

Dell Ready Bundle for Cloudera Hadoop

Architecture Guide

Version 5.10



Dell Converged Platforms and Solutions

Contents

List of Figures.....	v
List of Tables.....	vi
Trademarks.....	8
Glossary.....	9
Notes, Cautions, and Warnings.....	14
Chapter 1: Dell Ready Bundle for Cloudera Hadoop Overview.....	15
Introducing the Dell Ready Bundle for Cloudera Hadoop.....	16
Solution Use Case Summary.....	16
Solution Components.....	17
ETL Solution Components.....	19
Software Support.....	19
Cloudera Enterprise Software Overview.....	20
Hadoop for the Enterprise.....	20
Rethink Data Management.....	20
What's Inside?.....	21
Cloudera Enterprise Data Hub.....	21
Syncsort DMX-h Overview.....	22
Hadoop for Data Transformation.....	22
Chapter 2: Cluster Architecture.....	23
High-Level Node Architecture.....	24
Node Definitions.....	25
Network Fabric Architecture.....	26
Network Definitions.....	27
Cluster Sizing.....	28
Rack.....	28
Pod.....	28
Cluster.....	29
Sizing Summary.....	29
High Availability.....	30
Hadoop Redundancy.....	30
Network Redundancy.....	30
HDFS Highly Available NameNodes.....	30
Resource Manager High Availability.....	31
Database Server High Availability.....	31
Chapter 3: Hardware Architecture.....	32
Server Infrastructure Options.....	33
Dell PowerEdge R730xd Server.....	33
Dell PowerEdge FX Architecture.....	35

Chapter 4: Network Architecture.....	40
Cluster Networks.....	41
Physical Network Components.....	41
Server Node Connections.....	42
Pod Switches.....	44
Cluster Aggregation Switches.....	45
Core Network.....	47
Layer 2 and Layer 3 Separation.....	47
iDRAC Management Network.....	47
Network Equipment Summary.....	48
Chapter 5: Cloudera Enterprise Software.....	49
Cloudera Manager.....	50
Cloudera RTQ (Impala).....	50
Cloudera Search.....	50
Cloudera BDR.....	51
Cloudera Navigator.....	51
Cloudera Support.....	52
Chapter 6: Syncsort Software.....	53
Syncsort DMX-h Engine.....	54
Syncsort DMX-h Service.....	54
Syncsort DMX-h Client.....	54
Syncsort SILQ.....	54
Chapter 7: Deployment Methodology.....	55
Appendix A: Physical Rack Configuration - Dell PowerEdge R730xd.....	56
Dell PowerEdge R730xd Single Rack Configuration.....	57
Dell PowerEdge R730xd Initial Rack Configuration.....	58
Dell PowerEdge R730xd Additional Pod Rack Configuration.....	59
Appendix B: Physical Chassis Configuration - Dell PowerEdge FX2.....	61
Physical Chassis Configuration - Dell PowerEdge FX2.....	62
Appendix C: Physical Rack Configuration - Dell PowerEdge FX2.....	63
Dell PowerEdge FX2 Single Rack Configuration.....	64
Dell PowerEdge FX2 Initial and Second Pod Rack Configuration.....	65
Appendix D: Update History.....	68
Changes in Version 5.10.....	69

Appendix E: References.....	70
About Cloudera.....	71
About Syncsort.....	71
To Learn More.....	71

List of Figures

Figure 1: Dell Ready Bundle for Cloudera Hadoop Components.....	18
Figure 2: ETL Solution Components.....	19
Figure 3: Cluster Architecture.....	24
Figure 4: Cluster Network Fabric Architecture.....	27
Figure 5: Dell PowerEdge R730xd Servers – 2.5” and 3.5” Chassis Options.....	33
Figure 6: Dell PowerEdge FX2 Components.....	36
Figure 7: Dell PowerEdge FX2 Server Chassis with Dual Dell PowerEdge FC630 and Dual Dell PowerEdge FD332.....	36
Figure 8: Hadoop Network Connections.....	41
Figure 9: PowerEdge R730xd Node Network Ports.....	42
Figure 10: Dell PowerEdge FX2 Infrastructure Chassis Network Ports.....	43
Figure 11: Dell PowerEdge FX2 Worker Chassis Network Ports.....	43
Figure 12: Single Pod Networking Equipment.....	45
Figure 13: Dell Networking S6000-ON Multi-pod Networking Equipment.....	46
Figure 14: Multi-Pod View Using Dell Networking Z9100 Switches (Based on Layer-3 ECMP).....	47

List of Tables

Table 1: Big Data Solution Use Cases.....	16
Table 2: ETL Solution Use Cases.....	17
Table 3: Data processing and access components.....	19
Table 4: Dell Ready Bundle for Cloudera Hadoop Support Matrix.....	19
Table 5: Included Products.....	21
Table 6: Advanced Components.....	21
Table 7: DMX-h ETL Edition Key Features.....	22
Table 8: Cluster Node Roles.....	24
Table 9: Service Locations.....	25
Table 10: Networks.....	27
Table 11: Cloudera Distribution for Apache Hadoop Network Definitions.....	27
Table 12: Recommended Cluster Size.....	29
Table 13: Alternative Cluster Sizes.....	29
Table 14: Rack and Pod Density Scenarios.....	30
Table 15: Hardware Configurations – Dell PowerEdge R730xd Infrastructure Nodes....	34
Table 16: Hardware Configurations – Dell PowerEdge R730xd Worker Nodes.....	34
Table 17: Hardware Configurations – Dell PowerEdge FX2 FC630 Infrastructure Nodes.....	37
Table 18: Hardware Configurations – Dell PowerEdge FX2 FC630 Worker Nodes.....	38
Table 19: Cluster Networks.....	41
Table 20: Bond / Interface Cross Reference.....	43
Table 21: Per Rack Network Equipment.....	48
Table 22: Per Pod Network Equipment.....	48
Table 23: Per Cluster Aggregation Network Switches for Multiple Pods.....	48
Table 24: Per Node Network Cables Required – 10GbE Configurations.....	48
Table 25: Cloudera Support Features.....	52

Table 26: Single Rack Configuration – Dell PowerEdge R730xd.....	57
Table 27: Initial Pod Rack Configuration – Dell PowerEdge R730xd.....	58
Table 28: Additional Pod Rack Configuration – Dell PowerEdge R730xd.....	59
Table 29: Infrastructure Chassis Configuration – Dell PowerEdge FX2.....	62
Table 30: Worker Chassis Configuration – Dell PowerEdge FX2.....	62
Table 31: Single Rack Configuration – Dell PowerEdge FX2.....	64
Table 32: Initial and Second Pod Rack Configuration– Dell PowerEdge FX2.....	65

Trademarks

Copyright © 2011-2017 Dell Inc. or its subsidiaries. All rights reserved. Dell and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners.

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

Glossary

ASCII

American Standard Code for Information Interchange, a binary code for alphanumeric characters developed by ANSI®.

BMC

Baseboard Management Controller

BMP

Bare Metal Provisioning

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.

CMC

Chassis Management Controller

DBMS

Database Management System

DTK

Dell OpenManage Deployment Toolkit

EBCDIC

Extended Binary Coded Decimal Interchange Code, a binary code for alphanumeric characters developed by IBM®.

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

HVE

Hadoop Virtualization Extensions

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

PAM

Pluggable Authentication Modules, a centralized authentication method for Linux systems.

RPM

Red Hat Package Manager

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

Notes, Cautions, and Warnings



Note: A **Note** indicates important information that helps you make better use of your system.



Caution: A **Caution** indicates potential damage to hardware or loss of data if instructions are not followed.



Warning: 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.

Chapter 1

Dell Ready Bundle for Cloudera Hadoop Overview

Topics:

- [*Introducing the Dell Ready Bundle for Cloudera Hadoop*](#)
- [*Solution Use Case Summary*](#)
- [*Solution Components*](#)
- [*Cloudera Enterprise Software Overview*](#)
- [*Syncsort DMX-h Overview*](#)

The Dell Ready Bundle for Cloudera Hadoop 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.

Introducing the Dell Ready Bundle for Cloudera Hadoop

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 cluster.

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 Ready Bundle for Cloudera Hadoop 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 Enterprise and Dell PowerEdge and Dell Networking hardware. This solution includes components that span the entire solution stack:

- [Dell Ready Bundle for Cloudera Hadoop Architecture Guide](#) and best practices
- Optimized server configurations
- Optimized network infrastructure
- Cloudera Enterprise

Solution Use Case Summary

The Dell Ready Bundle for Cloudera Hadoop is designed to address the use cases described in [Table 1: Big Data Solution Use Cases](#) on page 16:

Table 1: Big Data 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.
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.

Use case	Description
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.
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.

The Dell Ready Bundle for Cloudera Hadoop with SyncSort is designed to address the use cases described in [Table 2: ETL Solution Use Cases](#) on page 17:

Table 2: ETL Solution Use Cases

Use case	Description
ETL offload	Offload ETL processing from a RDBMS 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.
Integration with data warehouse	Extract, transfer and load data into and out of Hadoop into a separate DBMS for advanced analytics.
High-Performance data transformations	Includes high-performance sort, joins, aggregations, multi-key lookup, advanced text processing, hashing functions, and source/record/field-level operations.
Mainframe data ingestion & translation	Reads files directly from the mainframe, parses and transforms the data – packed decimal, occurs depending on, EBCDIC/ASCII, multi-format records, and more — without installing any software on the mainframe and without writing any code.

Solution Components

[Figure 1: Dell Ready Bundle for Cloudera Hadoop Components](#) on page 18 illustrates the primary components in the Dell Ready Bundle for Cloudera Hadoop.

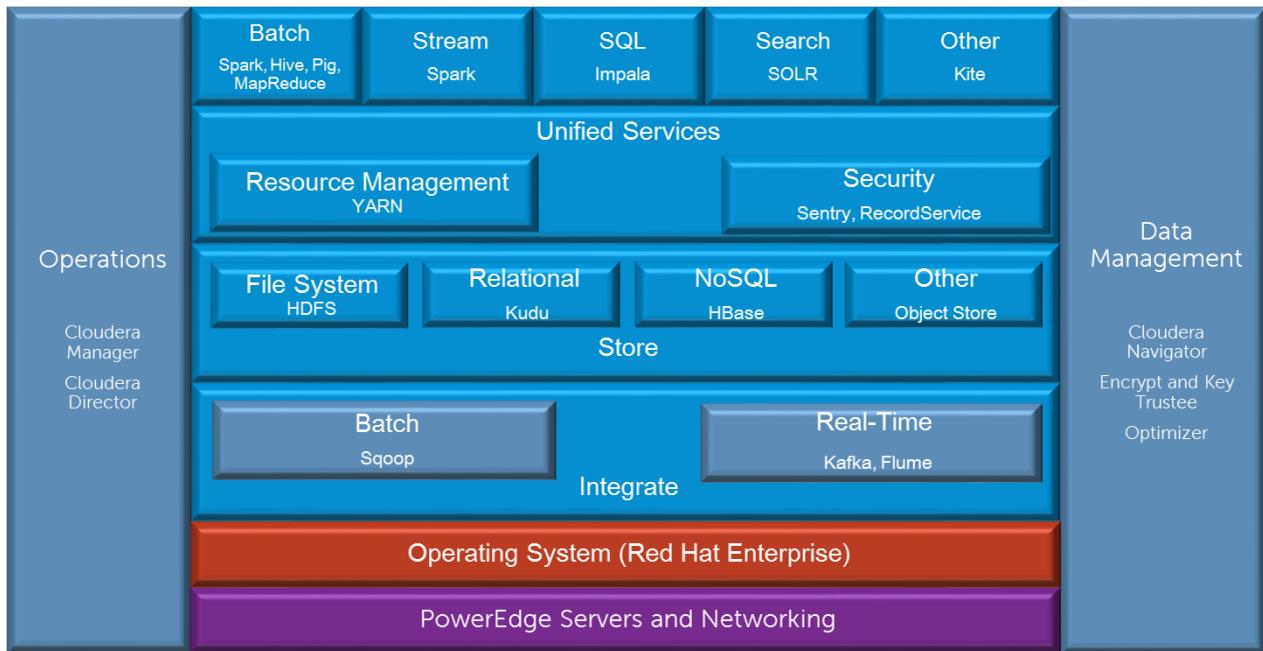


Figure 1: Dell Ready Bundle for Cloudera Hadoop Components

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

The Store layer components provide multiple layers of functionality on top of this foundation. 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 Kudu provides a columnar relational storage option, while Apache HBase provides NoSQL access to storage. Object storage is also available.

The Integrate layer shows the 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 and Apache Kafka are optimized for real-time processing event and log data. The HDFS API and tools can also be used to move data files to and from the Hadoop system.

YARN provides a resource management framework for running distributed applications under Hadoop. 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. Enterprise grade Security services are provided by Apache Sentry and RecordService.

The right side of the diagram shows the data management capabilities that are integrated across the entire system, while the left side shows the operational components for Hadoop administration and management provided by Cloudera Manager and Cloudera Director.

Sitting atop the Cloudera Enterprise core are multiple complementary processing and access alternatives.

- Batch data processing
- Stream data processing
- SQL query
- Data search
- Access API

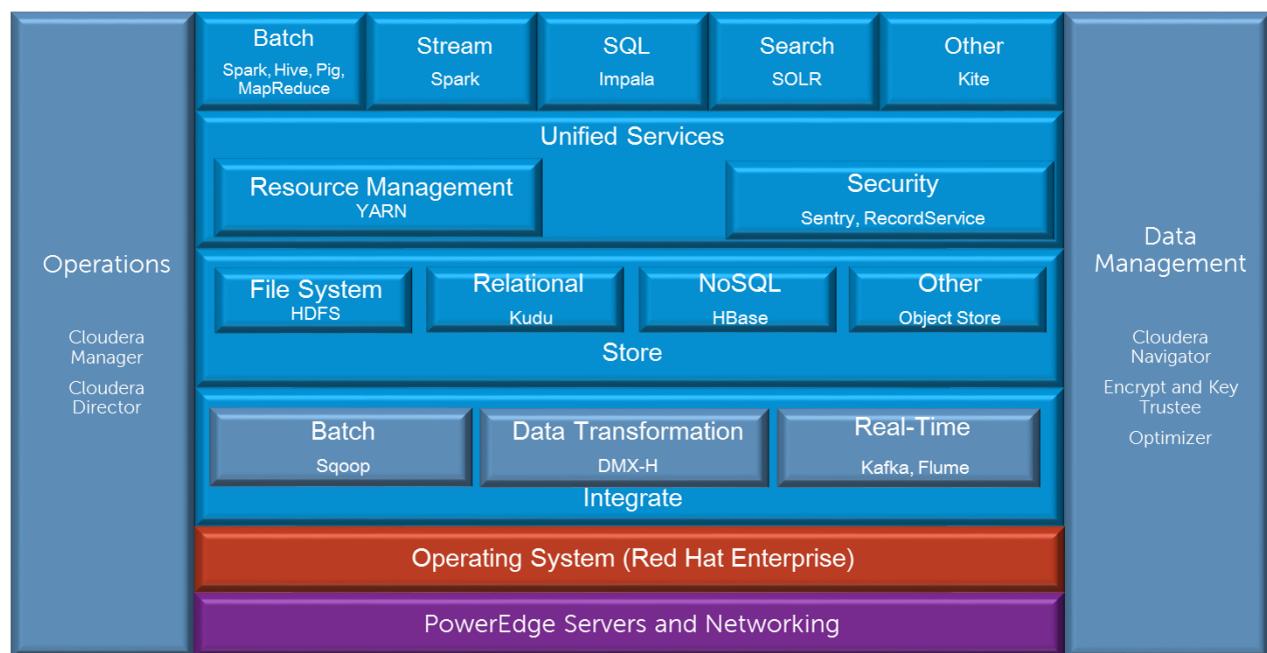
All of these layers can be used simultaneously or independently, depending on the workload and problems being solved.

Table 3: Data processing and access components

Access Layer	Description
Batch data processing	Spark, Hive, Pig, MapReduce provide access to the massively parallel Hadoop data processing framework.
Stream data processing	Spark can also be used for stream processing.
SQL query	Impala provides SQL query access to data.
Data search	Apache SOLR provides real-time search of indexed data.
General Access	Kite provides a high level data API for Hadoop.

ETL Solution Components

The ETL Solution is a variation of the architecture that adds Syncsort DMX-h to the installation. [Figure 2: ETL Solution Components](#) on page 19 illustrates this variation.

**Figure 2: ETL Solution Components**

The Syncsort DMX-h ETL engine runs under the YARN Resource management framework, and adds scalable ETL processing to the cluster. DMX-h can access mainframe and RDBMS data, while events can be handled using either Flume or Kafka. DMX-h coexists with all other available Hadoop components, and can be used in conjunction with them.

Software Support

[Table 4: Dell Ready Bundle for Cloudera Hadoop Support Matrix](#) on page 19 describes where you can obtain technical support for the various components of the Dell Ready Bundle for Cloudera Hadoop.

Table 4: Dell Ready Bundle for Cloudera Hadoop Support Matrix

Category	Component	Version	Available Support
Operating System	Red Hat Enterprise Linux Server	7.3	Red Hat Linux support

Category	Component	Version	Available Support
Operating System	CentOS	7.3	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 Enterprise	5.10	Cloudera support
Hadoop	Cloudera Manager	5.10	Cloudera support
Hadoop	Cloudera Navigator	2.9	Cloudera support
ETL Engine	Syncsort DMX-h	9.2	Syncsort 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 Manager 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:

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. Cloudera Navigator includes Active Data Optimization (Cloudera Navigator Optimizer), Governance & Data Management (Cloudera Navigator including auditing, lineage, discovery, & policy lifecycle management), Encryption & Key Management (Cloudera Navigator Encrypt & Key Trustee).

Syncsort DMX-h Overview

Syncsort DMX-h is high-performance software that turns Hadoop into a robust ETL solution, focused on delivering capabilities and use cases that are standard on traditional data integration platforms. DMX-h can accelerate your data integration initiatives and unleash Hadoop's potential with the only architecture that runs ETL processes natively within Hadoop.

Hadoop for Data Transformation

DMX-h is the Hadoop-enabled edition of DMExpress, providing the following Hadoop functionality:

- ETL Processing in Hadoop – Develop an ETL application entirely in the DMExpress GUI to run seamlessly in the Hadoop MapReduce framework, with no Pig, Hive, or Java programming required.
- Hadoop Sort Acceleration – Seamlessly replace the native sort within Hadoop MapReduce processing with the high-speed DMExpress engine sort, providing performance benefits without programming changes to existing MapReduce jobs.
- Apache Sqoop Integration – Use the Sqoop mainframe import connector to transfer mainframe data into HDFS.

Table 7: DMX-h ETL Edition Key Features

Feature	Description
High-Performance Data Transformations	Includes high-performance sort, joins, aggregations, multi-key lookup, advanced text processing, hashing functions, and source/record/field-level operations.
Rapid Development through Windows-based DMX-h Workstation	Lets you develop and test MapReduce ETL jobs locally in Windows through a graphical user interface, then deploy in Hadoop. Expression Builder helps define data transformations based on business rules.
Use Case Accelerators	Fast-tracks your Hadoop productivity with a library of fully-functional and reusable templates – including web log processing, change data capture, mainframe connectivity, joins, and more – to design your own data flows.
Data Source & Target Connectivity	Connects any source and target to Hadoop, including all major database management systems, flat files, XML files, mainframe and others.
Mainframe data ingestion & translation	Reads files directly from the mainframe, parses and transforms the data – packed decimal, occurs depending on, EBCDIC/ASCII, multi-format records, and more – without installing any software on the mainframe, and without writing any code.
Dynamic ETL Optimizer	Performs data transformations and functions at maximum speed based on hundreds of proprietary algorithms. The ETL optimizer automatically chooses the best algorithm to maximize performance of each node in Hadoop and adapts in real-time to system conditions.
File-based Metadata Capabilities	Provides greater transparency into impact analysis, data lineage, and execution flow without dependencies on third-party systems, such as relational databases.

Chapter 2

Cluster Architecture

Topics:

- *High-Level Node Architecture*
- *Network Fabric Architecture*
- *Cluster Sizing*
- *High Availability*

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 chapter summarizes the main aspects of the solution architecture.

High-Level Node Architecture

Figure 3: Cluster Architecture on page 24 displays the roles for the nodes in a basic cluster.

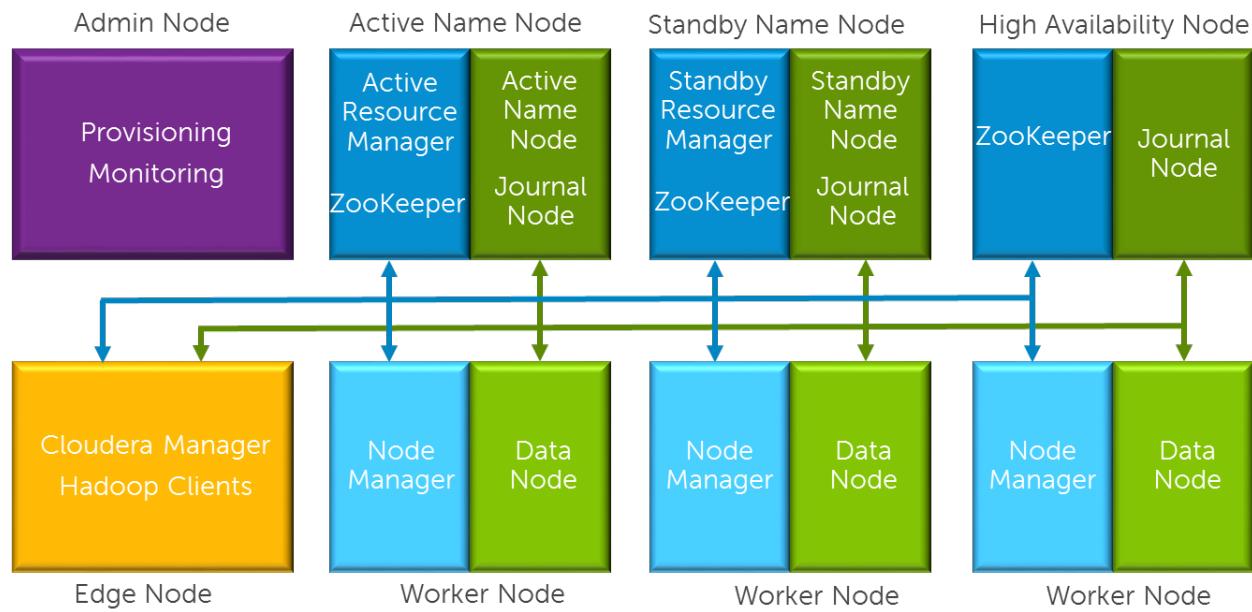


Figure 3: Cluster Architecture

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 - Worker 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 eight cluster nodes. The nodes have the following roles:

Table 8: Cluster Node Roles

Node Role	Required?	Hardware Configuration
Active Name Node	Required	Infrastructure
Standby Name Node	Required	Infrastructure
High Availability (HA) Node	Required	Infrastructure
Edge Node	Required	Infrastructure
Worker Node 1	Required	Worker
Worker Node 2	Required	Worker
Worker Node 3	Required	Worker
Worker Node 4	Required	Worker
Worker Node 5	Required	Worker

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.
- **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 the Spark History Server and 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 on the HA.
- **Edge Node** — provides an interface between the data and processing capacity available in the Hadoop cluster and a user of that capacity. An Edge Node has a an additional connection to the Edge Network, and is sometimes called a “gateway node.” At least one Edge Node is required.
- **Worker 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 five Worker Nodes are required, and larger clusters are scaled primarily by adding additional Worker Nodes. There are three types of services running on the Worker Nodes:
 - *DataNode* daemon (to support HDFS data storage)
 - *NodeManager* daemon (to support YARN job execution)
 - Services managed with Cloudera Manager service pools instead of YARN, such as *Impala* and *HBase*

Spark jobs also run on the Worker Nodes. However, there is no persistent service associated with Spark jobs.

[Table 9: Service Locations](#) on page 25 describes the node locations and functions of the cluster services.

Table 9: Service Locations

Physical Node	Software Function
Administration Node	Systems Management Services
First Edge Node	Hadoop Clients Cloudera Manager DMX-h DMEExpress Service (dmxd)

Physical Node	Software Function
Active Name Node	NameNode Resource Manager ZooKeeper Quorum Journal Node HMaster Impala State Store and Catalog Daemons
Standby Name Node	Yum Repositories Standby NameNode Standby Resource Manager (optional) Spark History Server Spark2 History Server ZooKeeper Quorum Journal Node
HA Node	ZooKeeper Quorum Journal Node Operational Databases (PostgreSQL)
Worker Node(N)	DataNode NodeManager HBase 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 4: Cluster Network Fabric Architecture](#) on page 27 displays details:

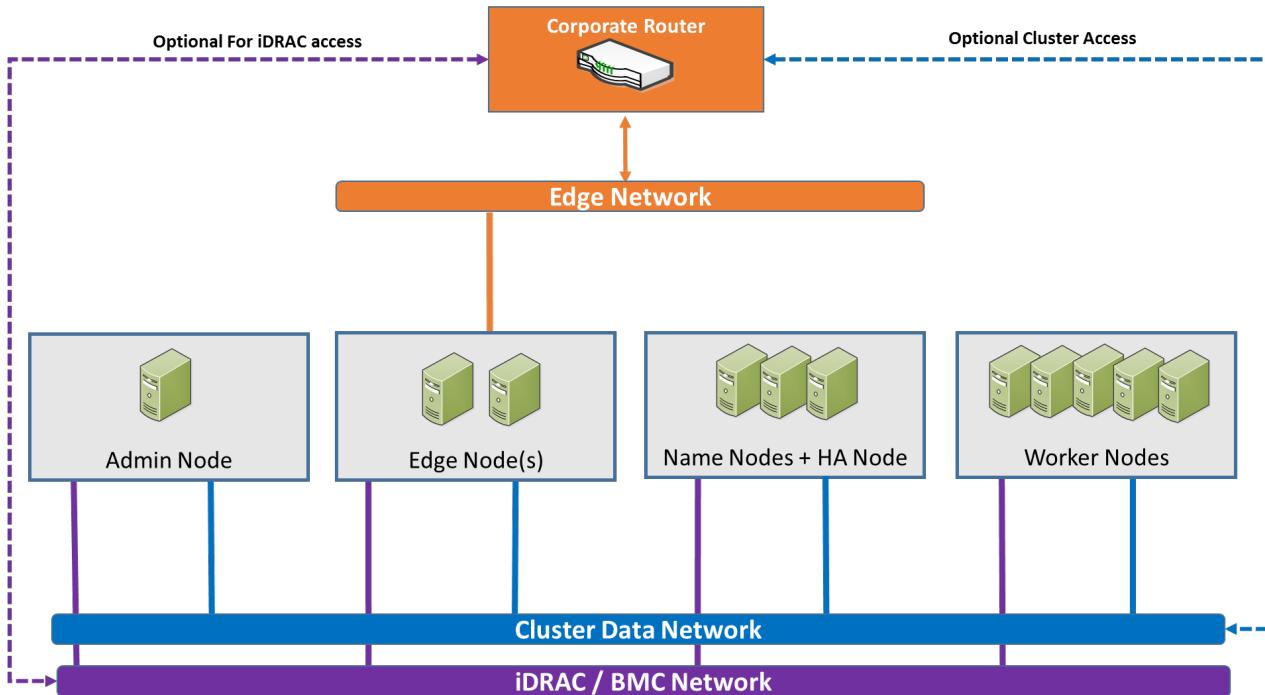


Figure 4: Cluster Network Fabric Architecture

Three distinct networks are used in the cluster:

Table 10: Networks

Logical Network	Connection	Switch
Cluster Data Network	Bonded 10GbE	Dual top of rack switches
iDRAC / BMC Network	1GbE	Switch per rack, dedicated or shared with Management network
Edge Network	10GbE, optionally bonded	Direct to edge network, or via pod or aggregation switch

Network Definitions

[Table 11: Cloudera Distribution for Apache Hadoop Network Definitions](#) on page 27 defines the Cloudera Distribution for Apache Hadoop networks.

Table 11: 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 pod, and pods 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.	<p>The Cloudera Enterprise services are available on this network.</p> <p>Note: The Cloudera Enterprise services do not support multi-homing, and are only accessible on the Cluster Data Network.</p>

Network	Description	Available Services
iDRAC / BMC Network	The BMC network connects the BMC or iDRAC ports and the out-of-band management ports of the switches, and is used for hardware provisioning and management. It is aggregated into a management switch in each rack.	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 the Edge Node(s) to an existing premises network, either directly, or via the pod or cluster aggregation switches.	SSH access is available on this network, and other application services may be configured and available.

Connectivity between the cluster and existing network infrastructure can be adapted to specific installations. Common scenarios are:

1. The cluster data network is isolated from any existing network and access to the cluster is via the edge network only.
2. The cluster data network is exposed to an existing network. In this scenario, the edge network is either not used, or is used for application access or ingest processing.

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 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 Worker Nodes. A rack is a physical unit, and its 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 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 using a scale out server like the Dell PowerEdge R730xd. For high density servers like the Dell PowerEdge FX2, a rack may contain 24 or more nodes. The node density of a rack does not affect overall cluster scaling and sizing, but it does affect fault zones in the cluster.

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 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 architecture, a cluster using Dell Networking S6000-ON switches can scale to 7 pods, and a maximum of 252 nodes. For larger clusters the Dell Networking Z9100 switch can be used.

Sizing Summary

The minimum configuration supported is eight nodes:

- Active Name Node
- Standby Name Node
- High Availability (HA) Node
- Five (5) Worker Nodes

Additionally, a minimum of one Edge Node is required per cluster. Larger clusters and clusters with high ingest volumes or rates may benefit from additional Edge Nodes. Cloudera recommends one Edge Node for every twenty Worker 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 Worker Nodes.

Table 12: Recommended Cluster Size on page 29 shows the recommended number of nodes per pod and pods per cluster, using the S4048-ON and S6000-ON switch models. *Table 13: Alternative Cluster Sizes* on page 29 shows some alternatives for cluster sizing with different bandwidth oversubscription ratios.

Table 12: Recommended Cluster Size

Nodes Per Rack	Nodes Per Pod	Pods Per Cluster	Nodes Per Cluster	Bandwidth Oversubscription
12	36	7	252	2.25 : 1

Table 13: Alternative Cluster Sizes

Nodes Per Rack	Nodes Per Pod	Pods Per Cluster	Nodes Per Cluster	Bandwidth Oversubscription
12	12	14	168	1.5 : 1
12	24	9	216	2 : 1
10	20	14	280	2.5 : 1
12	36	9	324	3 : 1

Power and cooling will typically be the primary constraints on rack density. However, a rack is a potential fault zone, and rack density will affect overall cluster reliability, especially for smaller clusters. *Table 14: Rack and Pod Density Scenarios* on page 30 shows some possible scenarios based on typical data center constraints.

Table 14: Rack and Pod Density Scenarios

Server Platform	Nodes Per Rack	Racks Per Pod	Comments
Dell PowerEdge R730xd	12	3	Typical configuration, requiring less than 10kW power per rack. Good rack level fault zone isolation.
Dell PowerEdge R730xd	10	2	Smaller rack and pod fault zones, with slightly higher bandwidth oversubscription of 2.5 : 1
Dell PowerEdge FX2 / FC630	12	3	Dell PowerEdge FX2 equivalent of Dell PowerEdge R730xd baseline configuration for compute density.
Dell PowerEdge FX2 / FC630	24	1.5	Can support two pods in three racks, requires approximately 20 kW power per rack. Rack level fault can potentially affect two pods.

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.

Chassis and Node Redundancy

For the Dell PowerEdge FX2 platform, multiple nodes can share the same chassis, which affects the potential failure zones. For this platform, we use the Hadoop Virtualization Extensions to provide additional information about the structure of the cluster to Hadoop. HVE is configured to identify the pairs of FC630 worker nodes in a single chassis as a node group. This configuration changes the Hadoop replication policy to ensure that replicas are distributed across more than one chassis. Full details of the recommended HVE configuration are in the [Dell Ready Bundle for Cloudera Hadoop Deployment Guide](#).

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 NameNodes

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

In quorum-based HA, there are typically two namenode 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 NameNode is simply acting as a slave, maintaining enough state to provide a fast failover if necessary.

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

The Standby NameNode is capable of reading the edits from the JournalNodes, and is constantly watching them for changes to the edit log. As the Standby NameNode sees the edits, it applies them to its own namespace. In the event of a failover, the Standby NameNode 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 NameNode has up-to-date information regarding the location of blocks in the cluster. In order to achieve this, the Worker Nodes are configured with the location of both NameNodes, and they send block location information and heartbeats to both.

There should be an odd number of (and at least three) JournalNode daemons, since edit log modifications must be written to a majority of JournalNodes. The JournalNode daemons run on the Active NameNode, Standby NameNode, and HA Node in this 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:** Resource Manager HA is not always implemented in production clusters.

Database Server High Availability

The architecture supports High Availability for the operational databases.

The database server used for the Cloudera Manager operational databases and the metadata databases stores its data on a RAID 10 partition to provide redundancy in the case of a drive failure.

 **Note:** Our default installation uses a single PostgreSQL instance, so there is a single point of failure. Database server High Availability can be implemented using one or more additional PostgreSQL instances on other nodes in the cluster, or by using an external database server.

Chapter

3

Hardware Architecture

Topics:

- *Server Infrastructure Options*

The Dell Ready Bundle for Cloudera Hadoop utilizes Dell's latest server solutions.

Server Infrastructure Options

The Dell Ready Bundle for Cloudera Hadoop includes the following server options:

- [Dell PowerEdge R730xd Server](#) on page 33
- [Dell PowerEdge FX Architecture](#) on page 35

Dell 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 product family, up to 24 DIMMS, PCI Express® (PCIe) 3.0 enabled expansion slots, and a choice of network interface technologies.

The Dell PowerEdge R730xd platform includes highly-expandable memory (up to 768GB) and impressive I/O capabilities to match. The Dell 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 Dell 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 5: Dell PowerEdge R730xd Servers – 2.5” and 3.5” Chassis Options

Dell PowerEdge R730xd Feature Summary

Dell PowerEdge R730xd features include:

- Intel Grantley platform and Intel Xeon E5-2600v4 (Broadwell) processors
- Up to 2400 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

Dell PowerEdge R730xd Hardware Configurations

The Dell Ready Bundle for Cloudera Hadoop supports the following Dell PowerEdge R730xd server configurations:

- [Table 15: Hardware Configurations – Dell PowerEdge R730xd Infrastructure Nodes](#) on page 34

- [Table 16: Hardware Configurations – Dell PowerEdge R730xd Worker Nodes](#) on page 34

Table 15: Hardware Configurations – Dell PowerEdge R730xd Infrastructure Nodes

Machine Function	Infrastructure Nodes - 3.5" Chassis Option	Infrastructure Nodes - 2.5" Chassis Option
Platform	Dell PowerEdge R730xd	Dell PowerEdge R730xd
Processor	2 x Intel Xeon E5-2650 v4 2.2GHz (12 core)	2 x E5-Intel Xeon E5-2690 v4 2.6GHz (14 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 600GB 15K RPM SAS 12Gbps 2.5-in	2 x 600GB 15K 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 16: Hardware Configurations – Dell PowerEdge R730xd Worker Nodes

Machine Function	Worker Nodes - 3.5" Chassis Option	Worker Nodes - 2.5" Chassis Option
Platform	Dell PowerEdge R730xd	Dell 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 v4 2.2GHz (12 core)	2 x E5-Intel Xeon E5-2690 v4 2.6GHz (14 core)
RAM (minimum)	256 GB	256 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 600GB 15K RPM SAS 12Gbps 2.5-in	2 x 600GB 15K RPM SAS 12Gbps 2.5-in
Storage Controller	PERC H730	PERC H730

Machine Function	Worker Nodes - 3.5" Chassis Option	Worker 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.

Dell 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 Worker 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 Ready Bundle for Cloudera Hadoop Deployment Guide](#).

The recommended rack layout for Dell PowerEdge R730xd clusters is illustrated here:

- [Physical Rack Configuration - Dell PowerEdge R730xd](#) on page 56

The full bills of material (BOM) listing for the Dell PowerEdge R730xd server configurations are listed in the [Dell Ready Bundle for Cloudera Hadoop Bill of Materials Guide](#).

Dell PowerEdge R730xd 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.

Dell PowerEdge FX Architecture

The Dell PowerEdge FX is Dell's revolutionary approach to converged infrastructure. The Dell PowerEdge FX converged architecture gives enterprises the flexibility to tailor their IT infrastructure to specific workloads, and the ability to scale and adapt that infrastructure as needs change.

The Dell PowerEdge FX2 is a hybrid rack-based computing platform that combines the density and efficiencies of blades with the simplicity and cost benefits of rack-based systems. With an innovative modular design that accommodates IT resource building blocks of various sizes — compute, storage, networking and management — the Dell PowerEdge FX2 enables data centers to construct their infrastructures with greater flexibility.

The foundation of the Dell PowerEdge FX architecture is a 2U rack enclosure — the Dell PowerEdge FX2 — that can accommodate various sized resource blocks of servers or storage. Resource blocks slide into the chassis, like a blade server node, and connect to the shared infrastructure through a flexible I/O fabric. Currently, Dell PowerEdge FX architecture components include the following server nodes and storage block:

1. Dell PowerEdge FM120x4: world's first enterprise-class microserver

2. Dell PowerEdge FC430: ultimate in shared infrastructure density
3. Dell PowerEdge FC630: shared infrastructure workhorse
4. Dell PowerEdge FD332: ultimate dense direct-attached storage with unprecedented flexibility.

The Dell PowerEdge FX2 enclosure enables servers and storage to share power, cooling, management and networking. It houses redundant power supply units (2000W, 1600W or 1100W) and eight cooling fans. With a compact highly flexible design, the Dell PowerEdge FX2 chassis lets you simply and efficiently add resources to your infrastructure when and where you need them, so you can let demand and budget determine your level of investment.

This architecture uses Dell PowerEdge FC630 server and Dell PowerEdge FD332 storage blocks for the optimum balance between compute and storage density.

Dell's agent-free integrated Dell Remote Access Controller (iDRAC) with Lifecycle Controller makes server deployment, configuration and updates automated and efficient. Using Chassis Management Controller (CMC), an embedded component that is part of every Dell PowerEdge FX2 chassis, you'll have the choice of managing server nodes individually or collectively via a browser-based interface. OpenManage Essentials provides enterprise-level monitoring and control of Dell and third-party data center hardware, and works with OpenManage Mobile to provide similar information on smart phones. OpenManage Essentials now also delivers server configuration management capabilities that automate bare-metal server and OS deployments, replication of configurations, and ensures ongoing compliance with system configurations.



Figure 6: Dell PowerEdge FX2 Components



Figure 7: Dell PowerEdge FX2 Server Chassis with Dual Dell PowerEdge FC630 and Dual Dell PowerEdge FD332

Dell PowerEdge FX2 Feature Summary

Dell PowerEdge FX2 features for the blocks used in this architecture are listed below.

Dell PowerEdge FD332 Storage Block

The Dell PowerEdge FD332 storage block is a densely packed storage module that allows FX-based infrastructures to rapidly and flexibly scale storage resources. It enables future-ready scale-out infrastructures that bring storage closer to compute for accelerated processing. Each 1U, half-width Dell PowerEdge FD332 storage block holds up to 16 small form factor, hot-plug storage drives, either HDD or SSD and can have up to two RAID controllers. This dense, highly scalable storage option means that up to 48 additional drives can be housed in an Dell PowerEdge FX2 chassis – resulting in massive direct attach capacity and an efficient pay-as-you-grow IT model

Specifications

- Designed as a half-width, 1U storage block that holds up to sixteen 2.5" hot plug SAS or SATA disk drives (SSD or HDD).
- Up to 48 drives in a single 2U Dell PowerEdge FX2 – leaving one halfwidth chassis slot to house an Dell PowerEdge FC630 for processing.
- Enables a pay-as-you-grow IT model.
- FX servers can be attached to a single Dell PowerEdge FD332, or multiple Dell PowerEdge FD332s, and can either attach to all 16 devices in the storage block or split access to the block and attach to 8 devices separately.
- Combine FX servers and storage in a wide variety of configurations to address specific processing needs – and Dell PowerEdge FD332 storage devices are independently serviceable while the Dell PowerEdge FX2 chassis is still in operation.

Dell PowerEdge FC630 server

The Dell PowerEdge FC630 server provides a strong foundation for corporate data centers and private clouds. The Dell PowerEdge FC630 readily handles demanding business applications like enterprise resource planning (ERP) and customer relationship management (CRM), and can also host large virtualization environments. With high-performance processors and large memory capacity, the two-socket Dell PowerEdge FC630 provides a resource rich environment that can support a wide range of mission critical workloads.

Specifications

- The half-width, two-socket server delivers an exceptional amount of computing power in a very small, easily scalable form factor, with the latest 22-core Intel® Xeon® E5-2600v4 processors and up to 24 dual in-line memory modules (DIMMs).
- Consider that a 2U Dell PowerEdge FX2 chassis fully loaded with four Dell PowerEdge FC630 servers can be scaled up to 176 cores and 96 DIMMs.
- The Dell PowerEdge FC630 offers two internal storage options - an eight 1.8 inch drive configuration and a two 2.5 inch drive configuration, the latter of which supports Express Flash NVME PCIe devices.

Dell PowerEdge FX2 Hardware Configurations

The Dell Ready Bundle for Cloudera Hadoop supports the following Dell PowerEdge FX2 server configurations:

- [Table 17: Hardware Configurations – Dell PowerEdge FX2 FC630 Infrastructure Nodes](#) on page 37
- [Table 18: Hardware Configurations – Dell PowerEdge FX2 FC630 Worker Nodes](#) on page 38

Table 17: Hardware Configurations – Dell PowerEdge FX2 FC630 Infrastructure Nodes

Machine Function	Infrastructure Nodes - Dell PowerEdge FC630
Platform	Dell PowerEdge FX2 FC630 Server
Processor	2 x Intel Xeon E5-2650 v4 2.2GHz (12 core)
RAM (minimum)	128GB, 2400 MT/s
Network Daughter Card	Intel X710 Quad Port, 10Gb KR Blade Network Daughter Card
Add in PCI Network Card	None
Onboard Storage	None
Storage Module	Dell PowerEdge FD332
Storage Controller	Single PERC H730

Machine Function	Infrastructure Nodes - Dell PowerEdge FC630
Disks	10 x 1TB 7.2K RPM Near-Line SAS 12Gbps 2.5in Hot-plug Hard Drive
Drive Configuration	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 18: Hardware Configurations – Dell PowerEdge FX2 FC630 Worker Nodes

Machine Function	Worker Nodes - Dell PowerEdge FC630
Platform	Dell PowerEdge FX2 FC630 Server
Processor	2 x Intel Xeon E5-2680 v4 2.4GHz (14 core)
RAM (minimum)	256 GB, 2400 MT/s
Network Daughter Card	Intel X710 Dual Port 10Gb KR Blade Network Daughter Card
Onboard Storage	1 x 400GB Solid State Drive SATA Mix Use MLC 6Gbps 2.5in Hot-plug Drive
Onboard Storage Controller	PERC S130
Storage Module	Dell PowerEdge FD332
Storage Controller	Dual PERC H730
Disks	16 x 2TB 7.2K RPM NLSAS 12Gbps 512n 2.5in Hot-plug Hard Drive
Drive Configuration	HBA - OS JBOD - data drives

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

Dell PowerEdge FX2 Configuration Notes

The Dell PowerEdge FX2 configurations use the Dell PowerEdge FC630 server. The Dell PowerEdge FC630 configuration provides two worker nodes in a 2U chassis, with 16 drives and dual 14 core processors in each.

Dell PowerEdge FX2 systems require a chassis module in addition to processor and storage modules. To simplify configuration, this architecture organizes the modules into two different chassis units that can be ordered independently. The grouping is:

1. Infrastructure Chassis, containing one infrastructure node and one Dell PowerEdge FC630 worker node
2. Worker Chassis, containing two Dell PowerEdge FC630 worker nodes

The converged nature of the Dell PowerEdge FX2 platform affects the node and rack level fault zone logic used by Hadoop. This architecture compensates by allocating infrastructure nodes across multiple chassis and enabling the Hadoop Virtualization Extensions (HVE) to provide the expected reliability and replication behaviors across both chassis and rack.

Full details of the drive configuration, filesystem layouts, and HVE configuration are in the [Dell Ready Bundle for Cloudera Hadoop Deployment Guide](#).

The recommended rack layout for Dell PowerEdge FX2 clusters is illustrated here:

- [*Physical Rack Configuration - Dell PowerEdge FX2*](#) on page 63

The full bills of material (BOM) listing for the Dell PowerEdge FX2 server configurations are listed in the [Dell Ready Bundle for Cloudera Hadoop Bill of Materials Guide](#).

Dell PowerEdge FX2 Storage Sizing Notes

Changes in drive capacity will impact performance. Smaller drives can be used, but may warrant a change in processors or memory size.



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

Chapter 4

Network Architecture

Topics:

- *Cluster Networks*
- *Physical Network Components*

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-ON 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.

Cluster Networks

Three distinct networks are used in the cluster:

Table 19: Cluster Networks

Logical Network	Connection	Switch
Cluster Data Network	Bonded 10GbE	Dual top of rack (Pod) switches and aggregation switches
BMC Network	1GbE	Dedicated switch per rack
Edge Network	10GbE, optionally bonded	Direct to edge network, or via pod or aggregation switch

Each network uses a separate VLAN and dedicated components when possible. [Figure 8: Hadoop Network Connections](#) on page 41 shows the logical organization of the network.

For more information on the actual configuration of the interfaces and switches, please see the [Dell Ready Bundle for Cloudera Hadoop Deployment Guide](#).

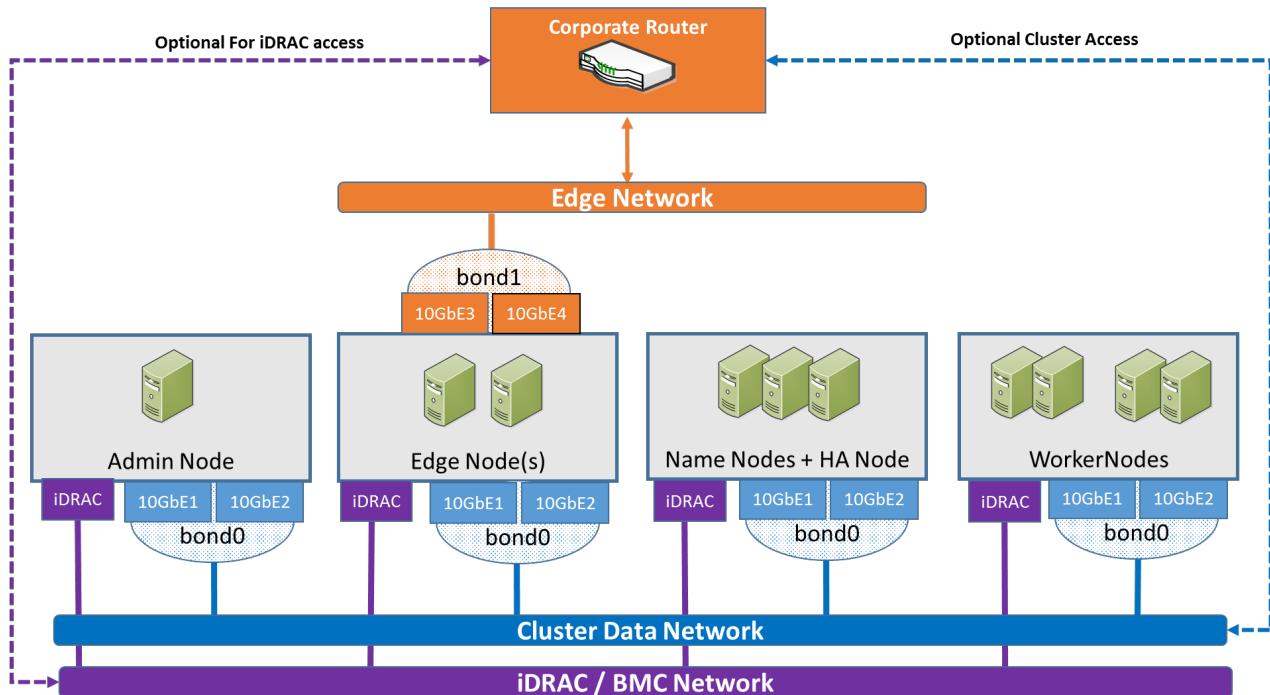


Figure 8: Hadoop Network Connections

Physical Network Components

The Dell Ready Bundle for Cloudera Hadoop physical networks consists of the following components:

- [Server Node Connections](#) on page 42
- [Pod Switches](#) on page 44
- [Cluster Aggregation Switches](#) on page 45

- [Core Network](#) on page 47
- [Layer 2 and Layer 3 Separation](#) on page 47
- [iDRAC Management Network](#) on page 47
- [Network Equipment Summary](#) on page 48

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, and the switches are configured as a high availability pair using VLT.

Connections to the BMC network use a single connection from the iDRAC port to a S3048-ON management switch in each rack.

Edge Nodes have an additional pair of 10GbE connections available. These connections facilitate high-performance cluster access between applications running on those nodes, and the optional edge network.

The mapping of bonds to individual interfaces is shown in [Table 20: Bond / Interface Cross Reference](#) on page 43.

The Dell PowerEdge FX2 architecture uses a converged iDRAC or CMC connection on the back of the chassis. All of the units in a Dell PowerEdge FX2 chassis use the same physical connector on the back of the unit for the physical network connection, and have separate IP addresses for each sub-unit. [Figure 11: Dell PowerEdge FX2 Worker Chassis Network Ports](#) on page 43 displays the connection for a single chassis iDRAC connection. The port next to it can be used to daisy chain the CMCs.

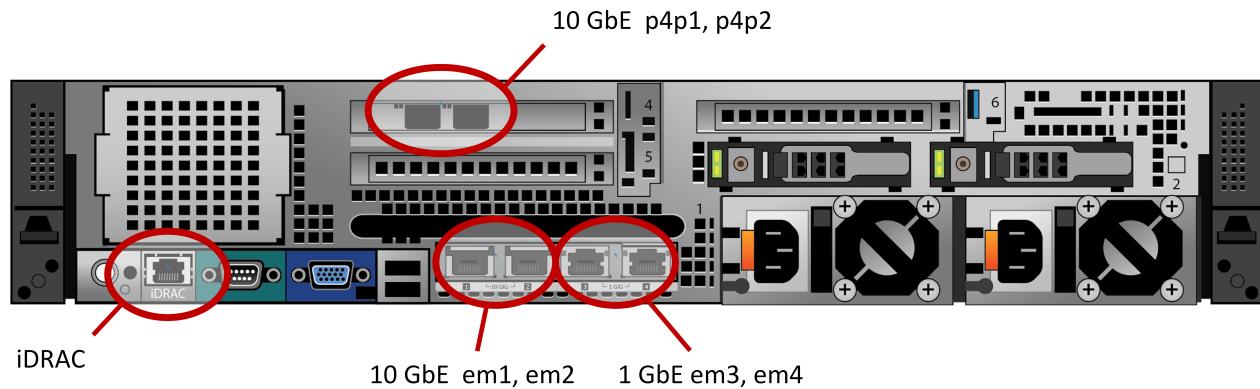


Figure 9: PowerEdge R730xd Node Network Ports

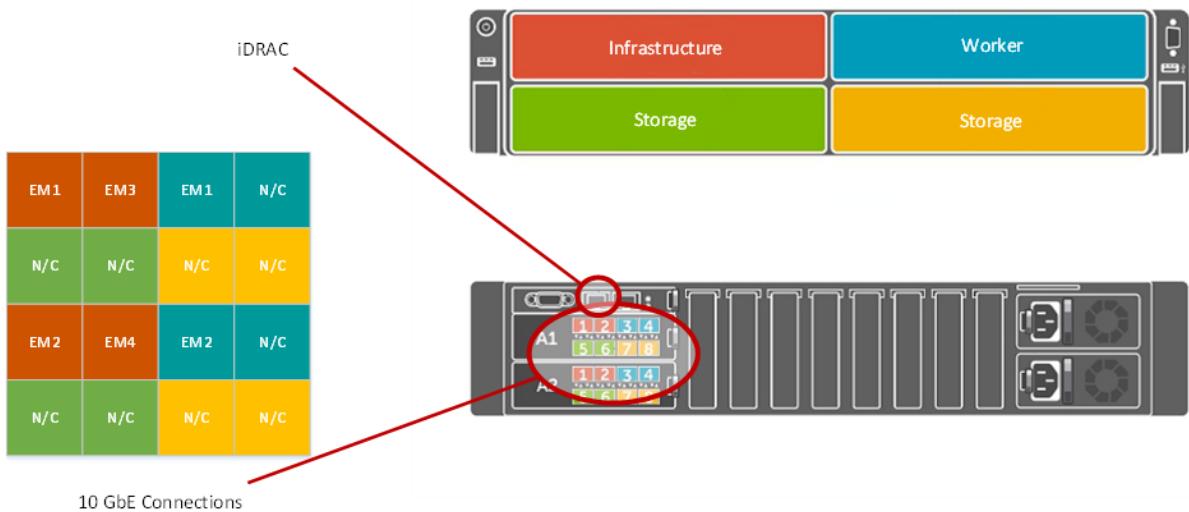


Figure 10: Dell PowerEdge FX2 Infrastructure Chassis Network Ports

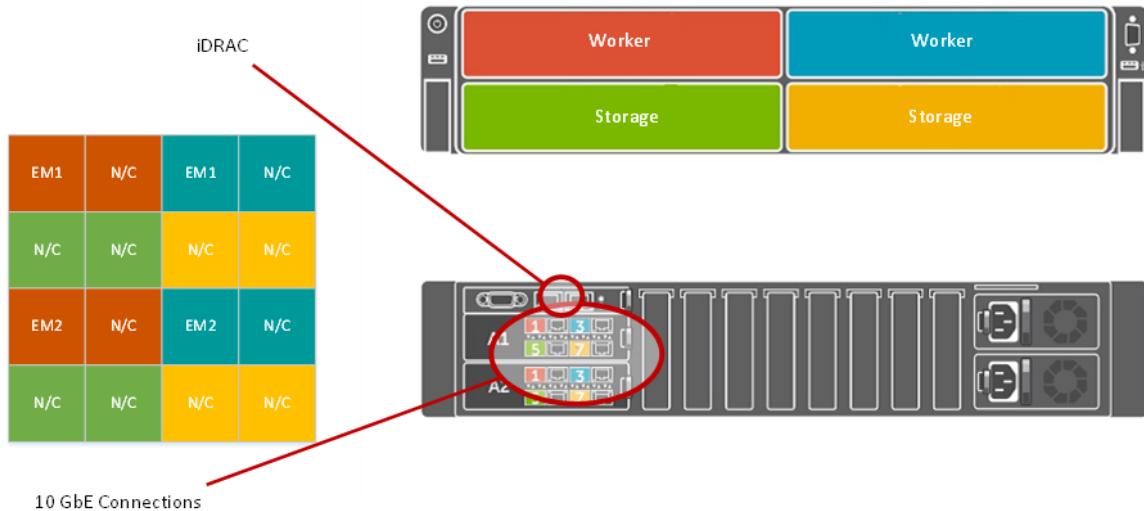


Figure 11: Dell PowerEdge FX2 Worker Chassis Network Ports

Note: The Dell PowerEdge FX2 has two iDRAC ports per chassis - an uplink port and a stacking port (STK). The uplink port is the main iDRAC port. The stacking port is only used when chassis are daisy-chained.

Table 20: Bond / Interface Cross Reference

Server Platform	Interface	Bond	Network
Dell PowerEdge R730xd	em1	bond0	Cluster Data
Dell PowerEdge R730xd	em2	bond0	Cluster Data
Dell PowerEdge R730xd	p4p1	bond1	Edge

Server Platform	Interface	Bond	Network
Dell PowerEdge R730xd	p4p2	bond1	Edge
Dell PowerEdge FX2	em1	bond0	Cluster Data
Dell PowerEdge FX2	em2	bond0	Cluster Data
Dell PowerEdge FX2	em3	bond1	Edge
Dell PowerEdge FX2	em4	bond1	Edge

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 12: Single Pod Networking Equipment on page 45 shows the single pod network configuration, with a pair of Dell Networking S4048-ON switches aggregating the pod traffic.

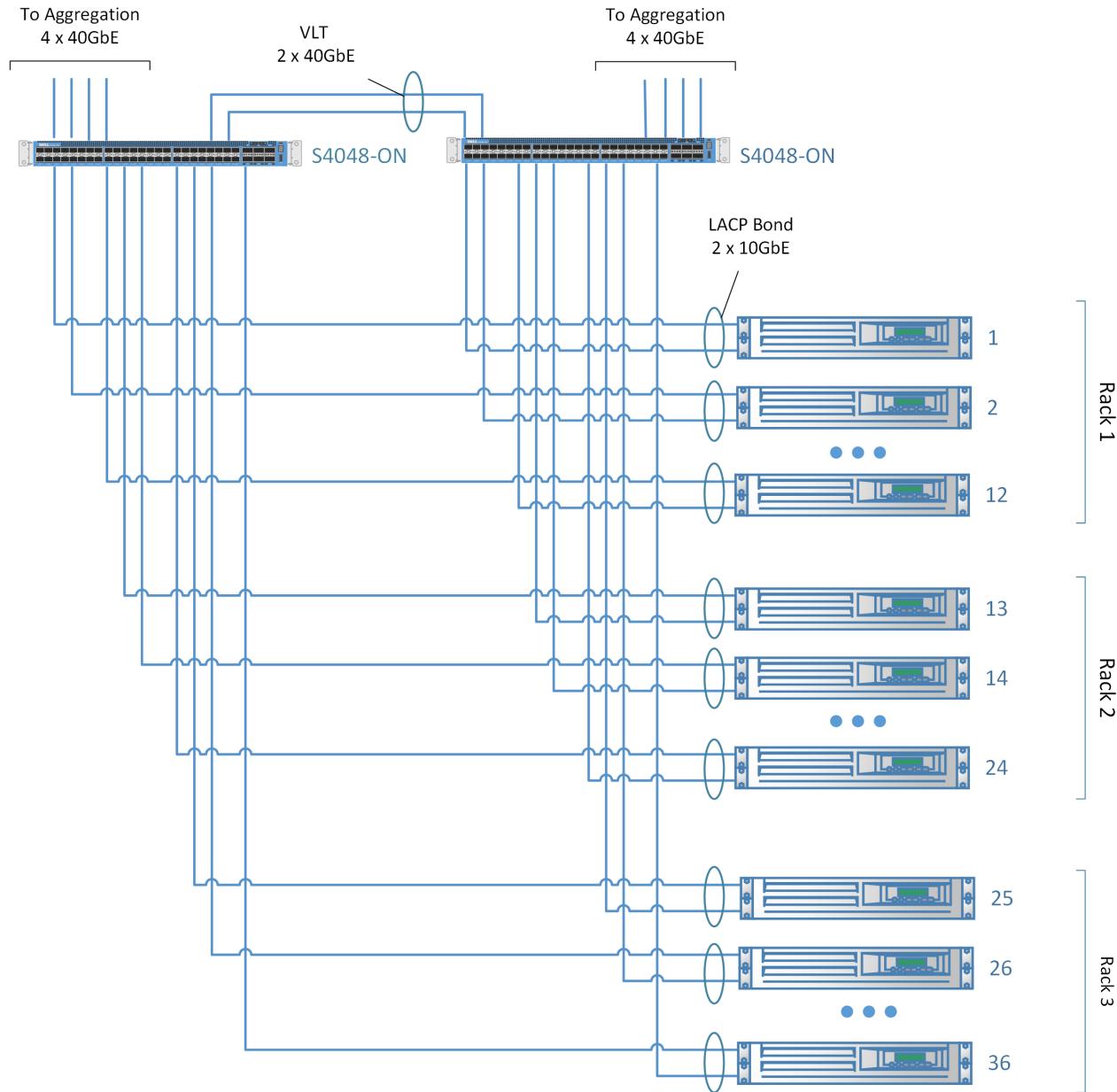


Figure 12: Single Pod Networking Equipment

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 more than one pod, the architecture uses either of the following models for aggregation switches:

- Dell Networking S6000-ON
- Dell Networking Z9100

The choice depends on the initial size and planned scaling. The Dell Networking S6000-ON is preferred for lower cost and medium scalability. This design can handle up to seven pods, as described in [Cluster Sizing](#) on page 28. The Dell Networking Z9100 is recommended for larger deployments.

Dell Networking S6000-ON Cluster Aggregation

Figure 13: Dell Networking S6000-ON Multi-pod Networking Equipment on page 46 illustrates the configuration for a multiple pod cluster using the S6000-ON 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 S6000-ONS connect to the aggregation pair, there is a collective bandwidth of 320 Gb available from each pod.

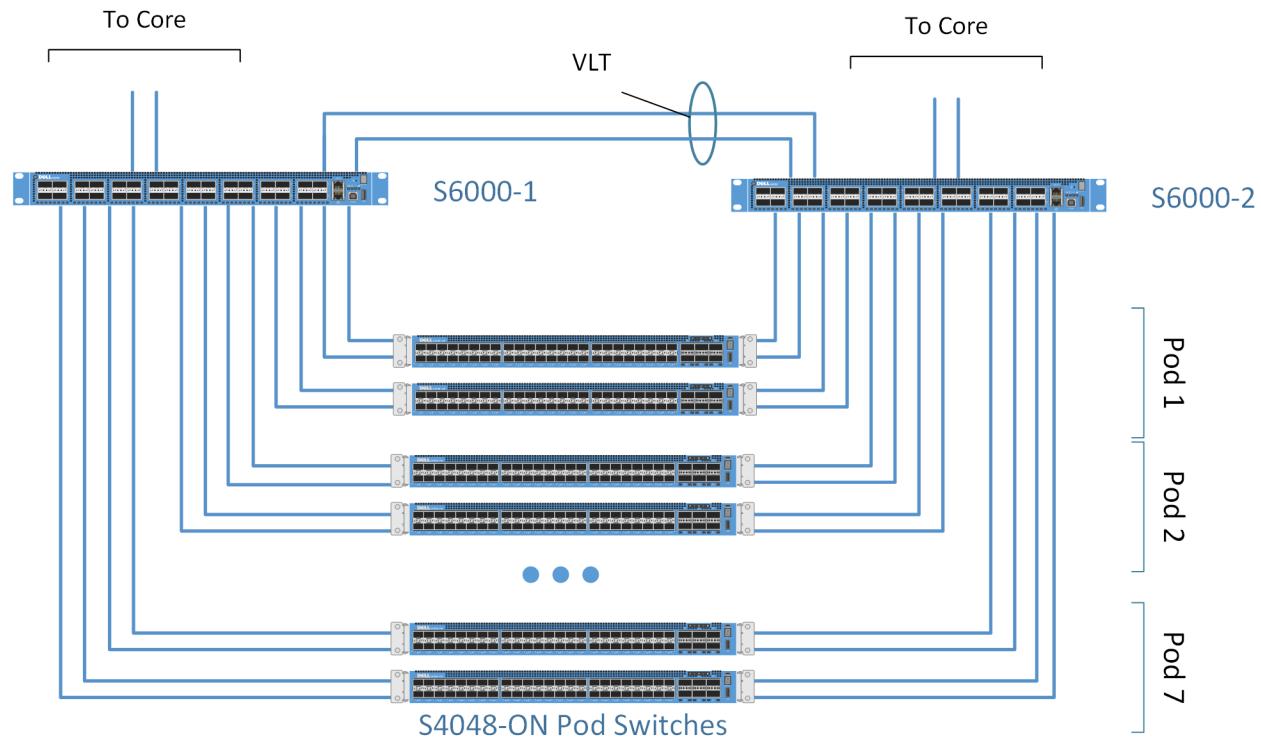


Figure 13: Dell Networking S6000-ON Multi-pod Networking Equipment

Dell Networking Z9100 Cluster Aggregation

Dell recommends the Dell Networking Z9100 core switch for:

- Larger initial deployments
- Deployments where extreme scale up is planned
- Instances where the cluster needs to be co-located with other applications in different racks

The Dell Networking Z9100 is a multi-rate 100GbE 1RU spine switch optimized for high performance, ultra-low latency data center requirements. The pod-to-pod bandwidth needed in Hadoop is best addressed by a non-blocking switch.

The Dell Networking Z9100 can provide a cumulative bandwidth of 7.4 Tbps of throughput at line-rate traffic from every port. It can be configured with up to:

- 32 ports of 100GbE (QSFP28)
- 64 ports of 50GbE (QSFP+)
- 32 ports of 40GbE (QSFP+)
- 128 ports of 25GbE (QSFP+)
- 128+2 ports of 10GbE

A straightforward modification of this architecture can use the Z9100 as an alternative to the S6000-ON 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 architecture, since there are enough pods and nodes in the cluster to minimize the impact of a network failure through 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 14: Multi-Pod View Using Dell Networking Z9100 Switches \(Based on Layer-3 ECMP\)](#) on page 47 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.

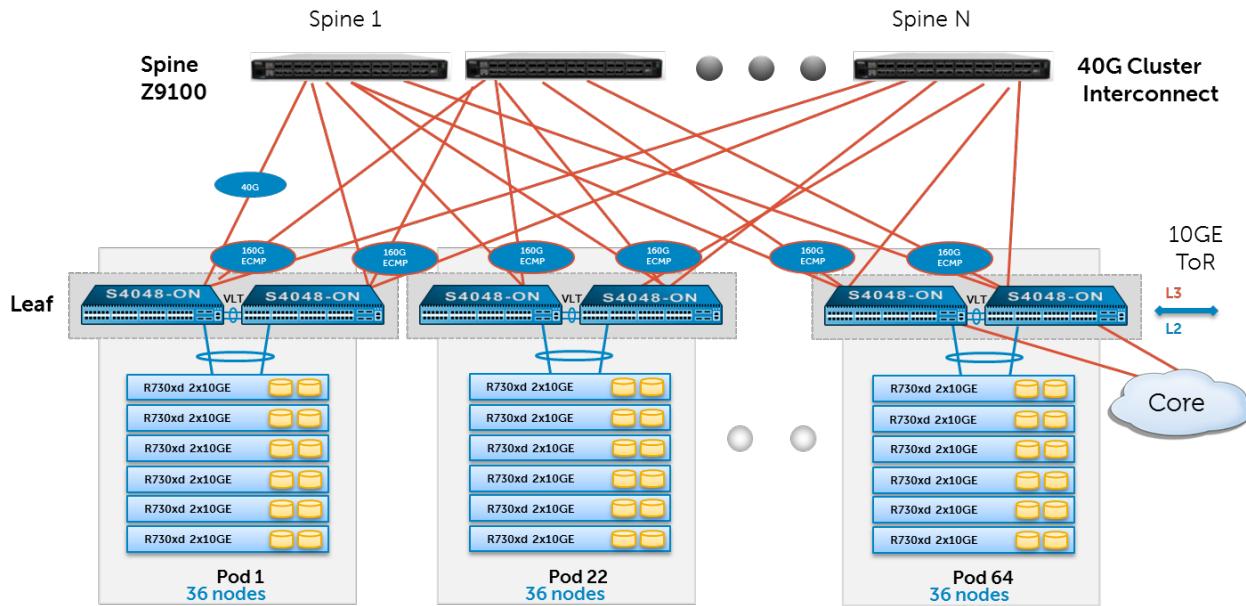


Figure 14: Multi-Pod View Using Dell Networking Z9100 Switches (Based on Layer-3 ECMP)

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, as displayed in [Figure 13: Dell Networking S6000-ON Multi-pod Networking Equipment](#) on page 46. When using the Dell Networking S6000-ON, 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 14: Multi-Pod View Using Dell Networking Z9100 Switches \(Based on Layer-3 ECMP\)](#) on page 47 represent the layer-2 and layer-3 boundaries. This document uses layer-2 as the reference up to the aggregation layer.

iDRAC Management Network

In addition to the cluster data network, a separate network is provided for cluster management - the iDRAC (or BMC) network.

The iDRAC management ports are all aggregated into a per rack Dell Networking S3048-ON switch, with dedicated VLAN. This provides a dedicated iDRAC / BMC network, for hardware provisioning and management. Switch management ports are also connected to this network.

The management switches can be connected to the core, or connected to a dedicated management network if out of band management is required.

Network Equipment Summary

[Table 21: Per Rack Network Equipment](#) on page 48, [Table 22: Per Pod Network Equipment](#) on page 48 and [Table 23: Per Cluster Aggregation Network Switches for Multiple Pods](#) on page 48 summarize the required cluster networking equipment. [Table 24: Per Node Network Cables Required – 10GbE Configurations](#) on page 48 summarizes the number of cables needed for a cluster.

Table 21: 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 22: 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 23: Per Cluster Aggregation Network Switches for Multiple Pods

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

Table 24: Per Node Network Cables Required – 10GbE Configurations

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

Chapter

5

Cloudera Enterprise Software

Topics:

- *Cloudera Manager*
- *Cloudera RTQ (Impala)*
- *Cloudera Search*
- *Cloudera BDR*
- *Cloudera Navigator*
- *Cloudera Support*

The Dell Ready Bundle for Cloudera Hadoop is based on Cloudera Enterprise, which includes Cloudera's distribution for Hadoop (CDH) 5.10 and Cloudera Manager.

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)

Apache 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 Cloudera Enterprise, 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 Cloudera Enterprise.

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 Cloudera Enterprise 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 Cloudera Enterprise.

Cloudera BDR makes it easy to configure and manage disaster recovery policies for data stored in Cloudera Enterprise. 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 Cloudera Enterprise.

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. Cloudera Navigator includes:

- Active Data Optimization (Cloudera Navigator Optimizer)
- Governance and Data Management (Cloudera Navigator including auditing, lineage, discovery, and policy lifecycle management)
- Encryption and Key Management (Cloudera Navigator Encrypt and Key Trustee).

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

- Store sensitive data in Cloudera Enterprise 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

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 25: 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

Chapter

6

Syncsort Software

Topics:

- [*Syncsort DMX-h Engine*](#)
- [*Syncsort DMX-h Service*](#)
- [*Syncsort DMX-h Client*](#)
- [*Syncsort SILQ*](#)

The data transformation functionality in the Dell Ready Bundle for Cloudera Hadoop is provided by Syncsort DMX-h. Syncsort DMX-h ETL Edition is high-performance ETL software that turns Hadoop into a robust and feature rich ETL solution, enabling users to maximize the benefits of MapReduce without compromising on capabilities, ease of use, and typical use cases of conventional data integration tools.

Syncsort DMX-h Engine

DMX-h is the only tool that runs ETL processes natively within Hadoop, via a pluggable sort enhancement (JIRA MAPREDUCE-2454), contributed by Syncsort and now part of Apache Hadoop. Other tools generate code (i.e., Java, Pig, HiveQL) that adds performance overhead and can become difficult to maintain and tune.

DMX-h is not a code generator. Instead, Hadoop MapReduce automatically invokes the highly-efficient DMX-h engine at runtime, which executes natively on all nodes as an integral part of the Hadoop framework. Once deployed, DMX-h automatically optimizes resource utilization – CPU, memory and I/O – on each node to deliver the highest levels of performance, with no tuning required. Simply stated, higher performance and efficiency per node means you can process more data in less time, with fewer servers.

Syncsort DMX-h Service

The DMX Service runs on an Edge Node in the Hadoop Cluster, and coordinates access to the DMX engine running under Hadoop. The Syncsort Client connects to the DMX Service to initiate jobs.

Syncsort DMX-h Client

The Syncsort DMX-h client consists of an intuitive graphical interface that allows users to design, execute and control data integration jobs.

DMX-h enables people with a much broader range of skills - not just MapReduce programmers - to create ETL tasks that execute within the MapReduce framework, replacing complex Java, Pig, or HiveQL code with a powerful, easy to use graphical development environment. DMX-h makes it easier to develop, maintain, and re-use applications running on Hadoop via comprehensive built-in transformations and built-in metadata capabilities, for greater reusability, impact analysis, and data lineage.

Syncsort SILQ

SILQ is a web based utility that helps convert complex 'ELT' style SQL jobs to DMX-h jobs running in Hadoop.

SILQ can read multiple SQL dialects, including BTEQ, NZ SQL, PL/SQL, and ANSI SQL-92. It generates graphical data flows, provides best-practices to develop DMX-h jobs , and automatically generates ETL jobs to run natively on Hadoop.

Chapter

7

Deployment Methodology

A suggested deployment workflow is documented in the [Dell Ready Bundle for Cloudera Hadoop Deployment Guide](#), which is a complement to the [Dell Ready Bundle for Cloudera Hadoop Architecture Guide](#).

Appendix

A

Physical Rack Configuration - Dell PowerEdge R730xd

Topics:

- [*Dell PowerEdge R730xd Single Rack Configuration*](#)
- [*Dell PowerEdge R730xd Initial Rack Configuration*](#)
- [*Dell PowerEdge R730xd Additional Pod Rack Configuration*](#)

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.

Dell PowerEdge R730xd Single Rack Configuration

Table 26: Single Rack Configuration – Dell 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	Empty
29	
28	Edge01: Dell PowerEdge R730xd
27	
26	Active Name Node:Dell PowerEdge R730xd
25	
24	Standby Name Node: Dell PowerEdge R730xd
23	
22	HA Node: Dell PowerEdge R730xd
21	
20	Empty
19	
18	Empty
17	
16	R1- Chassis08: Dell PowerEdge R730xd
15	
14	R1- Chassis07: Dell PowerEdge R730xd
13	
12	R1- Chassis06: Dell PowerEdge R730xd
11	

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

Dell PowerEdge R730xd Initial Rack Configuration

Table 27: Initial Pod Rack Configuration – Dell 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	Active Name Node:Dell PowerEdge R730xd	Edge01: Dell PowerEdge R730xd	R3 - Switch 1: Dell Networking S6000-ON
34			R3 - Switch 2: Dell Networking S6000-ON
33 32	Empty	Standby Name Node: Dell PowerEdge R730xd	HA Node: Dell PowerEdge R730xd
31 21	Empty	Empty	Empty

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

Dell PowerEdge R730xd Additional Pod Rack Configuration

Table 28: Additional Pod Rack Configuration – Dell 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

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

Appendix B

Physical Chassis Configuration - Dell PowerEdge FX2

Topics:

- *Physical Chassis Configuration - Dell PowerEdge FX2*

This appendix describes chassis configurations for the Dell PowerEdge FX2.

Physical Chassis Configuration - Dell PowerEdge FX2

Table 29: Infrastructure Chassis Configuration – Dell PowerEdge FX2

FC630 (Infrastructure)	FC630 (Worker)
FD332 (Single PERC)	FD332 (Dual PERC)

Table 30: Worker Chassis Configuration – Dell PowerEdge FX2

FC630 (Worker)	FC630 (Worker)
FD332 (Dual PERC)	FD332 (Dual PERC)

Appendix C

Physical Rack Configuration - Dell PowerEdge FX2

Topics:

- *Dell PowerEdge FX2 Single Rack Configuration*
- *Dell PowerEdge FX2 Initial and Second Pod Rack Configuration*

This appendix contains suggested rack layouts for single rack, single pod, and multiple pod installations. These layouts are based on power capacity of approximately 20 kW per rack. Actual rack layouts will vary depending on power, cooling, and loading constraints.

Dell PowerEdge FX2 Single Rack Configuration

Table 31: Single Rack Configuration – Dell PowerEdge FX2

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	Empty
34	Empty
33	Empty
29	
28	Empty
27	
26	Empty
25	
24	R1- Chassis12: Dell PowerEdge FX2 Worker -
23	Worker Node and Worker Node
22	R1- Chassis11: Dell PowerEdge FX2 Worker -
21	Worker Node and Worker Node
20	R1- Chassis10: Dell PowerEdge FX2 Worker -
19	Worker Node and Worker Node
18	R1- Chassis09: Dell PowerEdge FX2 Worker -
17	Worker Node and Worker Node
16	R1- Chassis08: Dell PowerEdge FX2 Worker -
15	Worker Node and Worker Node
14	R1- Chassis07: Dell PowerEdge FX2 Worker -
13	Worker Node and Worker Node

RU	RACK1
12	R1- Chassis06: Dell PowerEdge FX2 Worker - Worker Node and Worker Node
11	
10	R1- Chassis05: Dell PowerEdge FX2 Worker - Worker Node and Worker Node
9	
8	R1- Chassis04: Dell PowerEdge FX2 Infrastructure - Edge Node and Worker Node
7	
6	R1- Chassis03: Dell PowerEdge FX2 Infrastructure - HA Node and Worker Node
5	
4	R1- Chassis02: Dell PowerEdge FX2 Infrastructure - Standby Name Node and Worker Node
3	
2	R1- Chassis01: Dell PowerEdge FX2 Infrastructure - Active Name Node and Worker Node
1	

Dell PowerEdge FX2 Initial and Second Pod Rack Configuration

Table 32: Initial and Second Pod Rack Configuration– Dell PowerEdge FX2

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	Empty	R2 - Switch 1: Dell Networking S6000-ON	R3 - Switch 1: Dell Networking S6000-ON
34	Empty	R2 - Switch 2: Dell Networking S6000-ON	R3 - Switch 2: Dell Networking S6000-ON
33	Empty	Empty	Empty
25			

RU	RACK1	RACK2	RACK3
24 23	R1- Chassis12: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R2- Chassis12: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R3- Chassis12: Dell PowerEdge FX2 Worker - Worker Node and Worker Node
22 21	R1- Chassis11: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R2- Chassis11: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R3- Chassis11: Dell PowerEdge FX2 Worker - Worker Node and Worker Node
20 19	R1- Chassis10: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R2- Chassis10: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R3- Chassis10: Dell PowerEdge FX2 Worker - Worker Node and Worker Node
18 17	R1- Chassis09: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R2- Chassis09: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R3- Chassis09: Dell PowerEdge FX2 Worker - Worker Node and Worker Node
16 15	R1- Chassis08: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R2- Chassis08: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R3- Chassis08: Dell PowerEdge FX2 Worker - Worker Node and Worker Node
14 13	R1- Chassis07: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R2- Chassis07: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R3- Chassis07: Dell PowerEdge FX2 Worker - Worker Node and Worker Node
12 11	R1- Chassis06: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R2- Chassis06: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R3- Chassis06: Dell PowerEdge FX2 Worker - Worker Node and Worker Node
10 9	R1- Chassis05: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R2- Chassis05: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R3- Chassis05: Dell PowerEdge FX2 Worker - Worker Node and Worker Node
8 7	R1- Chassis04: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R2- Chassis04: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R3- Chassis04: Dell PowerEdge FX2 Worker - Worker Node and Worker Node
6 5	R1- Chassis03:Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R2- Chassis03: Dell PowerEdge FX2 Worker - Worker Node and Worker Node	R3- Chassis03: Dell PowerEdge FX2 Worker - Worker Node and Worker Node
4 3	R1- Chassis02: Dell PowerEdge FX2 Infrastructure - HA Node and Worker Node	R2- Chassis02: Dell PowerEdge FX2 Infrastructure - Edge Node and Worker Node	R3- Chassis02: Dell PowerEdge FX2 Worker - Worker Node and Worker Node

RU	RACK1	RACK2	RACK3
2	R1- Chassis01: Dell PowerEdge FX2 Infrastructure - Active Name Node and Worker Node	R2- Chassis01: Dell PowerEdge FX2 Infrastructure - Standby Name Node and Worker Node	R3- Chassis01: Dell PowerEdge FX2 Worker - Worker Node and Worker Node
1			

Appendix

D

Update History

Topics:

- [*Changes in Version 5.10*](#)

This appendix lists changes made to this document since the last published release.

Changes in Version 5.10

The following changes have been made since the 5.9 release:

- Update for Cloudera Enterprise 5.10
- Renamed to Dell Ready Bundle for Cloudera Hadoop.
- Renamed 'Data Nodes' to 'Worker Nodes' to be consistent with their usage and common industry terminology.

Appendix

E

References

Topics:

- *About Cloudera*
- *About Syncsort*
- *To Learn More*

Additional information can be obtained at <http://www.dell.com/en-us/work/learn/software-platforms-hadoop>.

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

About Cloudera

Cloudera is a key contributor to the Apache Hadoop project. The Cloudera Distribution for Apache Hadoop (CDH) is a highly-scalable open source platform for high-volume data management and analytics. CDH integrates with existing enterprise IT infrastructure, enabling data engineers and data scientists to quickly and easily develop and deploy Hadoop applications in a cost-efficient manner.

The Dell servers in this Architecture Guide are Cloudera Certified.



About Syncsort

Syncsort creates software that allows enterprises to collect, integrate, sort, and distribute large amounts of data quickly, with reduced resources usage, in a cost-effective manner.

Dell is a Syncsort-certified Technology Alliance Partner.



To Learn More

For more information on the Dell Ready Bundle for Cloudera Hadoop, visit <http://www.dell.com/en-us/work/learn/software-platforms-hadoop>.

Copyright © 2011-2017 Dell Inc. or its subsidiaries. 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 Inc. and its affiliates cannot be responsible for errors or omissions in typography or photography. Dell Inc.'s Terms and Conditions of Sales and Service apply and are available on request. Dell Inc. service offerings do not affect consumer's statutory rights.

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