Cisco UCS Integrated Infrastructure for Big Data and Analytics with Cisco ACI and SAP HANA Vora for In-Memory Analytics

Last Updated: May 12, 2017
About the Cisco Validated Design (CVD) Program

The CVD program consists of systems and solutions designed, tested, and documented to facilitate faster, more reliable, and more predictable customer deployments. For more information visit:


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 UCS Director Express, 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)

© 2017 Cisco Systems, Inc. All rights reserved.
# Table of Contents

**Executive Summary** .......................................................................................................................... 6  
**Cisco UCS and SAP HANA Vora Deliver a New Dimension to Big Data Analytics** ........................................ 6 

**Solution Overview** ............................................................................................................................... 7  
**Introduction** ........................................................................................................................................ 7  
**Audience** ............................................................................................................................................ 7  
**Solution** ................................................................................................................................................ 7  
**Solution Summary** ............................................................................................................................... 9  

**Technology Overview** .......................................................................................................................... 11 
**Cisco UCS Integrated Infrastructure for Big Data** .................................................................................... 11  
**Cisco UCS Fabric Interconnects** ........................................................................................................... 11  
**Cisco UCS C-Series Rack-Mount Servers** ............................................................................................... 12  
**Cisco UCS Virtual Interface Cards (VICs)** ............................................................................................... 13  
**Cisco UCS Manager** ............................................................................................................................. 14  

**Cisco UCS Director Express for Big Data** ............................................................................................... 15  
**Key Features of Cisco UCS Director (UCSD) Express for Big Data** ....................................................... 16  

**Cisco Application Centric Infrastructure (ACI) Overview** ....................................................................... 17  
**Architectural Benefits of Using Fabric Interconnect with Cisco ACI** ....................................................... 17  
**Centralized Management for the Entire Network** ................................................................................... 19  
**Dynamic Load Balancing** .................................................................................................................... 19  
**Multi-Tenant and Mixed Workload Support** ........................................................................................... 19  
**Easy Migration to 40 Gbps in the Network** ............................................................................................ 20  
**Cisco ACI Building blocks** .................................................................................................................... 20  
**Cisco Nexus 9000 Series Switches** ....................................................................................................... 20  
**Cisco Nexus 9508 Spine Switch** ........................................................................................................... 21  
**ACI Spine Line Card for Nexus 9508** ..................................................................................................... 21  
**Cisco Nexus 9396 Leaf Switch** ............................................................................................................. 22  
**Application Policy Infrastructure Controller (APIC)** ............................................................................... 23  
**Cisco ACI Topology** ............................................................................................................................. 24  

**Cisco UCS Datacenter Solution for SAP HANA** .................................................................................... 25  
**Reference Architecture: Flexpod Datacenter for SAP Solution with Cisco ACI** ...................................... 26  

**MapR Converged Data Platform 5.1** ...................................................................................................... 30  
**MapR Enterprise-Grade Platform Services** ............................................................................................ 31  
**MapR Open Source Technologies** .......................................................................................................... 33
Apache Spark .................................................................34
Spark Streaming ............................................................36
SAP HANA Vora ................................................................36
Solution Design ................................................................38
Requirements ....................................................................38
Physical Layout for the Solution ........................................38
Software Distributions and Versions ...............................39
Red Hat Enterprise Linux (RHEL) ..................................39
MapR ..............................................................................40
SAP HANA Vora ..............................................................40
Software Versions ............................................................40
Deployment Hardware and Software ............................42
System Architecture .........................................................42
Scaling the Architecture ..................................................45
Scaling the Architecture Further with Additional Spine Switches ..................................................45
SAP HANA and SAP HANA VORA scalability ...............46
Network Configuration .......................................................46
IP Address Assignment ......................................................46
Fabric Configuration ........................................................47
Server Configuration and Cabling for Cisco UCS C240 M4 Rack Server ..........................................47
Cisco UCS Fabric Configuration .......................................48
Configure Fabric Interconnect A ..................................48
Configure Fabric Interconnect B ..................................49
Logging into Cisco UCS Manager ................................50
Enabling Server Ports .....................................................50
Enabling Uplink Ports .......................................................51
Configuring Port-Channels ..............................................52
Adding Block of IP Addresses for KVM Access ...............54
Creating MAC Address Pools .......................................56
Creating Server Pool .......................................................57
Configuring Communication Services for Cisco UCS Manager ..................................................59
Configuring QoS System Class .....................................60
Install and Configure Hadoop, YARN, and Spark ..........61
Creating the Hadoop Cluster Using Cisco UCS Director Express for Big Data ..................................61
Create the IP Address Pools .............................................61
Creating a Hadoop Cluster.................................................................63
Monitoring the Hadoop Cluster Creation........................................67
Logging into the Hadoop UI................................................................72
Installing Spark ..................................................................................73
Installing Scala....................................................................................73
Create a dedicated Vora user..............................................................74
Installing Spark ..................................................................................74
Verifying Cluster Status .....................................................................76
Install and Configure SAP HANA Vora..............................................77
Preparing to Install SAP HANA Vora..................................................77
Preparing for SAP HANA Vora Distributed Log Server ..................77
Preparing for SAP HANA Vora Document Store Server..................78
Preparing for SAP HANA Vora Disk Engine Server .........................78
Preparing for SAP HANA Vora Cluster Manager.............................78
Configure Sudo Access ......................................................................78
Validate the cluster ..........................................................................79
Downloading SAP HANA Vora Installation Package .........................80
Generate an Initial Password for SAP HANA Vora ...........................80
Vora installation recommendation........................................................82
Configure the SAP HANA Vora Manager............................................83
Start the SAP HANA Vora Manager...................................................84
Deploy SAP HANA Vora Service........................................................84
Install the SAP HANA Vora Zeppelin Interpreter .............................87
Test the Zeppelin Interpreter .............................................................93
Preparing the connectivity between SAP HANA Vora and SAP HANA 94
Configuring the SAP HANA Spark Controller......................................95
Bill of Materials ..................................................................................98
Conclusion.........................................................................................102
Reference............................................................................................103
About the Authors..............................................................................104
Acknowledgements ..........................................................................104
Cisco UCS and SAP HANA Vora Deliver a New Dimension to Big Data Analytics

This Cisco Validated Design describes architecture and deployment procedures to create a SAP HANA Vora cluster on Cisco UCS Integrated Infrastructure for Big Data and Analytics, and Cisco Application Centric Infrastructure (ACI). The deployment creates a simple and linearly scalable architecture that is centrally managed. SAP HANA Vora adds contextual awareness and real-time analytics to big data deployments implemented on Cisco UCS Integrated Infrastructure for Big Data and Analytics. This solution provides access to more precise decision making, democratized data access and simplified big data ownership.

Cisco UCS Integrated Infrastructure for Big Data and Analytics with Cisco ACI and SAP HANA Vora helps businesses gain a new level of insight by bringing big data query results into the more static business data stored in SAP HANA. The following are just a few of the ways that the solution can help your staff get the information it needs:

- Optimize the supply chain and increase visibility
- Detect fraud
- Conduct targeted marketing campaigns
- Improve IT capacity planning activities
- Improve patient care.
- Proactively maintenance and improved visibility
- Manage adverse events and product recall activities
Solution Overview

Introduction

Information is most powerful when it is turned into real-time insight. Many organizations use Hadoop and Apache Spark to mine big data stores to identify trends and empower decision makers. Now it is possible to add contextual awareness to your big data deployments and run all of your big data and analytics operations on Cisco UCS Integrated Infrastructure for Big Data and Analytics. This solution gives you access to more precise decision making, democratized data access and simplified big data ownership.

While Hadoop can store and access vast amounts of detailed data at lower costs, it is not as well suited to the fast, drill-down nature of many of today’s business questions. Through data hierarchies that enable online analytical processing (OLAP) analysis of Hadoop data, enhancements in Spark SQL, and compiled queries for accelerated processing across nodes, SAP HANA Vora enables precision decision-making across all the data in enterprise applications, data warehouses, data lakes and edge sensors. SAP HANA Vora works with all major Hadoop distributions and applies the power of in-memory processing to massively distributed data stores. By helping to overcome the limitations of batch-oriented processing, it enables real-time, iterative access to data on Hadoop clusters. Companies can now discover new insights by combining traditional sources of data with valuable data arriving continually from outside the organization using enterprise-grade data management practices.

Although your enterprise data and big data have value separately, the capability to bring them together presents new opportunities for your data scientists and analysts. Running on the Apache Spark framework, SAP HANA Vora is an in-memory query engine that enables you to bring new insights easily into your SAP landscape. By combining your business information with data from other sources— including streaming, interactive queries, and machine learning—organizations accelerate and add context to their decision-making processes producing better business outcomes.

Audience

The intended audience for this document includes sales engineers, field consultants, professional services, IT managers, partner engineering and customers who want to deploy SAP HANA Vora on Cisco UCS Integrated Infrastructure for Big Data alongside their existing SAP HANA Enterprise Application landscape interconnected by Cisco ACI.

Solution

The Cisco UCS Integrated Infrastructure for Big Data and Analytics with Cisco ACI and SAP HANA Vora solution brings together enterprise applications and big data technologies to provide better business coherence for precise decision making with contextual awareness by combining business data with Hadoop data accessed via an in-memory processing engine. The components of the solution include:

- Cisco UCS Integrated Infrastructure for Big Data and Analytics
- Cisco UCS Director Express for Big Data
- Cisco ACI
- Apache Hadoop
- SAP HANA Vora

**Cisco UCS Integrated Infrastructure for Big Data and Analytics** includes computing, storage, connectivity and unified management capabilities to help companies manage the immense amount of data they collect today. It is built on the Cisco Unified Computing System™ (Cisco UCS) infrastructure, using Cisco UCS 6200 Series Fabric Interconnects, and Cisco UCS C-Series Rack Servers. This architecture is specifically designed for performance and linear scalability for big data workloads.

**Cisco UCS Director Express for Big Data** provides a single-touch solution that automates Apache Hadoop deployments on the Cisco UCS Integrated Infrastructure for Big Data and Analytics. It also provides a single management pane across both physical infrastructure and Apache Hadoop software. All elements of the infrastructure are handled automatically with little user input.

**Cisco Application Centric Infrastructure (Cisco ACI)** is a comprehensive SDN architecture. One of the core design principles behind Cisco ACI was to provide complete visibility into the infrastructure - physical and virtual. ACI is software-defined networking (SDN) and more. Most SDN models stop at the network. ACI extends the promise of SDN - namely agility and automation - to the applications themselves. Through a policy-driven model, the network can cater to the needs of each application, with security, network segmentation and automation at scale; and it can do so across physical and virtual environments, with a single pane of management.

**Figure 1 Solution Overview**

- Single point of management for Network
- Scale-Out Penalty-Free Overlay Network
Solution Overview

Solution Summary

This CVD describes the architecture and deployment procedures for setting up Cisco UCS C240 M4 servers, based on Cisco UCS Integrated Infrastructure for Big Data and Analytics and Cisco ACI, bringing together a highly scalable architecture designed to meet a variety of scale-out application demands with seamless data integration and management integration capabilities.

This CVD describes in detail the process of creating the Application Network Profile in the ACI for the Big Data application. The Application Network Profile is a collection of EPGs, their connections, and the policies that define those connections (described in detail later). Application Network Profiles are the logical representation of an application (in this case, Big Data) and its interdependencies in the network fabric. Application Network Profiles are designed to be modeled in a logical way that matches the way that applications are designed and deployed. The configuration and enforcement of policies and connectivity is handled by the system rather than manually by an administrator.

The current version of the Cisco UCS Integrated Infrastructure for Big Data and Analytics offers the configurations shown in Table 1

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Reference Architecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAP HANA Infrastructure</td>
<td>High Performance</td>
</tr>
<tr>
<td><strong>Refer to SAP Hana Design Zone</strong></td>
<td>2 Cisco UCS 6332 Fabric Interconnects</td>
</tr>
<tr>
<td></td>
<td>16 Cisco UCS C240 M4 Rack servers</td>
</tr>
<tr>
<td></td>
<td>• 2 Intel Xeon processors E5-2680 v4 CPUs</td>
</tr>
<tr>
<td></td>
<td>• 256 GB of memory</td>
</tr>
<tr>
<td></td>
<td>• Cisco 12-Gbps SAS Modular Raid Controller with 2GB flash-based write cache (FBWC)</td>
</tr>
<tr>
<td></td>
<td>• 16 1.6TB SSD drives (410TB total)</td>
</tr>
<tr>
<td></td>
<td>• 2 480GB 6Gbps 2.5inch Enterprise Value SATA SSDs for Boot</td>
</tr>
<tr>
<td></td>
<td>Cisco UCS VIC 1387 (with 2 40GE QSFP+ ports)</td>
</tr>
<tr>
<td></td>
<td><strong>Scaling:</strong></td>
</tr>
<tr>
<td></td>
<td>• Up to 24 servers per domain.</td>
</tr>
</tbody>
</table>
This CVD describes the install process of SAP HANA Vora with MapR 5.1 and MapR-Spark 1.6.1 on a 16-node cluster.
Technology Overview

This Cisco validated design brings together these main technologies:

2. Cisco UCS Director Express for Big Data.
3. Cisco UCS Datacenter Solution for SAP HANA (refer to Cisco Datacenter Solutions for SAP HANA).
4. Cisco Application Centric Infrastructure (ACI).

Cisco UCS Integrated Infrastructure for Big Data

The Cisco UCS solution for Hadoop is based on Cisco UCS Integrated Infrastructure for Big Data, a highly scalable architecture designed to meet a variety of scale-out application demands with seamless data integration and management integration capabilities built using the following components:

Cisco UCS Fabric Interconnects

Cisco UCS Fabric Interconnects provide high-bandwidth, low-latency connectivity for servers, with integrated, unified management provided for all connected devices by Cisco UCS Manager. Deployed in redundant pairs, Cisco fabric interconnects offer the full active-active redundancy, performance, and exceptional scalability needed to support the large number of nodes that are typical in clusters serving big data applications. Cisco UCS Manager enables rapid and consistent server configuration using service profiles, automating ongoing system maintenance activities, such as firmware updates, across the entire
cluster as a single operation. Cisco UCS Manager also offers advanced monitoring with options to raise alarms and send notifications about the health of the entire cluster.

<table>
<thead>
<tr>
<th>Generation</th>
<th>Cisco UCS Fabric Interconnect</th>
<th>Connectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd Generation</td>
<td>Cisco UCS 6332</td>
<td>32 x 40 Gigabit Ethernet ports</td>
</tr>
<tr>
<td>2nd Generation</td>
<td>Cisco UCS 6296 UP</td>
<td>96 x10 Gigabit Ethernet ports</td>
</tr>
<tr>
<td></td>
<td>Cisco UCS 6248 UP</td>
<td>48 x 10 Gigabit Ethernet ports</td>
</tr>
</tbody>
</table>

**Figure 2  Cisco UCS 6332 Fabric Interconnect**

26 x 40G QSFP+ 6 x 40G QSFP+

**Figure 3  Cisco UCS 6296UP Fabric Interconnect**

Cisco UCS C-Series Rack-Mount Servers

Cisco UCS C240 M4 High-Density Rack-Mount Servers (Small Form Factor Disk Drive Model) are enterprise-class systems that support a wide range of computing, I/O, and storage-capacity demands in compact designs. Cisco UCS C-Series Rack Servers are based on Intel Xeon E5-2680 v4 product family and 12-Gbps SAS throughput, delivering significant performance and efficiency gains over the previous generation of servers. The servers use dual Intel Xeon processor E5-2600 v4 series CPUs and supports up to 1.5 TB of main memory (256 GB is typical for big data applications) and a range of disk drive and SSD options – up to 24 Small Form Factor (SFF) disk drives. Cisco UCS Virtual Interface Cards (VICs) designed for Cisco UCS C-Series Rack Servers are optimized for high-bandwidth and low-latency cluster connectivity, with support for up to 256 virtual devices that are configured on demand through Cisco UCS Manager.
**Cisco UCS Virtual Interface Cards (VICs)**

Cisco UCS Virtual Interface Cards (VICs), unique to Cisco, incorporate next-generation converged network adapter (CNA) technology from Cisco, and offer dual ports (10-Gbps or 40-Gbps) designed for use with Cisco UCS C-Series Rack Servers. Optimized for virtualized networking, these cards deliver high performance and bandwidth utilization and support up to 256 virtual devices.

**Cisco UCS VIC 1227**

The Cisco UCS Virtual Interface Card (VIC) 1227 is a dual-port, Enhanced Small Form-Factor Pluggable (SFP+), 10 Gigabit Ethernet and Fiber Channel over Ethernet (FCoE)-capable, PCI Express (PCIe) modular LAN on motherboard (mLOM) adapter. It is designed exclusively for the M4 generation of Cisco UCS C-Series Rack Servers.

**Cisco UCS VIC 1387**

The Cisco UCS VIC 1387 (Figure 6) offers dual-port, Enhanced Quad, Small Form-Factor Pluggable (QSFP) 40 Gigabit Ethernet and Fiber 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.
Cisco UCS Manager

Cisco UCS Manager resides within the Fabric Interconnect, either the Cisco UCS 6200 Series or Cisco UCS 6300 Series. It makes the system self-aware and self-integrating, managing all of the system components as a single logical entity. Cisco UCS Manager can be accessed through an intuitive graphical user interface (GUI), a command-line interface (CLI), or an XML application-programming interface (API). Cisco UCS Manager uses service profiles to define the personality, configuration, and connectivity of all resources within Cisco UCS, radically simplifying provisioning of resources so that the process takes minutes instead of days. This simplification allows IT departments to shift their focus from constant maintenance to strategic business initiatives.
Cisco UCS Director Express for Big Data

Cisco UCS Director Express for Big Data provides a single-touch solution that automates deployment of Hadoop on Cisco UCS Integrated Infrastructure for Big Data and provides a single management pane across both physical infrastructure and Hadoop software. All elements of the infrastructure, from configuration of physical infrastructure - storage, compute, network, OS, Java packages, Hadoop installation, and provisioning of Hadoop services, are handled automatically with minimal user input. This allows organizations to have clusters up and running in a few hours and delivers the best performance out of the box. Cisco UCS Director Express is fully integrated with leading Hadoop distributions, providing a single pane of management to provide centralized visibility across the entire infrastructure, and executes core management tasks. It complements Hadoop managers by providing a system-wide perspective to Hadoop admins, enabling them to correlate Hadoop activity with network and compute activity on the Hadoop nodes.
Cisco Unified Computing System (UCS) has redefined the data center infrastructure by introducing the concept of stateless resource abstraction. By doing so, it has brought the ability to dynamically instantiate a server identity from the available pool of resources – compute, network and storage. Cisco UCS Service Profiles provide this abstraction, and together with Cisco UCS Manager, enable dynamic allocation of resources resulting in just-in-time deployment of nodes.

Cisco UCS Director Express for Big Data extends the concept of service profiles into the Hadoop application space, providing a reliable and consistent mechanism to define the infrastructure and the Hadoop services running on the cluster. This provides a tremendous flexibility for users to consistently and reliably deploy Hadoop clusters on the fly, and provision the requisite Hadoop services without manual intervention. Once the system is racked and stacked, the user can deploy Hadoop clusters on demand with a single GUI screen. It provides appliance-like capabilities, with flexibility for users to configure-on-demand and select the desired Hadoop distribution and services. Simplicity and agility are the two hallmarks of this solution. It reduces the time spent by IT teams to set up and maintain the Hadoop infrastructure, allowing time to focus on delivering business value from Big Data.

Key Features of Cisco UCS Director (UCSD) Express for Big Data

- **Faster and Easier Big Data Infrastructure Deployment**: Cisco UCS Director Express for Big Data extends the Cisco UCS Integrated Infrastructure for Big Data with one-click provisioning, installation, and configuration, delivering a consistent, repeatable, flexible, and reliable end-to-end Hadoop deployment.

- **Massive Scalability and Performance**: Cisco’s unique approach provides appliance-like capabilities for Hadoop with flexibility that helps ensure that resources are deployed right the first time and can scale without arbitrary limitations.
- **Centralized Visibility**: Cisco UCS Director Express for Big Data provides centralized visibility into the complete infrastructure to identify potential failures before they affect application and business performance.

- **Open and Powerful**: Provides open interfaces that allows further integration into third-party tools and services while allowing flexibility for your own add-on services.

**Cisco Application Centric Infrastructure (ACI) Overview**

Cisco ACI provides the network the ability to deploy and respond to the needs of applications, both in the data center and in the cloud. The network must be able to deliver the right levels of connectivity, security, compliance, firewalls, and load balancing, and it must be able to do this dynamically and on-demand.

This is accomplished through centrally defined policies and application profiles.

The profiles are managed by the new Application Policy Infrastructure Controller [APIC] and distributed to switches, like the Cisco Nexus 9000 Series. Cisco Nexus 9000 Series Switches, and the Cisco Application Policy Infrastructure Controller (APIC) are the building blocks for Cisco ACI.

Cisco ACI is software-defined networking (SDN) plus a whole lot more. Most SDN models stop at the network. Cisco ACI extends the promise of SDN—namely agility and automation—to the applications themselves. Through a policy-driven model, the network can cater to the needs of each application, with security, network segmentation, and automation at scale. And it can do so across physical and virtual environments, with a single pane of management.

The Cisco ACI fabric supports more than 64,000 dedicated tenant networks. A single fabric can support more than one million IPv4/IPv6 endpoints, more than 64,000 tenants, and more than 200,000 10G ports. The Cisco ACI fabric enables any service (physical or virtual) anywhere, with no need for additional software or hardware gateways, to connect between the physical and virtual services, and normalizes encapsulations for Virtual Extensible Local Area Network (VXLAN) / VLAN / Network Virtualization using Generic Routing Encapsulation (NVGRE).

The Cisco ACI fabric decouples the endpoint identity and associated policy from the underlying forwarding graph. It provides a distributed Layer 3 gateway that ensures optimal Layer 3 and Layer 2 forwarding. The fabric supports standard bridging and routing semantics without standard location constraints (any IP address anywhere), and removes flooding requirements for the IP control plane Address Resolution Protocol (ARP) / Generic Attribute Registration Protocol (GARP). All traffic within the fabric is encapsulated within VXLAN.

**Architectural Benefits of Using Fabric Interconnect with Cisco ACI**

The Cisco ACI fabric consists of discrete components that operate as routers and switches, but is provisioned and monitored as a single entity. The operation is like a single switch and router that provides advanced traffic optimization, security, and telemetry functions, stitching together virtual and physical workloads.
Cisco Application Centric Infrastructure (ACI) and Cisco Unified Computing System (Cisco UCS), working together, can cost-effectively scale capacity, and deliver exceptional performance for the growing demands of big data processing, analytics, and storage workflows. For larger clusters and mixed workloads, Cisco ACI uses intelligent, policy-based flowlet switching and packet prioritization to deliver:

- Centralized Management for the entire Network
- Dynamic load balancing
- Dynamic Packet Prioritization
- Multi-Tenant and Mixed Workload Support
- Deep Telemetry
Centralized Management for the Entire Network

Cisco ACI treats the network as a single entity rather than a collection of switches. It uses a central controller to implicitly automate common practices such as Cisco ACI fabric startup, upgrades, and individual element configuration. The Cisco Application Policy Infrastructure Controller (Cisco APIC) is the unifying point of automation and management for the Cisco Application Centric Infrastructure (ACI) fabric. This architectural approach dramatically increases the operational efficiency of networks, by reducing the time and effort needed to make modifications to the network and, also, for root cause analysis and issue resolution.

Dynamic Load Balancing

Cisco’s Application Centric Infrastructure is not only aware of the congestion points but is able to make dynamic decisions on how the traffic is switched/routed. This could be new flows that are about to start or existing long flows which could benefit from moving to a less congested route. Dynamic load balancing takes care of these decisions at run time automatically and helps utilize the links optimally – both the healthy and the congested links. This is useful in both congested link scenarios and scenarios where there are link failures. Even when there is no congestion this will maintain close to optimal distribution of traffic across the spines.

Dynamic Packet Prioritization (DPP), prioritizes short flows higher than long flows; a short flow is less than approximately 15 packets. Short flows are more sensitive to latency than long ones. Small and urgent data workloads, such as database queries, may suffer processing latency delays because larger data sets are being sent across the fabric ahead of them. This approach presents a challenge for instances in which database queries require near-real-time results.

Dynamic Packet Prioritization can improve overall application performance. Together these technologies enable performance enhancements to applications, including Big Data workloads.

Multi-Tenant and Mixed Workload Support

Cisco ACI is built to incorporate secure multi-tenancy capabilities. The fabric enables customers to host multiple concurrent Big Data clusters on a shared infrastructure. Cisco ACI provides the capability to enforce proper isolation and SLA’s for workloads of different tenants. These benefits extend beyond multiple Big Data workloads – Cisco ACI allows the same cluster to run a variety of different application workloads, not just Big Data, with the right level of security and SLA for each workload.

Deep Telemetry of Tenant and Application Network

One of the core design principles behind Cisco ACI is to provide complete visibility into the infrastructure – physical and virtual. Cisco APIC is designed to provide application and tenant health at a system level by using real-time metrics, latency details, atomic counters, and detailed resource consumption statistics.

If your application is experiencing performance issues, you can drill down easily into the lowest possible granularity – be it at a switch level, line card level, or port level.

The holistic approach to correlate virtual and physical and tie that intelligence to an application or tenant level ensures that troubleshooting becomes extremely simple across your infrastructure, through a single pane of glass.
Easy Migration to 40 Gbps in the Network

Cisco QSFP BiDi technology removes 40-Gbps cabling cost barriers for migration from 10-Gbps to 40-Gbps connectivity in data center networks. Cisco QSFP BiDi transceivers provide 40-Gbps connectivity with immense savings and simplicity compared to other 40-Gbps QSFP transceivers. The Cisco QSFP BiDi transceiver allows organizations to migrate the existing 10-Gbps cabling infrastructure to 40 Gbps at no cost and to expand the infrastructure with low capital investment. Together with Cisco Nexus 9000 Series Switches, which introduce attractive pricing for networking devices, Cisco QSFP BiDi technology provides a cost-effective solution for migration from 10-Gbps to 40-Gbps infrastructure.

Cisco ACI Building blocks

Cisco ACI consists of:

- Cisco Nexus 9000 Series Switches
- Centralized policy management and Cisco Application Policy Infrastructure Controller (APIC)

Cisco Nexus 9000 Series Switches

The Cisco Nexus 9000 Series Switches offer both modular (9500 switches) and fixed (9300 switches), 1/10/40/100 Gigabit Ethernet switch configurations designed to operate in one of two modes:

- Cisco NX-OS mode for traditional architectures and consistency across the Cisco Nexus portfolio.
- Cisco ACI mode to take full advantage of the policy-driven services and infrastructure automation features of ACI.

The ACI-Ready Cisco Nexus 9000 Series provides:

- Accelerated migration to 40G: zero cabling upgrade cost with Cisco QSFP+ BiDi Transceiver Module innovation.
- Switching platform integration: Nexus 9000 Series enables a highly scalable architecture and is software upgradable to ACI.
- Streamlined application management: drastically reduce application deployment time and get end-to-end application visibility.

This architecture consists of Cisco Nexus 9500 series switches acting as the spine, and Cisco Nexus 9300 series switches as leaves.
Cisco Nexus 9508 Spine Switch

The Cisco Nexus 9508 Switch offers a comprehensive feature set, high resiliency, and a broad range of 1/10/40 Gigabit Ethernet line cards to meet the most demanding requirements of enterprise, service provider, and cloud data centers. The Cisco Nexus 9508 Switch is an ACI modular spine device enabled by a non-blocking 40 Gigabit Ethernet line card, supervisors, system controllers, and power supplies.

The Cisco Nexus 9500 platform internally uses a Clos fabric design that interconnects the line cards with rear-mounted fabric modules. The Cisco Nexus 9500 platform supports up to six fabric modules, each of which provides up to 10.24-Tbps line-rate packet forwarding capacity. All fabric cards are directly connected to all line cards. With load balancing across fabric cards, the architecture achieves optimal bandwidth distribution within the chassis.

Figure 9  Cisco Nexus 9508 Switch

ACI Spine Line Card for Nexus 9508

There are multiple spine line cards supported on Cisco Nexus 9508. This architecture uses the N9K-X9736PQ: 40 Gigabit Ethernet ACI Spine Line Card.

- 36-port 40 Gigabit Ethernet QSFP+ line card
- Non-blocking
- Designed for use in an ACI spine switch role
- Works only in ACI mode
- Cannot mix with non-spine line cards
- Supported in 8-slot chassis
Cisco Nexus 9396 Leaf Switch

The Cisco Nexus 9396X Switch delivers comprehensive line-rate layer 2 and layer 3 features in a two-rack-unit (2RU) form factor. It supports line rate 1/10/40 GE with 960 Gbps of switching capacity. It is ideal for top-of-rack and middle-of-row deployments in both traditional and Cisco Application Centric Infrastructure (ACI)-enabled enterprise, service provider, and cloud environments.

Tenant: A tenant is a logical container or a folder for application policies. This container can represent an actual tenant, an organization, an application or can just be used for the convenience of organizing information. A tenant represents a unit of isolation from a policy perspective. All application configurations in Cisco ACI are part of a tenant. Within a tenant, you define one or more Layer 3 networks (VRF instances), one or more bridge domains per network, and EPGs to divide the bridge domains.

Application Profile: Modern applications contain multiple components. For example, an e-commerce application could require a web server, a database server, data located in a storage area network, and access to outside resources that enable financial transactions. An application profile models application requirements and contains as many (or as few) End Point Groups (EPGs) as necessary that are logically related to providing the capabilities of an application.

Bridge Domain: A bridge domain represents a L2 forwarding construct within the fabric. One or more EPG can be associated with one bridge domain or subnet. A bridge domain can have one or more subnets associated with it. One or more bridge domains together form a tenant network.

End Point Group (EPG): An End Point Group (EPG) is a collection of physical and/or virtual end points that require common services and policies. An End Point Group example is a set of servers or storage LIFs on a common VLAN providing a common application function or service. While the scope of an EPG definition is much wider, in the simplest terms an EPG can be defined on a per VLAN segment basis where all the servers or VMs on a common LAN segment become part of the same EPG.

Contracts: A service contract can exist between two or more participating peer entities, such as two applications running and communicating with each other behind different endpoint groups, or between providers and consumers, such as a DNS contract between a provider entity and a consumer entity. Contracts utilize filters to limit the traffic between the applications to certain ports and protocols.

Figure 12 illustrates the relationship between the ACI elements defined above. As shown in the figure, a Tenant can contain one or more application profiles and an application profile can contain one or more end point groups. The devices in the same EPG can talk to each other without any special configuration. Devices
in different EPGs can talk to each other using contracts and associated filters. A tenant can also contain one or more bridge domains and multiple application profiles and end point groups can utilize the same bridge domain.

Figure 12  AEP, Tenants, and Elements

Application Policy Infrastructure Controller (APIC)

The APIC is the unified point of automation, management, monitoring, and programmability for the Cisco Application Centric Infrastructure. The APIC supports the deployment, management, and monitoring of any application anywhere, with a unified operations model for physical and virtual components of the infrastructure. The APIC programmatically automates network provisioning and Control that is based on the application requirements and policies. It is the central control engine for the broader cloud network; it simplifies management and allows flexibility in how application networks are defined and automated. It also provides northbound REST APIs. The APIC is a distributed system that is implemented as a cluster of many controller instances.
Cisco ACI Topology

Cisco ACI topology is spine-leaf architecture. Each leaf is connected to each spine. It uses internal routing protocol; Intermediate System to Intermediate System (IS-IS) to establish IP connectivity throughout the fabric among all the nodes including spine and leaf. To transport tenant traffic across the IP fabric, integrated VxLAN overlay is used. The broadcast ARP traffic coming from the end point or hosts to the leaf are translated to unicast ARP in the fabric.

The forwarding is done as a host based forwarding. In the leaf layer the user information such as username, IP address, locations, policy groups etc., are decoupled from the actual forwarding path and encode them into the fabric VxLAN header and is forwarded to the desired destination.

Each spine has the complete forwarding information about the end hosts that are connected to the fabric and on every leaf have the cached forwarding information. The leaf only needs to know the hosts it needs to talk to. For example, if Server Rack-1 has to send some information to Server Rack-2, When packet comes in the ingress leaf (LEAF_1) it will encapsulate the information into the VxLAN header and forward that information to LEAF_2. If the LEAF_1 does not have information about the LEAF_2, it uses Spine as a proxy and since Spine has all the complete information about the entire end host connected to the fabric, it will resolve the egress leaf and forward the packet to the destination.

To the outside world, routing protocols can be used to learn outside prefixes or static routing can be used instead. The outside learned routes will be populated into the fabric or to the other leaves with Multiprotocol BGP (M-BGP). In M-BGP topology the spine nodes acts as route reflectors.

Figure 14 illustrates the Network topology of ACI.
The Cisco ACI infrastructure incorporates the following components:

- Two Cisco Nexus 9508 Spine Switch
- Cisco ACI Spine Line Card for Nexus 9508
- Cisco Nexus 9396 Leaf Switch for Data Traffic
- Cisco APIC-L1-Cluster with three APIC-L1 appliances

Cisco UCS Datacenter Solution for SAP HANA

Cisco and SAP have partnered to deliver an optimized UCS architecture for running SAP HANA, which provides fast transaction processing with real-time insights. Cisco UCS provides high-bandwidth connectivity between SAP HANA nodes and the persistency layer; this also allows SAP HANA deployments to scale more easily and transparently. Further, the Cisco UCS technology allows customers to scale dynamically as requirements and demand change.

Running SAP HANA on the Cisco UCS server platform offers the opportunity to reduce the hardware and maintenance costs associated with running multiple data warehouses, operational systems, and analytical systems. A principal design element of Cisco UCS is to break away from old static IT datacenter models and deliver on a new IT model that pools server, storage, and networking resources into a flexible physical and/or virtualized environment that can be provisioned (or reprovisioned) as workloads and business demands require.

This design guide provides an opportunity to integrate with any of the existing Cisco Datacenter Solution for SAP HANA CVD). The following are the Cisco UCS based design guides that can be used in conjunction with this CVD.
<table>
<thead>
<tr>
<th>SAP Solutions on Cisco UCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. FlexPod Datacenter for SAP solution with Cisco Nexus 9K</td>
</tr>
<tr>
<td>2. FlexPod Datacenter for SAP Solution with Cisco ACI (Used in this CVD for reference)</td>
</tr>
<tr>
<td>3. Cisco UCS Integrated Infrastructure for SAP Applications with EMC VNX</td>
</tr>
</tbody>
</table>

Reference Architecture: Flexpod Datacenter for SAP Solution with Cisco ACI

The FlexPod Datacenter solution for SAP HANA with NetApp FAS storage provides an end-to-end architecture with Cisco, NetApp and VMware technologies that demonstrate support for multiple SAP HANA workloads with high availability and server redundancy. The architecture uses Cisco UCS Manager with combined Cisco UCS B-Series and C-Series Servers with NetApp FAS 8000 series storage attached to the Nexus 9396PX switches for NFS access and iSCSI. The Cisco UCS C-Series Rack Servers are connected directly to Cisco UCS Fabric Interconnect with single-wire management feature. This infrastructure is deployed to provide PXE and iSCSI boot options for hosts with file-level and block-level access to shared storage. VMware vSphere 5.5 is used as server virtualization architecture.

Figure 15 shows the FlexPod Datacenter for SAP Solution with ACI. It highlights the FlexPod hardware components and the network connections for a configuration with IP-based storage.
Figure 15  FlexPod Datacenter for SAP Solution with ACI

Cisco UCS
5108 B-Series Blade Chassis
2208XP Chassis FEX Modules
B260 M4

Cisco UCS
5108 B-Series Blade Chassis
2208XP Chassis FEX Modules
B460 M4 B-Series Blades

Cisco UCS
6248UP Fabric Interconnects

Cisco Nexus
9336 Spine Switches

Cisco ACI
APIC Controllers

Cisco Nexus
9396 Leaf Switches

Cisco ACI
Controller(s)x3

NetApp FAS8040 Storage Controllers
w/ HA Backplane

Cisco Nexus 5596 Cluster Interconnects

NetApp
DS2240 Disk Shelves

Legend
--- SAS only  --- 10GbE Only  --- 40GbE Only
The reference hardware configuration includes:

**Cisco Unified Computing System**

- 2 x Cisco UCS 6248UP 48-Port or 6296UP 96-Port Fabric Interconnects
- 2 x Cisco UCS 5108 Blade Chassis with 2 x Cisco UCS 2204 Fabric Extenders with 4x 10 Gigabit Ethernet interfaces
- 2 x Cisco UCS B460 M4 High-Performance Blade Servers with 2x Cisco UCS Virtual Interface Card (VIC) 1280 and 2x Cisco UCS Virtual Interface Card (VIC) 1240
- 2 x Cisco UCS B260 M4 High-Performance Blade Servers with 1x Cisco UCS Virtual Interface Card (VIC) 1280 and 1x Cisco UCS Virtual Interface Card (VIC) 1240
- 1 x Cisco UCS C460 M4 High-Performance Rack-Mount Servers with 2x Cisco UCS Virtual Interface Card (VIC) 1225.
- 4 x Cisco UCS B200 M4 High-Performance Blade Servers with Cisco UCS Virtual Interface Card (VIC) 1340
- 1 x Cisco UCS C220 M4 High-Performance Blade Servers with Cisco UCS Virtual Interface Card (VIC) 1225
- 1 x Cisco UCS C240 M4 High-Performance Blade Servers with Cisco UCS Virtual Interface Card (VIC) 1225
- 2 x Cisco UCS C220 M3 for Management Servers with Cisco UCS Virtual Interface Card (VIC) 1225 and RAID controller with Internal Disks
- 2 x Cisco UCS 5108 Blade Chassis with 2 x Cisco UCS 2204 Fabric Extenders with 4x 10 Gigabit Ethernet interfaces
- 2 x Cisco UCS B460 M4 High-Performance Blade Servers with 2x Cisco UCS Virtual Interface Card (VIC) 1280 and 2x Cisco UCS Virtual Interface Card (VIC) 1240
- 2 x Cisco UCS B260 M4 High-Performance Blade Servers with 1x Cisco UCS Virtual Interface Card (VIC) 1280 and 1x Cisco UCS Virtual Interface Card (VIC) 1240
- 1 x Cisco UCS C460 M4 High-Performance Rack-Mount Servers with 2x Cisco UCS Virtual Interface Card (VIC) 1225.
- 4 x Cisco UCS B200 M4 High-Performance Blade Servers with Cisco UCS Virtual Interface Card (VIC) 1340
- 1 x Cisco UCS C220 M4 High-Performance Blade Servers with Cisco UCS Virtual Interface Card (VIC) 1225
- 1 x Cisco UCS C240 M4 High-Performance Blade Servers with Cisco UCS Virtual Interface Card (VIC) 1225
- 2 x Cisco UCS C220 M3 for Management Servers with Cisco UCS Virtual Interface Card (VIC) 1225 and RAID controller with Internal Disks
Cisco ACI

- 2 x Cisco Nexus 9396 Leaf Switch for 10 Gigabit Ethernet connectivity between the two UCS Fabric Interconnects
- 2 x Cisco Nexus 9508 Spine Switch for 40 Gigabit Ethernet connectivity for ACI fabric
- 3 x Cisco APIC Controllers for centralized management of ACI fabric

NetApp FAS8040 Storage

- NetApp FAS8040HA Storage Clustered Data ONTAP
- 4 x NetApp Disk Shelf DS2246 with 24x 600GB 10k 2,5" SAS Disks
- 2 x Cisco Nexus 5596 Switch for FAS 8000 Cluster Interconnect
- Server virtualization is achieved by VMware vSphere 5.5.

Although this is the base design, each of the components can be scaled easily to support specific business requirements. Additional servers or even blade chassis can be deployed to increase compute capacity without additional Network components. Two Cisco UCS 6248UP, 48 port Fabric interconnect can support up to:

- 20 Cisco UCS B-Series B460 M4 or 40 B260 M4 Server with 10 Blade Server Chassis
- 20 Cisco UCS C460 M4 Sever
- 40 Cisco UCS C220 M4/C240 M4 Server

For every eight Cisco UCS Servers, One NetApp FAS8040 HA pair with Clustered Data ONTAP is required to meet the SAP HANA storage requirement. While adding compute and storage for scaling, it is required to increase the network bandwidth between Cisco UCS Fabric Interconnect and Cisco Nexus 9000 switch. Addition of each NetApp Storage requires additional four 10 GbE connectivity from each Cisco UCS Fabric Interconnect to Cisco Nexus 9000 switches.

The number of Cisco UCS C-Series or Cisco UCS B-Series Servers and the NetApp FAS storage type depends on the number of SAP HANA instances. SAP specifies the storage performance for SAP HANA, based on a per server rule independent of the server size. In other words, the maximum number of servers per storage will remain the same if you want to use Cisco UCS B200 M4 with 192GB physical memory or Cisco UCS B460 M4 with 2TB physical memory.

Figure 16 shows a block diagram of a complete SAP Landscape built using the FlexPod architecture. It composed of multiple SAP HANA systems and SAP applications with shared infrastructure as illustrated in the figure. The FlexPod Datacenter reference architecture for SAP solutions supports SAP HANA system in both Scale-Up mode (bare metal/ virtualization) and Scale-Out mode with multiple servers with the shared infrastructures.

Virtualized SAP application servers with VMware vSphere 5.5 allows application servers to run on the same infrastructure as the SAP HANA database. The FlexPod datacenter solution manages the communication between the application server and the SAP HANA database. This approach enhances system performance.
by improving bandwidth and latency. It also improves system reliability by including the application server in the disaster-tolerance solution with the SAP HANA database.

**Figure 16  Shared Infrastructure Block Diagram**

FlexPod Datacenter for SAP Solution with Cisco ACI describes detailed procedures for the reference design and outlines the network, compute and storage configurations and deployment process for running SAP HANA on FlexPod platform.

**MapR Converged Data Platform 5.1**

As one of the technology leaders in Hadoop, the MapR Converged Data Platform provides enterprise-class big data solutions that are fast to develop and easy to administer. With significant investment in critical technologies, MapR offers one of the industry’s most comprehensive Hadoop platforms, fully optimized for performance and scalability. MapR’s distribution delivers more than a dozen tested and validated Hadoop software modules over a fortified data platform, offering exceptional ease of use, reliability and performance for big data solutions.
Features of MapR Converged Data Platform are:

- **Performance** – Ultra-fast performance and throughput
- **Scalability** – Up to a trillion files, with no restrictions on the number of nodes in a cluster
- **Standards-based API’s and tools** – Standard Hadoop API’s, ODBC, JDBC, LDAP, Linux PAM, and more
- **MapR Direct Access NFS** – Random read/write, real-time data flows, existing non-Java applications work seamlessly
- **Manageability** – Advanced management console, rolling upgrades, REST API support
- **Integrated security** – Kerberos and non-Kerberos options with wire-level encryption
- **Advanced multi-tenancy** – Volumes, data placement control, job placement control, queues, and more
- **Consistent snapshots** – Full data protection with point-in-time recovery
- **High availability** – Ubiquitous HA with a no-NameNode architecture, YARN HA, NFS HA
- **Disaster recovery** – Cross-site replication with mirroring
- **MapR-DB** – Integrated enterprise-grade NoSQL database
- **MapR Streams** – Global publish-subscribe event streaming system for Big Data

**MapR Enterprise-Grade Platform Services**

MapR Platform Services (Figure 17) are the core data handling capabilities of the MapR Converged Data Platform. Modules include MapR-FS, MapR-DB and MapR Streams. Its enterprise-friendly design provides a familiar set of file and data management services, including a global namespace, high availability, data protection, self-healing clusters, access control, real-time performance, secure multi-tenancy, and management and monitoring.

**Figure 17  MapR Enterprise-grade Platform Services**
Enterprise Storage

MapR-FS is an enterprise standard POSIX file system that provides high-performance read/write data storage for the MapR Converged Data Platform. MapR-FS includes important features for production deployments such as fast NFS access, access controls, and transparent data compression at a virtually unlimited scale.

Database

MapR-DB is an enterprise-grade, high performance, in-Hadoop NoSQL database management system. It is used to add real-time, operational analytics capabilities to applications built on the Hadoop or Spark ecosystems. Because it is integrated into the MapR Converged Data Platform, it inherits the protections and high performance capabilities.

Event Streaming

MapR Streams is a global publish-subscribe event streaming system for big data. It connects data producers and consumers worldwide in real-time, with unlimited scale. MapR Streams is the first big data-scale streaming system built into a converged data platform. It makes data available instantly to stream processing and other applications, and is the only big data streaming system to support global event replication reliably at IoT scale.

MapR Streams: Event Streaming on a Global Scale

Many big data sources are continuous flows of data in real time: sensor data, log files, transaction data to name just a few. Enterprises are struggling to deal with the high volume and high velocity of the data using existing bulk data-oriented tools.

MapR Streams (Figure 18) manages streaming data for real-time processing with enterprise-grade security and reliability at a global scale. It connects data producers and consumers worldwide in real time, with unlimited scale. MapR Streams scales to billions of events per second, millions of topics, and millions of producer and consumer applications. Geographically dispersed MapR clusters can be joined into a global fabric, passing event messages between producer and consumer applications in any topology, including one-to-one and many-to-many.

This centralized architecture provides real-time access to streaming data for batch or interactive processing on a global scale with enterprise features including secure access-control, encryption, cross data center replication, multi-tenancy and utility-grade uptime.
MapR Streams makes data available instantly to stream processing and other applications, providing:

- Kafka API for real-time producers and consumers for easy application migration.
- Out-of-the-box integration with popular stream processing frameworks like Spark Streaming, Storm and Flink.

MapR Streams globally replicates event data at IoT-scale with:

- Arbitrary topology supporting thousands of clusters across the globe. Topologies of connected clusters include one-to-one, one-to-many, many-to-one, many-to-many, star, ring and mesh. Topology loops are automatically handled to avoid data duplication.
- Global metadata replication. Stream metadata is replicated alongside data, allowing producers and consumers to failover between sites for high availability. Data is spread across geographically distributed locations via cross-cluster replication to ensure business continuity should an entire site-wide disaster occur.

**MapR Open Source Technologies**

MapR packages a broad set of Apache open source ecosystem projects that enable big data applications. The goal is to provide an open platform that provides the right tool for the job. MapR tests and integrates open source ecosystem projects such as Spark, Drill, Solr, HBase, among others. MapR is the only Hadoop vendor that supports multiple versions of key Apache projects providing more flexibility in updating the environment.
Figure 19 shows the Apache open source projects supported by the MapR Converged Data Platform. Features of some of the key technologies are highlighted below. In conjunction with the data ingestion capabilities provided by MapR Streams these technologies are building blocks for a system based on the Lambda Architecture.

Apache Spark

Apache Spark is a fast and general-purpose engine for large-scale data processing. By adding Apache Spark to the Hadoop deployment and analysis platform, and running it all on Cisco UCS Integrated Infrastructure for Big Data and Analytics, customers can accelerate streaming, interactive queries, machine learning, and batch workloads, and offer their user’s experiences that deliver more insights in less time.

Traditional servers are not designed to support the massive scalability, performance, and efficiency requirements of Big Data solutions. These outdated and siloed computing solutions are difficult to integrate with network and storage resources, and are time-consuming to deploy and expensive to operate. Cisco UCS Integrated Infrastructure for Big Data and Analytics with Apache Spark takes a different approach, combining computing, networking, storage access, and management capabilities into a unified, fabric-based architecture that is optimized for Big Data workloads.

Apache Spark enhances existing Big Data environments by adding new capabilities to Hadoop or other Big Data deployments. The platform unifies a broad range of capabilities—batch processing, real-time stream processing, advanced analytic capabilities, and interactive exploration that can intelligently optimize applications. Spark’s key advantage is speed, with an advanced DAG execution engine that supports cyclic data-flow and in-memory computing. It can run programs much faster than Hadoop/Map-Reduce. Applications can be developed using built-in, high-level Apache Spark operations, or they can interact with applications like Python, R, and Scala shells, or Java. These various options allow users to quickly and easily build new applications and explore data faster.
Apache Spark delivers the rapid response that is needed by real-time interactive applications, and experimentation environments. An important factor in the solution's performance is the way Apache Spark performs operations, most of which are done in memory. Spark provides programmers with any application interface, centered on a data structure called the resilient distributed dataset (RDD), a read-only multiset of data items distributed over a cluster of machines, that is maintained in a fault-tolerant way. Calculations are performed and results are delivered only when needed, and results can be configured to persist in memory, allowing Apache Spark to deliver a new level of computing efficiency and computation performance to Big Data deployments.

Apache Spark has a number of libraries:

- **Apache Spark SQL/DataFrame API** for querying structured data inside Spark programs.
- **Apache Spark Streaming** offers Spark's core API that is able to perform real-time processing of streaming data, including web server log files, social media, and messaging queues.
- **MLLib** to take advantage of machine-learning algorithms and accelerate application performance across clusters.
- **GraphX** unifies ETL, performs exploratory analysis, and accelerates iterative graphical computations in a single system.

Spark runs on Hadoop, Mesos, stand alone, or in the cloud. It can access diverse data sources including HDFS, Cassandra, HBase, and S3. Spark with YARN is an optimal way to schedule and run Spark jobs on a Hadoop cluster alongside a variety of other data-processing frameworks, leveraging existing clusters using queue placement policies, and enabling security by running on Kerberos-enabled clusters.

Some common use cases that are popular in the field with Apache Spark:

- **Simpler, faster, ETL** – Data can be processed into the required format by avoiding intermediate writes to disk, and cleaned and aggregated in-memory before the final disk write.
- **Real-Time Actions** – Anomalous behaviors detected in real-time, and downstream actions are processed accordingly. For example; credit card transactions occurring in a different location generating actions for fraud alert, IOT sensors transmitting device failure data, etc.
- **Data Enrichment** – Live data is enriched with more information by joining it with cached static datasets, allowing for a more comprehensive feature set in real-time.
- **Exploratory Analytics** – Events related to a specific time-window can be grouped together and analyzed. This sample data can be used by Data Scientists to update machine-learning models using tools like Python, etc. within Spark.
- **Streaming Data with Analytics** – The same code for streaming analytic operations can be used for batch, to compute over both the stream and historical data. This reduces moving parts and helps increase the productivity, consistency, and maintainability of analytic procedures. Spark is compatible with the rest of the streaming data ecosystem, supporting data sources including Flume, Kafka, ZeroMQ, and HDFS.
- **Model Building and Machine Learning** – Spark is a Big Data tool that data scientists find easy to use which makes it ideal for building models for analytical purposes. Offline model building which needs
data transfer from a Hadoop environment can be avoided now that Spark is used for model building and deployment.

- **Graph Analysis** - By incorporating the GraphX component, Spark brings all the benefits of its environment to graph computation; enabling use cases such as social network analysis, fraud detection, recommendations, and entity relationships, etc.

**Spark Streaming**

Spark Streaming brings Spark's **language-integrated API** to stream processing. The API is provided in Java, Scala, and Python. Spark’s single execution engine, and unified programming model for batch and streaming, lead to some unique benefits over other traditional streaming systems.

- Fast recovery from failures and stragglers.
- Better load balancing and resource usage.
- Combining streaming data with static datasets and interactive queries.
- Native integration with advanced processing libraries (SQL, machine learning, graph processing).

In Spark Streaming, batches of Resilient Distributed Datasets (RDDs) are passed to Spark Streaming, which processes these batches using the Spark Engine and returns a processed stream of batches. This processed stream can be written to the file system. Spark Streaming allows stateful computations, maintaining a state based on data coming in a stream. It also allows window operations (i.e., allows the developer to specify a time frame, and perform operations on the data flowing in that time window. The window has a sliding interval, which is the time interval of updating the window.

Each batch of data is a Resilient Distributed Dataset (RDD), which is the basic abstraction of a fault-tolerant dataset in Spark. This common representation allows batch and streaming workloads to interoperate seamlessly. Users can apply arbitrary Spark functions on each batch of streaming data: for example, it’s easy to join a DStream (key programming abstraction in Spark Streaming) with a precomputed static dataset (such as an RDD). Spark interoperability extends to rich libraries like MLlib (machine learning), SQL, and DataFrames.

Machine learning models generated offline with MLlib can be applied on streaming data. Fault tolerance in Spark Streaming is similar to fault tolerance in Spark. Like RDD partitions, DStream data is recomputed in case of a failure. The raw input is replicated in memory across the cluster. In case of a node failure, the data can be reproduced using the lineage. Spark Streaming is a streaming platform and allows reaching sub-second latency. The processing capability scales linearly with the size of the cluster; hence it is being used in production by many organizations.

**SAP HANA Vora**

SAP HANA Vora™ is an in-memory query engine that plugs into the Apache Spark execution framework to provide enriched interactive analytics on Hadoop. SAP HANA Vora extends the HANA-like analytics experience to ALL data. SAP HANA Vora plugs into the Apache SPARK framework which is part of Apache Hadoop, and allows us to bring OLAP-like analytics and business semantics of the data in and around the Hadoop ecosystem. This is important - to reach meaningful contextual information when new unstructured data such as data from IoT sensors, machine telemetry or from social media, comes together with business
data such as financial records, business goals, maintenance records, and employment data. It is only when these two different data sets meet, that business meaning is made. Meaningful business results require that we embrace ALL data, in a contextual way, to drive analytics driven outcomes.

SAP HANA Vora provides the following features:

- In-memory query engine running on Apache Spark execution framework
- Compiled queries for accelerated processing across nodes
- Enhanced Spark SQL semantics with hierarchies for analytical processing
- Enhanced mashup application programming interface (API) for easier access to enterprise application data for machine learning workloads

SAP HANA Vora can benefit customers in industries where highly interactive big data analytics in business process context is paramount, such as financial services, telecommunications, healthcare and manufacturing. Examples include:

- Mitigate risk and fraud by detecting new anomalies in financial transactions and customer history data.
- Optimize telecommunication bandwidth by analyzing traffic patterns to help avoid network bottlenecks and improve network quality of service (QoS).
- Deliver preventive maintenance and improve product re-call process by analyzing bill-of-material, services records and sensor data together.
Solution Design

Requirements

Physical Layout for the Solution

The physical layout for the solution is shown in Table 2. Each rack consists of two vertical PDUs. The solution consists of 2 Cisco R42610 racks. The Cisco Nexus 9396 leaf switch and the Fabric Interconnects are mounted on rack2, the APIC appliances are distributed across the racks. Similarly, the Cisco Nexus 9508 spine switch is mounted in rack2 for easier cabling between the spine and leaf switches. All the Switches and Cisco UCS Servers are dual connected to vertical PDUs for redundancy; thereby, ensuring high availability during power source failure.

<table>
<thead>
<tr>
<th>Table 2 Racks 1 and 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rack 1</strong></td>
</tr>
<tr>
<td>FI- B</td>
</tr>
<tr>
<td>FI- A</td>
</tr>
<tr>
<td>APIC- M1</td>
</tr>
<tr>
<td>APIC- M1</td>
</tr>
<tr>
<td>UCS C240M4</td>
</tr>
<tr>
<td>UCS C240M4</td>
</tr>
<tr>
<td>UCS C240M4</td>
</tr>
<tr>
<td>UCS C240M4</td>
</tr>
<tr>
<td>UCS C240M4</td>
</tr>
</tbody>
</table>
Software Distributions and Versions

The required versions of software distributions are listed below.

Red Hat Enterprise Linux (RHEL)

The operating system supported is Red Hat Enterprise Linux 7.2. For more information, please visit [http://www.redhat.com](http://www.redhat.com).
MapR

The MapR distribution used is MapR 5.1. For more information, please visit http://www.mapr.com.

SAP HANA Vora

The SAP HANA Vora used is version 1.3. For more information, please visit https://www.sap.com/product/data-mgmt/hana-vora-hadoop.html.

Software Versions

The software versions tested and validated in this document are shown in Table 3.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Software Versions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer</td>
<td>Component</td>
</tr>
<tr>
<td>Network</td>
<td>Cisco ACI OS</td>
</tr>
<tr>
<td></td>
<td>APIC OS</td>
</tr>
<tr>
<td>Compute</td>
<td>Cisco UCS 6296UP</td>
</tr>
<tr>
<td></td>
<td>Cisco UCS VIC1227 Firmware</td>
</tr>
<tr>
<td></td>
<td>Cisco UCS VIC1227 Driver</td>
</tr>
<tr>
<td>Storage</td>
<td>LSI SAS 3108</td>
</tr>
<tr>
<td>Software</td>
<td>Red Hat Enterprise Linux Server</td>
</tr>
<tr>
<td></td>
<td>Cisco UCS Manager</td>
</tr>
<tr>
<td></td>
<td>Cisco UCS Director Express for Big Data</td>
</tr>
<tr>
<td></td>
<td>MapR</td>
</tr>
<tr>
<td></td>
<td>MapR- Spark</td>
</tr>
<tr>
<td></td>
<td>SAP HANA Vora</td>
</tr>
</tbody>
</table>

The latest drivers can be downloaded from the link:
https://software.cisco.com/download/release.html?mdfid=283862063&flowid=25886&softwareid=283853158&release=1.5.7d&rellind=AVAILABLE&rellifecycle=&reltype=latest
The Latest Supported RAID controller Driver is already included with the RHEL 7.2 operating system.
System Architecture

The system architecture includes Cisco UCS C240 M4 servers, based on Cisco UCS Integrated Infrastructure for Big Data and Analytics.

The ACI fabric consists of three major components: The Application Policy Infrastructure Controller (APIC), spine switches, and leaf switches. These three components handle both the application of network policy and the delivery of packets.

The system architecture consists of a pair of FIs connecting to ACI having two N9508 switches acting as a Spine and two Cisco Nexus 9396 as the leaf switches and three APIC-M1 as an APIC appliance.

The following explains the system architecture:

- The 16 servers are rack mounted and are connected to a pair of FI representing a domain through 10GE link (dual 10GE link to a pair of FI)
- This Cisco UCS domain is connected to a pair of Cisco Nexus 9396 which is the ACI Fabric leaf nodes. Here 10GEx16 links from each FI are connected to Cisco Nexus 9396. This is done through a port-channel of 8 links connected to each of the Cisco Nexus 9396
- Cisco Nexus 9396 receives the 16x10GE from each pair of FI as a vPC (Virtual Port-Channel), i.e., 8 ports coming from each single FI as an uplink to the leaf. There are 2 vPC for this Cisco UCS domain in each of 9396 connecting to the pair of FIs.
- Each leaf is connected to Spines via 12 x 40 Gig connectivity cables.
- The three APIC's are connected to two leaves (Cisco Nexus 9396) via a 10 gig SFP cable.

Figure 20 shows the overall system architecture and physical layout of the solution.
Figure 20  System Architecture

Figure 21 show the connectivity between the leaf switches and fabric interconnect, where port channeling has been configured on Fabric Interconnect. This port channeling helps to aggregate the bandwidth towards the uplink leaf switches.
Figure 22 shows the connectivity between the leaf switches and fabric interconnect, where vPC has been configured on leaf switches through the APIC. These vPC ports are the same ports that were configured as port-channels in the fabric interconnect.

Figure 23 shows the connectivity between the one Cisco UCS C240 M4 servers and two Fabric.
Scaling the Architecture

The Cisco UCS Servers are directly connected to the Fabric Interconnect (FI), which connects to the Cisco Nexus 9K switches. This mode allows using the Cisco UCS Manager capabilities in FI for provisioning the servers.

Scaling the Architecture Further with Additional Spine Switches

The physical network of the Cisco Application Centric Infrastructure is built around leaf-spine architecture. It is possible to scale this infrastructure, immensely, by adding additional spine switches. The ACI infrastructure supports up to 12 spine switches.

Figure 23  Cisco UCS C240 M4 Server Connectivity

![Cisco UCS C240 M4 Server Connectivity Diagram]

Figure 24  Cisco ACI Fabric with Multiple Spine Switches

![Cisco ACI Fabric with Multiple Spine Switches Diagram]
With a 12-spine design, each leaf switch can be connected to up to 12 spine switches. Allowing for tens of thousands of servers to be part of this infrastructure – being interconnected by a non-blocking fabric.

**SAP HANA and SAP HANA VORA scalability**

The Base configuration 4 HANA Appliances (B460 M4/C460 M4 servers) + 16 C240 M4 (Cisco UCS Integrated Infrastructure for Big Data).

The recommended building block is made up of a set of 16 Cisco UCS C240 M4 servers for every two HANA servers.

The VORA tier can scale-out independent of the HANA tier if necessary.

<table>
<thead>
<tr>
<th>SAP HANA Tier (B460/C460)</th>
<th>SAP HANA Vora Tier (C240 M4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 servers</td>
<td>32 servers</td>
</tr>
<tr>
<td>8 servers</td>
<td>64 servers</td>
</tr>
<tr>
<td>16 servers</td>
<td>128 servers</td>
</tr>
</tbody>
</table>

**Network Configuration**

The network configuration using Cisco ACI can be found at: [http://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/UCS_CVDs/HaaS_on_Bare_Metal_with_UCS_DExpress_on_Cisco_UCS_Integrated_Infrastructure_for_Big_Data_and_ACI.html](http://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/UCS_CVDs/HaaS_on_Bare_Metal_with_UCS_DExpress_on_Cisco_UCS_Integrated_Infrastructure_for_Big_Data_and_ACI.html)

Follow the steps starting from the “Network Configuration” section until the “Fabric Configuration” section.

The link above is a reference guide to configure the network. The exact configuration may be different based on the deployment requirements. For this CVD, create only one tenant (the document referenced above describes how to create three tenants).

**IP Address Assignment**

The IP addresses for Cisco UCS and Cisco ACI management are configured as out of band management access through the management switch.

**APIC-1**
- 10.0.141.8/24 (Primary)

**APIC-2**
- 10.0.141.9/24

**APIC-3**
- 10.0.141.10/24

**Pod - 1**
- **UCSM** 10.0.141.20/24
- **FI-A** 10.0.141.21/24
- **FI-B** 10.0.141.22/24
- **KVM** 10.0.141.11/24 – 10.0.141.90/24
### Table 4 VLAN ID and IP Address

<table>
<thead>
<tr>
<th>VLAN ID</th>
<th>Tenant Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 (Mgmt)</td>
<td>172.16.10.0/24</td>
</tr>
</tbody>
</table>

### Fabric Configuration

This section provides the details for configuring a fully redundant, highly available Cisco UCS 6296 fabric configuration.

1. Initial setup of the Fabric Interconnect A and B.
2. Connect to UCS Manager using virtual IP address of using the web browser.
3. Launch Cisco UCS Manager.
4. Enable server and uplink ports.
5. Start discovery process.
6. Create pools and policies for service profile template.
7. Create Service Profile template and 64 Service profiles.
8. Associate Service Profiles to servers.

### Server Configuration and Cabling for Cisco UCS C240 M4 Rack Server

The Cisco UCS C240 M4 rack server is equipped with Intel Xeon E5-2680 v4 processors, 256 GB of memory, Cisco UCS Virtual Interface Card 1227, Cisco 12- Gbps SAS Modular Raid Controller with 2- GB FBWC, 24 1.8- TB 10K SFF SAS drives, 2 480- GB SATA SSD for Boot.

Figure 25 illustrates the port connectivity between the Fabric Interconnect and Cisco UCS C240 M4 server.
Cisco UCS Fabric Configuration

This section provides details for configuring a fully redundant, highly available Cisco UCS 6296 fabric configuration.

Configure Fabric Interconnect A

1. Connect to the console port on the first Cisco UCS 6296 Fabric Interconnect.
2. At the prompt to enter the configuration method, enter console to continue.
3. If asked to either perform a new setup or restore from backup, enter setup to continue.
4. Enter y to continue to set up a new Fabric Interconnect.
5. Enter y to enforce strong passwords.
6. Enter the password for the admin user.
7. Enter the same password again to confirm the password for the admin user.
8. When asked if this fabric interconnect is part of a cluster, answer y to continue.

10. Enter the cluster name for the system name.

11. Enter the Mgmt0 IPv4 address.

12. Enter the Mgmt0 IPv4 netmask.

13. Enter the IPv4 address of the default gateway.

14. Enter the cluster IPv4 address.

15. To configure DNS, answer y.

16. Enter the DNS IPv4 address.

17. Answer y to set up the default domain name.

18. Enter the default domain name.

19. Review the settings that were printed to the console, and if they are correct, answer yes to save the configuration.

20. Wait for the login prompt to make sure the configuration has been saved.

**Configure Fabric Interconnect B**

1. Connect to the console port on the second Cisco UCS 6296 Fabric Interconnect.

2. When prompted to enter the configuration method, enter console to continue.

3. The installer detects the presence of the partner Fabric Interconnect and adds this fabric interconnect to the cluster. Enter y to continue the installation.

4. Enter the admin password that was configured for the first Fabric Interconnect.

5. Enter the Mgmt0 IPv4 address.

6. Answer yes to save the configuration.

7. Wait for the login prompt to confirm that the configuration has been saved.

---

For more information on configuring Cisco UCS 6200 Series Fabric Interconnect, see:

Logging into Cisco UCS Manager

To log into Cisco UCS Manager, complete the following steps:

1. Open a Web browser and navigate to the Cisco UCS 6296 Fabric Interconnect cluster address.
2. Click the Launch link to download the Cisco UCS Manager software.
3. If prompted to accept security certificates, accept as necessary.
4. When prompted, enter admin for the username and enter the administrative password.
5. Click Login to log in to the Cisco UCS Manager.

Enabling Server Ports

1. Select the Equipment tab on the top left of the window.
3. On the Right window select all the ports that are connected to the UCS C240 server (1 per Server), right-click them, and select > Configure as Server Port.
Enabling Uplink Ports

1. Select the Equipment tab on the top left of the window.


3. On the Right window select all the ports that are connected to the Cisco Nexus 9396 leaf switch (16 per FI), right-click them, and select > Configure as uplink Port.

5. Expand the Unconfigured Ethernet Ports section.

6. Select all the ports that are connected to the Nexus 9396 leaf switch (16 per FI), right-click them, and select > Configure as uplink Port.

   The ports that are configured as uplink port should appear as Network under IF role.

Configuring Port-Channels

1. Select the LAN tab on top left window.

2. Expand the LAN Cloud > Fabric A.

3. On the right window select Create Port Channel.
4. On Set Port Channel Name window, perform the following actions:

   a. In the ID field, specify the ID “01” as the first port channel.
   b. In Name field, type P01 for Port-channel01 and click Next.

5. In the Add Ports window select all the ports that are connected to the Nexus 9396 Leaf Switch and click >>. This will add all the ports to the port channel created earlier.
6. Click Finish.

**Adding Block of IP Addresses for KVM Access**

To create a block of KVM IP addresses for server access in the Cisco UCS environment, complete the following steps:
1. Select the LAN tab at the top of the left window.


3. Right-click IP Pool ext-mgmt.

4. Select Create Block of IPv4 Addresses.

5. Enter the starting IP address of the block and number of IPs needed, as well as the subnet and gateway information.

6. Click OK to create the IP block.

7. Click OK in the message box.
Creating MAC Address Pools

To create MAC address pools, complete the following steps:

1. Select the LAN tab on the left of the window.
2. Select Pools > root.
3. Right-click MAC Pools under the root organization.
4. Select Create MAC Pool to create the MAC address pool. Enter ucs for the name of the MAC pool.
5. (Optional) Enter a description of the MAC pool.
6. Select Assignment Order Sequential.
7. Click Next.
8. Click Add.
9. Specify a starting MAC address.
10. Specify a size of the MAC address pool, which is sufficient to support the available server resources.
11. Click OK.
12. Click Finish.

13. When the message box displays, click OK.

Creating Server Pool

A server pool contains a set of servers. These servers typically share the same characteristics. Those characteristics can be their location in the chassis, or an attribute such as server type, amount of memory, local storage, type of CPU, or local drive configuration. You can manually assign a server to a server pool, or use server pool policies and server pool policy qualifications to automate the assignment
To configure the server pool within the Cisco UCS Manager GUI, complete the following:

1. Select the Servers tab in the left pane in the UCS Manager GUI.

2. Select Pools > root.

3. Right-click the Server Pools.

4. Select Create Server Pool.

5. Enter your required name (ucs) for the Server Pool in the name text box.

6. (Optional) enter a description for the organization

7. Click Next > to add the servers.

8. Select all the Cisco UCS C240M4SX servers to be added to the server pool you previously created (ucs), then Click >> to add them to the pool.

9. Click Finish.

10. Click OK and then click Finish.
To configure communication services for Cisco UCS Manager, complete the following steps:

1. Log in to UCS Manager and click the Admin tab.


3. In the right window, in Web Session Limits enter 256 for Maximum Sessions Per User.
Configuring QoS System Class

To set jumbo frames and enable QoS, complete the following steps:

1. Select the **LAN** tab in the left pane in the UCSM GUI.
2. Select **LAN Cloud > QoS System Class**.
3. In the right pane, select the **General** tab.
4. In the **Platinum** row, enter 9216 for MTU.
5. Check the **Enabled** check box next to **Platinum**.
6. In the **Best Effort** row, select **none** for weight.
7. In the **Fiber Channel** row, select **none** for weight.
8. Click **Save Changes**.
9. Click **OK**.
Install and Configure Hadoop, YARN, and Spark

Hadoop installation has been automated using cisco UCS Director Express for Big Data. More details on deploying the tool can be found here:

⚠️ Please follow the steps from section “Cisco UCS Director Express Management Server Configuration” through the “Configure the Bare Metal Agent’s DHCP Services” section.

When the Cisco UCS Director Express for Big Data has been deployed along with the necessary services and licenses, it can be accessed using the GUI (http://<UCSD-VM’s IP>/).

Creating the Hadoop Cluster Using Cisco UCS Director Express for Big Data

To create a Hadoop cluster of a distribution, the Cisco UCS Manager managing the target servers must be pre-configured to meet the following requirements. For performing these configurations, refer to any Cisco UCS Integrated Infrastructure for Big Data Cisco Validated Design found at http://www.cisco.com/go/bigdata_design.

1. The uplink ports fabric interconnects must be reachable by the UCSD-Express appliances management network (i.e. eth0).
2. The Cisco UCS Manager must be configured with a host firmware policy containing C-series rack mount server firmware packages.
3. The Cisco UCS Manager must be configured to discover the Rack Servers in its domain, and the respective ports are configured as server ports.
4. The server pool must be configured with appropriate set of physical servers that are part of the UCS domain.
5. The QOS System Classes Platinum and Best Effort must be configured and enabled.

Create the IP Address Pools

1. Using a web browser, visit the URL http://<UCSD-VM’s IP>/.
2. Login as user admin with the default password admin.
4. Click on the Big Data IP Pools Tab.
5. Click <+ Add>.
6. In the Create an IP Pool dialog box.

7. Enter the name Prod_Mgmt. Click Next to continue.

8. In the IPv4 Blocks table 9, click <+>.

9. In the Add Entry to IPv4 Blocks dialog box, enter the following.

10. In the Static IP Pool field, enter the Static IP Address pool range in the format A.B.C.X – A.B.C.Y.
11. In the Subnet Mask field, enter the appropriate subnet mask.

12. In the Default Gateway field, enter the IP address of the Gateway.

13. In the Primary DNS field, enter the IP address of the DNS server.

14. Click Submit.

![Add Entry to IPv4 Blocks](image)

The Default Gateway, Primary and Secondary DNS fields are optional.

15. Click Submit again to create the Big Data IP Pool.

Creating a Hadoop Cluster

1. Using a web browser, visit the URL [http://<UCSD-VM’s IP>/](http://<UCSD-VM’s IP>/).

2. Login as user admin with the default password admin.

3. Navigate to Solutions → Big Data Containers.

4. Select the QUICK_UCS template and click Edit.

![Containers](image)

5. In the edit window check the “Use one vNIC” and click next and click Submit.
6. Repeat the above step to make use of only one vNIC for template QUICK_UCS_MAPR.

7. Click the Cluster Deploy Templates Tab.

8. Click Create Instant Hadoop Cluster.

9. In the Instant Hadoop Cluster Creation dialog box, enter the following.

10. In Big Data Account Name field, enter a preferred name.

11. In the Cisco UCS Manager Policy Name Prefix field, enter a prefix that is less than equal to 5 letters long.

12. In the Hadoop Cluster Name field, enter a preferred name of the cluster – this will be the name assigned to the Hadoop cluster within the context of the selected Hadoop Manager.

13. In the Hadoop Node Count filed, enter the desired number of nodes.

14. In the password fields, enter the preferred passwords and confirm them.

15. Choose the RHEL7OS Version from the drop-down box. For C220 M4/C240 M4 rack servers.

16. In the Hadoop Distribution field, select the MapR from the drop-down list.
17. In the Hadoop Distribution Version field, select MapR-5.1.0 version from the drop down list.

18. In the Cisco UCS Manager Account, select the appropriate UCS-Manager account.

19. Select the organization.

21. In the new Edit window, enter 16 for the Node Count and select rack11;org-root for the server pool.
22. Click Submit.

23. For the vNIC Template, double-click on row eth0 and select appropriate Mgmt IP pool, MAC Address Pool and enter the MGMT VLAN id. Click Submit.

24. Click Submit to start provisioning the cluster.

Monitoring the Hadoop Cluster Creation

1. In the UCSD-Express web console, navigate to Organization > Service Requests.
2. Browse through the workflows. There are 3 types of workflows executed:
   a. There would be one Master Workflow, for example, Cisco UCS CPA Multi-UCSM Hadoop cluster WF, per Hadoop cluster creation request. Master workflow kick-starts one or more UCSM-specific workflows. This master workflow is also responsible for Hadoop cluster provisioning.
   b. UCSM specific workflows, for example, Single UCSM Server Configuration WF, would in turn kick start one or more UCS CPA Node Bare metal workflows.
   c. Cisco UCS CPA Bare metal workflows provision the UCS service profiles and perform OS installation and custom configuration per node.

3. Double-click one of the master workflows for example, UCS CPA Multi-UCSM Hadoop Cluster, to view the various steps undertaken to provision a Hadoop cluster.
If necessary click the Log tab to view the logs generated during the provisioning of the Hadoop Cluster.

4. Double-click one of the child workflows: for example, UCS CPA Node Bare metal.
<table>
<thead>
<tr>
<th>Service Request Status</th>
<th>Current status for the service request</th>
<th>Date/Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>Initiated by admin</td>
<td>10/26/2016 09:46:44</td>
</tr>
<tr>
<td>Advanced</td>
<td>Modify Workflow Priority (High)</td>
<td>10/26/2016 09:46:50</td>
</tr>
<tr>
<td>3.0</td>
<td>Acquire DB Lock on Cluster</td>
<td>10/26/2016 09:46:53</td>
</tr>
<tr>
<td>Complete</td>
<td>Record current SR id for UCS Server</td>
<td>10/26/2016 09:46:57</td>
</tr>
<tr>
<td>UCSMi8sa/rack-unit-51</td>
<td>Release DB lock on Cluster</td>
<td>10/26/2016 09:46:59</td>
</tr>
<tr>
<td>admin</td>
<td>Create UCS Service Profile from template</td>
<td>10/26/2016 09:47:08</td>
</tr>
<tr>
<td></td>
<td>Unbind UCS Service Profile from Template</td>
<td>10/26/2016 09:47:09</td>
</tr>
<tr>
<td></td>
<td>Modify UCS Service Profile Boot Policy</td>
<td>10/26/2016 09:47:12</td>
</tr>
<tr>
<td></td>
<td>Unbind eth0 vNIC Template</td>
<td>10/26/2016 09:47:16</td>
</tr>
<tr>
<td></td>
<td>Binding PXE vNIC Template to eth0</td>
<td>10/26/2016 09:47:18</td>
</tr>
<tr>
<td></td>
<td>Associate UCS Service Profile</td>
<td>10/26/2016 10:00:28</td>
</tr>
<tr>
<td></td>
<td>Acquire DB Lock on Cluster</td>
<td>10/26/2016 10:00:29</td>
</tr>
<tr>
<td></td>
<td>Reserve UCS Server</td>
<td>10/26/2016 10:00:32</td>
</tr>
<tr>
<td></td>
<td>Release DB lock on Cluster</td>
<td>10/26/2016 10:00:35</td>
</tr>
<tr>
<td></td>
<td>Setup PXE Boot [CentOSLive]</td>
<td>10/26/2016 10:00:51</td>
</tr>
<tr>
<td></td>
<td>Acquire DB Lock on Cluster</td>
<td>10/26/2016 10:00:54</td>
</tr>
<tr>
<td></td>
<td>Setup RAID Device</td>
<td>10/26/2016 10:04:07</td>
</tr>
</tbody>
</table>
Logging into the Hadoop UI

1. When the cluster creating is complete and to log in to the Hadoop UI, go to Solutions > Big Data > Accounts.

2. Select the Account that was created earlier and select Launch Hadoop Manager.

Or

3. Using the web browser log in to the UI using the IP address where the MapR Control System has been installed (https://<IP Address>:8443).
Installing Spark

SAP HANA Vora™ is an in-memory query engine that plugs into the Apache Spark execution framework to provide enriched interactive analytics on Hadoop. The admin node (rhel1) will be configured to serve as the Vora client node as well. Thus, the following packages are necessary.

- Scala programming language platform
- MapR-Spark 1.6.1
- SAP HANA Vora Extension packages.

This section provides the steps necessary to perform the first two parts. The Vora Extension packages will be installed at a later stage.

Installing Scala

1. Download the latest release of Scala RPM from the scala-lang.org website (http://www.scala-lang.org/download/all.html) using the command below

   \[ \text{wget } \text{http://downloads.lightbend.com/scala/2.11.8/scala-2.11.8.rpm} \]

2. Copy the rpm file to all the nodes using the command below

   \[ \text{clush -a -b -c scala-2.11.8.rpm} \]
3. Install the Scala language platform using the following command.

   ```
   clush -a -b "rpm -ivh scala-2.11.8.rpm"
   ```

Create a dedicated Vora user

Create a new user by name “vora” of group “vora”. When adding a user to the cluster nodes, make sure the user ID (UID) is always the same. The same applies to the Group ID (GID).

   ```
   clush -a -b groupadd vora -gid 44936
   clush -a -b useradd vora -uid 44936 -g vora
   ```

Installing Spark

To install the MapR- Spark, complete the following steps:

1. Log in to rhel1 (admin node) and install the MapR- Spark package on all the nodes using the clush command.

   ```
   clush -a "yum -y install mapr-spark"
   ```

2. Install the spark-history-server package on rhel1 server.

   ```
   yum -y install mapr-spark-historyserver
   ```
Install and Configure Hadoop, YARN, and Spark

3. Create the /apps/spark directory on MapR-FS and set the correct permissions on the directory as follows from the rhel1,

```bash
hadoop fs -mkdir /apps/spark
hadoop fs -chmod 777 /apps/spark
```

4. Edit the container-executer.cfg as shown below to allow user root to execute jobs only for testing purposes. Restore the default value once the testing is completed.

```bash
cd /opt/mapr/hadoop/hadoop-2.7.0/etc/hadoop
vi container-executer.cfg
min.user.id=0
allow.system.users=mapr,root
```

5. Copy the container-executer.cfg file on all the nodes

```bash
clush -a -b -c /opt/mapr/hadoop/hadoop-2.7.0/etc/hadoop/container-executer.cfg
```

6. Run the spark shell command to enter interactive mode for spark.
Install and Configure Hadoop, YARN, and Spark

7. Confirm the installation of spark-historyserver using following URL

http://rhel1:18080

Verifying Cluster Status

Verify Cluster Status using the Web Interface

1. Log in to the MapR Control System.

2. Under the Cluster group in the left pane, click Dashboard.

3. Check the Services pane and make sure each service is running the correct number of instances, according to the cluster plan.

Verifying Cluster Status Using the Command Line Interface

1. Log in to a cluster node.

2. Use the following command to list MapR services:

   $ maprcli service list
   $ maprcli license list
   $ maprcli disk list -host <name or IP address>
Install and Configure SAP HANA Vora

Preparing to Install SAP HANA Vora

The SAP HANA Vora SQL engine is a service that you add to your existing Hadoop installation. SAP HANA Vora instances hold data in memory and boost the performance of out-of-the-box Spark. To increase execution performance on the node level, you add an SAP HANA Vora instance to each compute node so that it contains the following:

- A Spark Worker
- An SAP HANA Vora engine

The SAP HANA Vora extension library allows SAP HANA Vora to be accessed through Spark. It also provides additional functionality, such as a hierarchy implementation, which allows you to build hierarchies and run hierarchical queries.

This Vora extension package must be installed on the same node on which Spark is installed.

Preparing for SAP HANA Vora Distributed Log Server

1. Log in to the admin node to install libaio package

```
clush -a -b yum -y install libaio
```

2. Increase the system file descriptor limit by adding or modifying the following line in /etc/sysctl.conf file and copy to all the nodes.

```
vi /etc/sysctl.conf
fs.file-max=16777216
```

```
clush -a -b -c /etc/sysctl.conf --dest=/etc/
```

It is generally recommended to set the limit to 65536 per 1 GB of RAM.
3. Run the following command to reload the new setting

```
crush -a -b "sysctl --load=/etc/sysctl.conf"
```

4. Add or modify the following line in the `/etc/security/limits.conf` file and copy to all the nodes

```
-nofile 1000000
```

```
crush -a -b -c /etc/security/limits.conf --dest=/etc/security/
```

5. Log out or reboot so that the ulimit change takes effect

**Preparing for SAP HANA Vora Document Store Server**

The SAP HANA Vora Document Store component requires the RPM package `numactl` to be installed on all nodes.

Log in to the admin node (`rhel1`) and install the package on all the nodes.

```
crush -a -b yum -y install numactl
```

**Preparing for SAP HANA Vora Disk Engine Server**

The SAP HANA Vora Disk Engine component requires the RPM packages `libtool` and `libaio` to be installed on all nodes.

Log in to the admin node (`rhel1`) and install the package `libtool` and `libtool-ltdl` on all the nodes.

```
crush -a -b yum -y install libtool libtool-ltdl
```

**Preparing for SAP HANA Vora Cluster Manager**

The SAP HANA Vora Manager component requires the `lsof` and `ifconfig` RPM packages to be installed on all nodes.

Log in to the admin node (`rhel1`) and install the package `lsof` and `net-tools` on all the nodes.

```
crush -a -b yum -y install lsof
```

```
crush -a -b yum -y install net-tools
```

**Configure Sudo Access**

To run scripts that use sudo, you need to ensure that the requiretty setting is disabled and that the user (except root) has sudo permission. Make the necessary changes in the `/etc/sudoers` file using the `visudo` command.

1. Open the `/etc/sudoers` file using the command `visudo`:

```
visudo
```

2. Disable requiretty by commenting out the line:

```
# Defaults requiretty
```
3. Enable a user to run sudo without a password by adding the following:

```bash
mapr ALL = NOPASSWD: /opt/mapr/vora/
```

4. Copy to all the nodes:

```bash
clush -a -b -c /etc/sudoers --dest=/etc/
```

**Validate the cluster**

1. Make sure the SPARK_HOME is set correctly, if not add the following line into .bash_profile file and save and exit.

```
[root@rhel1 ~]# vi .bash_profile
export SPARK_HOME=/opt/mapr/spark/spark-1.6.1/
[root@rhel1 ~]# echo $SPARK_HOME
/opt/mapr/spark/spark-1.6.1/
```

2. Make sure the HDFS is accessible by executing the command:

```
hadoop fs -ls /
```
3. Execute the following command to make sure spark is running properly:

```bash
/opt/mapr/spark/spark-1.6.1/bin/spark-submit --class/org.apache.spark.examples.SparkPi --master yarn-client --num-executors 2 --driver-memory 512m --executor-memory 512m --executor-cores 2 --queue default
$SPARK_HOME/lib/spark-examples*.jar 10 2>/dev/null
```

Output:

```
Pi is roughly 3.140064
```

### Downloading SAP HANA Vora Installation Package

1. On admin node download VORAMR03P_13-70002592.TGZ (from the SAP Software Download Center at https://support.sap.com/swdc to the admin node (rhel1).

    The actual filename downloaded from the SAP Software Download Center would be VORAMR03P_<version>.tgz.

2. Copy the file VORAMR03P_13-70002592.TGZ over to the root directory and untar the file.

    ```bash
    [root@rhel1 ]# tar -zxvf VORAMR03P_13-70002592.TGZ
    ```

### Generate an Initial Password for SAP HANA Vora

SAP HANA Vora is shipped with two UI tools; the SAP HANA Vora Manager, which is used to administer the SAP HANA Vora services, and the SAP HANA Vora Tools, which allow you to query data and create relational models. Both UIs require a username and password to log on.

As the administrator, you need to create the initial username and password for both UIs during the installation of SAP HANA Vora.

The password needs to be stored in an encrypted form in a file named htpasswd on the file system where either the SAP HANA Vora Tools or SAP HANA Vora Manager will run. You therefore need to distribute the htpasswd file to all nodes that have the master role (that is, where the SAP HANA Vora Manager will be installed as a master) or that will host the SAP HANA Vora Tools.

1. Execute the genpasswd.sh script as root user:

    ```bash
    ./genpasswd.sh
    ```
2. Enter the following username and password when prompted:

   Username=  vora  
   Password=  Cisco!123

3. Use the default directory [/etc/vora/datatools/] to store the htpasswd file.

4. Setup the htpasswd file on all hosts that will host the SAP HANA Vora manager service. In this CVD, vora-manager service is running on all the nodes.

5. As a root user, create the directory /etc/vora/manager

   clush -a -b mkdir -p /etc/vora/manager

6. Copy htpasswd from /etc/vora/datatools directory to /etc/vora/manager directory and copy to all the nodes

   cp /etc/vora/datatools/htpasswd /etc/vora/manager/
   cd /etc/vora/manager
   clush -a -b -c htpasswd --dest=/etc/vora/manager

7. Change the ownership of htpasswd to the user vora

   clush -a -b "chown vora /etc/vora/manager/htpasswd"

8. Change the permission to rw for vora

   clush -a -b "chmod 600 /etc/vora/manager/htpasswd"

9. Configure the htpasswd file on all hosts that will host the SAP HANA Vora Tools service. In this CVD, Vora Tools is running on rhel[1-3]

10. As a root user create a directory /etc/vora/datatools on all three nodes and copy the htpasswd file:

    clush -a -w rhel[1-3] "mkdir -p /etc/vora/datatools"
    cd /etc/vora/datatools
    clush -w rhel[1-3] -c htpasswd -dest=/etc/vora/datatools

11. Change the ownership of htpasswd to the user vora:

    clush -w rhel[1-3] -b "chown vora /etc/vora/datatools/htpasswd"

12. Change the permission to rw for vora:

    clush -w rhel[1-3] -b "chmod 600 /etc/vora/datatools/htpasswd"
Vora installation recommendation

Install the appropriate Vora package on the nodes based on the MapR roles installed on the nodes:

<table>
<thead>
<tr>
<th>Vora Packages</th>
<th>Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vora-base</td>
<td>All nodes</td>
</tr>
<tr>
<td>Mapr-vora-manager</td>
<td>All nodes</td>
</tr>
<tr>
<td>Mapr-vora-manager-master</td>
<td>Zookeeper, CLDB or Resource Manager nodes</td>
</tr>
<tr>
<td>Mapr-vora-manager-worker</td>
<td>All nodes</td>
</tr>
<tr>
<td>Vora-deps</td>
<td>All nodes</td>
</tr>
</tbody>
</table>

For the purpose of this CVD following roles have been assigned across the cluster:

<table>
<thead>
<tr>
<th>Role</th>
<th>Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>rhel[1-2]</td>
<td>CLDB</td>
</tr>
<tr>
<td>rhel[1-3]</td>
<td>Zookeeper</td>
</tr>
<tr>
<td>rhel[3-5]</td>
<td>Resource Manager</td>
</tr>
<tr>
<td>rhel[1-16]</td>
<td>Fileserver, nfs, nodemanager</td>
</tr>
<tr>
<td>rhel3</td>
<td>webserver</td>
</tr>
</tbody>
</table>

1. Log into rhel1 and copy all the rpms except vora-manager-master to all the nodes and install the necessary RPM based on the table above:

   clush -a -b -c vora-base-1.3.66.4_vora_1.3-GA.x86_64.rpm
   clush -a -b -c mapr-vora-manager-1.3.66.4_vora_1.3-GA.x86_64.rpm
   clush -a -b -c mapr-vora-manager-worker-1.3.66.4_vora_1.3-GA.x86_64.rpm
   clush -a -b -c vora-deps-1.3.66.4_vora_1.3-GA.redhat.x86_64.rpm

2. Copy mapr-vora-manager-master into rhel2 and rhel3:

   clush -w rhel[2-3] -c mapr-vora-manager-master-1.3.66.4_vora_1.3-GA.x86_64.rpm

3. Install vora-base on all the nodes

   clush -a -b yum -y install vora-base-1.3.66.4_vora_1.3-GA.x86_64.rpm

4. Install vora-deps on all the nodes:

   clush -a -b yum -y install vora-deps-1.3.66.4_vora_1.3-GA.redhat.x86_64.rpm

5. Install mapr-vora-manager-master on all CLDB and zookeeper nodes:

   clush -w rhel[1-3] yum -y install mapr-vora-manager-master-1.3.66.4_vora_1.3-GA.x86_64.rpm

6. Install mapr-vora-Manager on all the nodes:

   clush -a -b yum -y install mapr-vora-manager-1.3.66.4_vora_1.3-GA.x86_64.rpm
7. Install mapr-vora-manager-worker on all the nodes:

```
crush -a -b yum -y install mapr-vora-manager-worker-1.3.66.4_vora_1.3-GA.x86_64.rpm
clush -a -b "/opt/mapr/server/configure.sh -R -no-autostart"
```

Configure the SAP HANA Vora Manager

The SAP HANA Vora Manager configuration is contained in two configuration files:

- `/opt/mapr/conf/conf.d/vora_default_settings.sh`
- `/etc/vora/vora-env.sh`

The `vora_default_settings.sh` file contains all configuration parameters for the SAP HANA Vora services. It is implemented as a shell script and uses environment variables for interaction with the SAP HANA Vora Manager. You can change the parameters for the ports and log location in this file.

The `vora-env.sh` file contains environment variables for working with the SAP HANA Vora software.

1. Go to `/opt/mapr/conf/conf.d`:

   ```
   cd /opt/mapr/conf/conf.d/
   ```

2. Edit the file `vora_default_settings.sh` with following changes:

   ```
   export VORA_DISCOVERY_BIND_INTERFACE="enp5s0"
   ```

3. Upload the configuration file to the central configuration using the command below:
Install and Configure SAP HANA Vora

```
clush -a -b -c vora_default_settings.sh --dest=/opt/mapr/conf/conf.d/
```

Start the SAP HANA Vora Manager

1. Log in to rhel1 and use the command below to start the master:

   ```
   /opt/mapr/vora/service-control.sh manager-master start
   ```

   ![Output of command](image)

2. Start the vora manager worker:

   ```
   /opt/mapr/vora/service-control.sh manager-worker start
   ```

   ![Output of command](image)

Deploy SAP HANA Vora Service

Use the SAP HANA Vora Manager UI to configure and deploy the SAP HANA Vora service on the cluster.

The SAP HANA Vora Manager UI allows you to start and stop services as well as manage their configuration and node assignments.

When initially installed, the SAP HANA Vora services are not yet configured. Before starting the services, walk through the service list and for each service:
• Make sure that the configuration parameters are correctly set
• Assign the nodes on which the service should be deployed

To configure the services, complete the following steps:
1. Open a browser and point to <VORA MASTER HOST>:19000.

   ![Welcome to SAP HANA Vora](image)

   User can log into any master nodes if there are more than one.

   Username: vora
   Password: Cisco!123

2. Log in using the initial user and password defined earlier.

3. Choose services.

4. Choose Vora Catalog and in the right window enter “enp5s0” for Network interface name and binding.
5. Click Node Assignment tab and select the number of nodes for the particular role. The number of node selections are based on the service requirement and can vary across all the services.
Install and Configure SAP HANA Vora

6. Repeat steps 4 and 5 for all the services.

⚠️ Please make sure that the Vora Thriftserver and Vora Tools are running on the same machine for better performance because Vora Tools uses Vora Thriftserver to connect to Spark.

7. When all the services are configured, click the Start button to start all the services.

Install the SAP HANA Vora Zeppelin Interpreter

Zeppelin is a graphical user interface that allows you, as a data scientist, to interact easily with a cluster. The SAP HANA Vora Spark extension provides an interpreter for the Zeppelin user interface.
Install and Configure SAP HANA Vora

The SAP HANA Vora extension library has its own SQLContext class. A modified Zeppelin interpreter, spark.vora, is therefore required to allow Zeppelin to run in the modified context. To enable the interpreter, you need to register it with Zeppelin.

1. Download the Zeppelin 0.6.1 build from the web site [http://zeppelin.apache.org/download.html](http://zeppelin.apache.org/download.html).

2. Copy the binary to the root directory of admin node (rhel1).

3. Extract the file in using the command below:

   ```
tar -zxv zeppelin-0.6.1-bin-all.tgz
   
```

4. Copy the zeppelin-1.3.107.1-vora-1.3.jar file from /opt/vora/lib/vora-spark/zeppelin directory to /root/zeppelin-0.6.1-bin-all/interpreter/spark.

5. Extract the interpreter-setting.json file from zeppelin-1.3.107.1-vora-1.3.jar using the command below:

   ```
jar xf zeppelin-1.3.107.1-vora-1.3.jar interpreter-setting.json
   
```

6. Replace the interpreter-setting.json file from the old one in the zeppelin-spark_2.11-0.6.1.jar:

   ```
jar uf zeppelin-spark_2.11-0.6.1.jar interpreter-setting.json
   ```
7. Go to `/root/zeppelin-0.6.1-bin-all/conf` and make a copy of a file `zeppelin-site.xml.template` and `zeppelin-env.sh.template`:

```
cp zeppelin-site.xml.template .
zeppelin-site.xml

cp zeppelin-env.sh.template .
zeppelin-env.sh
```

8. Change the permission of the file:

```
chmod 0755 zeppelin-site.xml
chmod 0755 zeppelin-env.sh
```

9. Exit the `zeppelin-env.sh` file and add the following variables:

```
export MASTER=yarn-client
export HADOOP_CONF_DIR="/opt/mapr/hadoop/hadoop-2.7.0/etc/hadoop"
export ZEPPELIN_HOME="/opt/mapr/hadoop/hadoop-2.7.0/"
export ZEPPELIN_JAVA_OPTS="-Djava.security.auth.login.config=/opt/mapr/conf/mapr.login.conf"
export SPARK_HOME="/opt/mapr/spark/spark-1.6.1/
```

10. Add the interpreter class `sap.zeppelin.spark.SapSqlInterpreter` to the `zeppelin.interpreters` property in the `zeppelin-site.xml` file:

```
<property>
   <name>zeppelin.interpreters</name>
```

11. Edit the following properties and save and exit.

```
<property>
   <name>zeppelin.server.port</name>
   <value>9099</value>
   <description>Server port.</description>
</property>
```
12. Start the zeppelin server:

```bash
./zeppelin-daemon.sh start
```

13. Open the web browser and log into the zeppelin UI:

http://172.16.51.11:9099/#

14. Remove and re-add the spark Interpreter.

15. In the top right corner, click on the username and in the dropdown menu choose interpreter.
16. Scroll down and remove the spark interpreter

17. Re-add the spark interpreter, name it spark, and choose spark as the interpreter group.

18. Click the Create button and enter the name as spark and interpreter group spark
19. Scroll down and edit the master properties to yarn-client.

20. Scroll down and the following two values in the Dependencies:

```
/opt/mapr/hadoop/hadoop-2.7.0/share/hadoop/common/lib/zookeeper-3.4.5-mapr-1503.jar
```
Install and Configure SAP HANA Vora

Make sure that the zookeeper-3.4.5-mapr-1503.jar dependency is added before the spark-sap-datasource-assembly jar file

21. Save the changes.

22. The spark interpreter should be visible again and should now include spark.vora.

Test the Zeppelin Interpreter

1. Click the Notebook drop-down list and choose “+Create new note.”

2. Enter testnote as name.
Install and Configure SAP HANA Vora

3. Enter the following command in single line to execute and click READY button:

   `%spark.vora CREATE TABLE table_test (a1 double, a2 int, a3 string) USING com.sap.spark.vora`

4. Enter the following command and execute:

   `%spark.vora SHOW TABLES`

---

Preparing the connectivity between SAP HANA Vora and SAP HANA

Configure the Spark controller to use SAP HANA Vora. This allows you to connect from SAP HANA to SAP HANA Vora and query SAP HANA Vora tables.

1. Download the HANA Spark Controller, file named “HANASPARKCTRL00P_1-70001881.zip” from https://support.sap.com

2. Log into rhel1 and create a directory **spark_controller**.

   ```bash
   mkdir -p spark_controller
   ```

3. Copy the HANASPARKCTRL00P_1-70001881.zip file to spark_controller directory and unzip the file.
Install and Configure SAP HANA Vora

```
cd spark_controller
unzip HANASPARKCTRL00P_1-70001881.zip
```

4. Now install the “sap.hana.spark.controller-1.6.1-1.noarch.rpm” using the command below.

```
rpm –ivh sap.hana.spark.controller-1.6.1-1.noarch.rpm
```

Configuring the SAP HANA Spark Controller

Make sure that the SAP HANA Vora data sources JAR and the Spark assembly JAR available to the Spark controller

1. Use the command below to verify:

```
ls -1 $SPARK_HOME/lib
```

```
<table>
<thead>
<tr>
<th>root@rhole1 ~</th>
<th>ls -l $SPARK_HOME/lib</th>
</tr>
</thead>
<tbody>
<tr>
<td>rw-r--r-- 1</td>
<td>mapr root       366656 Dec 1 09:01 datanucleus-api-jdo-4.2.1.jar</td>
</tr>
<tr>
<td>rw-r--r-- 1</td>
<td>mapr root       2009163 Dec 1 09:01 datanucleus-core-4.1.6.jar</td>
</tr>
<tr>
<td>rw-r--r-- 1</td>
<td>mapr root       1888550 Dec 1 09:01 datanucleus-rdbms-4.1.7.jar</td>
</tr>
<tr>
<td>rw-r--r-- 1</td>
<td>mapr root       6920044 Dec 1 09:01 spark-1.6.1-mapr-1611-yarn-shuffle.jar</td>
</tr>
<tr>
<td>rw-r--r-- 1</td>
<td>mapr root       155718167 Dec 1 09:01 spark-assembly-1.6.1-mapr-1611-hadoop2.7.0-mapr-1602.jar</td>
</tr>
<tr>
<td>rw-r--r-- 1</td>
<td>mapr root       106509590 Dec 1 09:01 spark-examples-1.6.1-mapr-1611-hadoop2.7.0-mapr-1602.jar</td>
</tr>
</tbody>
</table>
```

2. Set the following environment variables for $VORA_SPARK_HOME in .bash_profile file.

```
export VORA_SPARK_HOME=/opt/vora/lib/vora-spark/
```

3. Verify vora data source jar file is available.

```
ls $VORA_SPARK_HOME/lib
```

```
<table>
<thead>
<tr>
<th>root@rhole1 ~</th>
<th>ls -l $VORA_SPARK_HOME/lib</th>
</tr>
</thead>
<tbody>
<tr>
<td>rw-r--r-- 1</td>
<td>vora vora      19590 Feb 18 23:00 data-migration.jar</td>
</tr>
<tr>
<td>rw-r--r-- 1</td>
<td>vora vora      184683601 Feb 18 23:00 spark-sap-datascources-1.3.107.1-vora-1.3-assembly.jar</td>
</tr>
</tbody>
</table>
```

4. Set the following environment variables in /usr/sap/spark/controller/conf/hana_hadoop-env.sh:

```
HANES_USER=vora
```

```
export HANA_SPARK_ASSEMBLY_JAR=/opt/mapr/spark/spark-1.6.1/lib/spark-assembly-1.6.1-mapr-1611-hadoop2.7.0-mapr-1602.jar
export HANA_SPARK_ADDITIONAL_JARS=/usr/sap/spark/controller/conf/spark-sap-datascources-1.3.107.1-vora-1.3-assembly.jar
```

5. Add the vora user to the sapsys user. This will allow the vora user to start the spark controller and write to the log file

```
usermod -a -G sapsys vora
```
6. Make sure that the Vora class path is loaded first, followed by the other class paths. To do so, modify the `/usr/sap/spark/controller/bin/hanaes` file as follows:

```bash
CLASSPATH="${HANA_SPARK_ADDITIONAL_JARS}:${HADOOP_CLASSPATH}:${HANA_SPARK_ASSEMBLY_JAR}"
```

7. Remove the ZooKeeper JAR from the `spark-sap-datasources-<VERSION>-assembly.jar` file:

   a. Extract the `$VORA_SPARK_HOME/lib/spark-sap-datasources-<VERSION>-assembly.jar` file, delete the ZooKeeper folder from the JAR, and repack the modified JAR file.
   b. Point the modified JAR file to the Spark controller by adding the `HANA_SPARK_ADDITIONAL_JARS` parameter in the `hana_hadoop-env.sh` file. For example:

   ```bash
   export HANA_SPARK_ADDITIONAL_JARS=/usr/sap/spark/controller/bin/tmp/extract/modified.jar
   ```

8. Configure the spark controller

   In the spark controller configuration file `/usr/sap/spark/controller/conf/hanessite.xml`, change the value of the property `sap.hana.hadoop.datastore` from hive to vora.

   ```xml
   <property>
   <name>sap.hana.hadoop.datastore</name>
   <value>vora</value>
   <final>true</final>
   </property>
   ```

   Make sure that the Spark-specific properties match the cluster's environment, that is, `spark.executor.memory` and `spark.executor.instances`. Otherwise the Spark controller may not be able to start up properly because of resource allocation issues

9. Restart the spark controller:
cd /usr/sap/spark/controller/bin

./hanaes stop

./hanaes start

10. Verify the configuration changes. To verify whether the configuration changes were successful, check the spark controller log file: /var/log/hanas/hana_controller.log
Bill of Materials

This section gives the BOM for the 16 node High Capacity Vora cluster.

The Bill of Materials (below) can be easily added to a CCW estimate using the smart play bundle UCS-SL-CPA3-P.

The generic Big Data Performance optimized bundle (UCS-SL-CPA3-P) is considered a high capacity Vora configuration.

Table 5  Bill of Materials

<table>
<thead>
<tr>
<th>Part Number</th>
<th>Description</th>
<th>Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCSC-C240-M4SX</td>
<td>UCS C240 M4 SFF 24 HD w/o CPU,mem,HD,PCIe,PS, railkt w/ expndr</td>
<td>16</td>
</tr>
<tr>
<td>CON-OSP-C240M4SX</td>
<td>SNCT-24X7X4OS UCS C240 M4 SFF 24 HD w/o CPU,mem</td>
<td>16</td>
</tr>
<tr>
<td>UCSC-RAILB-M4</td>
<td>Ball Bearing Rail Kit for C220 M4 and C240 M4 rack servers</td>
<td>16</td>
</tr>
<tr>
<td>UCSC- HS-C240M4</td>
<td>Heat sink for UCS C240 M4 rack servers</td>
<td>32</td>
</tr>
<tr>
<td>UCSC-SCCBL240</td>
<td>Supercap cable 250mm</td>
<td>16</td>
</tr>
<tr>
<td>UCSC-MRAID12G</td>
<td>Cisco 12G SAS Modular Raid Controller</td>
<td>16</td>
</tr>
<tr>
<td>UCSC-MRAID12G-2GB</td>
<td>Cisco 12Gbps SAS 2GB FBWC Cache module (Raid 0/1/5/6)</td>
<td>16</td>
</tr>
<tr>
<td>UCSC- PCI-1C-240M4</td>
<td>Right PCI Riser Bd (Riser 1) 2onbd SATA bootdrvs+2PCI slts</td>
<td>16</td>
</tr>
<tr>
<td>UCSC-MLOM-CSC-02</td>
<td>Cisco UCS VIC1227 VIC MLOM - Dual Port 10Gb SFP+</td>
<td>16</td>
</tr>
<tr>
<td>C1UCS-OPT-OUT</td>
<td>Cisco ONE Data Center Compute Opt Out Option</td>
<td>16</td>
</tr>
<tr>
<td>UCSC-CPU-E52680E</td>
<td>2.40 GHz E5-2680 v4/120W 14C/35MB Cache/DDR4 2400MHz</td>
<td>32</td>
</tr>
<tr>
<td>UCSC-MR-1X322RV-A</td>
<td>32GB DDR4-2400- MHz RDIMM/PC4-19200/dual rank/x4/1.2v</td>
<td>128</td>
</tr>
<tr>
<td>UCSC-PSU2-1400W</td>
<td>1400W AC Power Supply (200 - 240V) 2U &amp; 4U C Series Servers</td>
<td>32</td>
</tr>
<tr>
<td>UCSC-M4-V4-LBL</td>
<td>Cisco M4 - v4 CPU asset tab ID label (Auto-Expand)</td>
<td>16</td>
</tr>
<tr>
<td>UCSC-HD18TB10KS4K</td>
<td>1.8 TB 12G SAS 10K RPM SFF HDD (4K)</td>
<td>384</td>
</tr>
<tr>
<td>UCSC-SD480GBS4-EB</td>
<td>480 GB 2.5 inch Enterprise Value 6G SATA SSD (Boot)</td>
<td>32</td>
</tr>
<tr>
<td>CAB-C13-C14-2M</td>
<td>Power Cord Jumper, C13-C14 Connectors, 2 Meter Length</td>
<td>32</td>
</tr>
<tr>
<td>UCSC-FI-6296UP-UPG</td>
<td>UCS 6296UP 2RU Fabric Int/No PSU/48 UP/18p LIC</td>
<td>2</td>
</tr>
<tr>
<td>CON-OSP-FI6296UP</td>
<td>SNCT-24X7X4OS UCS 6296UP 2RU Fabric Int/2 PSU/4 Fans</td>
<td>2</td>
</tr>
<tr>
<td>UCSC-ACC-6296UP</td>
<td>UCS 6296UP Chassis Accessory Kit</td>
<td>2</td>
</tr>
<tr>
<td>N10-MGT014</td>
<td>UCS Manager v3.1</td>
<td>2</td>
</tr>
<tr>
<td>UCSC-FI-E16UP</td>
<td>UCS 6200 16- port Expansion module/16 UP/8p LIC</td>
<td>2</td>
</tr>
<tr>
<td>CON-OSP-FIE16UP</td>
<td>SNCT-24X7X4OS 16prt 10Gb UnifiedPrt/Expnsn mod UCS6200</td>
<td>2</td>
</tr>
<tr>
<td>Part Number</td>
<td>Description</td>
<td>Qty</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>UCS-FAN-6296UP</td>
<td>UCS 6296UP Fan Module</td>
<td>8</td>
</tr>
<tr>
<td>UCS-L-6200-10G-C</td>
<td>2nd Gen FI License to connect C - direct only</td>
<td>100</td>
</tr>
<tr>
<td>UCS-FI-E16UP</td>
<td>UCS 6200 16- port Expansion module/16 UP/ 8p LIC</td>
<td>2</td>
</tr>
<tr>
<td>CON-OSPF-FIE16UP</td>
<td>SNTC-24X7X4OS 16p 10Gb UnifiedPrt/Expnsn mod UCS6200</td>
<td>2</td>
</tr>
<tr>
<td>UCS-PSU-6296UP-AC</td>
<td>UCS 6296UP Power Supply/100-240VAC</td>
<td>4</td>
</tr>
<tr>
<td>UCS-FI-E16UP</td>
<td>UCS 6200 16- port Expansion module/16 UP/ 8p LIC</td>
<td>2</td>
</tr>
<tr>
<td>CON-OSPF-FIE16UP</td>
<td>SNTC-24X7X4OS 16p 10Gb UnifiedPrt/Expnsn mod UCS6200</td>
<td>2</td>
</tr>
<tr>
<td>SFP-10G-SR</td>
<td>10GBASE-SR SFP Module</td>
<td>8</td>
</tr>
<tr>
<td>CAB-C13-C14-AC</td>
<td>Power cord, C13 to C14 (recessed receptacle), 10A</td>
<td>4</td>
</tr>
<tr>
<td>RACK-UCS2</td>
<td>Cisco R42610 standard rack, w/side panels</td>
<td>1</td>
</tr>
<tr>
<td>CON-SNT-R42610</td>
<td>SNTC-8X5XNBD Cisco R42610 expansion rack, no side pan</td>
<td>1</td>
</tr>
<tr>
<td>RP208-30-1P-U-1</td>
<td>Cisco RP208-30-U-1 Single Phase PDU 2x C13, 4x C19</td>
<td>2</td>
</tr>
<tr>
<td>CON-SNT-RPDUX</td>
<td>SNTC-8X5XNBD Cisco RP208-30-U-X Single Phase PDU</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 6  Bill of Material for High Performance Cluster

<table>
<thead>
<tr>
<th>Part Number</th>
<th>Description</th>
<th>Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCSC-C240-M4SX</td>
<td>UCS C240 M4 SFF 24 HD w/o CPU,mem,HD,PCIe,PS,railkt w/expndr</td>
<td>16</td>
</tr>
<tr>
<td>CON-OSP-C240M4SX</td>
<td>SNTC-24X7X4OS UCS C240 M4 SFF 24 HD w/o CPU,mem</td>
<td>16</td>
</tr>
<tr>
<td>UCSC-RAILB-M4</td>
<td>Ball Bearing Rail Kit for C220 M4 and C240 M4 rack servers</td>
<td>16</td>
</tr>
<tr>
<td>UCSC-HS-C240M4</td>
<td>Heat sink for UCS C240 M4 rack servers</td>
<td>32</td>
</tr>
<tr>
<td>UCSC-SCCBL240</td>
<td>Supercap cable 250mm</td>
<td>16</td>
</tr>
<tr>
<td>UCSC-MRAID12G</td>
<td>Cisco 12G SAS Modular Raid Controller</td>
<td>16</td>
</tr>
<tr>
<td>UCSC-MRAID12G-2GB</td>
<td>Cisco 12Gbps SAS 2GB FBWC Cache module (Raid 0/1/5/6)</td>
<td>16</td>
</tr>
<tr>
<td>UCSC-PCI-1C-240M4</td>
<td>Right PCI Riser Bd (Riser 1) 2onbd SATA bootdrvs+2PCI slts</td>
<td>16</td>
</tr>
<tr>
<td>C1UCS-OPT-OUT</td>
<td>Cisco ONE Data Center Compute Opt Out Option</td>
<td>16</td>
</tr>
<tr>
<td>UCS-CPU-ES52680E</td>
<td>2.40 GHz E5-2680 v4/120W 14C/35MB Cache/DDR4 2400MHz</td>
<td>32</td>
</tr>
<tr>
<td>UCS-MR-1X322RV-A</td>
<td>32GB DDR4-2400- MHz RDIMM/PC4-19200/dual rank/x4/1.2v</td>
<td>128</td>
</tr>
<tr>
<td>UCSC-PSU2-1400W</td>
<td>1400W AC Power Supply (200 - 240V) 2U &amp; 4U C Series Servers</td>
<td>32</td>
</tr>
<tr>
<td>UCS-M4-V4-LBL</td>
<td>Cisco M4 - v4 CPU asset tab ID label (Auto- Expand)</td>
<td>16</td>
</tr>
<tr>
<td>UCS-SD480GBKS4-EB</td>
<td>480 GB 2.5 inch Enterprise Value 6G SATA SSD (Boot)</td>
<td>32</td>
</tr>
<tr>
<td>CAB-C13-C14-2M</td>
<td>Power Cord Jumper, C13-C14 Connectors, 2 Meter Length</td>
<td>32</td>
</tr>
<tr>
<td>UCSC-MLOM-C40Q-03</td>
<td>Cisco VIC 1387 Dual Port 40Gb QSFP CNA MLOM</td>
<td>16</td>
</tr>
<tr>
<td>UCS-SD16TBKS4-EV</td>
<td>1.6TB 2.5 inch Enterprise Value 6G SATA SSD</td>
<td>256</td>
</tr>
<tr>
<td>N20-BBLKD</td>
<td>UCS 2.5 inch HDD blanking panel</td>
<td>128</td>
</tr>
<tr>
<td>RACK-UCS2</td>
<td>Cisco R42610 standard rack, w/side panels</td>
<td>1</td>
</tr>
</tbody>
</table>
# Bill of Materials

<table>
<thead>
<tr>
<th>Part Number</th>
<th>Description</th>
<th>Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON- SNT- R42610</td>
<td>SNTC- 8X5XNBDD Cisco R42610 expansion rack, no side pan</td>
<td>1</td>
</tr>
<tr>
<td>RP208- 30- 1P- U- 1</td>
<td>Cisco RP208- 30- U- 1 Single Phase PDU 2x C13, 4x C19</td>
<td>2</td>
</tr>
<tr>
<td>CON- SNT- RPDUX</td>
<td>SNTC- 8X5XNBDD Cisco RP208- 30- U- X Single Phase PDU</td>
<td>2</td>
</tr>
<tr>
<td>UCS- FI- 6332- U</td>
<td>UCS 6332 1RU Fabric Interconnect/No PSU/32 QSFP+ports/8p Lic</td>
<td>2</td>
</tr>
<tr>
<td>CON- OSP- FI6332U</td>
<td>ONSITE 24X7X4, UCS 6332 1RU Fabric Interconnect/No PSU/32 QS</td>
<td>2</td>
</tr>
<tr>
<td>N10- MGT014</td>
<td>UCS Manager v3.1</td>
<td>2</td>
</tr>
<tr>
<td>UCS- FAN- 6332</td>
<td>UCS 6332 Fan Module</td>
<td>8</td>
</tr>
<tr>
<td>UCS- ACC- 6332</td>
<td>UCS 6332 Chassis Accessory Kit</td>
<td>2</td>
</tr>
<tr>
<td>UCS- LIC- 6300- 40GC</td>
<td>3rd Gen Fi Per port License to connect C- direct only</td>
<td>20</td>
</tr>
<tr>
<td>UCS- PSU- 6332- AC</td>
<td>UCS 6332 Power Supply/100- 240VAC</td>
<td>4</td>
</tr>
<tr>
<td>CAB- C13- C14- 2M</td>
<td>Power Cord Jumper, C13- C14 Connectors, 2 Meter Length</td>
<td>4</td>
</tr>
</tbody>
</table>

## Table 7 Red Hat Enterprise Linux License

<table>
<thead>
<tr>
<th>Part Number</th>
<th>Description</th>
<th>Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHEL- 2S2V- 3A</td>
<td>Red Hat Enterprise Linux</td>
<td>16</td>
</tr>
<tr>
<td>CON- ISV1- EL2S2V3A</td>
<td>3 year Support for Red Hat Enterprise Linux</td>
<td>16</td>
</tr>
</tbody>
</table>

## Table 8 Bill of Materials for Nexus Device and APIC

<table>
<thead>
<tr>
<th>Part Number</th>
<th>Description</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>N9K- C9508- B2</td>
<td>Nexus 9508 Chassis Bundle with 1 Sup, 3 PS, 2 SC, 6 FM, 3 FT</td>
<td>2</td>
</tr>
<tr>
<td>N9K- C9396PX</td>
<td>Nexus 9300 with 48p 1/10G SFP+ and 1 uplink module slot</td>
<td>2</td>
</tr>
<tr>
<td>N9k- X9736PQ</td>
<td>Spine Line- Card</td>
<td>2</td>
</tr>
<tr>
<td>APIC- L1</td>
<td>APIC Appliance</td>
<td>3</td>
</tr>
<tr>
<td>N9K POWERCABLES</td>
<td>Power Cables</td>
<td>3</td>
</tr>
<tr>
<td>CAB- C13- C14- AC</td>
<td>Power cord, C13 to C14 (recessed receptacle), 10A</td>
<td>4</td>
</tr>
<tr>
<td>QSFP- H40G- CU3M</td>
<td>40GBASE- CR4 Passive Copper Cable, 3m</td>
<td>24</td>
</tr>
<tr>
<td>Nexus 9372TX</td>
<td>Nx- OS mode switch for out of band Management</td>
<td>1</td>
</tr>
<tr>
<td>N9K- M12PQ</td>
<td>ACI Uplink Module for Nexus 9300, 12p 40G QSFP</td>
<td>3</td>
</tr>
<tr>
<td>N9K- C9500- RMK</td>
<td>Nexus 9500 Rack Mount Kit</td>
<td>2</td>
</tr>
</tbody>
</table>

---

100
<table>
<thead>
<tr>
<th>Part Number</th>
<th>Description</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAB-C19-CBN</td>
<td>Cabinet Jumper Power Cord, 250 VAC 16A, C20-C19 Connectors</td>
<td>6</td>
</tr>
<tr>
<td>N9K-C9500-LC-CV</td>
<td>Nexus 9500 Linecard slot cover</td>
<td>16</td>
</tr>
<tr>
<td>N9K-C9500-SUP-CV</td>
<td>Nexus 9500 Supervisor slot cover</td>
<td>2</td>
</tr>
<tr>
<td>N9K-PAC-3000W-B</td>
<td>Nexus 9500 3000W AC PS, Port-side Intake</td>
<td>6</td>
</tr>
<tr>
<td>N9K-SUP-A</td>
<td>Supervisor for Nexus 9500</td>
<td>2</td>
</tr>
<tr>
<td>N9K-SC-A</td>
<td>System Controller for Nexus 9500</td>
<td>4</td>
</tr>
<tr>
<td>N9K FABRIC</td>
<td>Fabric Module</td>
<td>2</td>
</tr>
<tr>
<td>N9300 RACK</td>
<td>Rack Mount Kit</td>
<td>3</td>
</tr>
<tr>
<td>N9K-C9300-RMK</td>
<td>Nexus 9300 Rack Mount Kit</td>
<td>3</td>
</tr>
</tbody>
</table>

**Table 9  Bill of Materials for Cisco UCS Director Express for Big Data**

<table>
<thead>
<tr>
<th>Part Number</th>
<th>Description</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUIC-SVR-OFFERS</td>
<td>Cisco UCS Director Server Offerings</td>
<td>1</td>
</tr>
<tr>
<td>CON-SAUSVROFFERS</td>
<td>Cisco UCS Director Server Offerings Software Application Sup</td>
<td>1</td>
</tr>
<tr>
<td>CUIC-Base-K9</td>
<td>Cisco UCS Director Software License</td>
<td>1</td>
</tr>
<tr>
<td>CON-SAUCUICBASE</td>
<td>SW APP SUPP + UPGR Cisco UCS Director Base Software</td>
<td>1</td>
</tr>
<tr>
<td>CUIC-TERM</td>
<td>Acceptance of Cisco UCS Director License Terms</td>
<td>1</td>
</tr>
<tr>
<td>C1-ECS-BIGDATA</td>
<td>C1 Enterprise Cloud Suite - Big Data Automation</td>
<td>1</td>
</tr>
<tr>
<td>C1-3Y-SVC-TRK</td>
<td>C1 Subscription - Service Contract Tracking 3YR</td>
<td>1</td>
</tr>
<tr>
<td>C1A2TEC5BDAK9</td>
<td>C1 - ECS - Big Data Automation - Per Server</td>
<td>16</td>
</tr>
<tr>
<td>C1A2-3Y-ECSBDA</td>
<td>C1 - ECS - Big Data Automation - 3YR</td>
<td>16</td>
</tr>
<tr>
<td>C1-CUIC-EBDS-T</td>
<td>UCSD Director Express for Big Data - 1 Server License</td>
<td>16</td>
</tr>
<tr>
<td>C1-3Y-SVC-TRK</td>
<td>C1 Subscription - Service Contract Tracking 3YR</td>
<td>16</td>
</tr>
</tbody>
</table>
Conclusion

Hadoop has become a popular data management platform across all verticals. Cisco UCS Integrated Infrastructure for Big Data and Analytics and Cisco Application Centric Infrastructure (ACI) offer a dependable deployment model for enterprise Apache Hadoop, Apache Spark and SAP HANA Vora that provides a fast and predictable path for businesses to unlock value in big data. This architecture allows using the UCS Manager capabilities in the Fabric Interconnect for provisioning the servers within a single domain while providing a facility to interconnect multiple Fabric Interconnect domains with ACI.

Cisco continues a long history of delivering innovative IT infrastructure for SAP landscapes with certified reference architectures that reduce cost and risk. The entire family of solutions - SAP Applications, SAP HANA, and now SAP HANA Vora – is designed to interoperate with the data center you have today. Cisco uses industry-standard architectures and best practices so no special IT processes are needed to incorporate or maintain the solutions in your data center.

The configuration detailed in this document can be extended to clusters of various sizes depending on what application demands as discussed in the Scalability section. Next generation Big Data Infrastructure needs to cater to the emerging trends in Big Data Applications to meet multiple Lines of Business SLAs. Cisco UCS Integrated Infrastructure for Big Data and Analytics and Cisco ACI bring numerous advantages to a big data deployment – fewer points of management for the network, enhanced performance, superior failure handling characteristics, unprecedented scalability. Further, ACI paves way for the next generation data center network accelerating innovation with its SDN capabilities in the Big Data space.
Reference

- Cisco Big Data design zone: http://www.cisco.com/go/bigdata_design
- SAP HANA Vora Community Network: https://help.sap.com/hana_vora_re
- Cisco SAP Solutions: http://www.cisco.com/go/sap
- Cisco ACI: http://www.cisco.com/go/aci
About the Authors

Amrit Kharel, Network Engineer, Data Center Solutions Group at Cisco Systems, Inc.

Amrit’s focus areas are solutions and emerging trends in big data related technologies and infrastructure in the Data Center.

Karthik Karupasamy, Product Manager and Technical Marketing Engineer in the Data Center Solutions Group at Cisco Systems, Inc.

Karthik’s main focus areas are architecture, solutions, and emerging trends in big data related technologies and infrastructure in the Data Center. He is also the Product Manager of Cisco UCS Director Express for Big Data.

Isan Sahoo, SAP

Isan is responsible for end-to-end support for the Vora software. He is one of the active participant of the SAP HANA Vora software development team.

Acknowledgements

For their support and contribution to the design, validation, and creation of this Cisco Validated Design, the authors would like to thank:

- Shane Handy, Big Data Solutions Architect, Cisco Systems, Inc.