Dell | Cloudera Solution Reference Architecture v5.0

A Dell Reference Architecture Guide

May 30, 2014

Summary

This document presents the reference architecture of the Dell™ | Cloudera™ Solution for Apache Hadoop, which Dell designed jointly with Cloudera.

The reference architecture introduces all the high-level components, hardware, and software that are included in the stack. Each high-level component is then described individually.

Table of Contents

Tables

Figures

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, and Dell PowerEdge and Force 10 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 following use cases:

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, transfer 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

Dell | Cloudera Hadoop Solution Components

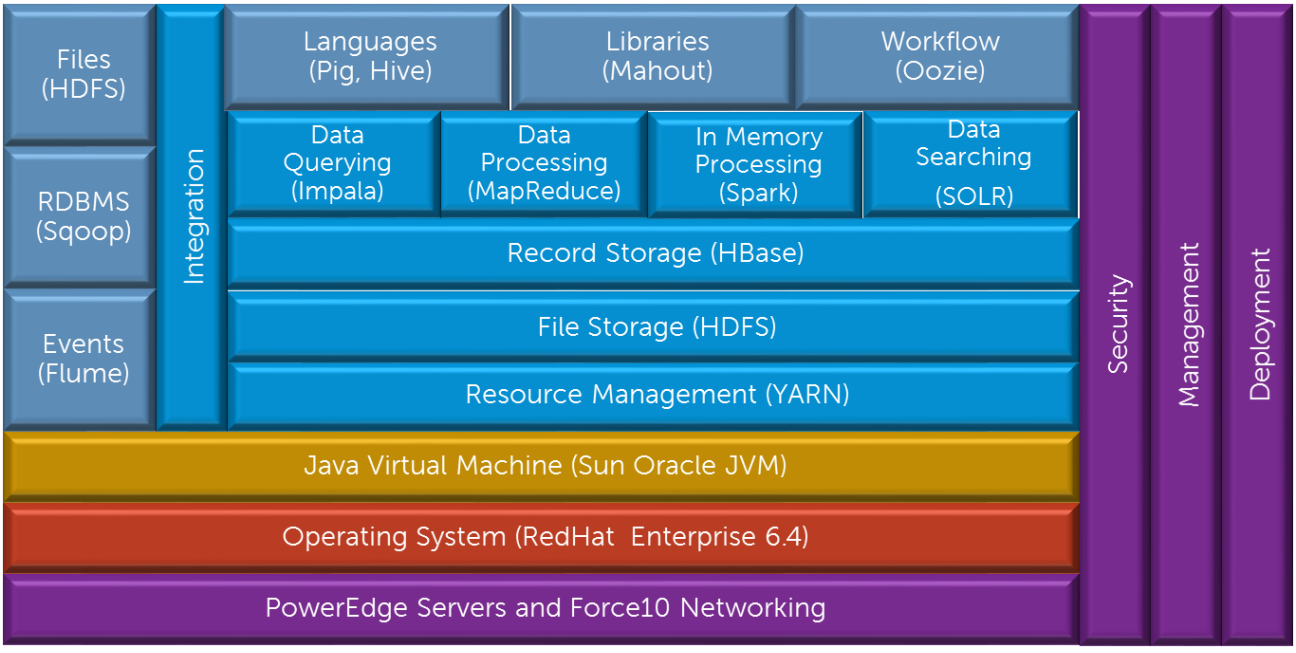

Figure 1: Dell | Cloudera Solution Components

Figure 1 illustrates the primary components in the Dell | Cloudera Solution.

The PowerEdge servers, Force10 networking, the operating system and the Java Virtual Machine make up the foundation on which the Hadoop software stack runs.

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 on top of these storage layers are four complementary access layers providing data processing, in-memory processing, data query and data search. MapReduce is the core processing framework in the Hadoop system, and provides a massively parallel data processing framework inspired by Google's MapReduce papers. Another processing framework is the real-time, in-memory processing framework called Spark. The Data Query layer provides real-time query access to data using Cloudera Impala. The Data Search layer provides real-time search of indexed data using Apache SOLR Cloud technology. All four of these layers can be used simultaneously or independently, depending on the workload and problems being solved.

Above these layers are a number of Hadoop end-user tools, providing a higher level of abstraction for data access and processing. Apache Pig and Apache Hive are data access and processing languages, while Apache Mahout provides machine learning capabilities. Apache Oozie provides a general workflow capability for coordinating complex sequences of production jobs, and Apache HUE provides a web interface for analyzing data.

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.

Support Matrix

The supported components and operating environments for the Dell | Cloudera® Apache Hadoop Solution are shown in Table 2.

Table 2: Solution Support Matrix

Category	Component	Version	Available Support
Operating System	Red Hat Enterprise Linux	6.4	Red Hat Linux support
Operating System	CentOS	6.4	Dell Hardware support
Java Virtual Machine	Sun Oracle JVM	Java 7 (1.7.0_25 minimum)	N/A
Hadoop	Cloudera Distribution for Apache Hadoop (CDH)	5.0	Cloudera support
Hadoop	Cloudera Manager	5.0	Cloudera support
Hadoop	Cloudera Navigator	1.2	Cloudera support

Cloudera Enterprise Software Overview

Hadoop for the Enterprise

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. 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?

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.

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:

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 (MPP) SQL engine built for Hadoop.

Search – Cloudera Search

Cloudera Search, based on SOLR, lets your users query and browse data in Hadoop just they would search Google or your 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 remaining sections of the document cover the details in depth.

High-level Node Architecture

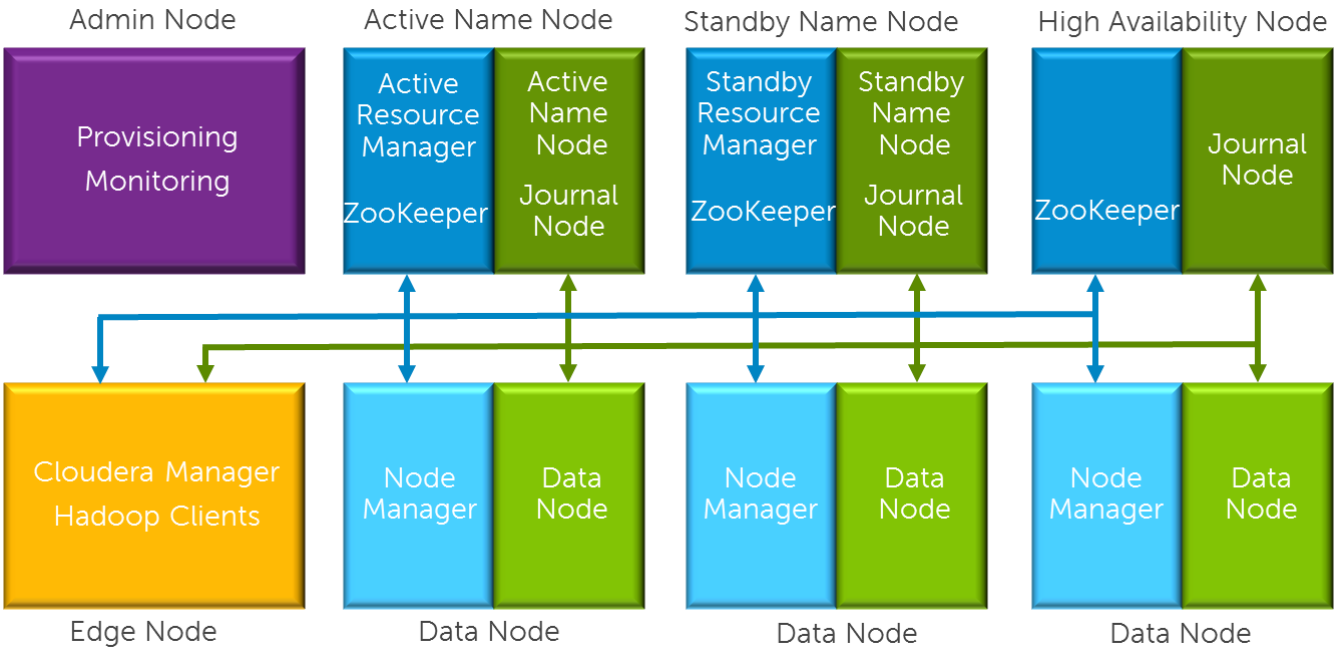

Figure 2: Cluster Architecture

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.

Figure 2 shows the roles for the nodes in a basic cluster.

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

Node Role		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

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.

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 master and secondary name nodes provide the first and second journal nodes. It also runs a third Zookeeper service.

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 three data nodes are required, and larger clusters are scaled primarily by adding additional data nodes There are two types of services running on the data nodes:

- NodeManager Daemon (to support YARN job execution)
- DataNode Daemon (to support HDFS data storage)

Table 3 Service Locations

Physical Node	Software Function
Administration Node	Operating System Provisioning Yum Repositories Monitoring Functions
Edge Node	Cloudera Manager
Active Name Node	NameNode Resource Manager Zookeeper Quorum Journal Node HMaster

Standby Name Node	Standby Namenode Standby Resource Manager Zookeeper Quorum Journal Node
HA Node	Zookeeper Quorum Journal Node
Data Node(x)	DataNode NodeManager RegionServer

Network Fabric Architecture

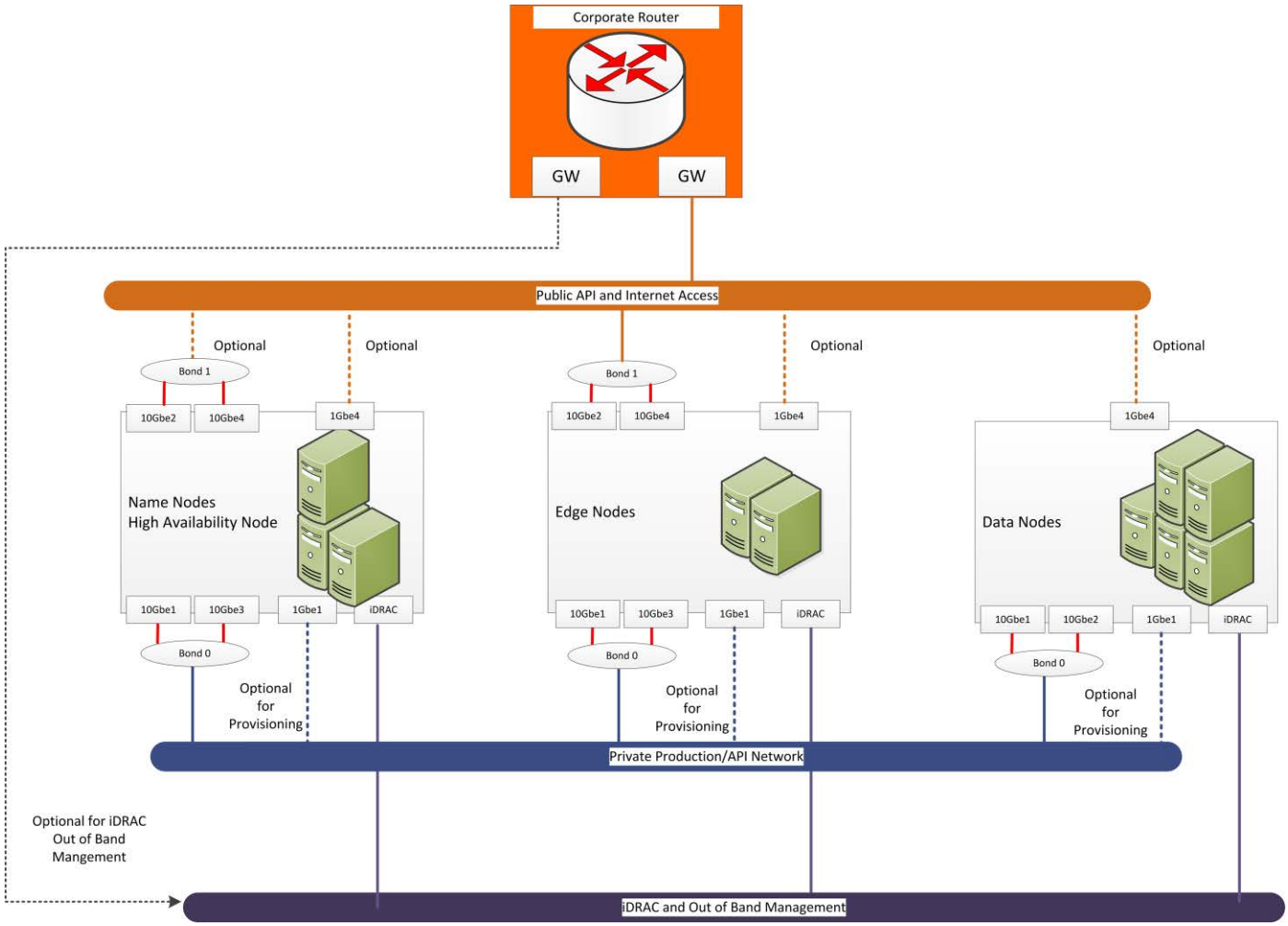

Figure 3: Cluster Network Fabric Architecture

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

Four distinct networks are used in the cluster:

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

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.

Management Network—the management network is used to provide cluster management and provisioning capabilities.

BMC / IPMI 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.

Edge Network—the Edge network provides connectivity from edge nodes to the existing core network via the top of rack or cluster switch.

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 and 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 all the necessary power, the network cabling and the two Ethernet switches necessary to support up to 20 data nodes. A rack should use its own power and space within the data center, separate from other racks, and should be treated as a fault zone.

Pod

A pod is an installation composed of three racks, based on server and network sizing. A pod is capable of supporting enough Hadoop server nodes and network switches for a minimum commercial scale installation.

Cluster

A cluster is a single Hadoop environment attached to a pair of distribution switches providing an aggregation layer for the entire cluster. A cluster can range in size from a single rack to a set of pods. A cluster shares the infrastructure nodes and management tools for operating the Hadoop environment. The size of the cluster can vary depending on the capacity of the aggregation network. For example, a Dell™ Force10™ Z9000 aggregation switch can run a larger cluster than the Dell™ Force10™ S4810 switches.

Sizing Constraints

The minimum configuration supported is six nodes:

- Master name node

- Secondary name node
- High availability (HA) node
- Three data nodes

The hardware configurations for the infrastructure nodes support clusters in the petabyte storage range. Beyond the infrastructure nodes, cluster size is primarily a function of the server platform and disk drives chosen, and the number of data nodes. Table 4 shows the approximate number of data nodes per rack, pod and cluster for the various server models. In practice the actual density per rack will be influenced by physical constraints like power and cooling as well as available network ports.

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.

Table 4: Cluster Sizes by Server Model

Server Model	Max Per Rack	Max Per Pod	Max Per Cluster
R720 Data Node	20	60	To be determined based on sizing criteria

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. 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.

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

Network Redundancy

The production network uses bonded connections to multiple switches in each rack. This 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 NameNode is responsible for all client operations in the cluster, while the Standby is simply acting as a slave, maintaining enough state to provide a fast failover if necessary.

In order for the Standby node to keep its state synchronized with the Active node in this implementation, both nodes communicate with a group of separate daemons called JournalNodes. When any namespace modification is performed by the Active node, it durably logs a record of the modification to a majority of these JournalNodes. The Standby node is capable of reading the edits from the JournalNodes, and is constantly watching them for changes to the edit log. As the Standby Node sees the edits, it applies them to its own namespace. In the event of a failover, the Standby 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 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.

This feature is not always implemented in production clusters.

Hardware Architecture

Server Infrastructure Options

The Dell | Cloudera Solution supports the Dell™ PowerEdge™ R720(xd) series for server infrastructure.

The following sections describe the supported server models and configurations required. Detailed part lists and rack layouts are included in the appendices.

PowerEdge R720 / R720xd Server

The PowerEdge R720 and R720xd servers are Dell's 12G PowerEdge mainstream dual socket 2U rack servers. They are designed to deliver the most competitive feature set, best performance and best value. In this generation, Dell offers a large storage footprint, best-in-class I/O capabilities and more advanced management features. The PowerEdge R720 and R720xd are technically similar except the R720xd has a backplane that can accommodate more drives (up to 24).

Figure 4: PowerEdge 720xd Server

PowerEdge R720xd feature summary:

- Intel® Romley platform and Intel® Xeon® E5-2600v2 processors
- 1600MHz DDR3
- Network daughter cards for customer choice of LOM speed, fabric and brand at point of sale
- PCIe SSD in a front-accessible, hot-plug format
- Internal GPGPU support
- Intel® Node Manager power management technology
- Software RAID
- Platinum efficiency power supplies, common across 600 and 700 series platforms

PowerEdge R720 / R720xd Hardware Configurations

Table 5: Hardware Configurations – PowerEdge R720 Infrastructure Nodes

Machine Function	Infrastructure Nodes
Platform	PowerEdge R720
CPU	2 x E5-2670v2 (10-core)
RAM (minimum)	128GB
LOM	4 x 1GbE
Add In Network	2 x Intel X520 DP 10Gb DA/SFP+ (for 10GbE networking)
DISK	8 x 1TB 7.2K SATA 3.5-in.
Storage Controller	PERC H710
RAID	RAID 10

Notes:

- *Be sure to consult your Dell account representative before changing the recommended disk sizes.*

Table 6: Hardware Configurations –PowerEdge R720xd Data Nodes

Machine Function	Data Nodes	Data Nodes
Platform	PowerEdge R720xd	PowerEdge R720xd
CPU	2 x E5-2670v2 (10-core)	2 x E5-2670v2 (10-core)
RAM (minimum)	64GB	64GB
LOM	4 x 1GbE	4 x 1GbE
DISK	12 x 4TB 7.2K RPM SATA 3Gbps 3.5in	24 x 1TB SATA 7.2K 2.5-in.
Add In Network	1 x Intel X520 DP 10Gb DA/SFP+ (for 10GbE networking)	1 x Intel X520 DP 10Gb DA/SFP+ (for 10GbE networking)
Storage Controller	LSI 9207	LSI 9207
RAID	JBOD	JBOD

Notes:

- *Be sure to consult your Dell account representative before changing the recommended disk sizes.*

PowerEdge R720xd Configuration Notes

Appendix A : Illustrates the recommended rack layout for R720 clusters.

Appendix B :, Appendix B :, and Appendix C : contain the full bill of materials (BOM) listing for the PowerEdge R720 and R720Xd server configurations.

The R720 and R720xd configurations can be used with 10GbE networking. To use 10GbE networking support, additional network cards are required. Infrastructure nodes require two dual-port cards, while data nodes require one dual-port card. The BOM listings include the required cards for 10GbE networking.

Data nodes can be configured with either the LSI 9207 or the PERC H710 disk controller. The LSI 9207 is recommended for new deployments. The PERC H170 is supported as an alternative, primarily for compatibility with existing clusters. Refer to the "JBOD versus single disk RAID 0 Configuration" section for more information.

Storage Sizing Notes

For drive capacities greater than 3TB or node storage density over 36TB, 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 increased to 128 kb or more. 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.

You 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 supports two options for networking: 1GbE and 10GbE. The 1GbE option uses Dell™ Force10™ S60 switches as the top-of-rack connectivity to all Hadoop-related nodes, while the 10GbE option uses Dell™ Force10™ S4810 switches. Hadoop applications are increasingly being deployed on 10GbE servers for the scale and price advantages they bring, and this is the recommended configuration for new clusters.

Four distinct networks are used in the cluster:

Logical Network	Connection	Switch
Cluster Data Network	Bonded 10GbE	Dual top of rack switches
Management Network	1GbE,	Dedicated switch per rack
BMC Network	1GbE	Dedicated switch per rack
Edge Network	Bonded	Top of rack or cluster switch

Each network uses a separate VLAN, and dedicated components when possible. Figure 5 shows the logical organization of the network.

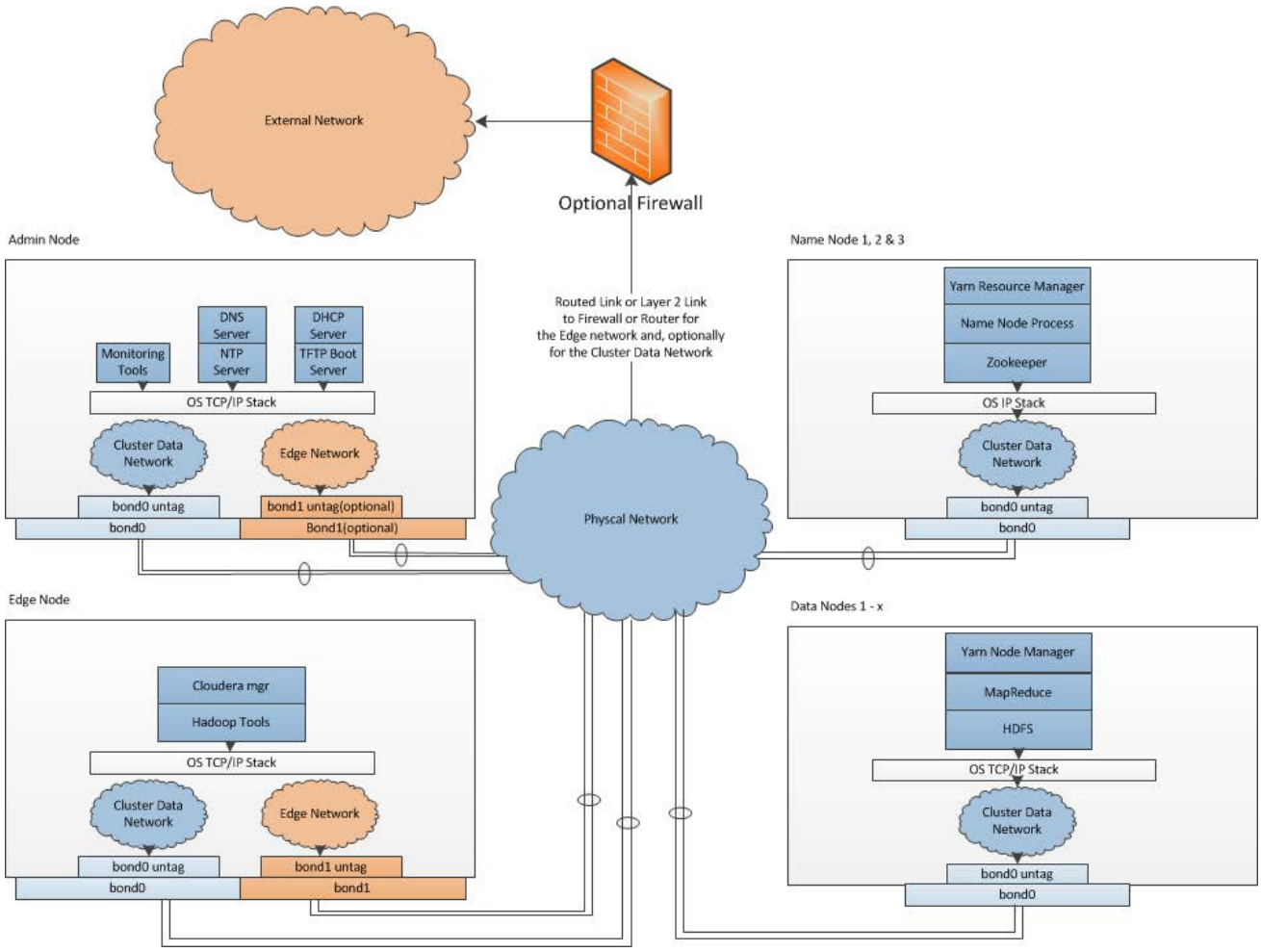

Figure 5: Hadoop Logical Network Diagram

Physical Network Components

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. (Under Linux, this is balanced-alb or mode 6 bonding) The connections are made to a pair of ToR 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 ToR switch. This connection facilitates high performance ingest and cluster access between applications running on those nodes, and the core datacenter network.

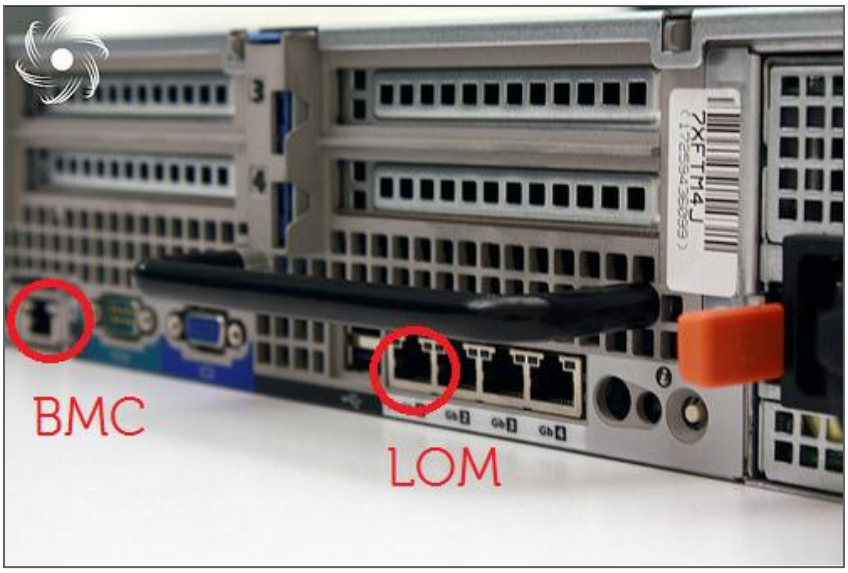

Figure 6: PowerEdge R720xd Node 1GbE Network Interconnects

Top of Rack (ToR) Switches

Each rack uses a pair of Force10 S4810's as top of rack 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 rack if a switch fails or needs maintenance, while providing active-active bandwidth utilization.

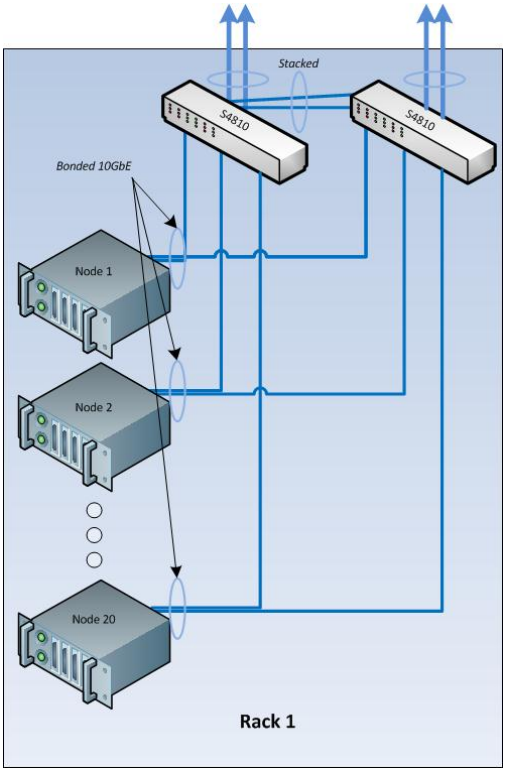

Figure 7: Single Rack Networking Equipment

Figure 7 shows the single rack network configuration, with a pair of Force10 S4810 switches aggregating the rack traffic.

For a single rack, the top of rack switches can act as the cluster aggregation layer. For larger clusters, a cluster aggregation layer is required.

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

Cluster Aggregation Switches

For clusters consisting of one more pods, the architecture uses either the Force10 S4810, or the Force10 Z9000 for aggregation switches. The choice depends on the initial size and planned scaling. The Force10 S4810-based aggregation design is preferred for lower cost and medium scalability. This design can handle up to six racks or two pods. The Z9000 is recommended for larger deployments.

Like the ToR switches, the aggregation switches are also connected in pairs using VLT. The uplink from each S4810 ToR switch to the aggregation pair is 80Gb, using a pair of 40G interfaces Since both S4810's connect to the aggregation pair, there is a collective bandwidth of 160G available from each rack.

S4810 Cluster Aggregation

Figure 8 illustrates the configuration for a multiple rack cluster using the S4810 as a cluster aggregation switch.

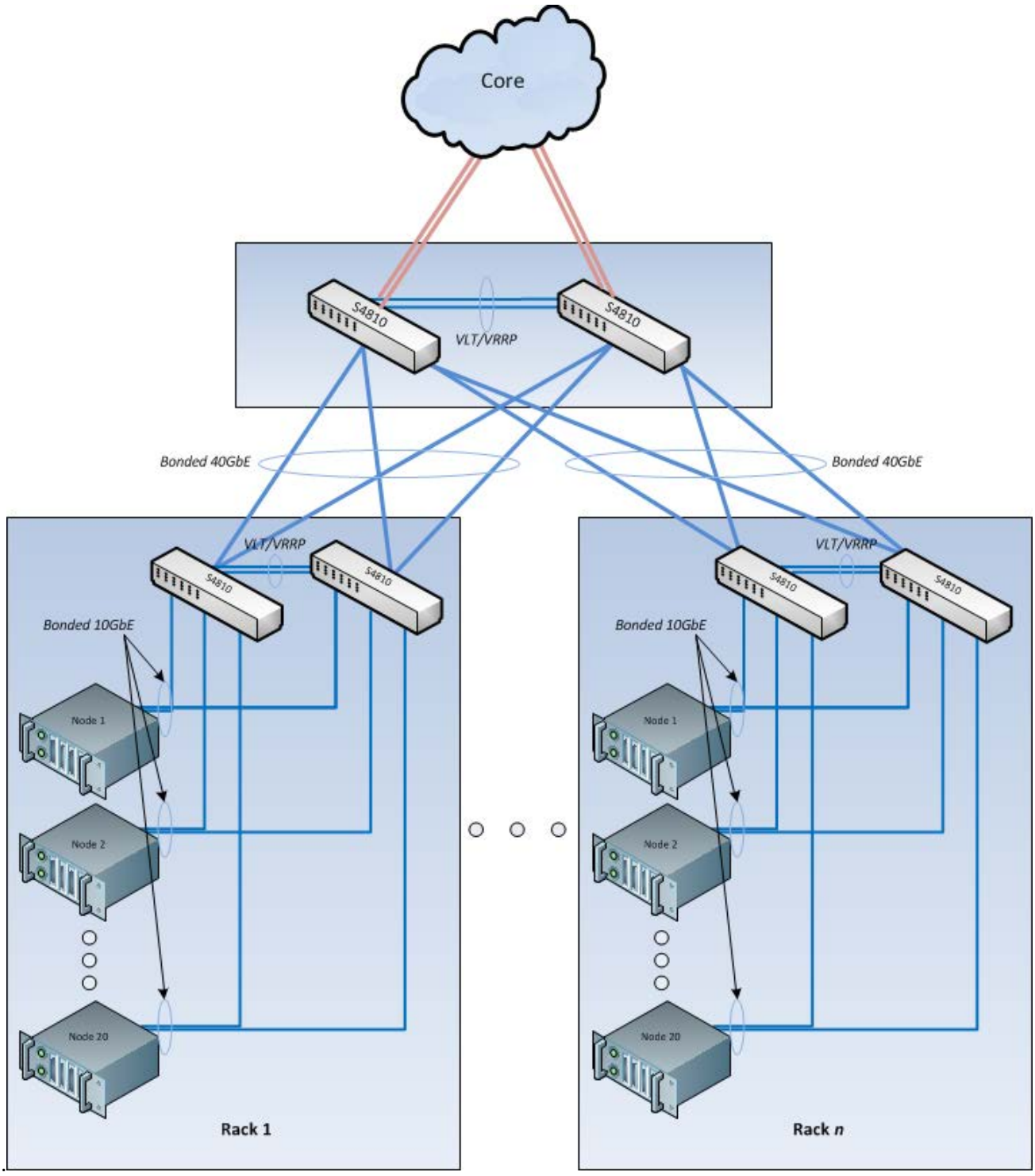

Figure 8: S4810 Multi-rack Networking Equipment

Force10 Z9000 Cluster Aggregation

For larger initial deployments, deployments where 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 Force10 Z9000 core switch. The Force10 Z9000 is a 32-port, 40G high-capacity switch. It can aggregate up to 15 racks of high-density PowerEdge C8000 servers. The rack-to-rack bandwidth needed in Hadoop is best addressed by a 40G-capable, non-blocking switch and the Force10 Z9000 can provide a cumulative bandwidth of 1.5TB of

throughput at line-rate traffic from every port. In many cases, The Force10 Z9000 does not need to connect into any other higher-tier core switches because the capacity is enough for a data center with hundreds of servers.

Figure 9 illustrates the configuration for a multiple rack cluster using the Z9000 as a cluster aggregation switch. This is an example of a Clos fabric that grows horizontally. This technique of network fabric deployment has been used in the data centers of some of the largest web companies, whose businesses range from social media to public cloud. Some of the largest recent Hadoop deployments also use this new approach to networking.

Each switch in Figure 9 forms a layer-2 LAG, This assumes that the Force10 Z9000 pair in the aggregation forms a VLT pair for HA. Now we have two tiers of VLT, one forming at the ToR for servers and another at the aggregation for the ToR switches.

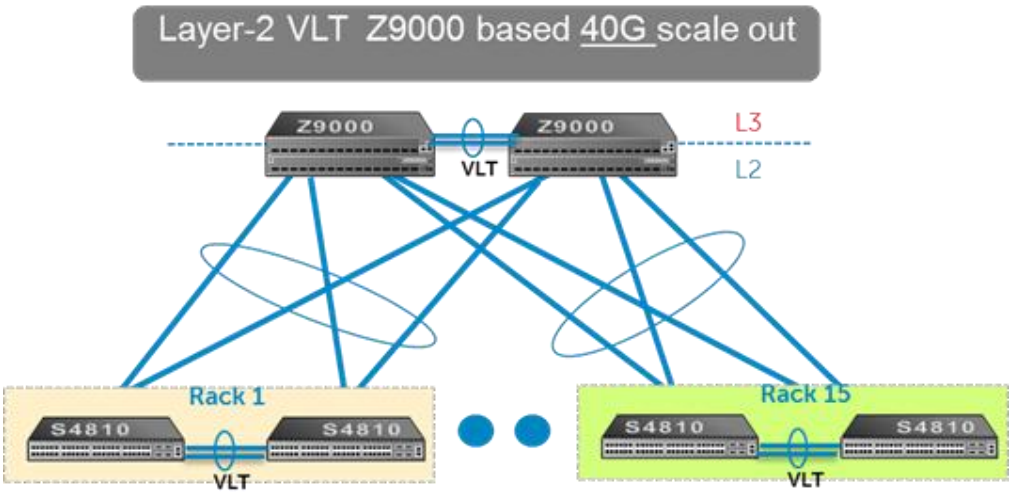

Figure 9: Multi-Rack View Using Force10 Z9000 Switches (Based on Layer-2)

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. Details of the connection are site specific, and need to be determined as part of the deployment planning.

Layer-2 and Layer-3

The layer-2 and layer-3 boundaries are separated at either the ToR or the aggregation layer. Either of the options is equally viable. The colors blue and red in Figure 9 represent the layer-2 and layer-3 boundaries. This document uses layer-2 as the reference up to the aggregation layer.

Management Network

The management network of all the servers and switches is aggregated into a Dell™ Force10™ S55 switch, which is located in each rack of the POD. It uplinks on a 10G link to the aggregation switches or the core directly, wherever the split for out-of-band is required.

Network Equipment Summary

Table 7 and Table 8 summarize the required cluster networking equipment. Table 9 summarizes the number of cables needed for a cluster.

Table 7: Per Rack Network Equipment	
Total Racks	1 (6-20 Nodes)
Top-of-rack switches	1 x Force10 S55

	2 x Force10 S4810
Aggregation switch	Not needed for a single rack
Switch Interconnect cables	2 x 40Gb QSFP+ Cables
Modules in each ToR	1x 12-2port Stacking, 1x 10G -2 port uplink

Table 8: Aggregation Network Switches for 3 or more racks

Total racks	3 to 15 Racks (1 – 5 pods)
Aggregation Layer Switches	2 x Force10 Z9000
Pod-interconnect cabling	4 x 40Gb QSFP+ Cables per Rack
Switch Interconnect Cables	4 x 40GB QSFP+ cables 1 M

Table 9: 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

Network Connectivity Summary

The network interconnects between various hardware components of the solution are depicted in Figure 10. For more information, please see the *Dell | Cloudera Apache Hadoop Solution Deployment Guide*.

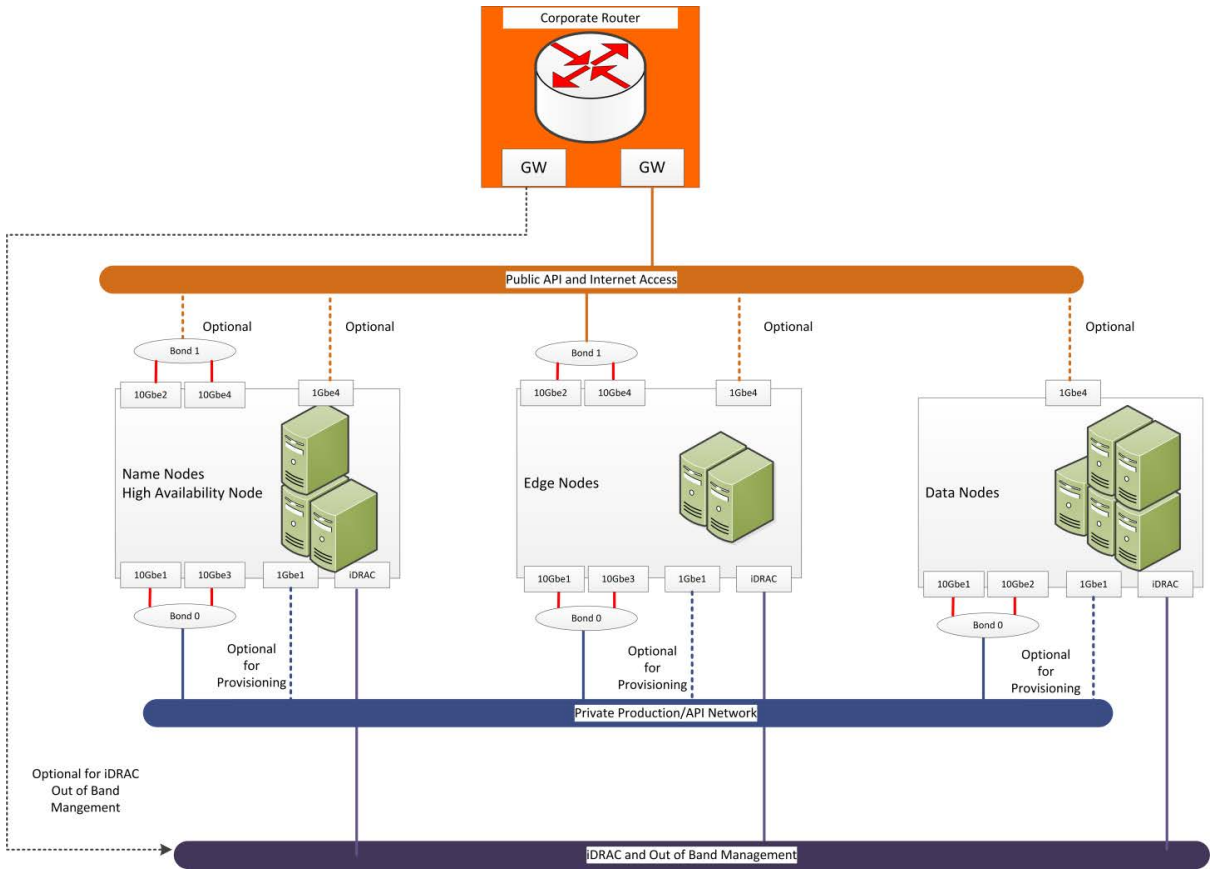

Figure 10: Network Connections for 10GbE

IPv6 Capabilities

At this time, the architecture does not support or allow for the use of IPv6 for network connectivity.

Cloudera Enterprise Software

The Dell | Cloudera Solution is based on Cloudera Enterprise, which includes Cloudera's distribution for Hadoop (CDH) 5.0 and Cloudera Manager.

Cloudera Manager

Cloudera Manager is designed to make administration of CDH 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. The first release of Cloudera Navigator (v1.0) was developed specifically to address data security concerns most typically associated with highly regulated industries, such as financial services, healthcare and government. It includes a full suite of auditing capabilities across all CDH components that store data.

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, queryable 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:

Flexible Support Windows

- Choose 8×5 or 24×7 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

Dell | Cloudera Solution Deployment Methodology

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

Appendix A : Physical Configuration — PowerEdge R720xd

Table 10: Rack Configuration – PowerEdge R720xd (or R720/R720xd)

RU	RACK1	RACK2	RACK3
42	R1- Switch 2: Force10 S4810	R2- Switch2: Force10 S4810	R3- Switch2: Force10 S4810
41	R1- Switch 1: Force10 S4810	R2- Switch1: Force10 S4810	R3- Switch1: Force10 S4810
40	Cable Management	Cable Management	Cable Management
39	Cable Management	Cable Management	Cable Management
38	Master Name Node:R720xd or R720	Edge01: R720xd or R720	R3 - Switch 1: Force10 S4810
37			R3 - Switch 2: Force10 S4810
36	Cable Management	Cable Management	Cable Management
35	Cable Management	Cable Management	Cable Management
34	Admin Node R720xd or R720	Secondary Name Node R720xd or R720	HA Node: R720xd or R720
33			
32	R1 - S55 iDRAC Mgmt switch	R2 - S55 iDRAC Mgmt switch	R3 - S55 iDRAC Mgmt switch
21-31	Empty	Empty	Empty
20	R1- Chassis10: R720xd	R2- Chassis10: R720xd	R3- Chassis10: R720xd
19			
18	R1- Chassis09: R720xd	R2- Chassis09: R720xd	R3- Chassis09: R720xd
17			
16	R1- Chassis08: R720xd	R2- Chassis08: R720xd	R3- Chassis08: R720xd
15			
14	R1- Chassis07: R720xd	R2- Chassis07: R720xd	R3- Chassis07: R720xd
13			
12	R1- Chassis06: R720xd	R2- Chassis06: R720xd	R3- Chassis06: R720xd
11			
10	R1- Chassis05: R720xd	R2- Chassis05: R720xd	R3- Chassis05: R720xd
9			
8	R1- Chassis04: R720xd	R2- Chassis04: R720xd	R3- Chassis04: R720xd
7			
6	R1- Chassis03: R720xd	R2- Chassis03: R720xd	R3- Chassis03: R720xd
5			
4	R1- Chassis02: R720xd	R2- Chassis02: R720xd	R3- Chassis02: R720xd
3			
2	R1- Chassis01: R720xd	R2- Chassis01: R720xd	R3- Chassis01: R720xd
1			

Appendix B : Bill of Materials – PowerEdge R720 Nodes

Table 11: Active and Standby Name, Admin, Edge and HA Nodes – PowerEdge R720

SKU	Component
331-3765	UEFI BIOS Setting
591-BBBP	PowerEdge R720 Motherboard, TPM
210-ABVP	PowerEdge R720, Intel Xeon E-26XX v2 Processors
342-3587	3.5" Chassis with up to 8 Hard Drives
331-4437	PowerEdge R720 Shipping
338-BDBG	Intel Xeon E5-2670v2 2.5GHz,25M Cache, 8.0GT/s QPI, Turbo, HT, 10C, 115W, Max Mem 1866MHz
331-4508	Heat Sink for PowerEdge R720and R720xd Quantity 2
338-BDBV	Intel Xeon E5-2670v2 2.5GHz,25M Cache, 8.0GT/s QPI, Turbo, HT, 10C, 115W, Max Mem 1866MHz,2nd Proc
317-8688	DIMM Blanks for Systems with2 Processors
331-4424	1600 MHz RDIMMS
331-4428	Performance Optimized
319-1812	16GB RDIMM, 1600MT/s, Low Volt, Dual Rank, x4 Data Width Quantity 8
331-4403	Unconfigured RAID for H710P/H710/H310 (1-16 HDDs)
342-3529	PERC H710 Integrated RAID Controller, 512MB NV Cache
341-8730	1TB 7.2K RPM SATA 3Gbps 3.5in Hot-plug Hard Drive Quantity 8
421-5339	iDRAC7 Enterprise
430-4447	Intel Ethernet I350 QP 1Gb Network Daughter Card
331-4440	Risers with up to 6, x8 PCIeSlots + 1, x16 PCIe Slot
430-4445	Intel X520 DP 10Gb DA/SFP+ Server Adapter Quantity 2
331-4605	Dual, Hot-plug, Redundant Power Supply (1+1), 750W
330-3151	Power Cord, C13 to C14, PDU Style, 12 Amps, 2 meter, Qty 2
330-5116	Power Saving Dell Active Power Controller
331-4433	ReadyRails Sliding Rails With Cable Management Arm
318-1375	Bezel
313-9092	DVD ROM, SATA, INTERNAL
310-5171	No System Documentation, No OpenManage DVD Kit
420-6320	No Operating System
421-5736	No Media Required
939-2768	Dell Hardware Limited Warranty Plus On Site Service Initial Year
936-4603	ProSupport: 7x24 HW / SW Tech Support and Assistance, 3 Year
939-2678	Dell Hardware Limited Warranty Plus On Site Service Extended Year
936-4593	ProSupport: Next Business Day Onsite Service After ProblemDiagnosis, 2 Year Extended
988-9281	ProSupport: Next Business Day Onsite Service After ProblemDiagnosis, Initial Year
900-9997	On-Site Installation Declined
926-2979	Proactive Maintenance Service Declined
331-3282	CLOUD COMPUTE NODE,DCS, INFO MOD,HADOOP

Appendix C : Bill of Materials – PowerEdge R720xd 3.5" Data Node

Table 12: Data node – PowerEdge R720xd

SKU	Component
331-3765	UEFI BIOS Setting
210-ABMY	PowerEdge R720xd, Intel XeonE-26XX v2 Processors
591-BBBP	PowerEdge R720 Motherboard, TPM
342-3567	Chassis with up to 12, 3.5" Hard Drives
331-4437	PowerEdge R720 Shipping
338-BDBG	Intel Xeon E5-2670v2 2.5GHz,25M Cache, 8.0GT/s QPI, Turbo, HT, 10C, 115W, Max Mem 1866MHz
317-8688	DIMM Blanks for Systems with2 Processors
331-4508	Heat Sink for PowerEdge R720and R720xd Quantity 2
338-BDBV	Intel Xeon E5-2670v2 2.5GHz,25M Cache, 8.0GT/s QPI, Turbo, HT, 10C, 115W, Max Mem 1866MHz,2nd Proc
331-4424	1600 MHz RDIMMS
331-4428	Performance Optimized
319-1811	8GB RDIMM, 1600MT/s, Low Volt, Dual Rank, x4 Data Width Quantity 8
331-4533	No RAID for H310 (1-16 HDDs)
428-BBBX	LSI 9207, Internal Passthrough Host Bus Adapter Card for R720 and R720 XD with 3.5in HDDs
342-5272	4TB 7.2K RPM SATA 3Gbps 3.5in Hot-plug Hard Drive Quantity 12
421-5339	iDRAC7 Enterprise
430-4447	Intel Ethernet I350 QP 1Gb Network Daughter Card
430-4445	Intel X520 DP 10Gb DA/SFP+ Server Adapter
331-4605	Dual, Hot-plug, Redundant Power Supply (1+1), 750W
330-3151	Power Cord, C13 to C14, PDU Style, 12 Amps, 2 meter, Qty2
330-5116	Power Saving Dell Active Power Controller
331-4433	ReadyRails Sliding Rails With Cable Management Arm
318-1375	Bezel
331-5914	Electronic System Documentation and OpenManage DVD Kit forR720 and R720xd
420-6320	No Operating System
421-5736	No Media Required
939-3398	Dell Hardware Limited Warranty Plus On Site Service Extended Year
989-3439	Dell ProSupport. For tech support, visit http://support.dell.com/ProSupport or call 1-800-945-3355
936-7243	ProSupport: Next Business Day Onsite Service After ProblemDiagnosis, 2 Year Extended
936-7263	ProSupport: 7x24 HW / SW Tech Support and Assistance, 3 Year
936-0967	Dell Hardware Limited Warranty Plus On Site Service Initial Year
989-2701	ProSupport: Next Business Day Onsite Service After ProblemDiagnosis, Initial Year
900-9997	On-Site Installation Declined
926-2979	Proactive Maintenance Service Declined
331-3282	CLOUD COMPUTE NODE,DCS, INFO MOD,HADOOP

Appendix D : Bill of Materials – PowerEdge R720xd 2.5" Data Node

Table 13: Data node – PowerEdge R720xd

SKU	Component
331-3765	UEFI BIOS Setting
210-ABMY	PowerEdge R720xd, Intel XeonE-26XX v2 Processors
591-BBBP	PowerEdge R720 Motherboard, TPM
342-3566	Chassis with up to 24, 2.5" Hard Drives
331-4437	PowerEdge R720 Shipping
331-4508	Heat Sink for PowerEdge R720and R720xd Quantity 2
338-BDBG	Intel Xeon E5-2670v2 2.5GHz,25M Cache, 8.0GT/s QPI, Turbo, HT, 10C, 115W, Max Mem 1866MHz
317-8688	DIMM Blanks for Systems with2 Processors
338-BDBV	Intel Xeon E5-2670v2 2.5GHz,25M Cache, 8.0GT/s QPI, Turbo, HT, 10C, 115W, Max Mem 1866MHz,2nd Proc
331-4424	1600 MHz RDIMMS
331-4428	Performance Optimized
319-1811	8GB RDIMM, 1600MT/s, Low Volt, Dual Rank, x4 Data Width Quantity 8
331-4533	No RAID for H310 (1-16 HDDs)
342-5964	LSI 9207, Internal Passthrough Host Bus Adapter Card for R720 and R720 XD with 2.5in HDDs
342-1998	1TB 7.2K RPM SATA 3Gbps 2.5in Hot-plug Hard Drive Quantity 24
421-5339	iDRAC7 Enterprise
430-4447	Intel Ethernet I350 QP 1Gb Network Daughter Card
430-4445	Intel X520 DP 10Gb DA/SFP+ Server Adapter
331-4605	Dual, Hot-plug, Redundant Power Supply (1+1), 750W
330-3151	Power Cord, C13 to C14, PDU Style, 12 Amps, 2 meter, Qty2
330-5116	Power Saving Dell Active Power Controller
331-4433	ReadyRails Sliding Rails With Cable Management Arm
318-1375	Bezel
331-5914	Electronic System Documentation and OpenManage DVD Kit forR720 and R720xd
420-6320	No Operating System
421-5736	No Media Required
989-3439	Dell ProSupport. For tech support, visit http://support.dell.com/ProSupport or call 1-800-945-3355
939-3398	Dell Hardware Limited Warranty Plus On Site Service Extended Year
936-0967	Dell Hardware Limited Warranty Plus On Site Service Initial Year
936-7243	ProSupport: Next Business Day Onsite Service After ProblemDiagnosis, 2 Year Extended
936-7263	ProSupport: 7x24 HW / SW Tech Support and Assistance, 3 Year
989-2701	ProSupport: Next Business Day Onsite Service After ProblemDiagnosis, Initial Year
900-9997	On-Site Installation Declined

926-2979	Proactive Maintenance Service Declined
331-3282	CLOUD COMPUTE NODE,DCS, INFO MOD,HADOOP

Appendix E : Part Numbers – Force10 Network Equipment

Table 14: Network Equipment – 1GbE – Dell Force10

SKU	Description
225-2446	Force10, Z9000, 2U, 32 x 40Gbe QSFP+ Ports, 1 AC Pwr Supply, Fan w/IO Panel to PSU (Normal) Airflow (Non-Redundant Pwr)
331-5996	Force10, Power Cord, 125V, 15A, 10 Feet, NEMA 5-15/C13, S-Series
331-5343	Force10, Z9000, AC Power Supply for Chassis with IO Panel to PSU (Normal) Airflow
430-4543	Force10, Transceiver, 40GE QSFP+ Short Reach Optics, 850nmWavelength, 100-150m Reach onOM3/OM4
331-7279	Force10, Z9000 Cable Management Kit
225-2477	Force10, S4810P, 48 x 10GbE SFP+, 4 x QSFP 40GbE, 1 x AC PSU, 2 x Fans, IO Panel to PSU Airflow
225-2479	Force10, S4810P, 48 x 10GbE SF P+, 4 x QSFP 40GbE, 1 x AC PSU , 2 x Fans, PSU to IO Panel Airflow
331-5103	Force10, S4810, AC Power Supply, IO Panel to PSU Airflow
331-5105	Force10, S4810, AC Power Supply, PSU to IO Panel Airflow
331-5258	Force10, Cable, SFP+ to SFP+, 10GbE, Copper Twinax Direct Attach Cable, 2 Meters
331-5996	Force10, Power Cord, 125V, 15A, 10 Feet, NEMA 5-15/C13, S-Series
421-6981	Force10, Software, L3 Latest Version, S4810
430-4543	Force10, Transceiver, 40GE QSFP+ Short Reach Optics, 850nmWavelength, 100-150m Reach onOM3/OM4
331-5274	Force10, Transceiver, SFP+, 10GbE, SR, 850nm Wavelength, 300m Reach
430-4543	Force10, Transceiver, 40GE QSFP+ Short Reach Optics, 850nmWavelength, 100-150m Reach onOM3/OM4
331-5393	Force10, Rear Rack Mounting Bracket, 4 Post, S4810
225-2450	Force10, S60, 44 x 10/100/1000 BASE-T, 4 x SFP, 2 Expansion Slots, 1 x AC PSU, 2 x fans, P SU to IO Panel Airflow
331-5233	Force10, SFP+ Expansion Module , 2 x 10 GbE Ports, S60 Series (SFP+ optics required)
331-5996	Force10, Power Cord, 125V, 15A , 10 Feet, NEMA 5-15/C13, S-Series
331-5226	Force10, S60, AC Power Supply, PSU to IO Panel Airflow
331-5398	Force10, Rear Rack Mounting Bracket, Metal, 4 Post, S60
	Force10 S60 2 port, 12G, Stacking module
	Force10 S60 12 Gig 60cms stacking cable

Table 15: Network Equipment – 10GbE – Dell Force10	
SKU	Description
	Cluster Network
331-5274	Dell Networking, Transceiver, SFP+, 10GbE, SR, 850nm Wavelength, 300m Reach
330-8723	SFP+, Short Range, Optical Transceiver, LC Connector, 10Gb and 1Gb compatible(Intel 10G SFP+)
225-2477	Force10, S4810P, 48 x 10GbE SFP+, 4 x QSFP 40GbE, 1 x AC PSU, 2 x Fans, IO Panel to PSU Airflow
331-5996	Force10, Power Cord, 125V, 15A, 10 Feet, NEMA 5-15/C13, S-Series
331-5272	Dell Networking, Transceiver, SFP, 1000BASE-LX, 1310nm Wavelength, 10km Reach
331-5393	Force10, Rear Rack Mounting Bracket, 4 Post, S4810
331-6279	Force10, User Documentation for S4810, DAO/BCC
935-0103	SW Support,Force10 Software ,3 Years
935-0143	ProSupport: 7x24 HW / SW Tech Support and Assistance, 3 Years
931-3856	ProSupport: 4-Hour 7x24 Parts Only After Problem Diagnosis, Initial Year
989-3439	Dell ProSupport. For tech support, visit http://support.dell.com/ProSupport or call 1-800-945-3355
996-2760	Dell Hardware Limited Warranty Extended Year(s)
935-0123	ProSupport: 4-Hour 7x24 Parts Only After Problem Diagnosis, 2 Year Extended
996-2670	Dell Hardware Limited Warranty Initial Year
900-9997	On-Site Installation Declined
996-3080	ProSupport for, Force10,Layer 3 Enablement, 1 Year
331-9460	Force10, Software, iSCSI-Optimized Configuration, S4810
331-5217	Customer Kit, Dell Networking, Cable, QSFP+, 40GbE SFP+ Passive Copper Direct Attach Cable, 1 Meter
331-3282	CLOUD COMPUTE NODE,DCS, INFO MOD,HADOOP
	Administration Network
225-2503	Force10, S55, 44 x 10/100/1000 BASE-T, 4 x SFP, 2 Expansion Slots, 1 x AC PSU, 2 x Fans, IO Panel to PSU Airfl (225-2503)
331-5233	Forcd10 SFP+ Expansion Module 2x10 Gbe Ports
331-5243	Force10, S55, AC Power Supply, IO Panel to PSU Airflow (331-5243)
331-5996	Force10, Power Cord, 125V, 15A, 10 Feet, NEMA 5-15/C13, S-Series (331-5996)
331-5252	Force10, Rear Rack Mounting Bracket, 4 Post, S55 (331-5252)
331-9233	No Returns Allowed on Dell Force10 Switches (331-9233)
331-6271	Force10, User Documentation for S55/S60, DAO/BCC (331-6271)
935-1367	Dell Hardware Limited Warranty Initial Year (935-1367)
938-7578	Dell Hardware Limited Warranty Extended Year(s) (938-7578)

989-3439	Dell ProSupport. For tech support, visit http://support.dell.com/ProSupport or call 1-800-945-3355 (989-3439)
995-0592	ProSupport: Next Business Day Parts Delivery, 2 Year Extended (995-0592)
995-0622	ProSupport: 7x24 HW / SW Tech Support and Assistance, 3 Years (995-0622)
995-9649	SW Support,Force10 Software ,5 Years (995-9649)
996-0530	ProSupport: Next Business Day Parts Delivery, Initial Year (996-0530)
996-0540	Force10, 5 Year Return To Depot Service, Base Warranty (996-0540)
990-9997	On-Site Installation Declined (900-9997)
331-3282	CLOUD COMPUTE NODE,DCS, INFO MOD,HADOOP

Networking Equipment notes

Theses SKU's are provided for reference. The actual quantities of switches and connections required will depend on the cluster size, and the final rack layout.

The above list of SKUs includes switches that have specific air flow options. There are both I/O to PSU SKU numbers and PSU to I/O side options available for reverse air flow. Redundant FANs (other than the minimum supplied with chassis) should also be same direction as the base switch. The airflow cannot be reversed in the field at this time.

The above list shows the AC power supplies only. All switch models are available in DC as well.

The above list includes the necessary cables for the connections between the switches for uplinks and interconnects.

The BOMs do not include the cables required for connecting the individual servers into the cluster, since the exact cables required depend on the final chosen rack layout, and choice of cable is often based on customer preference. Refer to Table 9 for the required cable quantities.

Server Racks and Power

The above list of SKUs for the servers includes many items. However, they do not include racks or power distribution units, as they are generally site specific. The PowerEdge C8000 server line requires 240V power and other servers are dual voltage (110/240). The physical dimensions and power requirements need to be reviewed, as the PowerEdge C8000 requires extra space for front-side cable management and rear power distribution, in addition to extra depth. The PowerEdge R720 requires rear cable management and power distribution.

Appendix F : Bill of Materials – Software and Support

Software, training and support SKUs change regularly, and are related to specific global regions. Please refer to the "Hadoop Solution SKUs" document on Dell SalesEdge (Dell internal link) or contact your Dell account representative for the latest information.

The Sample Bill of Materials appendices include service and support SKUs for the United States. These SKUs need to be changed for other regions.

Appendix G : JBOD versus single disk RAID 0 Configuration

The Hadoop community's strong advocacy for the "non-RAIDed" drives configuration known as "Just a Bunch of Disks," or JBOD, has caused some confusion for readers of our reference architecture. We fully endorse this approach but feel a need for clarification because there are multiple valid ways to achieve this configuration.

Normally, the optimum disk configuration for Hadoop data nodes is considered to be JBOD mode rather than RAID. This is because HDFS provides its own data replication, eliminating the need for the redundancy provided by RAID levels 1-6. HDFS also implements efficient round robin parallel I/O across multiple drives, eliminating the need for the parallelism provided by the striping capabilities of RAID 0.

The LSI 9207 controller is a SAS + SATA controller, and provides JBOD capabilities as a standard hard disk bus adapter (HBA.)

Some drive controllers, such as the PERC H710, support only RAID mode, and so can't be used in a plain host bus adapter (HBA) mode for JBOD. For these situations, configuring the controllers to use the disks as multiple RAID 0 "arrays" allows HDFS to own them as a single drive. In this configuration, the controller is effectively operating just like a standard HBA in JBOD mode, and the RAID 0 and JBOD performance characteristics are comparable. While having a RAID controller adds a minor latency, the latency is offset by adaptive read-ahead caching on the controller.

Appendix H : Abbreviations

Abbreviation	Definition
BMC	Baseboard management controller
CDH	Cloudera Distribution for Hadoop
DBMS	Database management system
EDW	Enterprise data warehouse
EoR	End-of-row switch/router
HDFS	Hadoop File System
IPMI	Intelligent Platform Management Interface
NIC	Network interface card
LOM	Local area network on motherboard
OS	Operating system
ToR	Top-of-rack switch/router

Update History

Changes in Version 5.0

The following changes have been made to this guide since the 2.4 release:

- Re-organize document to include an overall architecture summary section
- Network architecture has been condensed to remove duplicate content
- Update to Cloudera Enterprise Version 5.0, including YARN.
- Removed configuration parameters, since these are documented in the deployment guide.
- Administration node and deployment software is now optional instead of required.
- Removed C8000 configurations.
- Added additional guidance for nodes with large storage density.

To Learn More

For more information on the Dell | Cloudera Solution, visit:
www.Dell.com/Hadoop