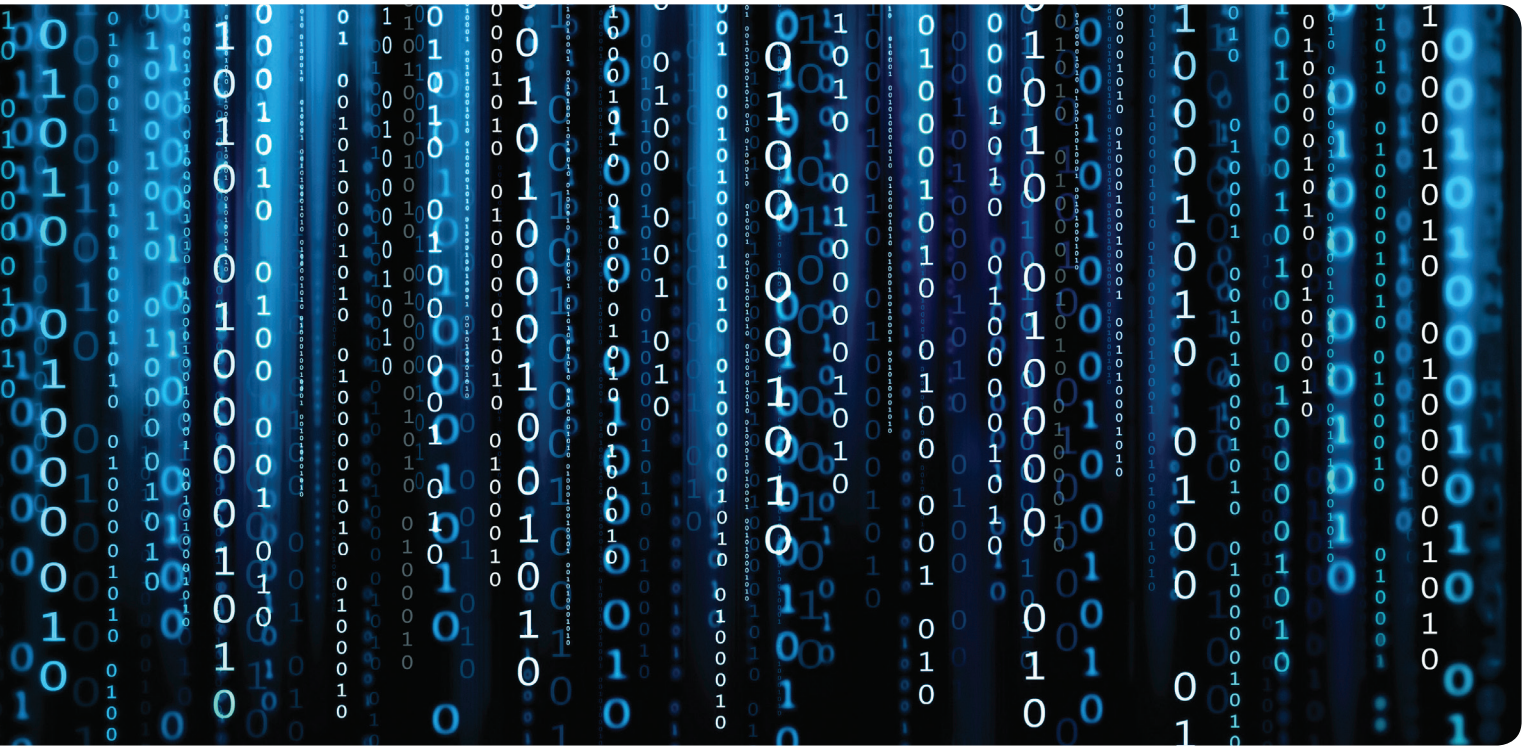
High-Level Overview

Recommended Configurations for Dell PowerEdge Servers Running Splunk

A Dell Big Data White Paper

By Armando Acosta, Hadoop Planning and Product Manager



Executive summary

This white paper provides a high-level overview of recommend configurations for Dell™ PowerEdge™ servers running Splunk Enterprise. The intended audiences for this document are customers and solution architects looking for information on selecting and configuring servers for Splunk Enterprise installations. The document focuses only on server configuration; it does not focus on architecture topologies, performance tuning or best practices for Splunk Enterprise.

Dell believes in the value of big data, starting with our first Hadoop reference architecture in 2011. Big data equals machine data for organizations where devices drive their business. Machine data is becoming even more critical in today's data economy, driving use cases like log aggregation. Log aggregation allows organizations to process and analyze logs from their connected devices to help predict unexpected failures.

Manufacturing is a vertical market where log aggregation is a good fit. Consider the manufacturing floor with a variety machines or devices that an organization uses to build and assemble products. If the manufacturing floor is impacted by machine or device failure, this can affect an organization's ability to produce products, resulting in lost revenue. IT Operations wants to make sure this is prevented; Splunk Enterprise running on Dell PowerEdge servers can help your organization make sure that doesn't happen. Splunk Enterprise allows users to monitor your end-to-end infrastructure to avoid service degradation or outages. You gain operational intelligence with real-time visibility and critical insights into the customer experience, transactions and other key business metrics.

This is just one example; log aggregation is a viable solution in other vertical markets like healthcare, transportation, finance and retail. Additionally, with the rise of the Internet of Things (IoT), organizations will place a greater emphasis on machine data as the proliferation of sensors and smart devices increases in the coming years. Dell's experience in the big data space and expertise in helping customers build big data environments can help your organization simplify the time, effort and resources to get faster time to value.

Dell PowerEdge servers

The Dell PowerEdge portfolio offers a variety form factors and density options. For IT organizations looking to standardize on a 2U server platform, the Dell™ PowerEdge™ R730XD is the perfect building block for a single instance or distributed deployment for a Splunk Enterprise installation. The Dell PowerEdge R730XD server, based on Intel® Xeon® processor technology, provides a flexible building block due to the variety of configuration options, allowing organizations to use a single platform for multiple Splunk configurations.

The PowerEdge R730XD has exceptionally flexible and scalable two-socket 2U rack servers that deliver high-performance processing and a broad range of workload-optimized storage. The extensive configurability makes it an ideal candidate for indexers, search heads and forwarders. The PowerEdge R730XD offers the latest Intel® Xeon® processor E5 2600 v3 product family, 24 DIMMs of high-performance DDR4 memory and a broad range of local storage options, including SSDs.

For IT organizations looking to gain economies of scale and density, the Dell PowerEdge modular FX platform offers a good fit. Dell's modular PowerEdge FX2 platform allows system administrators to mix and match compute, storage and networking "blocks" within a single 2U chassis to meet workload-specific needs. Based off the proven shared-infrastructure designs of the Dell PowerEdge M1000e blade platform and the Dell PowerEdge VRTX, the FX2 platform plays well into converged infrastructure needs.

The PowerEdge FX2 platform supports combinations of quarter-width Dell PowerEdge FC430 server blocks, half-width PowerEdge FC630 server blocks, full-width PowerEdge FC830 server blocks or half-width PowerEdge FD332 storage blocks.

Splunk core components: Splunk indexers, search heads and forwarders

Splunk indexers provide data processing and storage for local and remote data and host the primary Splunk data store.

A **search head** is a Splunk Enterprise instance that distributes searches to indexers (referred to as "search peers" in this context). Search heads can be either dedicated or not, depending on whether they also perform indexing. Dedicated search heads don't have any indexes of their own, other than the usual internal indexes. Instead, they consolidate and display results that originate from remote search peers.

Forwarders are Splunk instances that forward data to remote indexers for data processing and storage. In most cases, they do not index data themselves. Fowarders are typically only used in a distributed Splunk deployment.

Additionally, a Splunk Enterprise instance can also serve as a deployment server. The **deployment server** is a tool for distributing configurations, apps and content updates to groups of Splunk Enterprise instances. You can use it to distribute updates to most types of Splunk components: forwarders, non-clustered indexers and non-clustered search heads.

Index replication and indexer clusters

An **indexer cluster** is a group of indexers configured to replicate each other's data, so that the system keeps multiple copies of all data. This process is known as index replication. By maintaining multiple, identical copies of data, indexer clusters prevent data loss while promoting data availability for searching.

Splunk Enterprise clusters feature automatic failover from one indexer to the next. This means that, if one or more indexers fail, incoming data continues to get indexed and indexed data continues to be searchable. In addition to enhancing data availability, clusters have other features that you should consider when you are scaling a deployment—for example, a capability to coordinate configuration updates easily across all indexers in the cluster. Clusters also include a built-in distributed search capability.

Splunk Enterprise deployment considerations

A Splunk Enterprise deployment has many dimensions. These scenarios determine whether a single reference machine can handle indexing and search load.

In some cases, a single reference machine can collect, store and search data efficiently. In other cases, consider adding machines to your Splunk Enterprise deployment to increase performance.

Here is a list of items that can have a significant impact on Splunk Enterprise performance.

- **Amount of incoming data.** The more data you send to Splunk Enterprise, the more time it needs to process the data into events that you can search, report and generate alerts on.
- **Amount of indexed data.** As the amount of data stored in a Splunk Enterprise index increases, so does the I/O bandwidth needed to store data and provide results for searches.
- **Number of concurrent users.** If more than one person at a time uses an instance of Splunk Enterprise, that instance requires more resources for those users to perform searches and create reports and dashboards.

- **Number of saved searches.** If you plan to invoke a lot of saved searches, Splunk Enterprise needs capacity to perform those searches promptly and efficiently. A higher search count over a given period of time requires more resources.
- **Types of search you use.** Almost as important as the number of saved searches is the types of search that you run against a Splunk Enterprise instance. There are several types of search, each of which affects how the indexer responds to search requests.
- **Whether or not you run Splunk apps.** Splunk apps and solutions can have unique performance, deployment and configuration considerations. If you plan to run apps, consider the resource requirements of the apps that you are using.

Splunk Enterprise hardware considerations

Splunk Enterprise can deploy in a single instance or it can be distributed across multiple servers using multiple indexers and search heads. The distributed deployment can support clustering or non-clustering. When deploying in a distributed deployment it is recommended to add a deployment server, cluster manager (clustering) and license master along with multiple indexers and search heads.

Basic sizing guidelines

CPUs

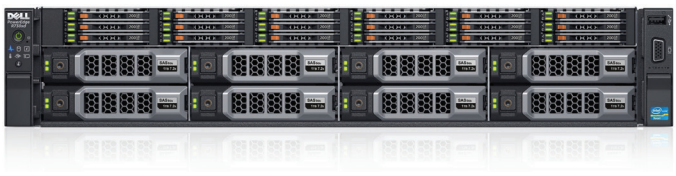
- Search process utilizes up to 1 CPU core
- Indexers still need to do the heavy lifting (search exists on indexer and search head)
- Limited benefit for indexing (up to 4 CPU cores for indexing)

Memory

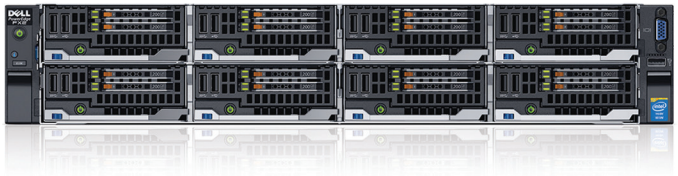
- More memory is good for search heads and indexers

Disks

- Faster is better (15k rpm) or SSD
- More disks in RAID 10 = faster
- SSDs can provide benefits for rare term searches and many concurrent jobs.



PowerEdge R730XD



Dell FX2

Table 1. Single Instance

	High Capacity	High Performance
Indexer, Search Head	**PowerEdge R730XD**	**PowerEdge R730XD**
CPU	2 x Intel Xeon processor E5-2680 v3	2 x Intel Xeon processor E5-2680 v3
Memory	256GB	256GB
Controller	PERC H730	PERC H730
Storage	OS Drives: 2 x 120GB SSD 2.5" Flex Bay Production Data: 24 x 1.2TB SAS 10K 2.5" RAID 10	OS Drives: 2 x 120GB SSD 2.5" Flex Bay Production Data: 6 x 800GB SSD 2.5" RAID10
NIC (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)
Use Case	Fast Performance, High Capacity Retention	Extreme Performance, Fast Response Times, Many Concurrent Searches

Table 2. Single Instance High Density

	High Capacity	High Performance
Indexer, Search Head	**FX-FC630 & 2 x FD332 Single Chassis 2U**	**PowerEdge FC630 Single Chassis 2U**
CPU	2 x Intel Xeon processor E5-2680 v3	2 x Intel Xeon processor E5-2680 v3
Memory	256GB	256GB
Controller	PERC H730	PERC H730
Storage	OS Drives: 2 x 120GB SSD 2.5" (FC630) Production Data: 32 x 1TB SAS 10K 2.5" RAID 10 (FD332)	OS Drives/ Production Data: 8 x 800GB SSD 1.8" RAID10
NIC (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)
Use Case	Fast Performance, High Capacity Retention	Extreme Performance, Fast Response Times, Many Concurrent Searches

Table 3. Distributed

	High Capacity	High Performance
Indexer Deploy in Multiples	**PowerEdge R730XD**	**PowerEdge R730XD**
CPU	2 x Intel Xeon processor E5-2680 v3	2 x Intel Xeon processor E5-2680 v3
Memory	256GB	256GB
Controller	PERC H730	PERC H730
Storage	OS Drives: 2 x 120GB SSD 2.5" Flex Bay Production Data: 24 x 1.2TB SAS 10K 2.5" RAID 10	OS Drives: 2 x 120GB SSD 2.5" Flex Bay Production Data: 6 x 800GB SSD 2.5" RAID10
NIC (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)
Search Heads Deploy in Multiples	**PowerEdge R730XD**	
CPU	2 x Intel Xeon processor E5-2680 v3	
Memory	256GB	
Controller	PERC H730	
Storage	OS Drives/Production Data: 2 x 600GB SAS 2.5" RAID 1	
NIC (Daughter Card)	Intel X520 DP 10Gb DA/SFP+,+ I350 DP 1Gb Ethernet (2 x 10GbE, 2x 1GbE)	
Administration, Deployment, Master, and Fowarders	**PowerEdge R730XD**	
CPU	2 x Intel Xeon processor E5-2650 v3	
Memory	128GB	
Controller	PERC H730	
Storage	OS Drives/Production Data: 2 x 600GB SAS 2.5" RAID 1	
NIC (Daughter Card)	Intel X520 DP 10Gb DA/SFP+,and I350 DP 1Gb Ethernet (2 x 10GbE, 2x 1GbE)	

Table 4. Distributed High Density

	High Capacity	High Performance
Indexer Deploy in Multiples	**PowerEdge FC630 (2) PowerEdge FD332 (2) Single Chassis 2U**	**FC630 (4) Single Chassis 2U**
CPU	2 x Intel Xeon processor E5-2680 v3	2 x Intel Xeon processor E5-2680 v3
Memory	256GB	256GB
Controller	PERC H730	PERC H730
Storage	OS Drives: 2 x 120GB SSD 2.5" Flex Bay (FC630) Production Data: 16 x 1TB SAS 10K 2.5" RAID 10 (FD332)	OS Drives/Production Data: 8 x 800GB SSD 1.8" RAID10
NIC (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)
Search Heads Deploy in Multiples	**PowerEdge FC430 (8) Single Chassis 2U**	
CPU	2 x Intel Xeon processor E5-2680 v3	
Memory	256GB	
Controller	PERC S130	
Storage	OS Drives/Production Data: 2 x 800GB SSD 1.8" RAID 1	
NIC (Daughter Card)	Intel X520 DP 10Gb DA/SFP+,+ I350 DP 1Gb Ethernet (2 x 10GbE, 2x 1GbE)	
Administration, Deployment, Master, and Fowarders	**PowerEdge FC430 (4) Single Chassis 2U**	
CPU	2 x Intel Xeon processor E5-2650 v3	
Memory	128GB	
Controller	PERC S130	
Storage	OS Drives/Production Data: 2 x 800GB SSD 1.8" RAID 1	
NIC (Daughter Card)	Intel X520 DP 10Gb DA/SFP+,and I350 DP 1Gb Ethernet (2 x 10GbE, 2x 1GbE)	

Table 5. Rules of Thumb

How Many Indexers
1 per 250 GB/day
Leaves room for: Daily peaks, light searching and reporting for about 5 concurrent users
Need more indexers for: Heavy reporting, more users, slower disks, slower CPUs and fewer CPUs
How Many Search Heads
1 per 20 to 40 concurrent jobs
Limit is concurrent queries
Search Query may utilize up to 1 CPU core
Only add first search head if ≥3 indexers
Don't add search heads, add indexers; indexers do most of the work
How Many Deployment Servers
1 per 3000 polls/minute
Just use one deployment server, and adjust the polling

Table 6. Performance and Sizing Tips

System Change	Search Speed	Indexing Speed
Faster Disks	●	●
Add an Indexer	●	●
Add a Search Head	●	
Report Acceleration/Summaries	●	
Optimize Searches	●	
Optimize Field Extraction	●	
Optimize Input Parsing		●
Faster CPU	●	●

Resources:

Architecting and Sizing your Splunk Deployment—Simeon Yep, Senior Manager Business Development Technical Services, Splunk; Karandeep Bains, Senior Sales Engineer, Splunk

Splunk Enterprise 6.2.0 Capacity Planning Manual

To learn more, visit Dell.com/BigData

April 2016 | Version 1.0