Dell PowerEdge FX

Recommended Configurations

A Dell Technical White Paper

Kris Applegate
Solution Architect
Dell Solution Centers

Bryan Gilcrease
Solution Architect
Dell Solution Centers

Robert Wilbert
Solution Architect
Dell Solution Centers

Executive Summary

Data volumes are exploding and resources are becoming more constrained. Trying to get a handle on both of these rapidly evolving problems can be a challenge. Implementing Hadoop on Dell's PowerEdge FX2 platform enables running today's most data-intensive workloads inside a flexible converged-infrastructure chassis. This allows for more enhanced flexibility, simplified networking, and denser compute environments. Dell's long history with Hadoop and proven Big Data solutions make it a strong platform partner on which to deliver these solutions for the enterprise.

August 2015

Table of contents

Tables

Figures

Dell Solution Centers

Complex solutions such as Hadoop and other Big Data technologies often require the assistance from subject-matter experts. Along with a knowledgeable sales force, Dell has created a global team of solution architects that can help have these conversations – all free of charge.

The Dell Solution Centers are a global network of connected labs that allow Dell to help customers architect, validate and build solutions. With multiple sites in every region, the Dell Solution Centers help customers understand anything from Dell's various hardware and software offerings to more complex solutions. Each Dell Solution Center is staffed with experienced subject-matter experts that can provide customers with a 30-60 minute technical briefing on a particular topic, a half-day architectural design workshop, or a proof-of-concept that allows customers to define and validate their solution. Customers may engage with their account team to take advantage of one of these free services.

Background

Hadoop

Hadoop is an open-source software platform that enables collecting, managing, storing and processing ever-increasing amounts of structured, unstructured, and semi-structured data. Hadoop's linearly-scalable processing and storage capabilities make it a popular platform for building Big Data solutions to support modern workloads. Companies like Cloudera, Hortonworks, and MapR offer their own Hadoop distributions, as well as training and support services to enterprises that allow them to build production-ready Big Data solutions for a variety of use-cases.

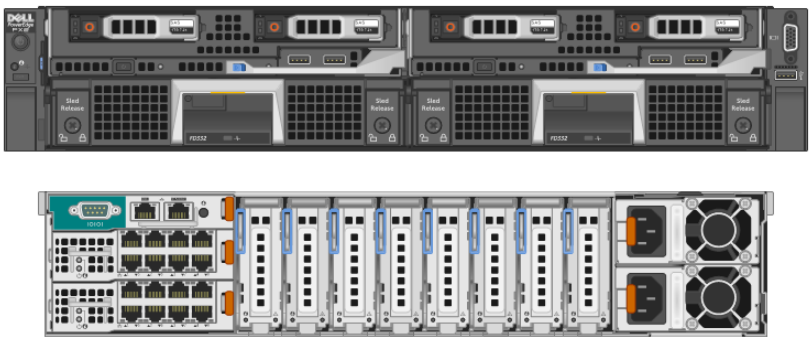
Dell | Cloudera Apache Hadoop Solutions

Dell offers several Hadoop solutions built around Dell PowerEdge servers and Cloudera's distribution of Hadoop. These solutions couple massive amounts of local storage with multiple processing cores to form an excellent building block for Hadoop's scale-out architecture. Whether getting started with a small five-node QuickStart Hadoop cluster, implementing Dell's general-purpose Hadoop Reference Architecture, offloading your ETL processes to Hadoop, or deploying Hadoop clusters at-scale with the Dell In-Memory Appliance for Cloudera Enterprise, Dell offers proven solutions to meet virtually every workload or use-case.

PowerEdge FX

Dell's modular PowerEdge FX2 platform allows customers to mix and match compute, storage, and networking "blocks" within a single 2U chassis to meet workload-specific needs. Based off the proven Dell shared-infrastructure designs of the PowerEdge M1000e blade platform and the PowerEdge VRTX, the FX2 platform plays well into the customer's converged infrastructure needs.

The PowerEdge FX2 platform supports combinations of quarter-width PowerEdge FC430 server blocks, half-width PowerEdge FC630 server blocks, full-width PowerEdge FC830 server blocks, or half-width PowerEdge FD332 storage blocks. The following image shows a PowerEdge FX2 chassis with two PowerEdge FC630 server blocks, each attached to a single PowerEdge FD332 storage block.



This guide will propose some options for selecting the optimal configuration for Hadoop on the PowerEdge FX2 platform. While the PowerEdge FX2 platform supports numerous configurations, only a few configuration options are covered in this document. Customers are encouraged to engage the Dell Solution Centers to further discuss the various PowerEdge FX2 platform options that may work best for their specific workloads and use-cases.

Recommended Configurations

PowerEdge FX2 Infrastructure Nodes

Most Hadoop clusters require additional servers besides the data nodes to serve as infrastructure nodes (name nodes, edge node, HA node, etc.). The PowerEdge FX2 platform offers a lot of flexibility for infrastructure nodes since the infrastructure nodes are not as performance sensitive as the data nodes.

Component	PowerEdge FX2 Infrastructure Nodes
Chassis	(1) PowerEdge FX2S chassis
Networking (per chassis)	(2) PowerEdge FX2 10GbE pass-through modules
Server Blocks (per chassis)	(4) PowerEdge FC430 with on-board 10GbE NIC
Storage Blocks (per chassis)	(2) PowerEdge FD332 with dual PERC in RAID mode
Processors (per server block)	(2) Intel Xeon E5-2650 v3, 2.3GHz, 10-core
Memory (per server block)	128GB DDR4 RDIMMs (8x16GB) @ 2133MT/s
Disks (per server block)	(2) Intel S3610 1.8" 200GB uSATA mixed-use SSD
NICs (per server block)	On-board 10GbE NIC Intel X710 dual-port 10GbE PCIe NIC (edge node)
Disks (per storage block)	(16) 1TB 7.2K 2.5" NL-SAS

Table 1. PowerEdge FX2 Infrastructure Nodes

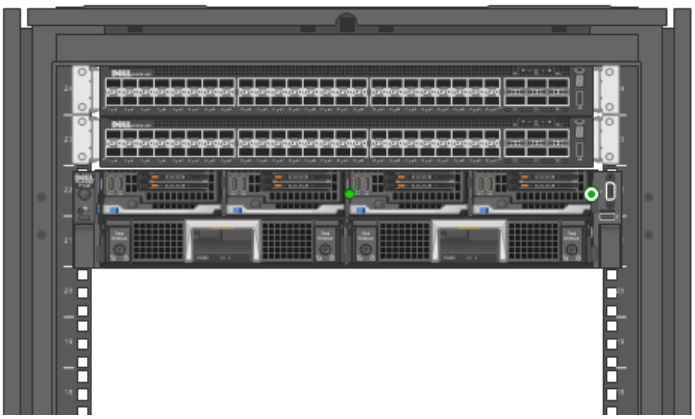


Figure 1. PowerEdge FX2 Infrastructure Nodes Rack Diagram

PowerEdge FX2 General-Purpose Data Nodes

Many Hadoop deployments run a wide-variety of workloads – such as streaming workloads, batch processing and ad-hoc analysis – each with their own execution characteristics and service-level agreements (SLAs). Selecting the appropriate infrastructure to run all of those disparate workloads may prove challenging.

The PowerEdge FC630 server block, in conjunction with the PowerEdge FD332 storage block, provide a flexible and robust building block for general-purpose Hadoop data nodes. The PowerEdge FC630 server block allows greater flexibility for processors and memory configurations, and supports three different chassis configurations. The PowerEdge FC630 server block supports either two 2.5" hard drives, two PCIe solid-state drives (SSDs), or eight 1.8" SSDs. With recent versions of Hadoop enabling heterogeneous storage tiers, the eight 1.8" SSDs on the PowerEdge FC630 server block may be used as a flash storage tier with the spinning disks on the PowerEdge FD332 storage block acting as the standard disk storage tier.

Component	PowerEdge FX2 General-Purpose Data Nodes
Chassis	(1) PowerEdge FX2S chassis
Networking (per chassis)	(2) PowerEdge FX2 10GbE pass-through modules
Server Blocks (per chassis)	(2) PowerEdge FC630
Storage Blocks (per chassis)	(2) PowerEdge FD332 with dual PERC in RAID mode
Processors (per server block)	(2) Intel Xeon E5-2670 v3, 2.3GHz, 12-core
Memory (per server block)	128GB DDR4 RDIMM (4x32GB) @ 2133MT/s
Disks (per server block)	(2) 300GB 10K 2.5" NL-SAS
NICs (per server block)	(1) Intel X710 dual-port 10GbE NDC
Disks (per storage block)	(16) 1TB 7.2K 2.5" NL-SAS

Table 2. PowerEdge FX2 General-Purpose Data Nodes

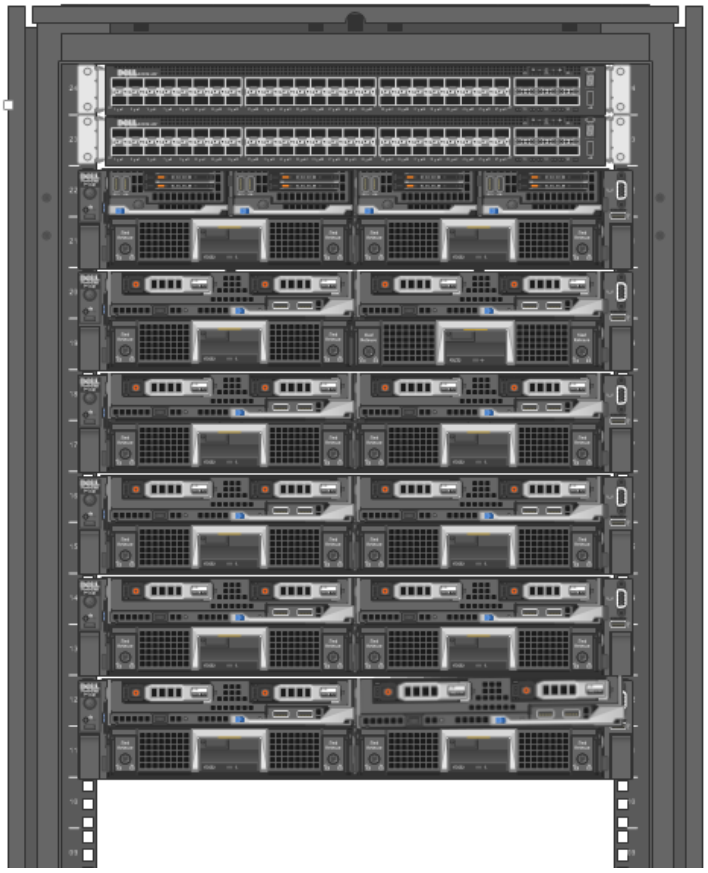


Figure 2. PowerEdge FX2 General-Purpose Data Nodes Rack Diagram

PowerEdge FX2 Compute-Dense Data Nodes

For compute-intensive Hadoop workloads, the PowerEdge FC430 storage block provides the best processor density within the PowerEdge FX2 platform. For storage, two PowerEdge FC430 server blocks can attach to a single PowerEdge FD332 storage block. The PowerEdge FD332 storage block supports up to sixteen 2.5" hard drives and up to two Dell PowerEdge RAID controllers (PERC) with up to eight hard drives attached to each storage controller. In this configuration, each PowerEdge FC430 server block connects to eight of the sixteen drives within a single PowerEdge FD332 storage block via a dedicated storage controller.

Component	PowerEdge FX2 Compute-Dense Data Nodes
Chassis	(1) PowerEdge FX2S chassis
Networking (per chassis)	(2) PowerEdge FX2 10GbE pass-through modules
Compute Blocks (per chassis)	(4) PowerEdge FC430 with on-board 10GbE NIC
Storage Blocks (per chassis)	(2) PowerEdge FD332 with dual PERC
Processors (per server block)	(2) Intel Xeon E5-2670 v3, 2.3GHz, 12-core
Memory (per server block)	256GB DDR4 RDIMMs (8x32GB) @ 2133MT/s
Disks (per server block)	(2) Intel S3610 1.8" 200GB uSATA mixed-use SSD
NICs (per server block)	On-board 10GbE NIC
Disks (per storage block)	(16) 1TB 7.2K 2.5" NL-SAS

Table 3. PowerEdge FX2 Computer-Dense Data Nodes

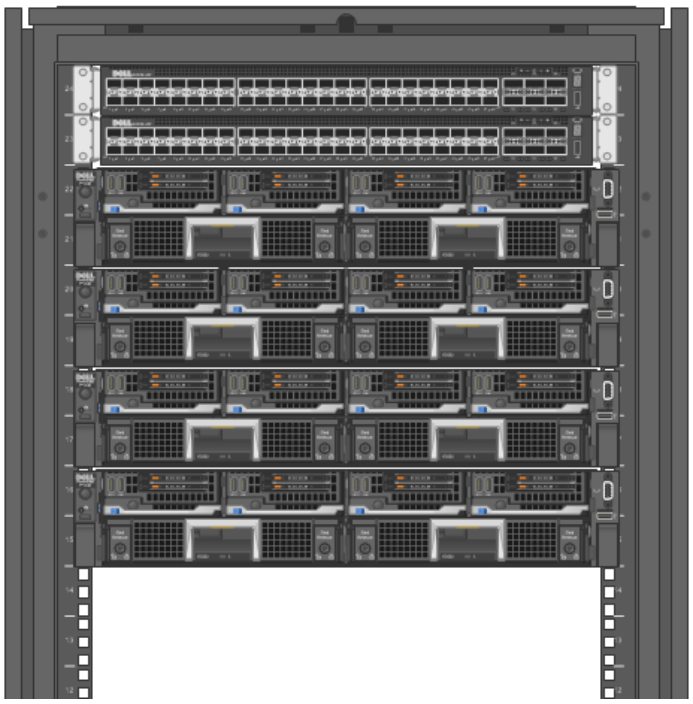


Figure 3. PowerEdge FX2 Compute-Dense Data Nodes Rack Diagram

Networking Considerations

Dell's FX2 chassis comes with multiple input/output (I/O) configuration options including a network pass-through module or an I/O network aggregator, as well as eight low-profile PCIe slots.

While the PowerEdge FC430 server blocks includes a choice of either an onboard dual-port 1GbE or 10GbE network adapter, the PowerEdge FC630 server block supports a Network Daughter Card (NDC) that provides additional networking options, such as a quad-port 10GbE network adapter for up to 40Gbps network throughput. Other network options include adding additional PCIe network adapters and mapping those network adapters to individual PowerEdge FC server blocks.

Depending on the workload and the specific PowerEdge FX2 configuration, it may be beneficial to utilize network pass-through modules and route all network traffic through a top-of-rack (TOR) switch, or it may be more efficient to keep east/west network traffic among the server blocks within a PowerEdge FX2 chassis through the use of an I/O aggregator module.

For most Hadoop workloads, the I/O pass-through module is recommended. The network pass-through module "passes through" two 10GbE ports from each server block to the TOR switch, giving every server block up to 20Gbps throughput to the TOR switch. On the PowerEdge FX2 chassis where the Hadoop administrator and edge nodes resides, an Intel X710 PCIe network adaptor should be added to connect those blocks to a separate IT/edge network.

Since the bottleneck on most Hadoop clusters tends to be the network, Dell recommends data center switches – such as the Dell Networking S4048 and S6000 switches – that utilize deep, per-port packet buffers. Campus switches – such as the Dell Networking N-series switches – utilize a single packet buffer that is shared across all of the network ports on the switch. Without deep, per-port packet buffers, a single Hadoop data node can easily consume the entire packet buffer on the switch and limit the network throughput of the other Hadoop data nodes.

The choice and configuration of the TOR switch depends greatly on the size and density of the Hadoop cluster. For most Hadoop clusters, the Dell Networking S4048 switch works well as the TOR switch. Dell recommends using two Dell Networking S4048 switches configured for high availability using virtual-link trunking (VLT). Larger or denser Hadoop clusters may require the Dell Networking S6000 switch as the TOR switch. Additionally, the Dell Networking S6000 switch should be used to aggregate multi-rack Hadoop clusters.

Server management is performed trough the Chassis Management Controller (CMC) on each PowerEdge FX2 chassis. The CMC enables chassis management as well as access to the integrated Dell Remote Access Controllers (iDRAC) on each server block within the chassis. Each PowerEdge FX2 chassis contains a single 1GbE port for both the CMC and iDRACs and may be connected to a Dell Networking S55 switch or any other capable 1GbE switch.

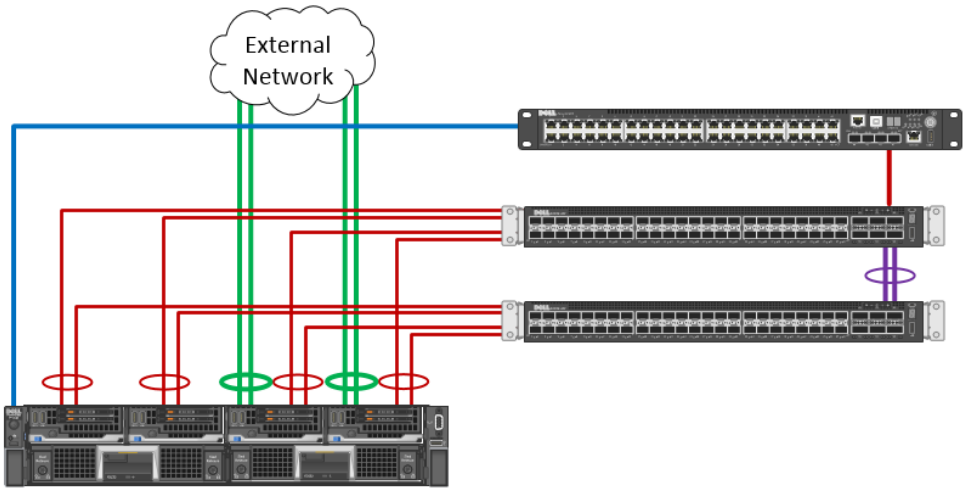


Figure 4. Infrastructure Node Configuration with (4) PowerEdge FC430 Server Blocks and (2) PowerEdge FD332 Storage Blocks

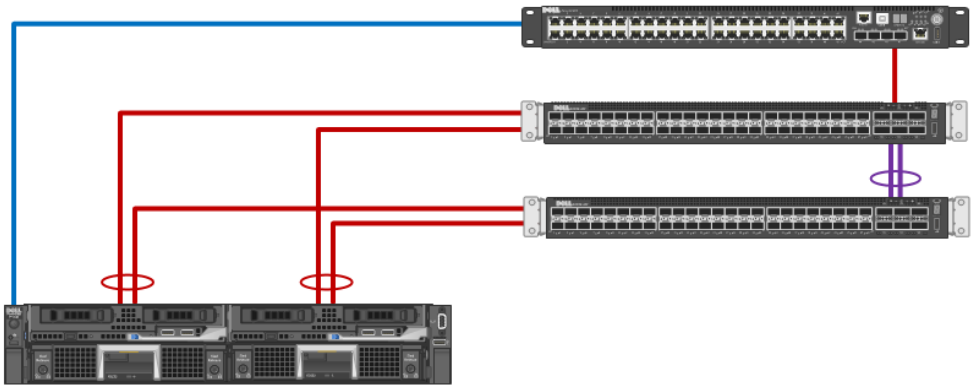


Figure 5. General-Purpose Data Node Configuration with (2) PowerEdge FC630 Server Blocks and (2) PowerEdge FD332 Storage Blocks

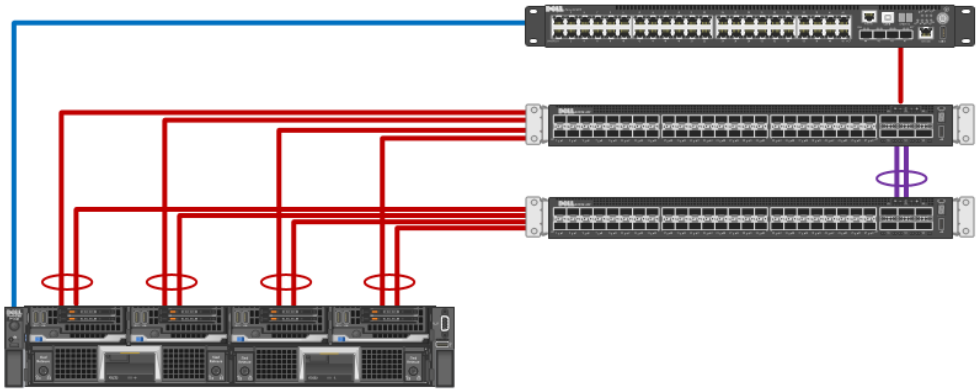


Figure 6. Compute-Dense Configuration with (4) PowerEdge FC430 Server Blocks and (2) PowerEdge FD332 Storage Blocks

Power Consumption Considerations

As platform compute and storage densities increase, so do the power requirements per rack. With the PowerEdge R730/R730XD platform, up to twenty Hadoop nodes can fit within a standard 42U rack. With the PowerEdge FX2 platform, up to eighty Hadoop nodes (or more) can fit into the same space. Although Dell engineers the PowerEdge server line to maximize power efficiency, greater server densities require additional power per the same rack unit space. Additionally, the PowerEdge FX2 platform consumes some additional power for the shared networking and system management. As a result, the PowerEdge FX2 system consumes slightly more power than a similarly equipped standalone system or a share-nothing, hyper-converge platform such as the PowerEdge C6320.

When deploying a Hadoop cluster on the PowerEdge FX2 platform, it is important to consider the power requirements for the PowerEdge FX2 configuration along with the available power at the rack and data center levels. The following table compares both the per-chassis and per-node power consumption across multiple workloads between similarly configured PowerEdge FX2 and PowerEdge R720XD Hadoop data nodes.[1] All power consumption values were calculated using the Dell Energy Smart Solution Advisor (ESSA).[2] While actual power consumption will vary from one workload to another, the Dell Solution Centers have found that, on average, actual power consumption ranges from 10-20% less than what ESSA calculates for similar workloads. However, the power consumption values in ESSA should be used for any power and capacity planning.

	8 Data Drive Configurations			16 Data Drive Configurations		24 Data Drive Configuration
	Hardware Configuration					
Platform	**FC430 + FD332**	**FC630 + FD332**	**R730XD**	**FC630 + FD332**	**R730XD**	**R730XD**
Processors	Dual Intel Xeon E5-2670 v3 (2.3GHz, 12-core, 120W)					
Memory	256GB DDR4 RDIMM (8x32GB)					
RAID Controller	Onboard SATA + Dual FD332 PERC	PERC H330 + Single FD332 PERC	PERC H730P	PERC H330 + Dual FD332 PERC	PERC H730P	
Storage (OS)	(2) 200GB SSD SATA	(2) 300GB 10K SAS		(2) 300GB 10K SAS		
Storage (Data)	(8) 1TB 7.2K NL-SAS			(16) 1TB 7.2K NL-SAS		(24) 1TB 7.2K NL-SAS
Networking	Onboard dual-port 10GbE	Intel dual-port X710 10GbE	Intel dual-port X710 10GbE + I350 1GbE	Intel dual-port X710 10GbE	Intel dual-port X710 10GbE + I350 1GbE	
Power Supplies	Dual 1600W (FX2 chassis)		Dual 1100W	Dual 1600W (FX2 chassis)	Dual 1100W	
Rack Density	4 nodes per 2U	2 nodes per 2U	1 node per 2U	2 nodes per 2U	1 node per 2U	
	Power Consumption per Chassis					
Idle	627W	497W	211W	587W	266W	321W
50% Transactional	1,373W	1,120W	303W	1,208W	357W	413W
100% Transactional	1,906W	1,365W	435W	1,452W	488W	544W
Computational	2,091W	1,372W	439W	1,459W	492W	548W
Memory Intensive	1,787W	1,378W	432W	1,445W	485W	541W
	Power Consumption per Node					
Idle	157W	249W	211W	294W	266W	322W
50% Transactional	343W	560W	303W	604W	358W	414W
100% Transactional	477W	683W	435W	726W	490W	546W
Computational	523W	686W	439W	730W	494W	549W
Memory Intensive	447W	689W	432W	723W	487W	542W

[1] Power consumption values calculated assuming 220V AC input voltage and 77°F data center temperature

[2] Dell Energy Smart Solution Advisor tool – http://essa.us.dell.com/DellStarOnline/DCCP.aspx

Failure Domain Considerations

Hadoop can tolerate a relatively high degree of node failures when each data node resides on a single physical server. One of the realities with hyper-converged infrastructure, such as the PowerEdge FX2 platform, is that a higher number of Hadoop nodes become impacted when a single chassis goes offline. While the PowerEdge FX2 features redundancy throughout the chassis to prevent total chassis failure, any planned or unplanned event on a single chassis may impact multiple Hadoop nodes.

Hadoop itself provides resiliency at the infrastructure level by spreading infrastructure services across multiple infrastructure nodes. While it is possible to place all of the infrastructure nodes on server blocks within a single PowerEdge FX2 chassis, in the event that that chassis goes offline, the entire Hadoop cluster will go offline as well. To increase the resiliency for the Hadoop cluster, it is recommended to spread the infrastructure nodes across multiple PowerEdge FX2 chassis.

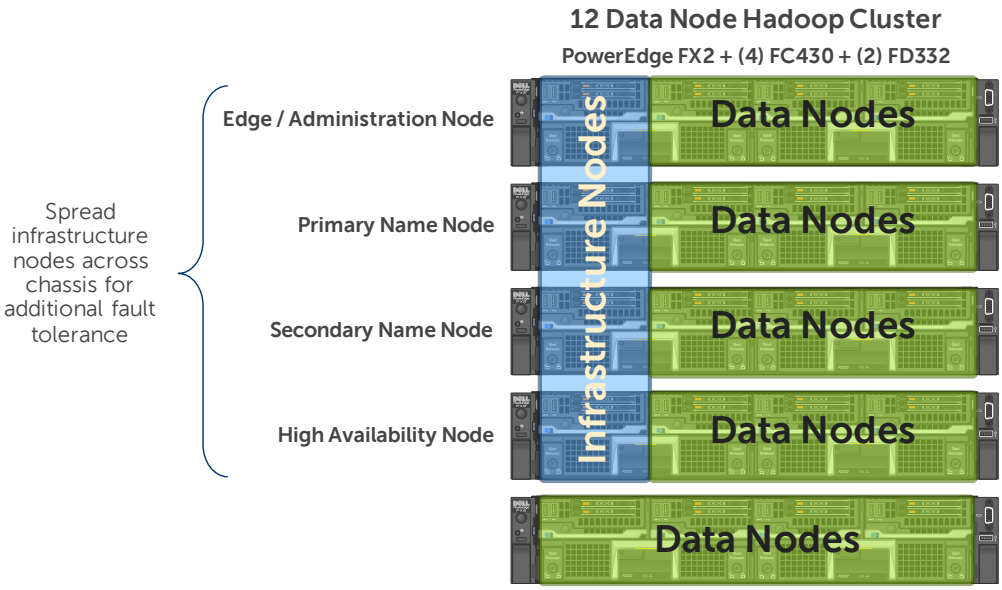


Figure 7. PowerEdge FX2 Hadoop Cluster Configuration for Infrastructure Node Resiliency

To improve data durability, Hadoop may be configured so that data is spread across multiple, logical racks. In the event that an entire rack goes offline, copies of the data are still available on data nodes within another rack. By default, Hadoop creates three copies (or replicas) of each data block within the cluster. Hadoop places the first copy on a random data node on the rack. Hadoop then places the second replica on a random data node in the same rack, and places the third replica on a different data node within another rack. If the entire cluster resides on the same logical rack (e.g. no rack awareness configured), Hadoop will place all three replicas on three different data nodes, selected randomly.

Without rack awareness, it may be possible that all three replicas reside on the same PowerEdge FX2 chassis. Data loss would occur if that particular chassis goes offline. To prevent potential data loss, it is recommended to group one or more PowerEdge FX2 chassis into separate, logical racks even if all of the PowerEdge FX2 chassis reside within the same physical rack.