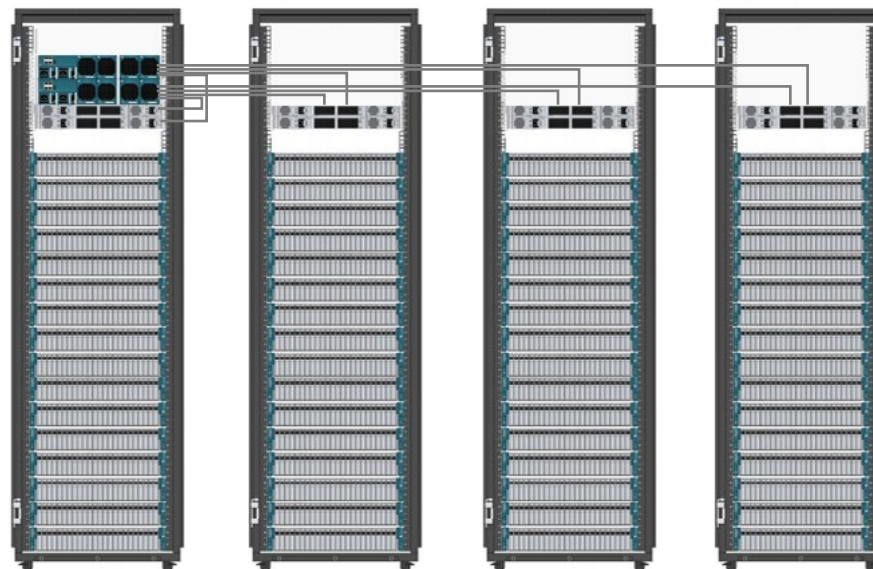- End-to-end installation for Intel distribution 2.5

ARCHITECTURE

2x Cisco UCS Fabric Interconnect 6296UP

2x Cisco UCS Fabric Extender 2232PP

16X Cisco UCS C240M3 Server

8 x 10 Gigabit Links

intel®

CVD: http://www.cisco.com/en/US/docs/unified_computing/ucs/UCS_CVDs/CPA_for_BigData_with_Hortonworks.html

Cisco UCS CPAv2 for Big Data with Pivotal HD and HAWQ

TECHNICAL HIGHLIGHTS

- 64 UCS C240 M3 LFF servers with scale-out option of up to 10 racks in a single switching domain
- Apache Hadoop 2.0 distribution with YARN
- Detailed RAID configuration and operating system tuning for Big Data application performance

BUSINESS CHALLENGES

- Require highly scalable architecture capable of meeting scale-out application (Big Data) demands with seamless data integration and management
- Need an enterprise-grade, hardened Apache Hadoop distribution with support
- Hadoop 2.0 with Yarn and HAWQ, a 100 percent compliant SQL query engine on Hadoop

SUMMARY

- HAWQ with its parallel SQL processing is fully SQL compliant on Hadoop
- End-to-end installation for Pivotal HD 1.1 and HAWQ 1.1
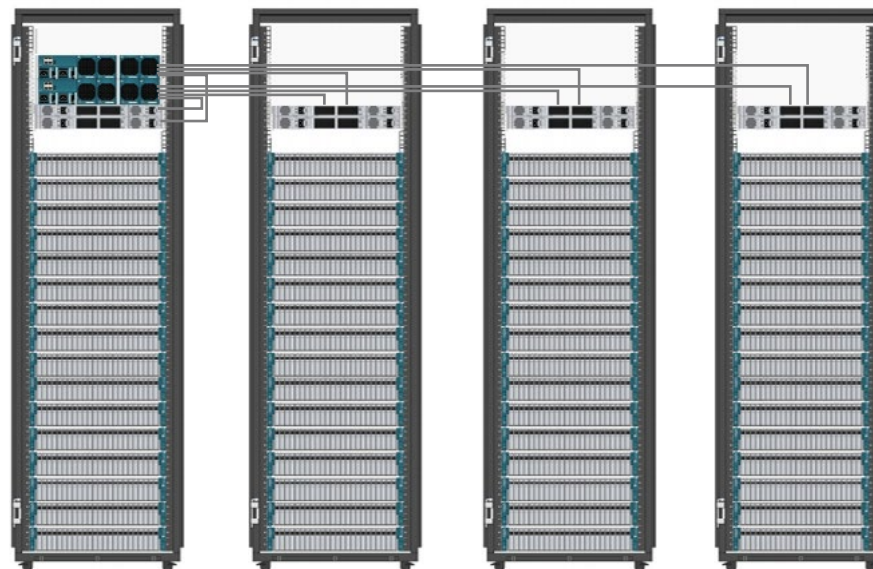
ARCHITECTURE

2x Cisco UCS Fabric Interconnect 6296UP

2x Cisco UCS Fabric Extender 2232PP

16X Cisco UCS C240M3 Server

8 x 10 Gigabit Links

Pivotal™

Version 2.2

Overview

Information

Design Zone

CISCO

Library View

Sitemap View

Outline View

Alphabetical

Information page

Additional Resources:

Design Zone: **www.cisco.com/go/cvd**

Data Center Design Zone: **www.cisco.com/go/dcdesignzone**

Virtualized Desktop Design Zone: **www.cisco.com/go/vdidesigns**

Cisco Validated Designs (CVDs) support many Cisco Data Center and UCS Solutions. For more information on Cisco Solutions broader than CVDs, use the following URLs:

- **www.cisco.com/go/bigdata**
- **www.cisco.com/go/flexpod**
- **www.cisco.com/go/hyperflex**
- **www.cisco.com/go/microsoft**
- **www.cisco.com/go/oracle**
- **www.cisco.com/go/openstack**
- **www.cisco.com/go/sap**
- **www.cisco.com/go/smartstack**
- **www.cisco.com/go/vdi**
- **www.cisco.com/go/versastack**
- **www.cisco.com/go/vspex**

For more information on UCS and UCS Solutions, use the following URL: **www.cisco.com/go/ucs**

Version:

Version 2.2 | November 2016

Viewing:

This playbook is best viewed in Adobe Reader Version XI (11.0.06).
If you do not have Adobe Reader, you can download it for free at **http://get.adobe.com/reader**

Version 2.2

Overview

Information

Design Zone

CISCO

Library View

Sitemap View

Outline View

Alphabetical

Thank you.