

Dell | Cloudera Apache Hadoop Solution Reference Architecture Guide - Version 5.5



© 2011-2015 Dell Inc.

Contents

Trademarks.....	4
Notes, Cautions, and Warnings.....	5
Glossary.....	6
Dell Cloudera Apache Hadoop Solution Overview.....	9
Solution Use Case Summary.....	9
Solution Components.....	10
Software Support.....	12
Cloudera Enterprise Software Overview.....	13
Hadoop for the Enterprise.....	13
Rethink Data Management.....	13
What's Inside?.....	13
Cloudera Enterprise Data Hub.....	14
Cluster Architecture.....	15
High-Level Node Architecture.....	15
Node Definitions.....	16
Network Fabric Architecture.....	17
Network Definitions.....	18
Cluster Sizing.....	19
Rack.....	19
Pod.....	20
Cluster.....	20
Sizing Summary.....	20
High Availability.....	21
Hadoop Redundancy.....	21
Network Redundancy.....	21
HDFS Highly Available Name Nodes.....	21
Resource Manager High Availability.....	22
Hardware Architecture.....	23
Server Infrastructure Options.....	23
PowerEdge R730xd Server.....	23
Network Architecture.....	26
Physical Network Components.....	27

Server Node Connections.....	27
Pod Switches.....	27
Cluster Aggregation Switches.....	28
Core Network.....	30
Layer 2 and Layer 3 Separation.....	30
Management and BMC Networks.....	30
Network Equipment Summary.....	31
Cloudera Enterprise Software.....	32
Cloudera Manager.....	32
Cloudera RTQ (Impala).....	32
Cloudera Search.....	33
Cloudera BDR.....	33
Cloudera Navigator.....	33
Cloudera Support.....	34
Deployment Methodology.....	36
Appendix A: Physical Rack Configuration - PowerEdge R730xd.....	37
Appendix B: Bill of Materials – PowerEdge R730xd 3.5" Infrastructure Node.....	41
Appendix C: Bill of Materials – PowerEdge R730xd 3.5" Data Node.....	43
Appendix D: Bill of Materials – PowerEdge R730xd 2.5" Infrastructure Node.....	45
Appendix E: Bill of Materials – PowerEdge R730xd 2.5" Data Node.....	47
Update History.....	49
Changes in Version 5.5.....	49
References.....	50
To Learn More.....	50

Trademarks

THIS DOCUMENT IS FOR INFORMATIONAL PURPOSES ONLY, AND MAY CONTAIN TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.

© 2011-2015 Dell Inc. All rights reserved. Reproduction of this material in any manner whatsoever without the express written permission of Dell Inc. is prohibited. For more information, contact Dell. Dell, the Dell logo, Dell Networking, OpenManage, PowerEdge, and the Dell badge, are trademarks of Dell Inc.

Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell disclaims proprietary interest in the marks and names of others. This document is for informational purposes only. Dell reserves the right to make changes without further notice to the products herein. The content provided is as-is and without expressed or implied warranties of any kind.

Notes, Cautions, and Warnings

-  A **Note** indicates important information that helps you make better use of your system.
-  A **Caution** indicates potential damage to hardware or loss of data if instructions are not followed.
-  A **Warning** indicates a potential for property damage, personal injury, or death.

This document is for informational purposes only and may contain typographical errors and technical inaccuracies. The content is provided as is, without express or implied warranties of any kind.

Glossary

BMC

Baseboard management controller

CDH

Cloudera Distribution for Apache Hadoop

Clos

A multi-stage, non-blocking network switch architecture. It reduces the number of required ports within a network switch fabric.

DBMS

Database management system

DTK

Dell OpenManage Deployment Toolkit

ECMP

Equal Cost Multi-Path

EDW

Enterprise data warehouse

EoR

End-of-row switch/router

ETL

Extract, Transform, Load is a process for extracting data from various data sources; transforming the data into proper structure for storage; and then loading the data into a data store.

HBA

Host Bus Adapter

HDFS

Hadoop Distributed File System

IPMI

Intelligent Platform Management Interface

JBOD

Just a Bunch of Disks

LACP

Link Aggregation Control Protocol

LAG

Link Aggregation Group

LOM

Local area network on motherboard

NIC

Network interface card

NTP

Network Time Protocol

OS

Operating system

RSTP

Rapid Spanning Tree Protocol

RTO

Recovery Time Objectives

SIEM

Security Information and Event Management

SLA

Service Level Agreement

THP

Transparent Huge Pages

ToR

Top-of-rack switch/router

VLT

Virtual Link Trunking

VRRP

Virtual Router Redundancy Protocol

YARN

Yet Another Resource Negotiator

Dell | Cloudera Apache Hadoop Solution Overview

The Dell™ | Cloudera™ Apache™ Hadoop® Solution lowers the barrier to adoption for organizations intending to use Apache Hadoop in production.

Hadoop is an Apache project being built and used by a global community of contributors, using the Java programming language. Yahoo!, has been the largest contributor to this project, and uses Apache Hadoop extensively across its businesses. Core committers on the Hadoop project include employees from Cloudera, eBay, Facebook, Getopt, Hortonworks, Huawei, IBM, InMobi, INRIA, LinkedIn, MapR, Microsoft, Pivotal, Twitter, UC Berkeley, VMware, WANDisco, and Yahoo!, with contributions from many more individuals and organizations.

Although Hadoop is popular and widely used, installing, configuring, and running a production Hadoop cluster involves multiple considerations, including:

- The appropriate Hadoop software distribution and extensions
- Monitoring and management software
- Allocation of Hadoop services to physical nodes
- Selection of appropriate server hardware
- Design of the network fabric
- Sizing and Scalability
- Performance

These considerations are complicated by the need to understand the type of workloads that will be running on the cluster, the fast-moving pace of the core Hadoop project and the challenges of managing a system designed to scale to thousands of nodes in a single instance.

Dell's customer-centered approach is to create rapidly deployable and highly optimized end-to-end Hadoop solutions running on hyperscale hardware. Dell listened to its customers and designed a Hadoop solution that is unique in the marketplace, combining optimized hardware, software and services to streamline deployment and improve the customer experience

The Dell | Cloudera Apache Hadoop Solution was jointly designed by Dell and Cloudera, and embodies all the hardware, software, resources and services needed to run Hadoop in a production environment. This end-to-end solution approach means that you can be in production with Hadoop in a shorter time than is typically possible with homegrown solutions.

The solution is based on the Cloudera Distribution for Apache Hadoop (CDH), and Dell PowerEdge and Dell Networking hardware. This solution includes components that span the entire solution stack:

- Reference architecture and best practices
- Optimized server configurations
- Optimized network infrastructure
- Cloudera Distribution for Apache Hadoop

Solution Use Case Summary

The Dell | Cloudera Apache Hadoop Solution is designed to address the use cases described in [Table 1: Solution Use Cases](#) on page 10:

Table 1: Solution Use Cases

Use case	Description
Big data analytics	Ability to query in real time at the speed of thought on petabyte scale unstructured and semi structured data using HBase and Hive.
ETL Offload	Offload the Extract, Transform, Load (ETL) process from a relational management database or enterprise data warehouse into a Hadoop cluster.
Data Warehouse Optimization	Augment the traditional relational management database or enterprise data warehouse with Hadoop. Hadoop acts as single data hub for all data types.
Data storage	Collect and store unstructured and semi-structured data in a secure, fault-resilient scalable data store that can be organized and sorted for indexing and analysis.
Batch processing of unstructured data	Ability to batch-process (index, analyze, etc.) tens to hundreds of petabytes of unstructured and semi-structured data.
Data archive	Active archival of medium-term (12–36 months) data from EDW/DBMS to expedite access, increase data retention time, or meet data retention policies or compliance requirements.
Integration with data warehouse	Extract, transform, and load data in and out of Hadoop into separate DBMS for advanced analytics.
Big data visualization	Capture, index and visualize unstructured and semi structured big data in real time.
Search and predictive analytics	Crawl, extract, index and transform semi structured and unstructured data for search and predictive analytics.

Solution Components

Figure 1: Dell | Cloudera Apache Hadoop Solution Components on page 11 illustrates the primary components in the Dell | Cloudera Apache Hadoop Solution.

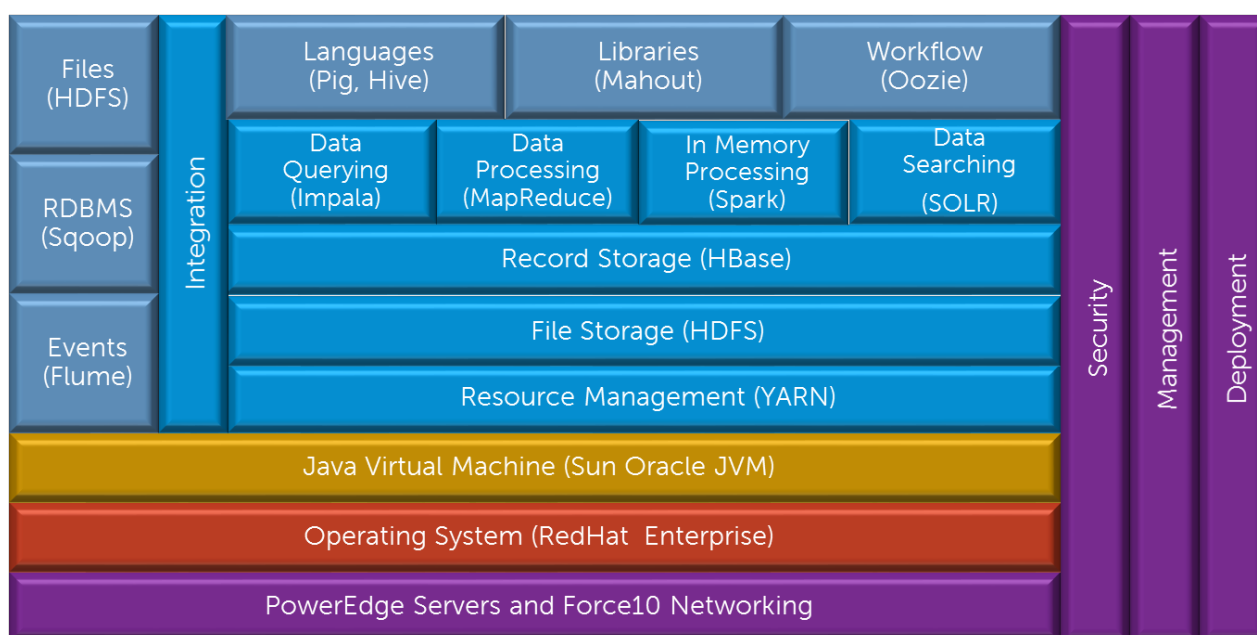


Figure 1: Dell | Cloudera Apache Hadoop Solution Components

The Dell PowerEdge servers, Dell Networking switches, the operating system, and the Java Virtual Machine make up the foundation on which the Hadoop software stack runs.

The left side of the diagram shows the integration components that can be used to move data in and out of the Hadoop system. Apache Sqoop provides data transfer to and from relational databases while Apache Flume is optimized for processing event and log data. The HDFS API and tools can be used to move data files to and from the Hadoop system.

The right side of the diagram shows the capabilities that are integrated across the entire system. Hadoop administration and management is provided by Cloudera Manager while enterprise grade security (via Apache Sentry) is integrated through the entire stack.

The Hadoop components provide multiple layers of functionality on top of this foundation. Apache Zookeeper provides a coordination layer for the distributed processing in the Hadoop system. The Hadoop Distributed File System (HDFS) provides the core storage for data files in the system. HDFS is a distributed, scalable, reliable and portable file system. Apache HBase is a layer that provides record-oriented storage on top of HDFS. HBase can be used as an alternative to direct data file access, optimized for real time data serving environments, and co-exists with direct data file access.

YARN provides a resource management framework for running distributed applications under Hadoop, without being tied to MapReduce. The most popular distributed application is Hadoop's MapReduce, but other applications also run under YARN, such as Apache Spark, Apache Hive, Apache Pig, etc.

Sitting atop these storage layers are four complementary access layers, providing:

- Data processing
- In-memory processing
- Data query
- Data search

All four of these layers can be used simultaneously or independently, depending on the workload and problems being solved.

Table 2: Access Layers

Access Layer	Description
Data processing	MapReduce is the core processing framework in the Hadoop system, and provides a massively parallel data processing framework inspired by Google's MapReduce papers.
In-memory processing	Another processing framework is the real-time, in-memory processing framework called Spark.
Data query	The Data Query layer provides real-time query access to data using Cloudera Impala.
Data search	The Data Search layer provides real-time search of indexed data using Apache SOLR Cloud technology.

Above these layers are a number of Hadoop end-user tools, providing a higher level of abstraction for data access and processing:

Table 3: Data Abstraction Tools

Tool	Description
Apache Pig	Data access and processing language
Apache Hive	Data access and processing language
Apache Mahout	Provides machine learning capabilities
Apache Oozie	Provides a general workflow capability for coordinating complex sequences of production jobs
Apache HUE	Provides a web interface for analyzing data

Software Support

[Table 4: Dell | Cloudera Apache Hadoop Solution Support Matrix](#) on page 12 describes where you can obtain technical support for the various components of the Dell | Cloudera Apache Hadoop Solution.

Table 4: Dell | Cloudera Apache Hadoop Solution Support Matrix

Category	Component	Version	Available Support
Operating System	Red Hat Enterprise Linux Server	6.6	Red Hat Linux support
Operating System	CentOS	6.6	Dell Hardware support
Java Virtual Machine	Sun Oracle JVM	Java 7 (1.7.0_67) Java 8 (1.8.0_60)	N/A
Hadoop	Cloudera Distribution for Apache Hadoop (CDH)	5.5	Cloudera support
Hadoop	Cloudera Manager	5.5	Cloudera support
Hadoop	Cloudera Navigator	2.4	Cloudera support

Cloudera Enterprise Software Overview

Cloudera Enterprise helps you become information-driven by leveraging the best of the open source community with the enterprise capabilities you need to succeed with Apache Hadoop in your organization.

Hadoop for the Enterprise

Designed specifically for mission-critical environments, Cloudera Enterprise includes CDH, the world's most popular open source Hadoop-based platform, as well as advanced system management and data management tools plus dedicated support and community advocacy from our world-class team of Hadoop developers and experts. Cloudera is your partner on the path to big data.

Cloudera Enterprise, with Apache Hadoop at the core, is:

- Unified – one integrated system, bringing diverse users and application workloads to one pool of data on common infrastructure; no data movement required
- Secure – perimeter security, authentication, granular authorization, and data protection
- Governed – enterprise-grade data auditing, data lineage, and data discovery
- Managed – native high-availability, fault-tolerance and self-healing storage, automated backup and disaster recovery, and advanced system and data management
- Open – Apache-licensed open source to ensure your data and applications remain yours, and an open platform to connect with all of your existing investments in technology and skills

Rethink Data Management

- One massively scalable platform to store any amount or type of data, in its original form, for as long as desired or required
- Integrated with your existing infrastructure and tools
- Flexible to run a variety of enterprise workloads - including batch processing, interactive SQL, enterprise search and advanced analytics
- Robust security, governance, data protection, and management that enterprises require

With Cloudera Enterprise, today's leading organizations put their data at the center of their operations, to increase business visibility and reduce costs, while successfully managing risk and compliance requirements.

What's Inside?

Table 5: Included Products

Product	Description
CDH	At the core of Cloudera Enterprise is CDH, which combines Apache Hadoop with a number of other open source projects to create a single, massively scalable system where you can unite storage with an array of powerful processing and analytic frameworks.
Automated Cluster Management - Cloudera Manager	Cloudera Enterprise includes Cloudera Manager to help you easily deploy, manage, monitor, and diagnose issues with your cluster. Cloudera is critical for operating clusters at scale.

Product	Description
Cloudera Support	Get the industry's best technical support for Hadoop. With Cloudera Support, you'll experience more uptime, faster issue resolution, better performance to support your mission critical applications, and faster delivery of the platform features you care about.

Cloudera Enterprise Data Hub

Cloudera Enterprise also offers support for several advanced components that extend and complement the value of Apache Hadoop:

Table 6: Advanced Components

Component	Description
Online NoSQL – HBase	HBase is a distributed key-value store that helps you build real-time applications on massive tables (billions of rows, millions of columns) with fast, random access.
Analytic SQL – Impala	Impala is the industry's leading massively-parallel processing (MPP) SQL engine built for Hadoop.
Search – Cloudera Search	Cloudera Search, based on Apache Solr, lets your users query and browse data in Hadoop just as they would search Google or their favorite e-commerce site.
In-Memory Machine Learning and Stream Processing – Apache Spark	Spark delivers fast, in-memory analytics and real-time stream processing for Hadoop.
Data Management – Cloudera Navigator	Cloudera Navigator provides critical enterprise data audit, lineage, and data discovery capabilities that enterprises require.

Cluster Architecture

The overall architecture of the solution addresses all aspects of a production Hadoop cluster, including the software layers, the physical server hardware, the network fabric, as well as scalability, performance, and ongoing management.

This Cluster Architecture section summarizes the main aspects of the solution architecture. The following topics cover the details in depth:

- [High-Level Node Architecture](#) on page 15
- [Network Fabric Architecture](#) on page 17
- [Cluster Sizing](#) on page 19
- [High Availability](#) on page 21

High-Level Node Architecture

[Figure 2: Cluster Architecture](#) on page 15 displays the roles for the nodes in a basic cluster.

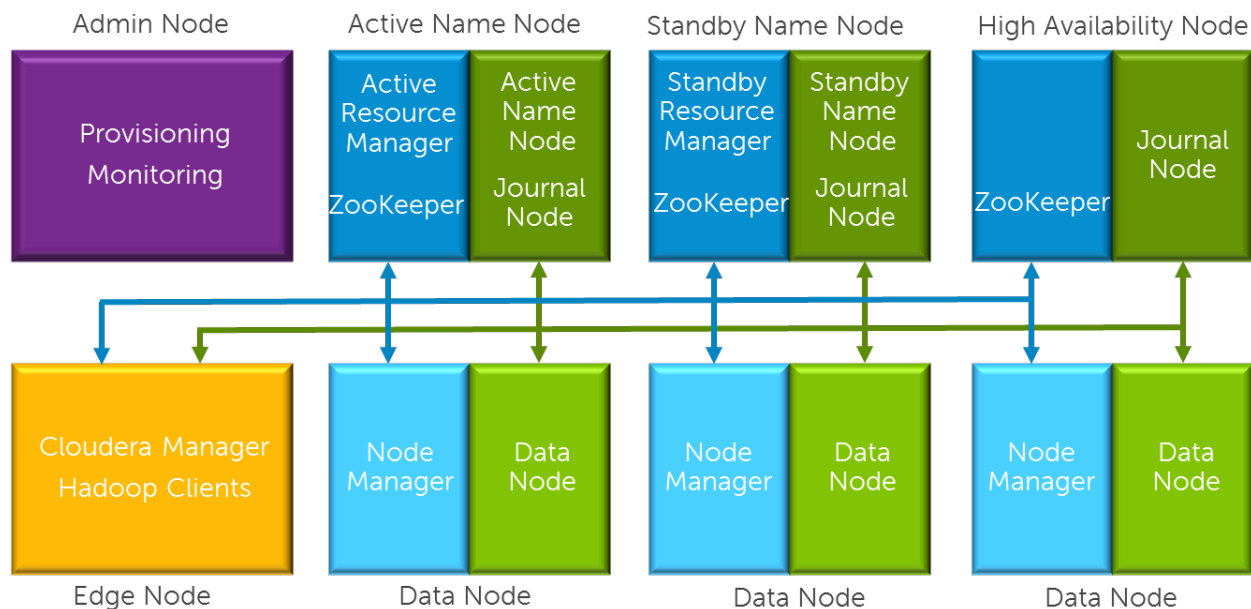


Figure 2: Cluster Architecture

The cluster environment consists of multiple software services running on multiple physical server nodes. The implementation divides the server nodes into several roles, and each node has a configuration optimized for its role in the cluster. The physical server configurations are divided into two broad classes - Data Nodes, which handle the bulk of the Hadoop processing, and Infrastructure Nodes, which support services needed for the cluster operation. A high performance network fabric connects the cluster nodes together, and separates the core data network from management functions.

The minimum configuration supported is six nodes, although at least seven are recommended. The nodes have the following roles:

Table 7: Cluster Node Roles

Node Role	Required?	Hardware Configuration
Administration Node	Optional	Infrastructure
Active Name Node	Required	Infrastructure
Standby Name Node	Required	Infrastructure
High Availability (HA) Node	Required	Infrastructure
Edge (or Gateway) Node	Recommended	Infrastructure
Data Node 1	Required	Data
Data Node 2	Required	Data
Data Node 3	Required	Data

Node Definitions

- **Administration Node** — provides cluster deployment and management capabilities. The Administration Node is optional in cluster deployments, depending on whether existing provisioning, monitoring, and management infrastructure will be used. This reference architecture does not specify the configuration for an administration node, since it is typically site-specific.
- **Active Name Node** — runs all the services needed to manage the HDFS data storage and YARN resource management. This is sometimes called the “master name node.” There are four primary services running on the Active Name Node:
 - Resource Manager (to support cluster resource management, including MapReduce jobs)
 - NameNode (to support HDFS data storage)
 - Journal Manager (to support high availability)
 - Zookeeper (to support coordination)
- **Standby Name Node** — when quorum-based HA mode is used, this node runs the standby namenode process, a second journal manager, and an optional standby resource manager. This node also runs a second Zookeeper service.
- **High Availability (HA) Node** — this node provides the third journal node for HA—the Active Name Nodes and Standby Name Nodes provide the first and second journal nodes. It also runs a third Zookeeper service. The operational databases required for Cloudera Manager and additional metastores are also on this node.
- **Edge Node** — provides an interface between the data and processing capacity available in the Hadoop cluster and a user of that capacity. The Edge Node is connected to the main access LAN, and is sometimes called a “gateway node.” Edge Nodes are optional, but highly recommended.
- **Data Node** — runs all the services required to store blocks of data on the local hard drives and execute processing tasks against that data. A minimum of four Data Nodes are required, and larger clusters are scaled primarily by adding additional Data Nodes. There are three types of services running on the Data Nodes:
 - DataNode Daemon (to support HDFS data storage)
 - NodeManager Daemon (to support YARN job execution)
 - Standalone Daemons like Impalad and HBase Region Server (for services that are not run under YARN.)

[Table 8: Service Locations](#) on page 17 describes the node locations and functions of the cluster services.

Table 8: Service Locations

Physical Node	Software Function
Edge Node	Cloudera Manager Operating System Provisioning Yum Repositories
Active Name Node	NameNode Resource Manager Zookeeper Quorum Journal Node HMaster Impala State Store and Catalog Daemons
Standby Name Node	Standby NameNode Standby Resource Manager Zookeeper Quorum Journal Node
HA Node	Zookeeper Quorum Journal Node Operational Databases (PostgreSQL)
Data Node(x)	DataNode NodeManager RegionServer ImpalaDaemon

Network Fabric Architecture

The cluster network is architected to meet the needs of a high performance and scalable cluster, while providing redundancy and access to management capabilities. [Figure 3: Cluster Network Fabric Architecture](#) on page 18 displays details:

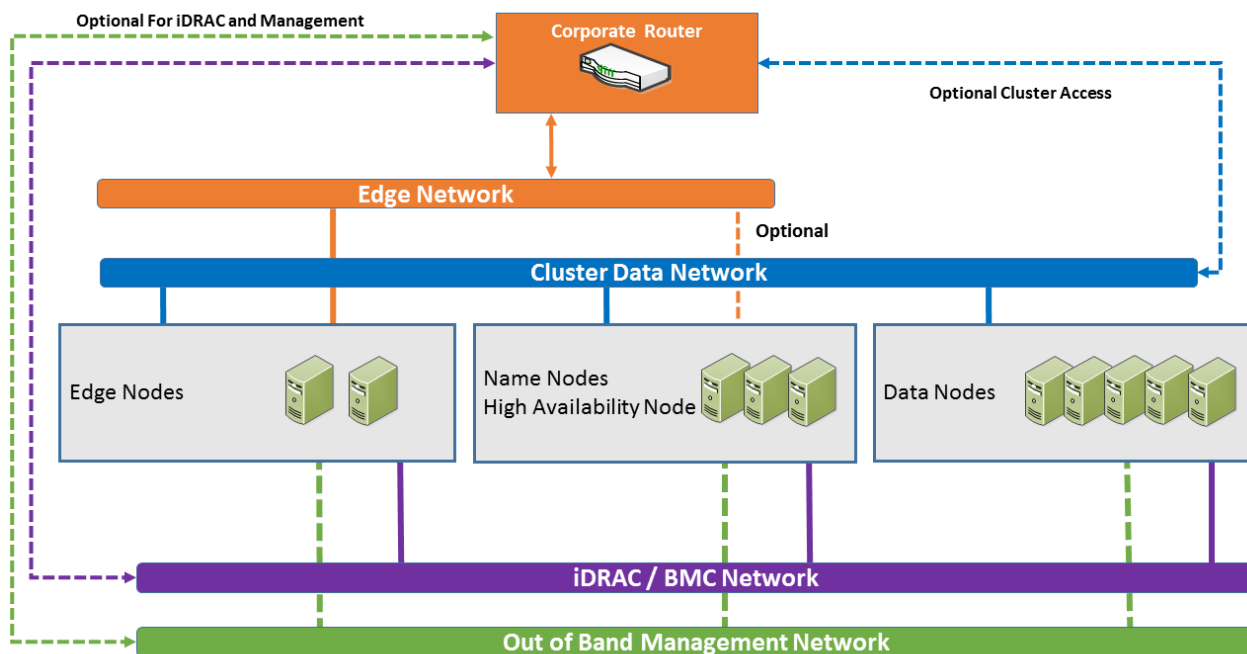


Figure 3: Cluster Network Fabric Architecture

Four distinct networks are used in the cluster:

Table 9: Networks

Logical Network	Connection	Switch
Cluster Data Network	Bonded 10GbE	Dual top of rack switches
Out of Band Management Network	1GbE	Switch per rack, dedicated or shared with BMC network
iDRAC / BMC Network	1GbE	Switch per rack, dedicated or shared with Management network
Edge Network	10GbE, optionally bonded	Top of rack or aggregation switch

Network Definitions

[Table 10: Cloudera Distribution for Apache Hadoop Network Definitions](#) on page 19 defines the Cloudera Distribution for Apache Hadoop networks.

Table 10: Cloudera Distribution for Apache Hadoop Network Definitions

Network	Description	Available Services
Cluster Data Network	The Data network carries the bulk of the traffic within the cluster. This network is aggregated within each rack, and racks are aggregated into the cluster switch. Dual connections with active load balancing are used from each node. This provides increased bandwidth and redundancy when a cable or switch fails.	The CDH services are available on this network.
Out of Band Management Network	The Management network is used to provide cluster management and provisioning capabilities.	This network provides SSH access to cluster nodes for administration. The CDH services are not available on this network, although Cloudera Manager is available for administrative functions.
iDRAC / BMC Network	The BMC network connects the BMC or iDRAC ports and the out-of-band management ports of the switches. It is aggregated into a dedicated switch in each rack, and optionally connected to the top of rack or cluster switches with dedicated vLAN.	This network provides access to the BMC and iDRAC functionality on the servers. It also provides access to the management ports of the cluster switches.
Edge Network	The Edge network provides connectivity from Edge Nodes to the existing core network via the top of rack or cluster switch.	SSH access is available on this network. Some CDH services may also be configured depending on the cluster configuration.

Connectivity between the cluster and existing network infrastructure can be adapted to specific installations. Normally, the cluster Data Nodes are isolated from any existing network but they can be exposed, and optionally routed through an application gateway or firewall.

Cluster Sizing

The architecture is organized into three units for sizing as the Hadoop environment grows. From smallest to largest, they are:

- [Rack](#)
- [Pod](#)
- [Cluster](#)

Each has specific characteristics and sizing considerations documented in this reference architecture. The design goal for the Hadoop environment is to enable you to scale the environment by adding the additional capacity as needed, without the need to replace any existing components.

Rack

A rack is the smallest size designation for a Hadoop environment. A rack consists of the power, network cabling and a management switch to support a group of Data Nodes. A rack is a physical unit, and it's capacity is defined by physical constraints including available space, power, cooling, and floor loading. A rack should use its own power within the data center, independent from other racks, and should

be treated as a fault zone. In the event of a rack level failure in a multiple rack cluster, the cluster will continue to function with reduced capacity.

This reference architecture uses 12 nodes as the nominal size of a rack, but higher or lower densities are possible. Typically, a rack will contain about 12 nodes, with an upper limit of around 20.

Pod

A pod is the set of nodes connected to the first level of network switches in the cluster, and consists of one or more racks. A pod can include a smaller number of nodes initially, and expand up to these maximums over time. A pod is a second level fault zone above the rack level. In the event of a pod level failure in a multiple pod cluster, the cluster will continue to function with reduced capacity. A pod is capable of supporting enough Hadoop server nodes and network switches for a minimum commercial scale installation.

In this reference architecture, a pod supports up to 36 nodes (nominally 3 racks). This size results in a bandwidth oversubscription of 2.25:1 between pods in a full cluster. The size of a pod can vary from this baseline recommendation. Changing the pod size affects the bandwidth oversubscription at the pod level, the size of the fault zones, and the maximum cluster size.

Cluster

A cluster is a single Hadoop environment attached to a pair of network switches providing an aggregation layer for the entire cluster. A cluster can range in size from a pod consisting of a single rack up to a many pods. A single pod cluster is a special case, and can function without an aggregation layer. This scenario is typical for smaller clusters before additional pods are added.

In this reference architecture, a cluster using Dell Networking S6000 switches can scale to 7 pods, and a maximum of 252 nodes. For larger clusters the Dell Networking Z9500 switch can be used.

Sizing Summary

The minimum configuration supported is six nodes:

- Active Name Node
- Standby Name Node
- High Availability (HA) node
- Three (3) Data Nodes

Additionally, a minimum of one Edge Node is recommended per cluster. Larger clusters and clusters with high ingest volumes or rates may benefit from additional Edge Nodes.

The hardware configurations for the Infrastructure Nodes support clusters in the petabyte storage range. Beyond the Infrastructure Nodes, cluster capacity is primarily a function of the server platform and disk drives chosen, and the number of Data Nodes.

[Table 11: Cluster Sizes by Server Model](#) on page 20 shows the recommended number of Data Nodes per rack, pod and cluster for the PowerEdge R730xd servers, using the S4048-ON and S6000 switch models. [Table 12: Alternative Cluster Sizes by Server Model](#) on page 21 shows some alternatives for cluster sizing with different bandwidth oversubscription ratios.

Table 11: Cluster Sizes by Server Model

Server Model	Nodes Per Rack	Nodes Per Pod	Pods Per Cluster	Nodes Per Cluster	Bandwidth Oversubscription
PowerEdge R730xd Data Node	12	36	7	252	2.25 : 1

Table 12: Alternative Cluster Sizes by Server Model

Server Model	Nodes Per Rack	Nodes Per Pod	Pods Per Cluster	Nodes Per Cluster	Bandwidth Oversubscription
PowerEdge R730xd Data Node	12	12	14	168	1.5 : 1
PowerEdge R730xd Data Node	12	24	9	216	2 : 1
PowerEdge R730xd Data Node	10	20	14	240	2.5 : 1
PowerEdge R730xd Data Node	12	36	9	324	3 : 1

High Availability

The architecture implements High Availability at multiple levels through a combination of hardware redundancy and software support.

Hadoop Redundancy

The Hadoop distributed filesystem implements redundant storage for data resiliency, and is aware of node and rack locality.

Data is replicated across multiple nodes, and across racks. This provides multiple copies of data for reliability in the case of disk failure or node failure, and can also increase performance. The number of replicas defaults to three, and can easily be changed. Hadoop will automatically maintain replicas when a node fails – the bonded network provides enough bandwidth to handle replication traffic as well as production traffic.



Note: The Hadoop job parallelism model can scale to larger and smaller numbers of nodes, allowing jobs to run when parts of the cluster are offline.

Network Redundancy

The production network uses bonded connections to pairs of switches in each pod, and the switch pairs are configured using VLT. This configuration provides increased bandwidth capacity, and allows operation at reduced capacity in the event of a network port, network cable, or switch failure.

HDFS Highly Available Name Nodes

The architecture implements High Availability for the HDFS directory through a quorum mechanism that replicates critical name node data across multiple physical nodes. Production clusters normally implement name node HA.

In quorum-based HA, there are typically two name node processes running on two physical servers. At any point in time, one of the NameNodes is in an Active state, and the other is in a Standby state. The Active Name Node is responsible for all client operations in the cluster, while the Standby Name Node is simply acting as a slave, maintaining enough state to provide a fast failover if necessary.

In order for the Standby Name Node to keep its state synchronized with the Active Name Node in this implementation, both nodes communicate with a group of separate daemons called JournalNodes. When any namespace modification is performed by the Active Name Node, it durably logs a record of the modification to a majority of these JournalNodes.

The Standby Name Node is capable of reading the edits from the JournalNodes, and is constantly watching them for changes to the edit log. As the Standby Name Node sees the edits, it applies them to its own namespace. In the event of a failover, the Standby Name Node will ensure that it has read all of the edits from the JournalNodes before promoting itself to the Active state. This ensures that the namespace state is fully synchronized before a failover occurs.

In order to provide a fast failover, it is also necessary that the Standby Name Node has up-to-date information regarding the location of blocks in the cluster. In order to achieve this, the DataNodes are configured with the location of both NameNodes, and they send block location information and heartbeats to both.

There should be an odd number (and at least three) JournalNode daemons, since edit log modifications must be written to a majority of JournalNodes. The JournalNode daemons run on the master, secondary, and HA nodes in this reference architecture.

Resource Manager High Availability

The architecture supports High Availability for the Hadoop YARN resource manager.

Without resource manager HA, a Hadoop resource manager failure causes currently executing jobs to fail. When resource manager HA is enabled, jobs can continue running in the event of a resource manager failure.

Furthermore, upon failover the applications can resume from their last check-pointed state; for example, completed map tasks in a MapReduce job are not rerun on a subsequent attempt. This allows events such as machine crashes or planned maintenance to be handled without any significant performance effect on running applications.

Resource manager HA is implemented by means of an Active/Standby pair of resource managers. On start-up, each resource manager is in the standby state: the process is started, but the state is not loaded. When transitioning to active, the resource manager loads the internal state from the designated state store and starts all the internal services. The stimulus to transition-to-active comes from either the administrator or through the integrated failover controller when automatic failover is enabled.



Note: This feature is not always implemented in production clusters.

Hardware Architecture

The Dell | Cloudera Apache Hadoop Solution utilizes Dell's latest server solutions.

Server Infrastructure Options

The Dell | Cloudera Apache Hadoop Solution includes the following server options:

- [PowerEdge R730xd Server](#) on page 23

PowerEdge R730xd Server

The Dell PowerEdge R730xd is Dell's latest 13th Generation 2-socket, 2U rack server designed to run complex workloads using highly scalable memory, I/O capacity, and flexible network options. It features the Intel® Xeon® processor E5- 2600 v3 product family (Haswell-EP), up to 24 DIMMS, PCI Express® (PCIe) 3.0 enabled expansion slots, and a choice of network interface technologies.

The PowerEdge R730xd platform includes highly-expandable memory (up to 768GB) and impressive I/O capabilities to match. The PowerEdge R730xd can readily handle very demanding workloads, such as data warehouses, e-commerce, virtual desktop infrastructure (VDI), databases, and high-performance computing (HPC).

In addition, the PowerEdge R730xd offers extraordinary storage capacity, making it well suited for data-intensive applications that require storage and I/O performance, like medical imaging and email servers.



Figure 4: PowerEdge R730xd Servers – 2.5" and 3.5" Chassis Options

PowerEdge R730xd Feature Summary

Dell PowerEdge R730xd features include:

- Intel Grantley platform and Intel Xeon E5-2600v3 (Haswell-EP) processors
- Up to 2133 MT/s DDR4 memory
- 24 DIMM slots
- iDRAC8 with Lifecycle Controller
- Network daughter cards for customer choice of LOM speed, fabric and brand
- Front accessible hot-plug drives
- Platinum efficiency power supplies

PowerEdge R730xd Hardware Configurations

The Dell | Cloudera Apache Hadoop Solution supports the following server configurations:

- [Table 13: Hardware Configurations – PowerEdge R730xd Infrastructure Nodes](#) on page 24
- [Table 14: Hardware Configurations – PowerEdge R730xd Data Nodes](#) on page 24

Table 13: Hardware Configurations – PowerEdge R730xd Infrastructure Nodes

Machine Function	Infrastructure Nodes - 3.5" Chassis Option	Infrastructure Nodes - 2.5" Chassis Option
Platform	PowerEdge R730xd	PowerEdge R730xd
Processor	2 x Intel Xeon E5-2650 v3 2.3GHz (10 core)	2 x E5-Intel Xeon E5-2690 v3 2.6GHz (12 core)
RAM (minimum)	128GB	128GB
Network Daughter Card	Intel X520 DP 10Gb DA/SFP+, + I350 DP 1Gb Ethernet (2 x 10GbE, 2x 1GbE)	Intel X520 DP 10Gb DA/SFP+, + I350 DP 1Gb Ethernet (2 x 10GbE, 2x 1GbE)
Add in PCI Network Card	Intel X520 DP 10Gb DA/SFP+ (2 x 10GbE)	Intel X520 DP 10Gb DA/SFP+ (2 x 10GbE)
Disk	8 x 1TB 7.2K SATA 3.5-in.	8 x 1.2TB 10K RPM SAS 12Gbps 2.5-in
Flex Bay Disk	2 x 300GB 10K RPM SAS 12Gbps 2.5-in	2 x 300GB 10K RPM SAS 12Gbps 2.5-in
Storage Controller	PERC H730	PERC H730
Drive Configuration	Combination of RAID 1, RAID 10, and dedicated spindles.	Combination of RAID 1, RAID 10, and dedicated spindles.



Note: Be sure to consult your Dell account representative before changing the recommended disk sizes.

Table 14: Hardware Configurations – PowerEdge R730xd Data Nodes

Machine Function	Data Nodes - 3.5" Chassis Option	Data Nodes - 2.5" Chassis Option
Platform	PowerEdge R730xd	PowerEdge R730xd
Chassis	Up to 12 x 3.5-in Hard Drives	Up to 24 x 2.5-in Hard Drives
Processor	2 x Intel Xeon E5-2650 v3 2.3GHz (10 core)	2 x E5-Intel Xeon E5-2690 v3 2.6GHz (12 core)
RAM (minimum)	128 GB	128 GB
Network Daughter Card	Intel X520 DP 10Gb DA/SFP+, + I350 DP 1Gb Ethernet (2 x 10GbE, 2x 1GbE)	Intel X520 DP 10Gb DA/SFP+, + I350 DP 1Gb Ethernet (2 x 10GbE, 2x 1GbE)
Disk	12 x 4TB 7.2K RPM NLSAS 6Gbps 3.5-in	24 x 1.2TB 10K RPM SAS 12Gbps 2.5-in
Flex Bay Disk	2 x 300GB 10K RPM SAS 12Gbps 2.5-in	2 x 300GB 10K RPM SAS 12Gbps 2.5-in
Storage Controller	PERC H730	PERC H730

Machine Function	Data Nodes - 3.5" Chassis Option	Data Nodes - 2.5" Chassis Option
Drive Configuration	RAID 1 - OS JBOD - data drives	RAID 1 - OS JBOD - data drives



Note: Be sure to consult your Dell account representative before changing the recommended disk sizes.

PowerEdge R730xd Configuration Notes

Two chassis options are supported, using either 3.5" drives, or 2.5" drives. The 3.5" chassis supports higher storage density, while the 2.5" chassis provides more spindles for higher I/O throughput.

This architecture uses the same chassis configuration for Infrastructure Nodes and Data Nodes to simplify maintenance and operations. The only differences are in the number of drives installed in the chassis, which allows a full chassis swap in the event of a node failure.

Full details of the drive configuration and filesystem layouts are in the *Dell | Cloudera Apache Hadoop Solution Deployment Guide*

[Appendix A: Physical Rack Configuration - PowerEdge R730xd](#) on page 37 illustrates the recommended rack layout for PowerEdge R730xd clusters.

[Appendix B: Bill of Materials – PowerEdge R730xd 3.5" Infrastructure Node](#) on page 41, [Appendix C: Bill of Materials – PowerEdge R730xd 3.5" Data Node](#) on page 43, [Appendix D: Bill of Materials – PowerEdge R730xd 2.5" Infrastructure Node](#) on page 45, and [Appendix E: Bill of Materials – PowerEdge R730xd 2.5" Data Node](#) on page 47, contain the full bills of material (BOM) listing for the PowerEdge R730xd server configurations.

Storage Sizing Notes

For drive capacities greater than 4TB or node storage density over 48TB, special consideration is required for HDFS setup. Configurations of this size are close to the limit of Hadoop per-node storage capacity. At a minimum, the HDFS block size should be no less than 128MB and can be as large as 1024MB. Since number of files, blocks per file, compression, and reserved space all factor into the calculations, the configuration will require an analysis of the intended cluster usage and data.

Large per-node density also has an impact on cluster performance in the event of node failure. The bonded 10GbE configuration is recommended for large node densities to minimize performance impacts in this case.



Note: Your Dell representative can assist with these estimates and calculations.

Network Architecture

The cluster network is architected to meet the needs of a high performance and scalable cluster, while providing redundancy and access to management capabilities.

The architecture is a leaf / spine model based on 10GbE network technology, and uses Dell Networking S4048-ON switches for the leaves, and Dell Networking S6000 switches for the spine. IPv4 is used for the network layer. At this time, the architecture does not support or allow for the use of IPv6 for network connectivity.

Four distinct networks are used in the cluster:

Table 15: Cluster Networks

Logical Network	Connection	Switch
Cluster Data Network	Bonded 10GbE	Dual top of rack (Pod) switches and aggregation switches
Out of Band Management Network	1GbE	Dedicated switch per rack
BMC Network	1GbE	Dedicated switch per rack
Edge Network	10GbE, optionally bonded	Top of rack or cluster switch

Each network uses a separate VLAN, and dedicated components when possible. [Figure 5: Hadoop Network Connections](#) on page 26 shows the logical organization of the network.

For more information on the actual configuration of the interfaces and switches, please see the *Dell | Cloudera Apache Hadoop Solution Deployment Guide*.

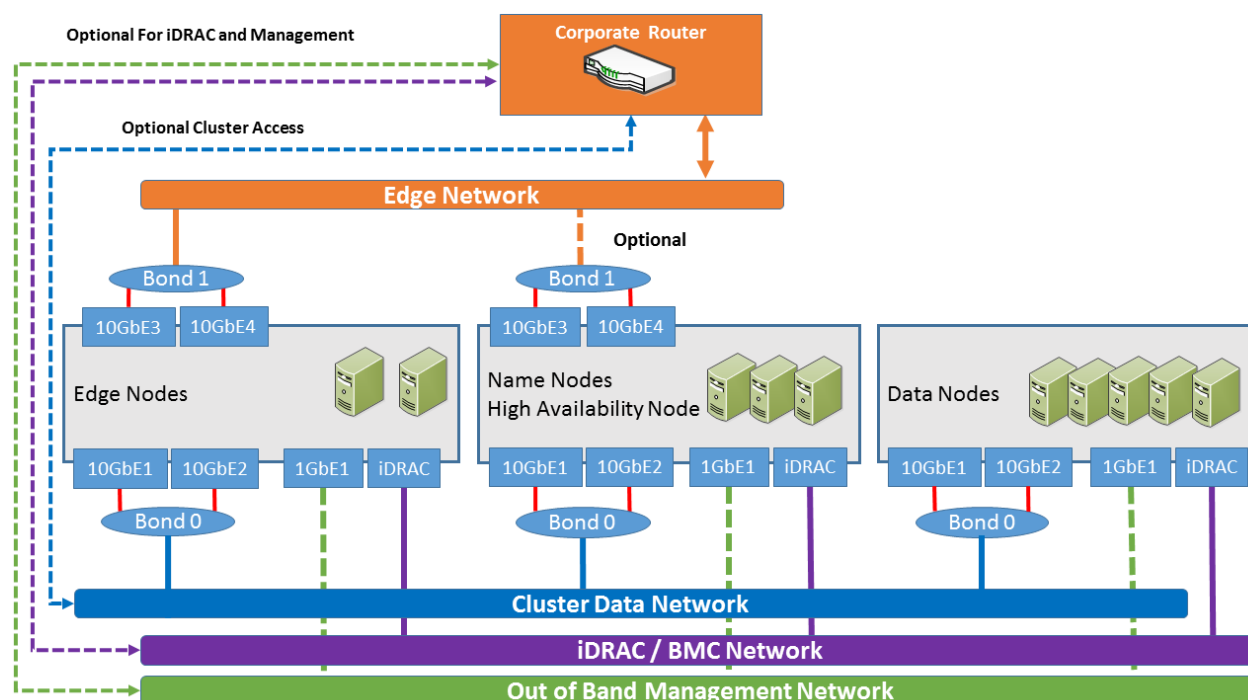


Figure 5: Hadoop Network Connections

Physical Network Components

The Dell | Cloudera Apache Hadoop Solution physical networks consists of the following components:

- [Server Node Connections](#) on page 27
- [Pod Switches](#) on page 27
- [Cluster Aggregation Switches](#) on page 28
- [Core Network](#) on page 30
- [Layer 2 and Layer 3 Separation](#) on page 30
- [Management and BMC Networks](#) on page 30
- [Network Equipment Summary](#) on page 31

Server Node Connections

Server connections to the network switches for the data network are bonded, and use an Active-Active LAN aggregation group (LAG) in a load-balance configuration using IEEE 802.3 Link Aggregation Control Protocol (LACP) . (Under Linux®, this is referred to as 802.3ad or mode 4 bonding).

The connections are made to a pair of Pod switches, to provide redundancy in the case of port, cable, or switch failure. The switch ports are configured as a LAG. Each server has an additional 1GbE connection to the management network to facilitate server management and provisioning.

Connections to the BMC network use a single connection from the BMC port to a dedicated switch in each rack.

Edge Nodes have an additional pair of 10GbE connections to the Pod switch. These connections facilitate high-performance ingest and cluster access between applications running on those nodes, and the core datacenter network.

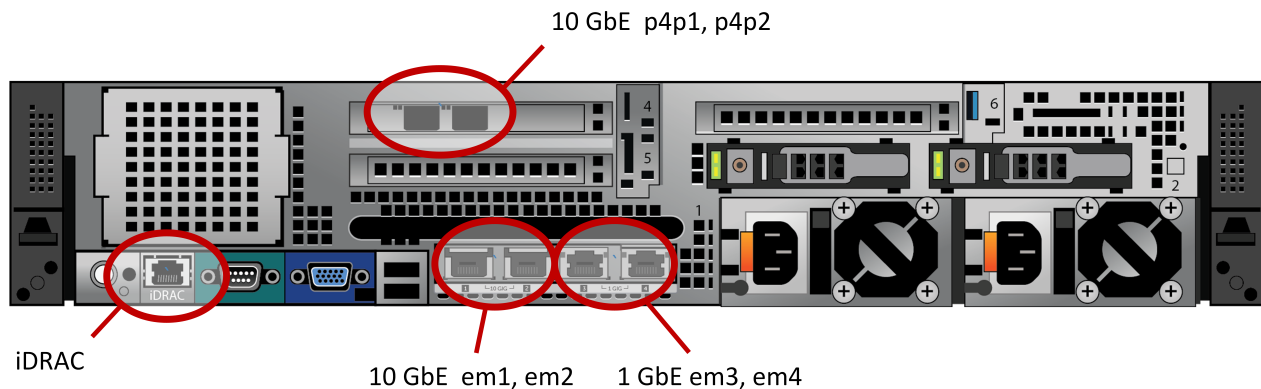


Figure 6: PowerEdge R730xd Node Network Ports

Pod Switches

Each pod uses a pair of Dell Networking S4048-ONs as the first layer switches. These switches are configured for high availability using the Virtual Link Trunking (VLT) feature. VLT allows the servers to terminate their LAG interfaces into two different switches instead of one. This provides redundancy within the pod if a switch fails or needs maintenance, while providing active-active bandwidth utilization. (The pod switches are often referred to as Top of Rack (ToR) switches, although this architecture splits a physical rack from a logical pod.)

[Figure 7: Single Pod Networking Equipment](#) on page 28 shows the single pod network configuration, with a pair of Dell Networking S4048-ON switches aggregating the pod traffic.

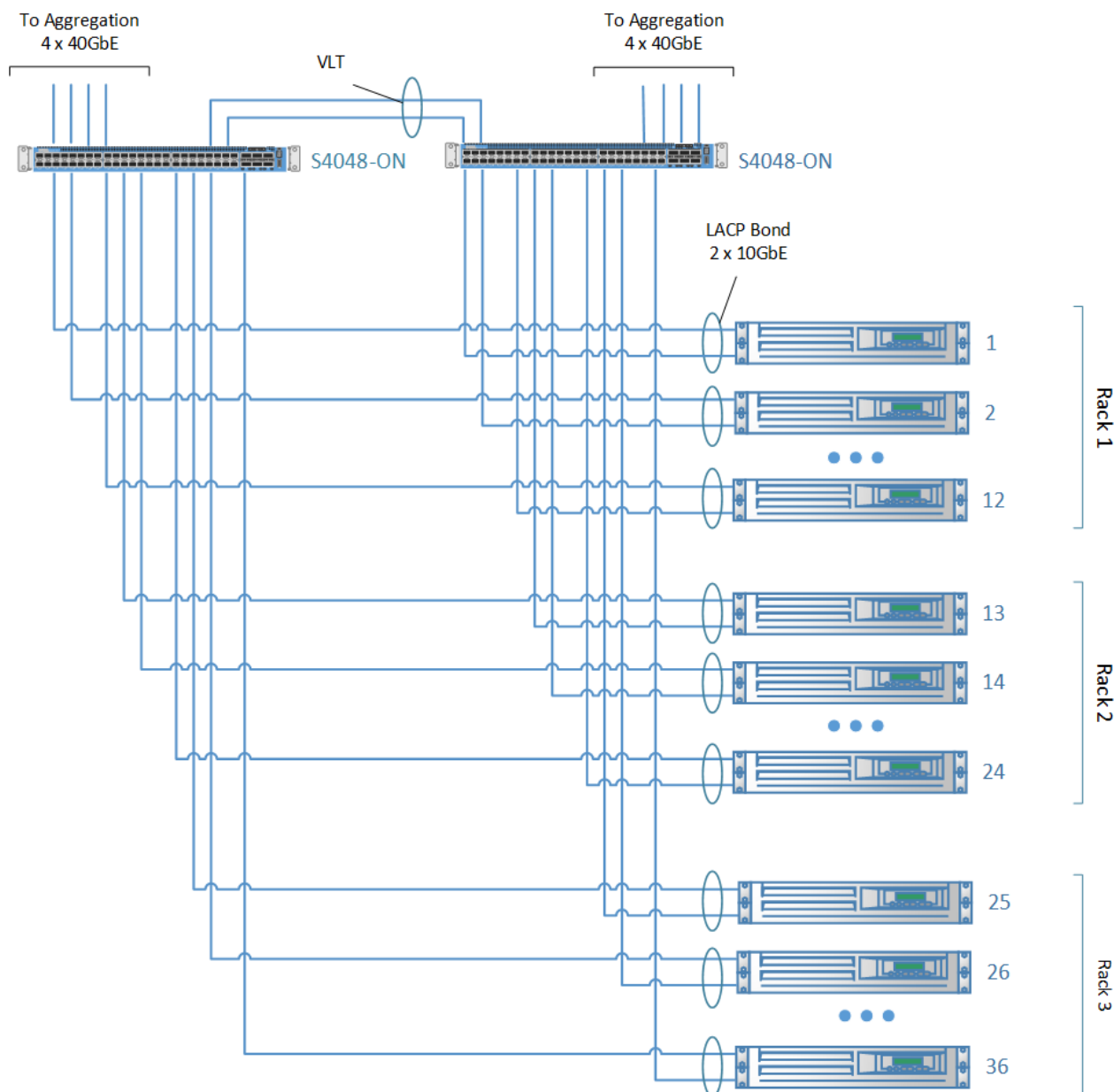


Figure 7: Single Pod Networking Equipment

For a single pod, the pod switches can act as the aggregation layer for the entire cluster. For multi-pod clusters, a cluster aggregation layer is required.

In this architecture, each pod is managed as a separate entity from a switching perspective, and pod switches connect only to the aggregation switches.

Cluster Aggregation Switches

For clusters consisting of one more pods, the architecture uses either of the following models for aggregation switches:

- Dell Networking S6000
- Dell Networking Z9500

The choice depends on the initial size and planned scaling. The Dell Networking S6000 is preferred for lower cost and medium scalability. This design can handle up to seven pods, as described in [Cluster Sizing](#) on page 19. The Dell Networking Z9500 is recommended for larger deployments.

Dell Networking S6000 Cluster Aggregation

Figure 8: Dell Networking S6000 Multi-pod Networking Equipment on page 29 illustrates the configuration for a multiple pod cluster using the S6000 as a cluster aggregation switch.

Like the pod switches, the aggregation switches are connected in pairs using VLT. The uplink from each S4048-ON pod switch to the aggregation pair is 160 Gb, using four 40Gb interfaces. Since both S6000s connect to the aggregation pair, there is a collective bandwidth of 320 Gb available from each pod.

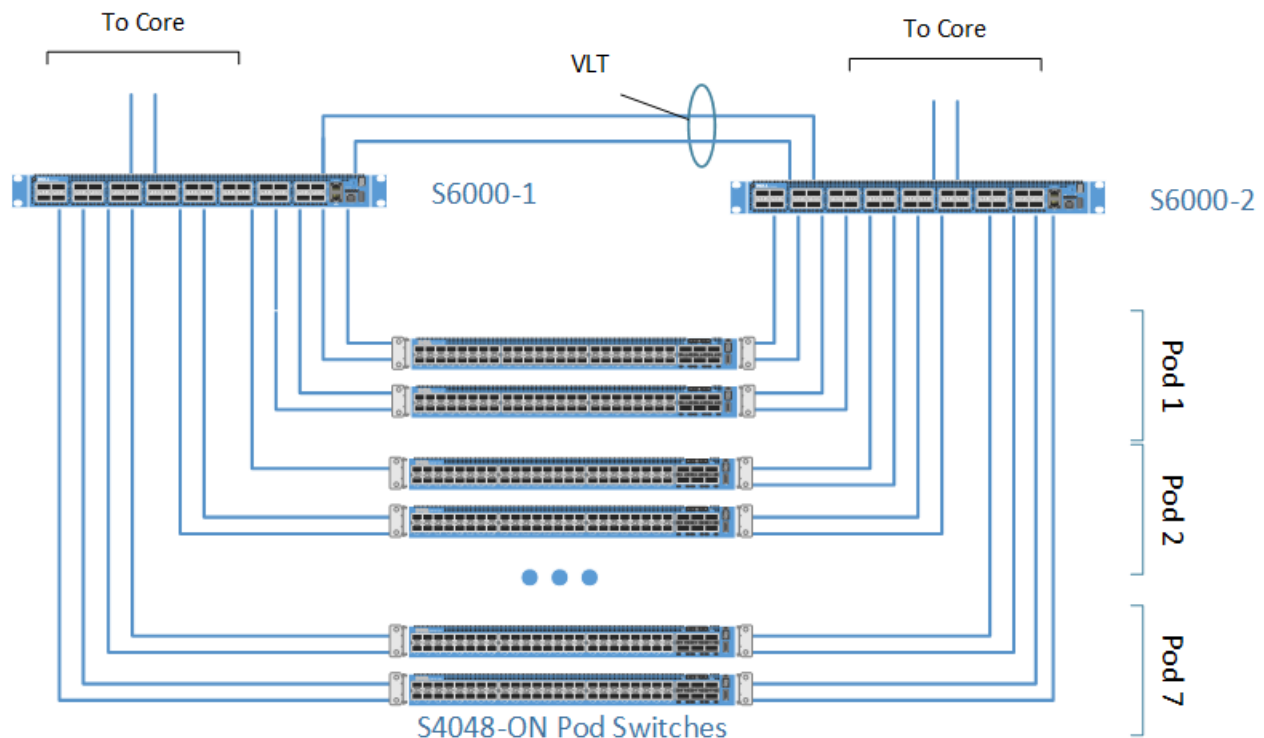


Figure 8: Dell Networking S6000 Multi-pod Networking Equipment

Dell Networking Z9500 Cluster Aggregation

For larger initial deployments, deployments where extreme scale up is planned, or instances where the cluster needs to be co-located with other applications in different racks, the recommended option is the Dell Networking Z9500 core switch.

The Dell Networking Z9500 is a 132-port, 40GbE high-capacity switch, that can be configured as 528 10GbE ports. The pod-to-pod bandwidth needed in Hadoop is best addressed by a 40G-capable, non-blocking switch and the Dell Networking Z9500 can provide a cumulative bandwidth of 10.4 Tbps of throughput at line-rate traffic from every port.

A straightforward modification of this reference architecture can use the Z9500 as an alternative to the S6000 for the cluster aggregation switch. In this configuration, a cluster can scale to 32 pods, or a maximum of 1280 nodes.

In practice, clusters larger than a few hundred nodes often eliminate the redundant dual network connectivity in this reference architecture, since there are enough pods and nodes in the cluster to minimize the impact of a network failure though the natural redundancy built into Hadoop. Also, network oversubscription ratios are often relaxed for clusters of this size. As a result, we will normally use a different network architecture for clusters of this size, based on Layer 3 networking. For example, *Figure 9: Multi-Pod View Using Dell Networking Z9500 Switches (Based on Layer-3 ECMP)* on page 30 illustrates an alternative configuration for a multiple pod cluster using Layer 3 and ECMP routing. In this configuration, a cluster can scale to 64 pods, or a maximum of 2560 nodes with a 2.5 : 1 oversubscription per pod.

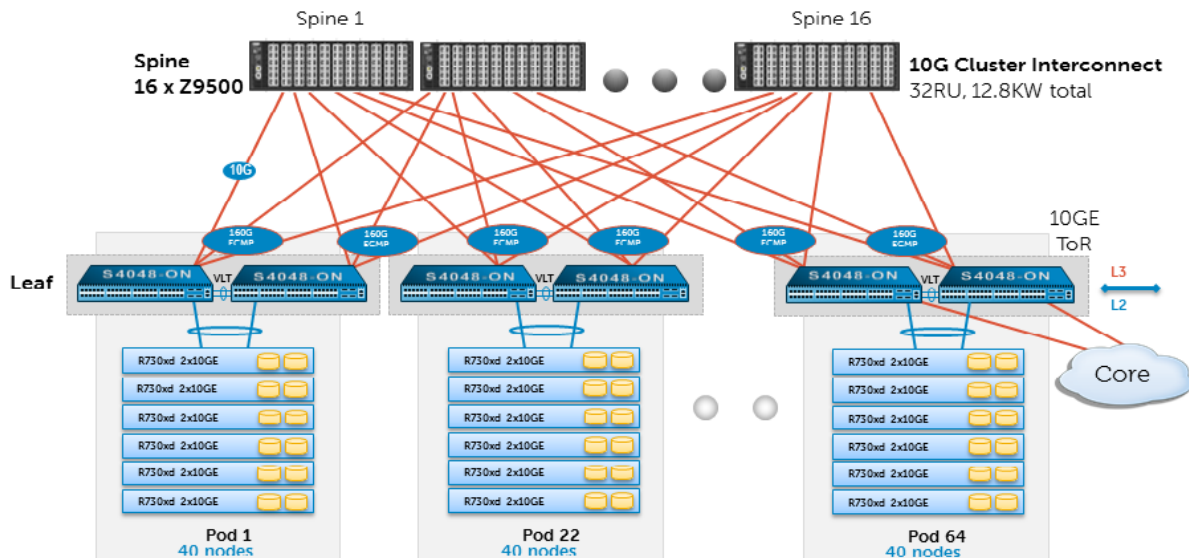


Figure 9: Multi-Pod View Using Dell Networking Z9500 Switches (Based on Layer-3 ECMP)

Core Network

The aggregation layer functions as the network core for the cluster. In most instances, the cluster will connect to a larger core within the enterprise, represented by the cloud in [Figure 8: Dell Networking S6000 Multi-pod Networking Equipment](#) on page 29. When using the Dell Networking S6000, four 40GbE ports are reserved at the aggregation level for connection to the core. Details of the connection are site specific, and need to be determined as part of the deployment planning.

Layer 2 and Layer 3 Separation

The layer-2 and layer-3 boundaries are separated at either the pod or the aggregation layer, and either option is equally viable. This architecture is based on layer 2 for switching within the cluster. The colors blue and red in [Figure 9: Multi-Pod View Using Dell Networking Z9500 Switches \(Based on Layer-3 ECMP\)](#) on page 30 represent the layer-2 and layer-3 boundaries. This document uses layer-2 as the reference up to the aggregation layer.

Management and BMC Networks

In addition to the cluster data network, two networks are provided for cluster management - the out of band management network and the iDRAC (or BMC) network.

The iDRAC and switch management ports are all aggregated into a per rack Dell Networking S3048-ON switch, with a dedicated vLAN. This provides a dedicated iDRAC / BMC network.

In addition, a 1GbE port from each server is assigned a separate vLAN, and tied into the same switch. This provides a separate management network for server provisioning and management tasks. Note, the Cloudera Enterprise services are not multi-homed, and are not available on the management interface.

The management switches can be connected to the core, or connected to a dedicated management network if out of band management is required. In most instances, the vLAN separation provides adequate isolation between the management networks. For more secure installations, independent physical switches can be used in each rack.

Network Equipment Summary

[Table 16: Per Rack Network Equipment](#) on page 31, [Table 17: Per Pod Network Equipment](#) on page 31 and [Table 18: Per Cluster Aggregation Network Switches for Multiple Pods](#) on page 31 summarize the required cluster networking equipment. [Table 19: Per Node Network Cables Required – 10GbE Configurations](#) on page 31 summarizes the number of cables needed for a cluster.

Table 16: Per Rack Network Equipment

Component	Quantity
Total Racks	1 (12 nodes nominal)
Management Switch	1 x Dell Networking S3048-ON
Switch Interconnect Cables	1 x 1 GbE Cables (to next rack management switch)

Table 17: Per Pod Network Equipment

Component	Quantity
Total Racks	3 (36 Nodes)
Top-of-rack Switches	2 x Dell Networking S4048-ON
Switch Interconnect Cables (for VLT)	2 x 40Gb QSFP+ Cables
Pod Uplink Cables (To Aggregate Switch)	8 x 40Gb QSFP+ Cables

Table 18: Per Cluster Aggregation Network Switches for Multiple Pods

Component	Quantity
Total Pods	7
Aggregation Layer Switches	2 x Dell Networking S6000
Switch Interconnect Cables	2 x 40GB QSFP+ Cables

Table 19: Per Node Network Cables Required – 10GbE Configurations

Description	1GbE Cables Required	10GbE Cables with SFP+ Required
Name and HA Nodes	2 x Number of Nodes	2 x Number of Nodes
Edge Nodes	2 x Number of Nodes	4 x Number of Nodes
Data Nodes	2 x Number of Nodes	2 x Number of Nodes

Cloudera Enterprise Software

The Dell | Cloudera Apache Hadoop Solution is based on Cloudera Enterprise, which includes Cloudera's distribution for Hadoop (CDH) 5.5 and Cloudera Manager.

Cloudera Enterprise software components include:

- [Cloudera Manager](#) on page 32
- [Cloudera RTQ \(Impala\)](#) on page 32
- [Cloudera Search](#) on page 33
- [Cloudera BDR](#) on page 33
- [Cloudera Navigator](#) on page 33
- [Cloudera Support](#) on page 34

Cloudera Manager

Cloudera Manager is designed to make Hadoop administration simple and straightforward, at any scale. With Cloudera Manager, you can easily deploy and centrally operate the complete Hadoop stack. The application automates the installation process, reducing deployment time from weeks to minutes; gives you a cluster-wide, real-time view of nodes and services running; provides a single, central console to enact configuration changes across your cluster; and incorporates a full range of reporting and diagnostic tools to help you optimize performance and utilization.

Cloudera Manager is available as part of both the Cloudera Standard and Cloudera Enterprise product offerings. With Cloudera Standard, you get a full set of functionality to deploy, configure, manage, monitor, diagnose and scale your cluster—the most comprehensive and advanced set of management capabilities available from any vendor. When you upgrade to Cloudera Enterprise, you get additional capabilities for integration, process automation and disaster recovery that are focused on helping you operate your cluster successfully in enterprise environments.

Cloudera RTQ (Impala)

Cloudera Impala is an open source Massively Parallel Processing (MPP) query engine that runs natively in Apache Hadoop. The Apache-licensed Impala project brings scalable parallel database technology to Hadoop, enabling users to issue low-latency SQL queries to data stored in HDFS and Apache HBase without requiring data movement or transformation.

Impala is integrated from the ground up as part of the Hadoop ecosystem and leverages the same flexible file and data formats, metadata, security and resource management frameworks used by MapReduce, Apache Hive™, Apache Pig™, and other components of the Hadoop stack.

Designed to complement MapReduce, which specializes in large-scale batch processing, Impala is an independent processing framework optimized for interactive queries. With Impala, analysts and data scientists now have the ability to perform real-time, “speed of thought” analytics on data stored in Hadoop via SQL or through business intelligence (BI) tools.

The result is that large-scale data processing and interactive queries can be done on the same system using the same data and metadata — removing the need to migrate data sets into specialized systems and/or proprietary formats simply to perform analysis.

Cloudera Search

Cloudera Search delivers full-text, interactive search to CDH, Cloudera's 100% open source distribution including Apache Hadoop. Powered by Apache Solr, Cloudera Search enriches the Hadoop platform and enables a new generation of search – Big Data search – through scalable indexing of data within HDFS and Apache HBase™.

Cloudera Search gains the same fault tolerance, scale, visibility, and flexibility provided to other Hadoop workloads, due to its integration with CDH.

Apache Solr has been the enterprise standard for open source search since its release in 2006. Its active and mature community drives wide adoption across verticals and industries, and its APIs are feature-rich and extensible. Cloudera Search extends the value of Apache Solr by tightly integrating and optimizing it to run on CDH and Cloudera Manager.

Cloudera BDR

BDR is an add-on subscription to Cloudera Enterprise that provides end-to-end business continuity. When you add BDR to your Cloudera Enterprise subscription, you'll get the management capabilities and support you need to get maximum value from the powerful disaster recovery features available in CDH.

Cloudera BDR makes it easy to configure and manage disaster recovery policies for data stored in CDH. With BDR you can:

- Centrally configure and manage disaster recovery workflows for files (HDFS) and metadata (Hive) through an easy-to-use graphical interface
- Consistently meet or exceed service level agreements (SLAs) and recovery time objectives (RTOs) through simplified management and process automation

BDR includes:

- Centralized management for HDFS replication through Cloudera Manager
- Centralized management for Hive replication through Cloudera Manager
- 8x5 or 24x7 Cloudera Support

Key features of BDR:

- Define file and directory-level replication policies
- Schedule replication jobs
- Monitor progress through a centralized console
- Identify discrepancies between primary and secondary system(s)

Cloudera Navigator

Navigator is an add-on subscription to Cloudera Enterprise that provides the first fully integrated data management tool for Cloudera Enterprise. It's designed to provide all of the capabilities required for administrators, data managers and analysts to secure, govern, and explore the large amounts of diverse data that land in CDH.

Cloudera Navigator is the only complete data governance solution for Hadoop, offering critical capabilities such as data discovery, continuous optimization, audit, lineage, metadata management, and policy enforcement.

The Navigator subscription gives you access to all of the capabilities of the Cloudera Navigator application. With Navigator, you can:

- Store sensitive data in CDH while maintaining compliance with regulations and internal audit policies
- Verify access permissions to files and directories
- Maintain a full audit history of HDFS, Hive and HBase data access
- Report on data access by user and type
- Integrate with third-party SIEM tools

Navigator includes:

- Centralized audit management and reporting for HDFS, Hive and HBase
- 8x5 or 24x7 Cloudera Support

Key features of Cloudera Navigator:

- Configuration of audit information for HDFS, HBase and Hive
- Centralized view of data access and permissions
- Simple, query-able interface with filters for type of data or access patterns
- Export of full or filtered audit history for integration with third-party SIEM tools

Cloudera Support

As the use of Hadoop grows and an increasing number of groups and applications move into production, your Hadoop users will expect greater levels of performance and consistency. Cloudera's proactive production-level support gives your administrators the expertise and responsiveness they need.

Cloudera Support includes:

Table 20: Cloudera Support Features

Feature	Description
Flexible Support Windows	Choose 8x5 or 24x7 to meet SLA requirements.
Configuration Checks	Verify that your Hadoop cluster is fine-tuned for your environment.
Escalation and Issue Resolution	Resolve support cases with maximum efficiency.
Comprehensive Knowledge Base	Expand your Hadoop knowledge with hundreds of articles and tech notes.
Support for Certified Integration	Connect your Hadoop cluster to your existing data analysis tools.
Proactive Notification	Stay up-to-speed on new developments and events.

With Cloudera Enterprise, you can leverage your existing team's experience and Cloudera's expertise to put your Hadoop system into effective operation. Built-in predictive capabilities anticipate shifts in the Hadoop infrastructure to support reliable function.

Cloudera Enterprise makes it easy to run open source Hadoop in production, by:

- Simplifying and accelerating Hadoop deployment
- Reducing the costs and risks of adopting Hadoop in production
- Reliably operating Hadoop in production with repeatable success

- Applying SLAs to Hadoop
- Increasing control over Hadoop cluster provisioning and management

Deployment Methodology

A suggested deployment workflow is documented in the *Dell | Cloudera Apache Hadoop Solution Deployment Guide*, which is a complement to this reference architecture.

Appendix A: Physical Rack Configuration - PowerEdge R730xd

This appendix contains suggested rack layouts for single rack, single pod, and multiple pod installations. Actual rack layouts will vary depending on power, cooling, and loading constraints.

Table 21: Single Rack Configuration – PowerEdge R730xd

RU	RACK1
42	R1- Switch 1: Dell Networking S4048-ON
41	R1- Switch 2: Dell Networking S4048-ON
40	Cable Management
39	Cable Management
38	R1 - Dell Networking S3048-ON iDRAC Mgmt switch
37	Cable Management
36	Cable Management
35 29	Empty
28 27	Edge01: PowerEdge R730xd
26 25	Master Name Node:PowerEdge R730xd
24 23	Secondary Name Node PowerEdge R730xd
22 21	HA Node: PowerEdge R730xd
20 19	Empty
18 17	Empty
16 15	R1- Chassis08: PowerEdge R730xd
14 13	R1- Chassis07: PowerEdge R730xd

RU	RACK1
12 11	R1- Chassis06: PowerEdge R730xd
10 9	R1- Chassis05: PowerEdge R730xd
8 7	R1- Chassis04: PowerEdge R730xd
6 5	R1- Chassis03: PowerEdge R730xd
4 3	R1- Chassis02: PowerEdge R730xd
2 1	R1- Chassis01: PowerEdge R730xd

Table 22: Initial Pod Rack Configuration – PowerEdge R730xd

RU	RACK1	RACK2	RACK3
42	Empty	R2- Switch 1: Dell Networking S4048-ON	Empty
41	Empty	R2- Switch 2: Dell Networking S4048-ON	Empty
40	Cable Management	Cable Management	Cable Management
39	Cable Management	Cable Management	Cable Management
38	R1 - Dell Networking S3048-ON iDRAC Mgmt switch	R2 - Dell Networking S3048-ON iDRAC Mgmt switch	R3 - Dell Networking S3048-ON iDRAC Mgmt switch
37	Cable Management	Cable Management	Cable Management
36	Cable Management	Cable Management	Cable Management
35	Master Name Node:PowerEdge R730xd	Edge01: PowerEdge R730xd	R3 - Switch 1: Dell Networking S6000
34			R3 - Switch 2: Dell Networking S6000
33 32	Admin Node PowerEdge R730xd	Secondary Name Node PowerEdge R730xd	HA Node: PowerEdge R730xd
31 21	Empty	Empty	Empty

RU	RACK1	RACK2	RACK3
20 19	R1- Chassis10: PowerEdge R730xd	R2- Chassis10: PowerEdge R730xd	R3- Chassis10: PowerEdge R730xd
18 17	R1- Chassis09: PowerEdge R730xd	R2- Chassis09: PowerEdge R730xd	R3- Chassis09: PowerEdge R730xd
16 15	R1- Chassis08: PowerEdge R730xd	R2- Chassis08: PowerEdge R730xd	R3- Chassis08: PowerEdge R730xd
14 13	R1- Chassis07: PowerEdge R730xd	R2- Chassis07: PowerEdge R730xd	R3- Chassis07: PowerEdge R730xd
12 11	R1- Chassis06: PowerEdge R730xd	R2- Chassis06: PowerEdge R730xd	R3- Chassis06: PowerEdge R730xd
10 9	R1- Chassis05: PowerEdge R730xd	R2- Chassis05: PowerEdge R730xd	R3- Chassis05: PowerEdge R730xd
8 7	R1- Chassis04: PowerEdge R730xd	R2- Chassis04: PowerEdge R730xd	R3- Chassis04: PowerEdge R730xd
6 5	R1- Chassis03: PowerEdge R730xd	R2- Chassis03: PowerEdge R730xd	R3- Chassis03: PowerEdge R730xd
4 3	R1- Chassis02: PowerEdge R730xd	R2- Chassis02: PowerEdge R730xd	R3- Chassis02: PowerEdge R730xd
2 1	R1- Chassis01: PowerEdge R730xd	R2- Chassis01: PowerEdge R730xd	R3- Chassis01: PowerEdge R730xd

Table 23: Additional Pod Rack Configuration – PowerEdge R730xd

RU	RACK1	RACK2	RACK3
42	Empty	R2- Switch 1: Dell Networking S4048-ON	Empty
41	Empty	R2- Switch 2: Dell Networking S4048-ON	Empty
40	Cable Management	Cable Management	Cable Management
39	Cable Management	Cable Management	Cable Management
38	R1 - Dell Networking S3048-ON iDRAC Mgmt switch	R2 - Dell Networking S3048-ON iDRAC Mgmt switch	R3 - Dell Networking S3048-ON iDRAC Mgmt switch
37	Cable Management	Cable Management	Cable Management
36	Cable Management	Cable Management	Cable Management

RU	RACK1	RACK2	RACK3
33 25	Empty	Empty	Empty
24 23	R1- Chassis12: PowerEdge R730xd	R2- Chassis12: PowerEdge R730xd	R1- Chassis12: PowerEdge R730xd
22 21	R1- Chassis11: PowerEdge R730xd	R2- Chassis11: PowerEdge R730xd	R3- Chassis11: PowerEdge R730xd
20 19	R1- Chassis10: PowerEdge R730xd	R2- Chassis10: PowerEdge R730xd	R3- Chassis10: PowerEdge R730xd
18 17	R1- Chassis09: PowerEdge R730xd	R2- Chassis09: PowerEdge R730xd	R3- Chassis09: PowerEdge R730xd
16 15	R1- Chassis08: PowerEdge R730xd	R2- Chassis08: PowerEdge R730xd	R3- Chassis08: PowerEdge R730xd
14 13	R1- Chassis07: PowerEdge R730xd	R2- Chassis07: PowerEdge R730xd	R3- Chassis07: PowerEdge R730xd
12 11	R1- Chassis06: PowerEdge R730xd	R2- Chassis06: PowerEdge R730xd	R3- Chassis06: PowerEdge R730xd
10 9	R1- Chassis05: PowerEdge R730xd	R2- Chassis05: PowerEdge R730xd	R3- Chassis05: PowerEdge R730xd
8 7	R1- Chassis04: PowerEdge R730xd	R2- Chassis04: PowerEdge R730xd	R3- Chassis04: PowerEdge R730xd
6 5	R1- Chassis03: PowerEdge R730xd	R2- Chassis03: PowerEdge R730xd	R3- Chassis03: PowerEdge R730xd
4 3	R1- Chassis02: PowerEdge R730xd	R2- Chassis02: PowerEdge R730xd	R3- Chassis02: PowerEdge R730xd
2 1	R1- Chassis01: PowerEdge R730xd	R2- Chassis01: PowerEdge R730xd	R3- Chassis01: PowerEdge R730xd

Appendix B: Bill of Materials – PowerEdge R730xd 3.5" Infrastructure Node

Table 24: Active Name Node, Standby Name Node, Admin Node, Edge Node and HA Nodes – PowerEdge R730xd

Quantity	SKU	Component
1	379-BBWM	INFO QS, 13G HADOOP BUNDLE
1	591-BBCH	PowerEdge R730/R730xd Motherboard
1	210-ADBC	PowerEdge R730xd Server
1	350-BBEW	Chassis with up to 12, 3.5" Hard Drives and 2, 2.5" Flex Bay Hard Drives
1	340-AKPM	PowerEdge R730xd Shipping
1	338-BFFF	Intel Xeon E5-2650 v3 2.3GHz,25M Cache,9.60GT/s QPI,Turbo,HT,10C/20T (105W) Max Mem 2133MHz
1	374-BBGM	Upgrade to Two Intel Xeon E5-2650 v3 2.3GHz,25M Cache,9.60GT/s QPI,Turbo,HT,10C/20T (105W)
1	330-BBCR	R730/xd PCIe Riser 1, Right
1	330-BBCO	R730/xd PCIe Riser 2, Center
1	370-ABUF	2133MT/s RDIMMs
1	370-AAIP	Performance Optimized
8	370-ABUG	16GB RDIMM, 2133 MT/s, Dual Rank, x4 Data Width
1	780-BBLH	No RAID for H330/H730/H730P (1-24 HDDs or SSDs)
1	405-AAEG	PERC H730 Integrated RAID Controller, 1GB Cache
2	400-AJPR	300GB 10K RPM SAS 12Gbps 2.5in Flex Bay Hard Drive
8	400-AEGJ	4TB 7.2K RPM SATA 6Gbps 3.5in Hot-plug Hard Drive,13G
1	540-BBBB	Intel X520 DP 10Gb DA/SFP+, + I350 DP 1Gb Ethernet, Network Daughter Card
2	407-BBEQ	SFP+, Short Range, Optical Transceiver, LC Connector, 10Gb and 1Gb compatible for Intel and Broadcom
1	540-BBCT	Intel X520 DP 10Gb DA/SFP+ Server Adapter
1	350-BBEJ	Bezel
1	770-BBBQ	ReadyRails Sliding Rails Without Cable Management Arm
1	384-BBBL	Performance BIOS Settings
1	450-ADWS	Dual, Hot-plug, Redundant Power Supply (1+1), 750W
2	492-BBDH	C13 to C14, PDU Style, 12 AMP, 2 Feet (.6m) Power Cord, North America

Quantity	SKU	Component
1	631-AAJG	Electronic System Documentation and OpenManage DVD Kit, PowerEdge R730/xd
1	619-ABVR	No Operating System
1	421-5736	No Media Required
1	800-BBDM	UEFI BIOS
1	332-1286	US Order
2	374-BBHM	Standard Heatsink for PowerEdge R730/R730xd
1	370-ABWE	DIMM Blanks for System with 2 Processors
1	385-BBHO	iDRAC8 Enterprise, integrated Dell Remote Access Controller, Enterprise
1	976-9030	ProSupport Plus: 7x24 HW/SW Tech Support and Assistance, 3 Year
1	976-9007	Dell Hardware Limited Warranty Plus On Site Service
1	976-9029	ProSupport Plus: 7x24 Next Business Day Onsite Service, 3 Year
1	951-2015	Thank you for choosing Dell ProSupport Plus. For tech support, visit http://www.dell.com/contactdell
2	900-9997	On-Site Installation Declined

Appendix C: Bill of Materials – PowerEdge R730xd 3.5" Data Node

Table 25: Data Node – PowerEdge R730xd 3.5"

Quantity	SKU	Component
1	379-BBWM	INFO QS, 13G HADOOP BUNDLE
1	210-ADBC	PowerEdge R730xd Server
1	591-BBCH	PowerEdge R730/R730xd Motherboard
1	350-BBEW	Chassis with up to 12, 3.5" Hard Drives and 2, 2.5" Flex Bay Hard Drives
1	310-AKPM	PowerEdge R730xd Shipping
1	338-BFFF	Intel Xeon E5-2650 v3 2.3GHz,25M Cache,9.60GT/s QPI,Turbo,HT,10C/20T (105W) Max Mem 2133MHz
1	371-BBGM	Upgrade to Two Intel Xeon E5-2650 v3 2.3GHz,25M Cache,9.60GT/s QPI,Turbo,HT,10C/20T (105W)
1	330-BBCR	R730/xd PCIe Riser 1, Right
1	330-BBCO	R730/xd PCIe Riser 2, Center
1	370-ABUF	2133MT/s RDIMMs
1	370-AAIP	Performance Optimized
8	370-ABUG	16GB RDIMM, 2133 MT/s, Dual Rank, x1 Data Width
1	780-BBLH	No RAID for H330/H730/H730P (1-21 HDDs or SSDs)
1	105-AAEG	PERC H730 Integrated RAID Controller, 1GB Cache
12	100-AEGJ	1TB 7.2K RPM SATA 6Gbps 3.5in Hot-plug Hard Drive,13G
2	100-AJPR	300GB 10K RPM SAS 12Gbps 2.5in Flex Bay Hard Drive
1	510-BBBB	Intel X520 DP 10Gb DA/SFP+, + I350 DP 1Gb Ethernet, Network Daughter Card
1	350-BBEJ	Bezel
1	770-BBBQ	ReadyRails Sliding Rails Without Cable Management Arm
1	381-BBBL	Performance BIOS Settings
1	150-ADWS	Dual, Hot-plug, Redundant Power Supply (1+1), 750W
2	192-BBDH	C13 to C11, PDU Style, 12 AMP, 2 Feet (.6m) Power Cord, North America
1	631-AAJG	Electronic System Documentation and OpenManage DVD Kit, PowerEdge R730/xd
1	619-ABVR	No Operating System
1	121-5736	No Media Required
1	800-BBDM	UEFI BIOS

Quantity	SKU	Component
1	332-1286	US Order
2	371-BBHM	Standard Heatsink for PowerEdge R730/R730xd
1	370-ABWE	DIMM Blanks for System with 2 Processors
1	385-BBHO	iDRAC8 Enterprise, integrated Dell Remote Access Controller, Enterprise
1	976-9030	ProSupport Plus: 7x21 HW/SW Tech Support and Assistance, 3 Year
1	976-9007	Dell Hardware Limited Warranty Plus On Site Service
1	976-9029	ProSupport Plus: 7x21 Next Business Day Onsite Service, 3 Year
1	951-2015	Thank you for choosing Dell ProSupport Plus. For tech support, visit http://www.dell.com/contactdell
2	900-9997	On-Site Installation Declined

Appendix D: Bill of Materials – PowerEdge R730xd 2.5" Infrastructure Node

Table 26: Active Name Node, Standby Name Node, Admin Node, Edge Node and HA Nodes – PowerEdge R730xd

Quantity	SKU	Component
1	379-BBWM	INFO QS, 13G HADOOP BUNDLE
1	210-ADBC	PowerEdge R730xd Server
1	591-BBCH	PowerEdge R730/R730xd Motherboard
1	350-BBFE	Chassis with up to 24, 2.5 Hard Drives and 2, 2.5" Flex Bay Hard Drives
1	340-AKPM	PowerEdge R730xd Shipping
1	338-BFFL	Intel Xeon E5-2690 v3 2.6GHz,30M Cache,9.60GT/s QPI,Turbo,HT,12C/24T (135W) Max Mem 2133MHz
1	374-BBGS	Upgrade to Two Intel Xeon E5-2690 v3 2.6GHz,30M Cache,9.60GT/s QPI,Turbo,HT,12C/24T (135W)
1	330-BBCR	R730/xd PCIe Riser 1, Right
1	330-BBCO	R730/xd PCIe Riser 2, Center
1	370-ABUF	2133MT/s RDIMMs
1	370-AAIP	Performance Optimized
8	370-ABUG	16GB RDIMM, 2133 MT/s, Dual Rank, x4 Data Width
1	780-BBLH	No RAID for H330/H730/H730P (1-24 HDDs or SSDs)
1	405-AAEG	PERC H730 Integrated RAID Controller, 1GB Cache
2	400-AJRD	300GB 15K RPM SAS 12Gbps 2.5in Flex Bay Hard Drive
8	400-AJON	1.2TB 10K RPM SAS 12Gbps 2.5in Hot-plug Hard Drive
1	540-BBBB	Intel X520 DP 10Gb DA/SFP+, + I350 DP 1Gb Ethernet, Network Daughter Card
1	540-BBCT	Intel X520 DP 10Gb DA/SFP+ Server Adapter
1	350-BBEJ	Bezel
1	770-BBBQ	ReadyRails Sliding Rails Without Cable Management Arm
1	384-BBBL	Performance BIOS Settings
1	450-ADWS	Dual, Hot-plug, Redundant Power Supply (1+1), 750W
2	492-BBDH	C13 to C14, PDU Style, 12 AMP, 2 Feet (.6m) Power Cord, North America
1	631-AAJG	Electronic System Documentation and OpenManage DVD Kit, PowerEdge R730/xd
1	619-ABVR	No Operating System

Quantity	SKU	Component
1	421-5736	No Media Required
1	800-BBDM	UEFI BIOS
1	332-1286	US Order
2	374-BBHM	Standard Heatsink for PowerEdge R730/R730xd
1	370-ABWE	DIMM Blanks for System with 2 Processors
1	385-BBHO	iDRAC8 Enterprise, integrated Dell Remote Access Controller, Enterprise
1	951-2015	Thank you for choosing Dell ProSupport Plus. For tech support, visit http://www.dell.com/contactdell
1	976-9030	ProSupport Plus: 7x24 HW/SW Tech Support and Assistance, 3 Year
1	976-9007	Dell Hardware Limited Warranty Plus On Site Service
1	976-9029	ProSupport Plus: 7x24 Next Business Day Onsite Service, 3 Year
2	900-9997	On-Site Installation Declined

Appendix E: Bill of Materials – PowerEdge R730xd 2.5" Data Node

Table 27: Data Node – PowerEdge R730xd 2.5"

Quantity	SKU	Component
1	379-BBWM	INFO QS, 13G HADOOP BUNDLE
1	210-ADBC	PowerEdge R730xd Server
1	591-BBCH	PowerEdge R730/R730xd Motherboard
1	350-BBFE	Chassis with up to 24, 2.5 Hard Drives and 2, 2.5" Flex Bay Hard Drives
1	340-AKPM	PowerEdge R730xd Shipping
1	338-BFFL	Intel Xeon E5-2690 v3 2.6GHz,30M Cache,9.60GT/s QPI,Turbo,HT,12C/24T (135W) Max Mem 2133MHz
1	374-BBGS	Upgrade to Two Intel Xeon E5-2690 v3 2.6GHz,30M Cache,9.60GT/s QPI,Turbo,HT,12C/24T (135W)
1	330-BBCR	R730/xd PCIe Riser 1, Right
1	330-BBCO	R730/xd PCIe Riser 2, Center
1	370-ABUF	2133MT/s RDIMMs
1	370-AAIP	Performance Optimized
8	370-ABUG	16GB RDIMM, 2133 MT/s, Dual Rank, x4 Data Width
1	780-BBLH	No RAID for H330/H730/H730P (1-24 HDDs or SSDs)
1	405-AAEG	PERC H730 Integrated RAID Controller, 1GB Cache
2	400-AJRD	300GB 15K RPM SAS 12Gbps 2.5in Flex Bay Hard Drive
24	400-AJON	1.2TB 10K RPM SAS 12Gbps 2.5in Hot-plug Hard Drive
1	540-BBBB	Intel X520 DP 10Gb DA/SFP+, + I350 DP 1Gb Ethernet, Network Daughter Card
1	350-BBEJ	Bezel
1	770-BBBQ	ReadyRails Sliding Rails Without Cable Management Arm
1	384-BBBL	Performance BIOS Settings
1	450-ADWS	Dual, Hot-plug, Redundant Power Supply (1+1), 750W
2	492-BBDH	C13 to C14, PDU Style, 12 AMP, 2 Feet (.6m) Power Cord, North America
1	631-AAJG	Electronic System Documentation and OpenManage DVD Kit, PowerEdge R730/xd
1	619-ABVR	No Operating System
1	421-5736	No Media Required
1	800-BBDM	UEFI BIOS

Quantity	SKU	Component
1	332-1286	US Order
2	374-BBHM	Standard Heatsink for PowerEdge R730/R730xd
1	370-ABWE	DIMM Blanks for System with 2 Processors
1	385-BBHO	iDRAC8 Enterprise, integrated Dell Remote Access Controller, Enterprise
1	951-2015	Thank you for choosing Dell ProSupport Plus. For tech support, visit http://www.dell.com/contactdell
1	976-9030	ProSupport Plus: 7x24 HW/SW Tech Support and Assistance,3 Year
1	976-9007	Dell Hardware Limited Warranty Plus On Site Service
1	976-9029	ProSupport Plus: 7x24 Next Business Day Onsite Service, 3 Year
2	900-9997	On-Site Installation Declined

Update History

Changes in Version 5.5

The following changes have been made since the 5.4 release:

- Update for Cloudera Enterprise 5.5.
- Update to use Dell Networking S4048-ON and S6000 network switches.
- Network architecture has been updated to clarify the rack, pod, and cluster designations and related switch usage. Node level network connections now use LACP bonding instead of Linux® mode 6 ALB.
- Server configurations have been updated to use the PowerEdge R730xd for both infrastructure and data nodes.
- Infrastructure node drive configuration has been changed to increase storage capacity and optimize performance.

References

Additional information can be obtained at <http://www.dell.com/hadoop>.

If you need additional services or implementation help, please contact your Dell sales representative.

To Learn More

For more information on the Dell | Cloudera Apache Hadoop Solution, visit <http://www.dell.com/hadoop>.

© 2011-2015 Dell Inc. All rights reserved. Trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Specifications are correct at date of publication but are subject to availability or change without notice at any time. Dell and its affiliates cannot be responsible for errors or omissions in typography or photography. Dell's Terms and Conditions of Sales and Service apply and are available on request. Dell service offerings do not affect consumer's statutory rights.

Dell, the DELL logo, the DELL badge, and PowerEdge are trademarks of Dell Inc.