Master The Dark Arts

Demystifying Splunk Architecture

Cory Minton | Principal Engineer in Dell EMC's Global Technology Office

October 2018 | Version 1.0

J. Cory Minton

Principal Engineer

- Global Technology Office
- 8+ years at Dell EMC
- Founder: Dell EMC's Splunk Ninjas
- Splunk SE and Hadoop Certified
- App Consulting Background
- BS Engineering and MBA
- www.BigDataBeard.com
- Check out our new podcast!



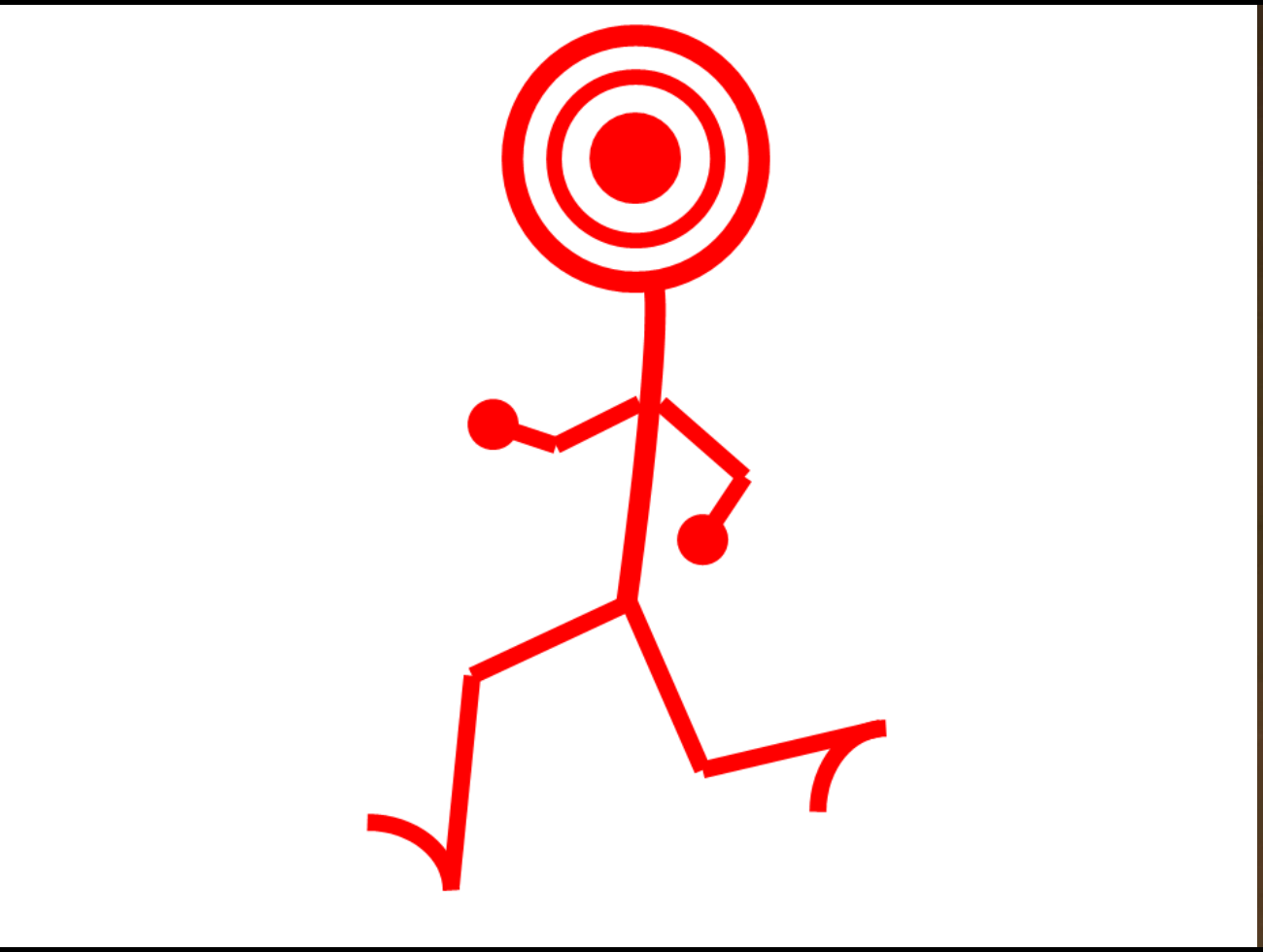



splunk> .conf18

Key takeaways

- Size the infrastructure for a Splunk deployment
- Understand infrastructure impacts from small changes in Splunk
- Learn design concepts that will scale
- Hear how Dell EMC is doing it internally
- An easier way…

Problem…

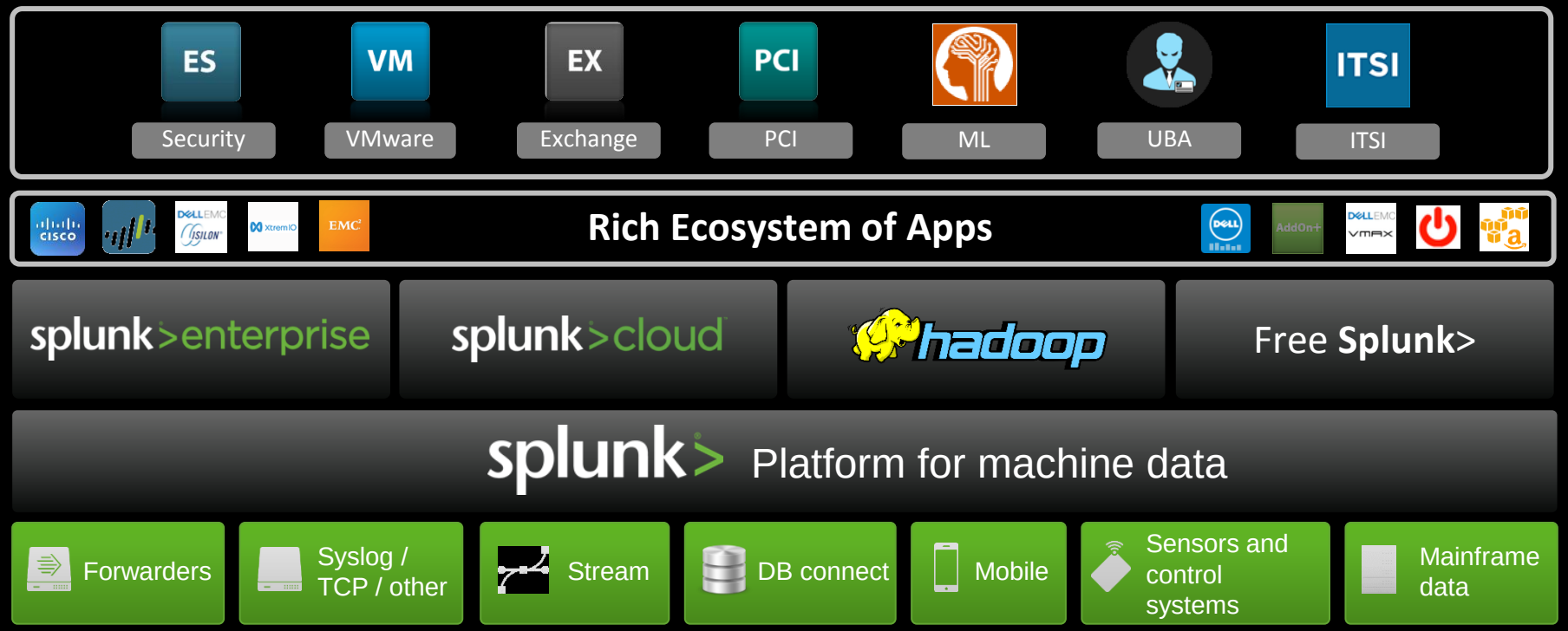


Provide fundamentals for sizing a Splunk deployment and share learned best practices.

90% empirical
+ 10% experience
≠ 100% perfect every time

Assumption #1

General understanding of Splunk platform



Assumption #2

General understanding of Splunk infrastructure

Search Heads

Query information across indexers and are usually CPU and memory intensive.

Indexers

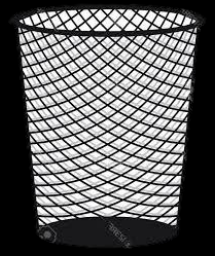
Write data to disk and are both CPU and I/O intensive.

Forwarders

Collect and forward data; usually lightweight and not resource intensive.

Assumption #3

General understanding of Splunk data management.



OR

FROZEN

HOT – Newest buckets of data that are still open for write

WARM – Recent data but closed for writing (read only)

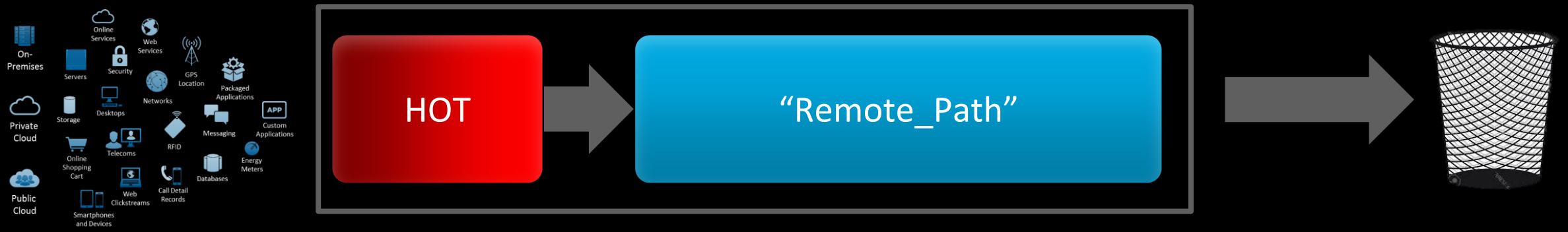
COLD – Oldest data, commonly on cheaper, slower storage

FROZEN – No longer searchable, commonly archived or deleted data

splunk> .conf18

Assumption #4

You've heard of Splunk S2 Architecture….



HOT – Newest buckets of data that are still open for write

REMOTE_PATH – Data aged from HOT to an S3-compatible object storage platform

splunk> .conf18

Big & Fast

What makes Splunk grow?

Performance

- ✓Volume Of Ingest
- ✓Search Performance
- ✓More Users
- ✓Big Apps

Capacity

- ✓Volume Of Ingest
- ✓Index Retention Periods
- ✓Indexer Clustering
- ✓Big Apps

Sizing Fundamentals

How many servers do I need?

splunk> .conf18

Machine Requirements

Indexers

Reference Minimum

- 12 cores
- 12GB RAM
- 800 IOPS

Mid-Range

- 24 cores
- 64GB RAM
- 1200 IOPS

High-Performance

- 48 cores
- 128GB RAM
- SSD

Others

Search Head

- 16 cores
- 12GB RAM
- 300 IOPS

Heavy Forwarder

- 16 cores
- 12GB RAM
- 300 IOPS

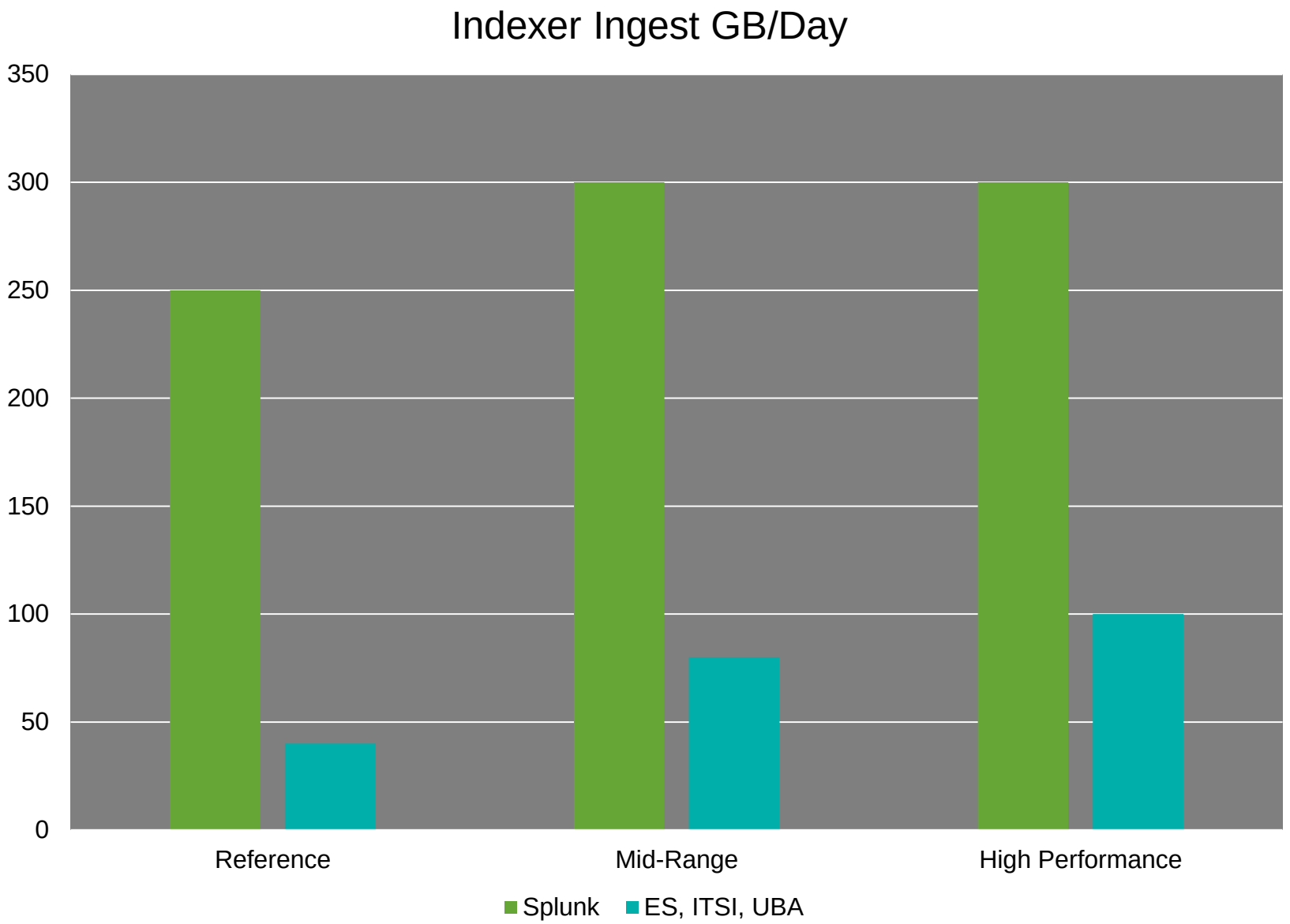
Utility

- 8 cores
- 8GB RAM
- 300 IOPS

Dark truth: Choose wisely…or scalability will suffer later.

Indexer Sizing

- vCPU = CPU
- Hyperthreading ≠ CPU
- When in doubt, 100



Search Heads

- **Dedicate**
- **When in doubt, 1 per 5**
- **Indexers > Search**

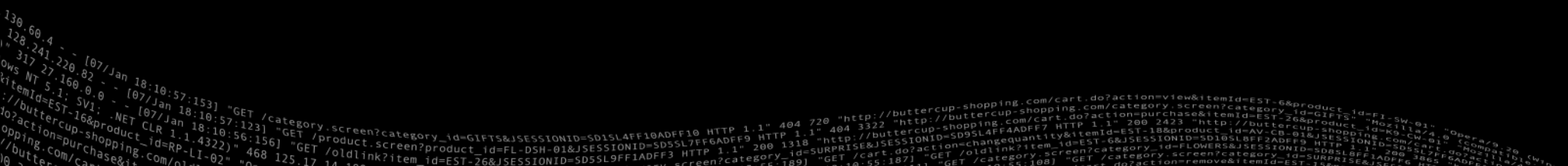
Daily Indexing Volume						
	< 2GB/day	2 to 300 GB/day	300 to 600 GB/day	600GB to 1TB/day	1 to 2TB/day	2 to 3TB/day
Total Users: less than 4	1 combined instance	1 combined instance	1 Search Head, 2 Indexers	1 Search Head, 3 Indexers	1 Search Head, 7 Indexers	1 Search Head, 10 Indexers
Total Users: up to 8	1 combined instance	1 Search Head, 1 Indexers	1 Search Head, 2 Indexers	1 Search Head, 3 Indexers	1 Search Head, 8 Indexers	1 Search Head, 12 Indexers
Total Users: up to 16	1 Search Head, 1 Indexers	1 Search Head, 1 Indexers	1 Search Head, 3 Indexers	2 Search Heads, 4 Indexers	2 Search Heads, 10 Indexers	2 Search Heads, 15 Indexers
Total Users: up to 24	1 Search Head, 1 Indexers	1 Search Head, 2 Indexers	2 Search Heads, 3 Indexers	2 Search Heads, 6 Indexers	2 Search Heads, 12 Indexers	3 Search Heads, 18 Indexers
Total Users: up to 48	1 Search Head, 2 Indexers	1 Search Head, 2 Indexers	2 Search Heads, 4 Indexers	2 Search Heads, 7 Indexers	3 Search Heads, 14 Indexers	3 Search Heads, 21 Indexers

Utility Servers

Handy helpers…

- Heavy Forwarder
- License Master
- DMC
- Cluster Master
- Deployment

1:3



splunk> .conf18

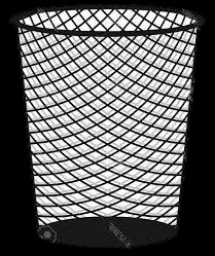
Sizing Fundamentals

How much storage do I need?

splunk> .conf18

Assumption #3

- General understanding of Splunk data management.



OR

FROZEN

HOT – can be DAS in server or SAN (Flash is best)

WARM – same as Hot

COLD – adds option for NAS

FROZEN – No longer searchable, so object stores are option here (last resort)

Myth about bucket sizing…

- # of buckets x bucket size
- Not days…

indexes.conf

```
# volume definitions

[volume:hotwarm_cold]
path = /mnt/fast_disk
maxVolumeDataSizeMB = 3984589

# index definition (calculation is based on a single index)

[main]
homePath = volume:hotwarm_cold/defaultdb/db
coldPath = volume:hotwarm_cold/defaultdb/colddb
thawedPath = $SPLUNK_DB/defaultdb/thaweddb
homePath.maxDataSizeMB = 512000
coldPath.maxDataSizeMB = 2560000
maxWarmDBCount = 4294967295
frozenTimePeriodInSecs = 2592000
maxDataSize = auto_high_volume
coldToFrozenDir = /mnt/big_disk/defaultdb/frozendb
```

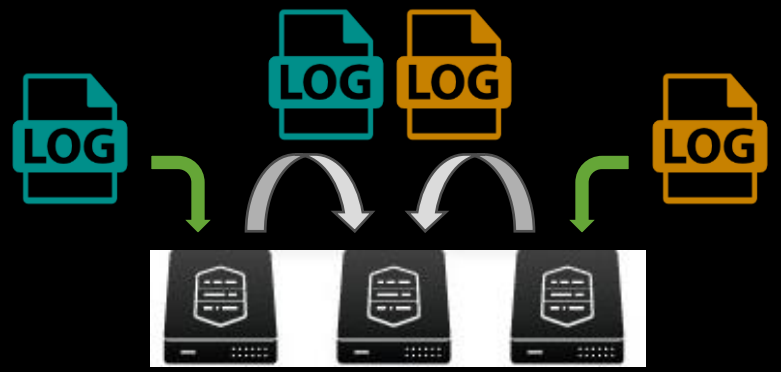
Indexer Deployment Options

Distributed Deployment

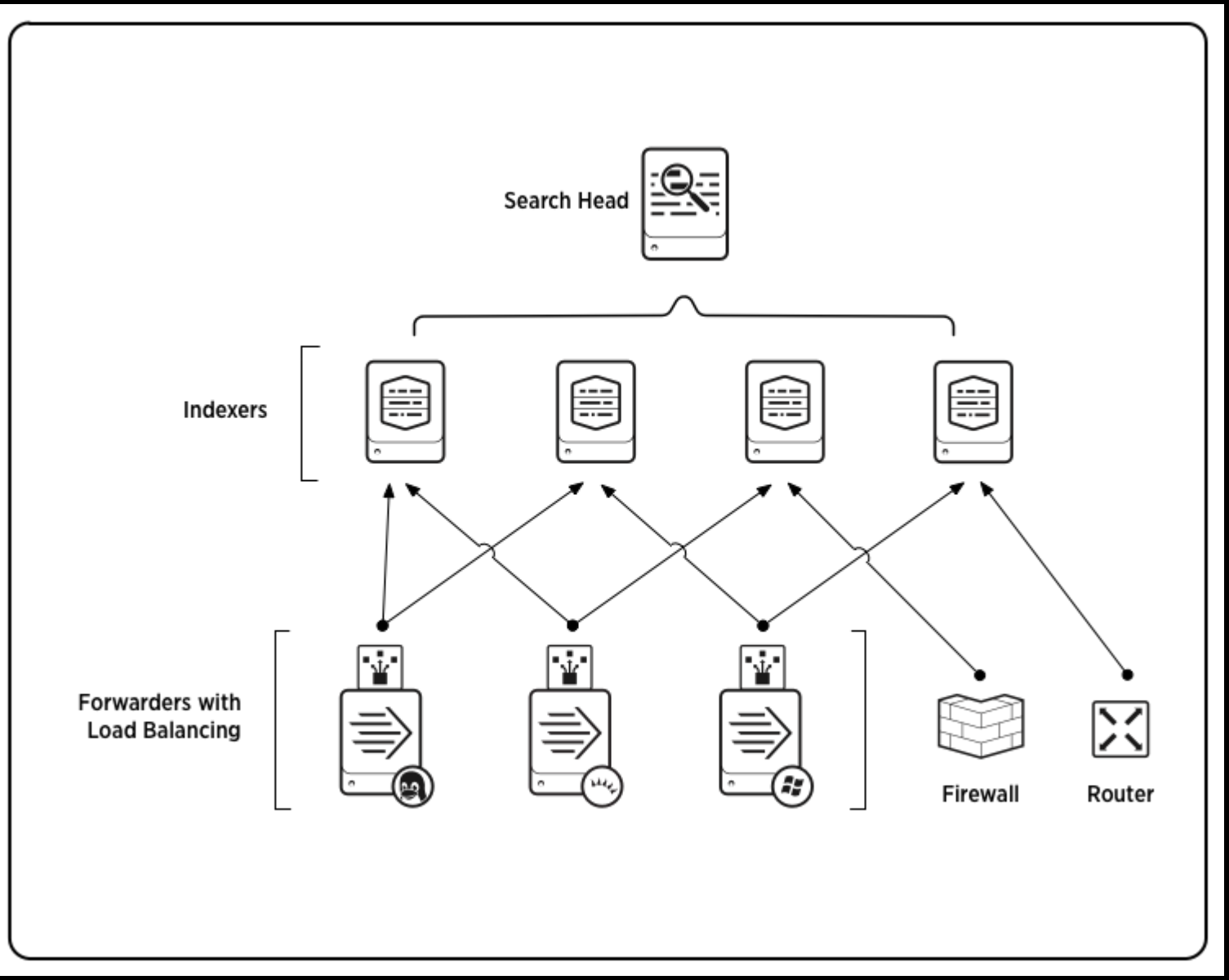
Indexer data is stored once and distributed across available indexers

Clustered Deployment

A group of indexers are configured to replicate each other's data

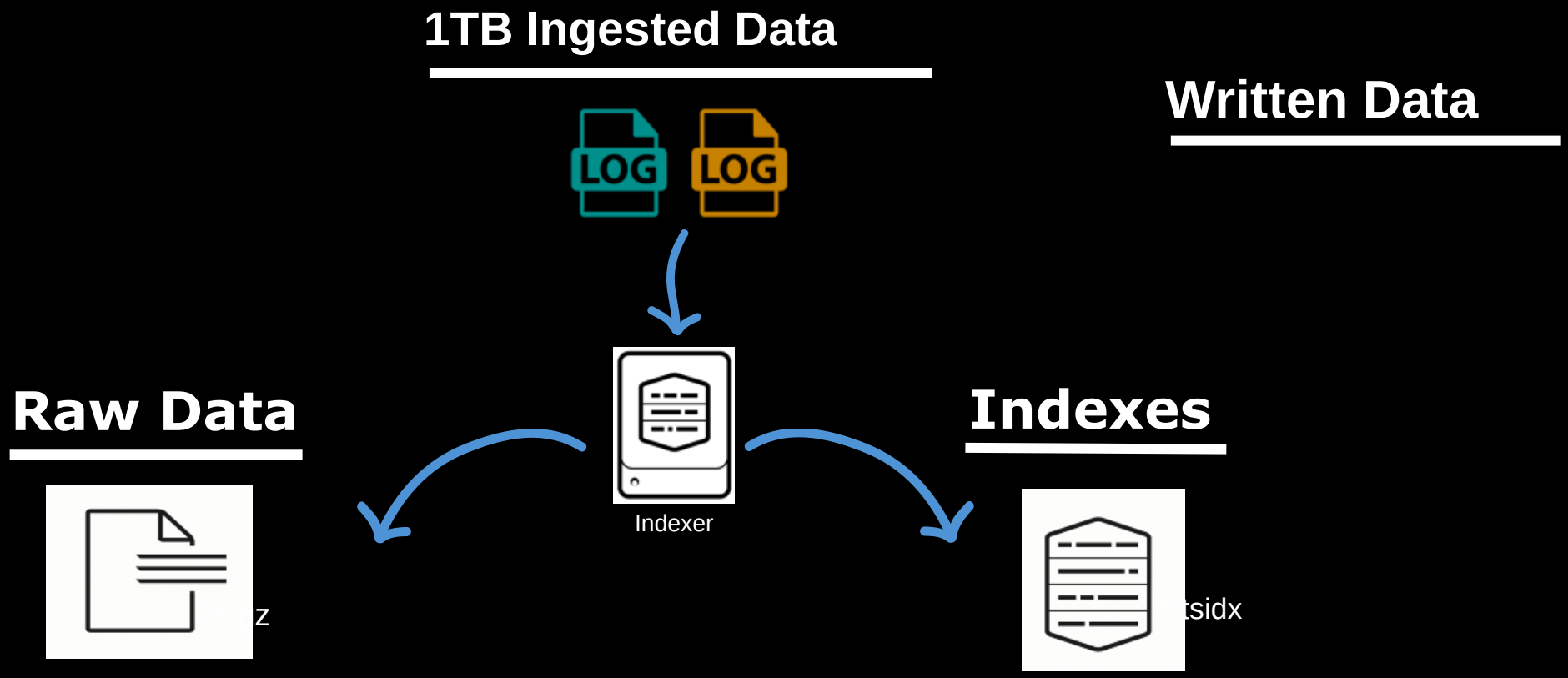


Distributed Deployment



- Single copy of data
- Small
- Starter
- Storage-bound

Indexer Storage Capacity



Compressed Raw data
30% of written data
→ 150GB

Uncompressed ‘indexes’
70% of written data
→ 350GB

How much storage you need?

= Daily indexing rate
x ½
x Retention policy

1TB Ingested Data

= 1TB x ½ x 60 days = 30TB


Indexer

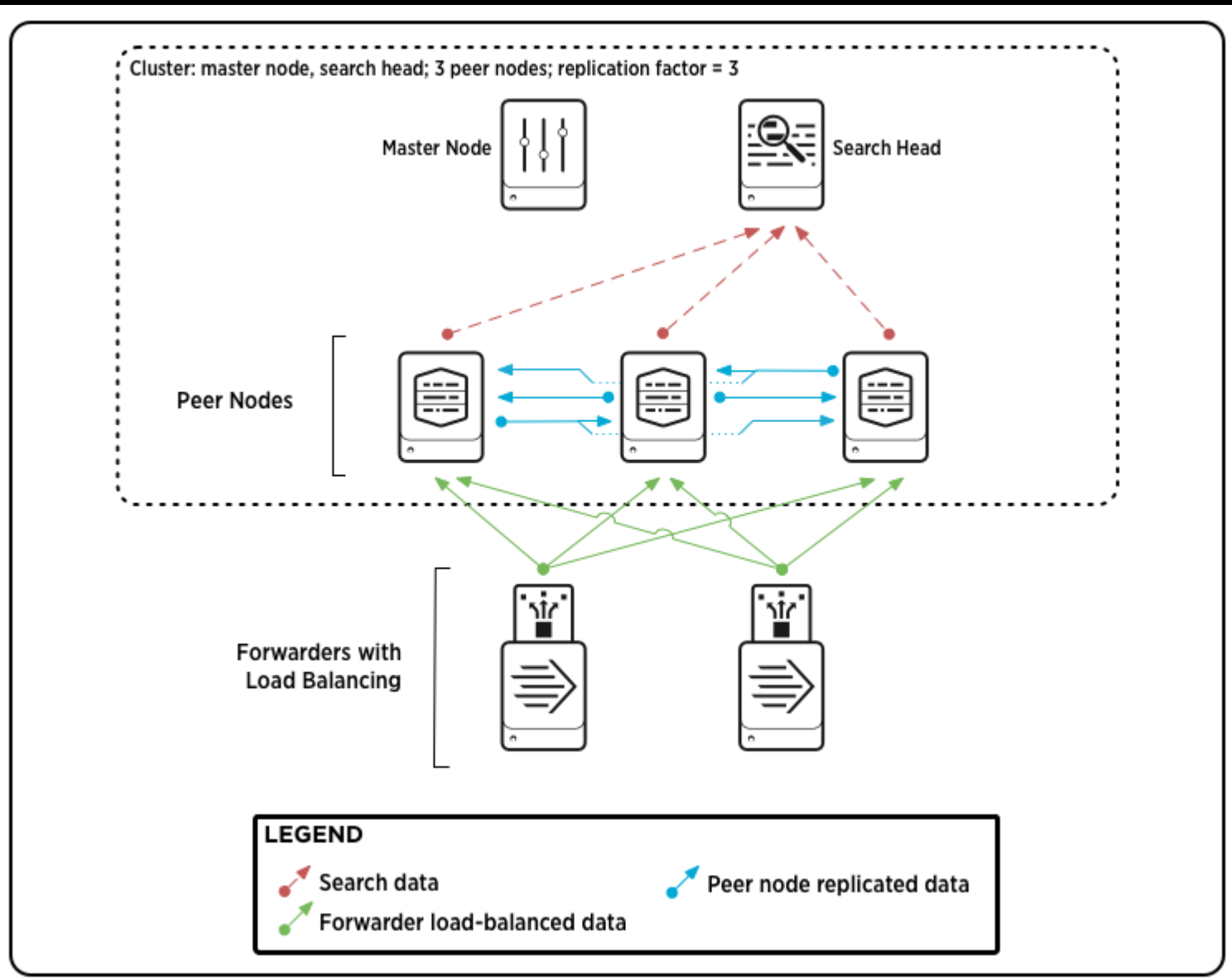
Raw Data


9TB

Indexes


21TB

Indexer Clustering

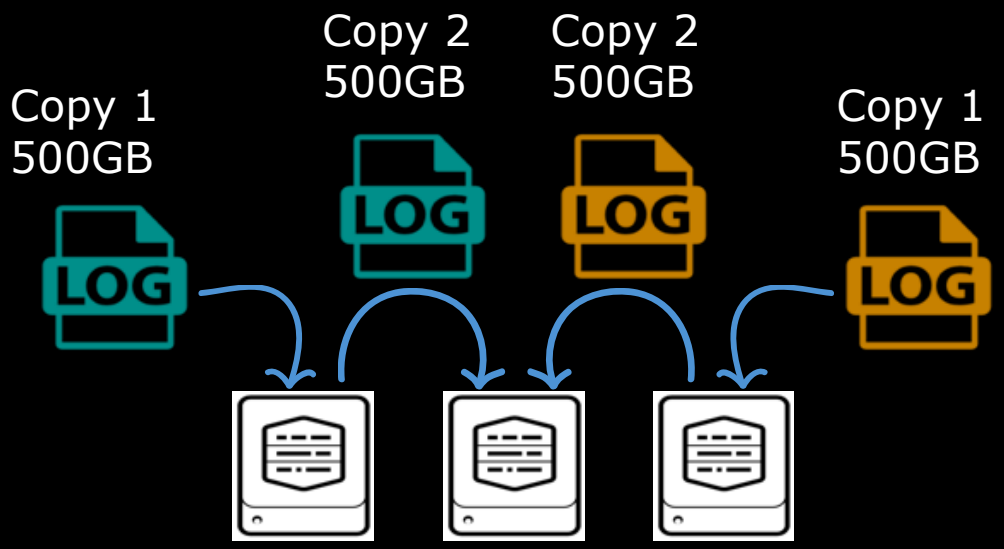


- High Availability for Indexes
- Indexer Clustering Settings
  - Replication Factor = copies of raw data
  - Search Factor = copies of indexes

SPLUNK INDEXER AVAILABILITY

Multiple copies of index and raw data

- Index → # copies of indexes → Search factor (SF)
- Raw Data -> # of of copies of raw data → Replication factor (RF)

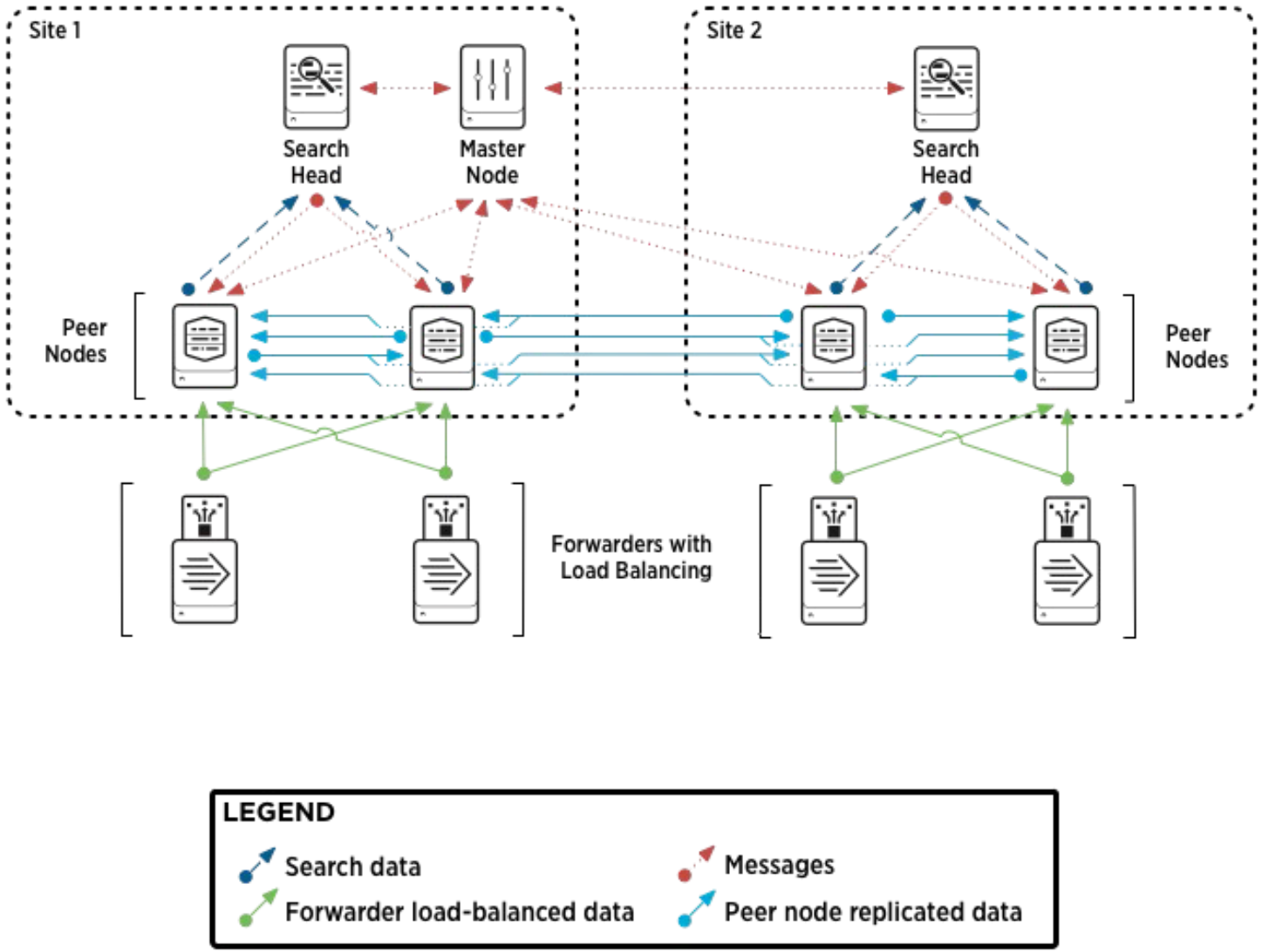


1TB Ingested Data
SF=2 / RF=2
500GB written → 500GB replicated

1TB * 60 days x ½ x **2**
= 60TB (RF/SF=2) ** **doubled** **

1TB * 60 days x ½ x **3**
= 90TB (RF/SF=3) ** **tripled** **

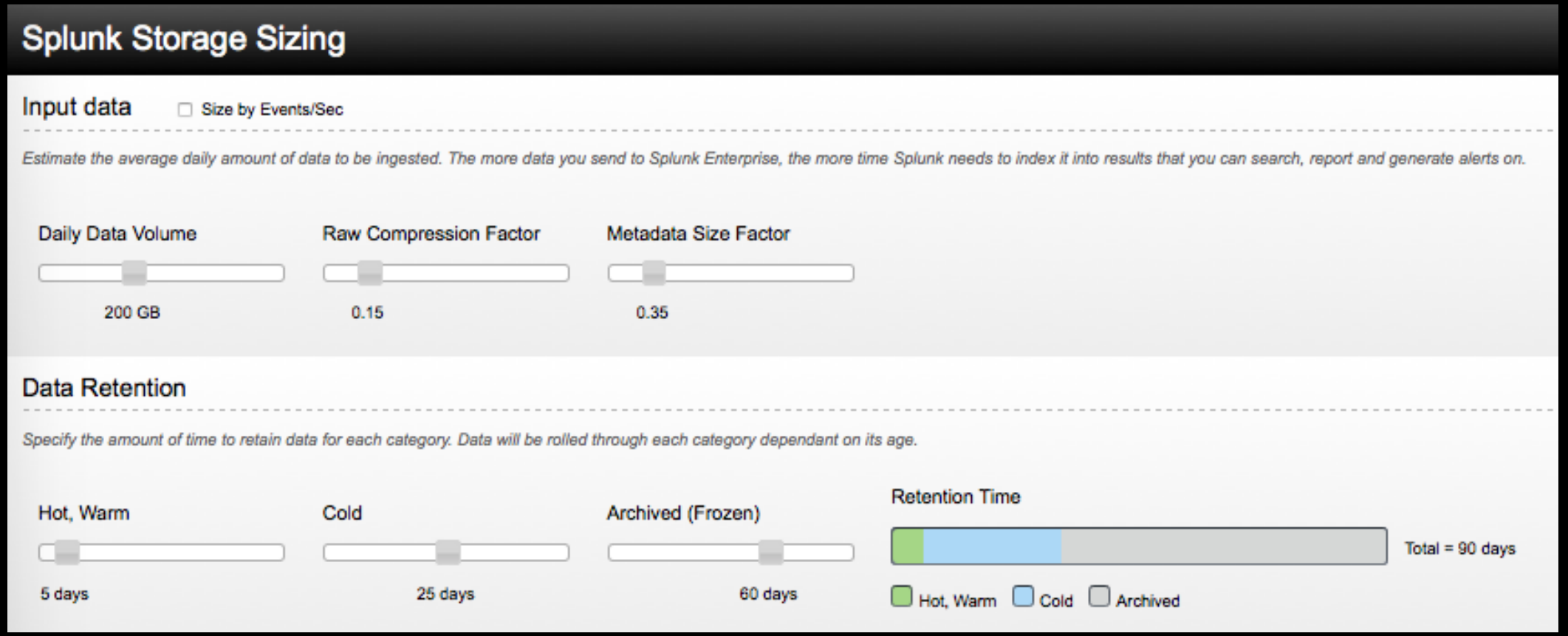
Multisite Indexer Clustering



- Protects indexes across disparate locations
- Enables Search Affinity
- Site specific RF/SF settings

- Sizing = each site + site protected

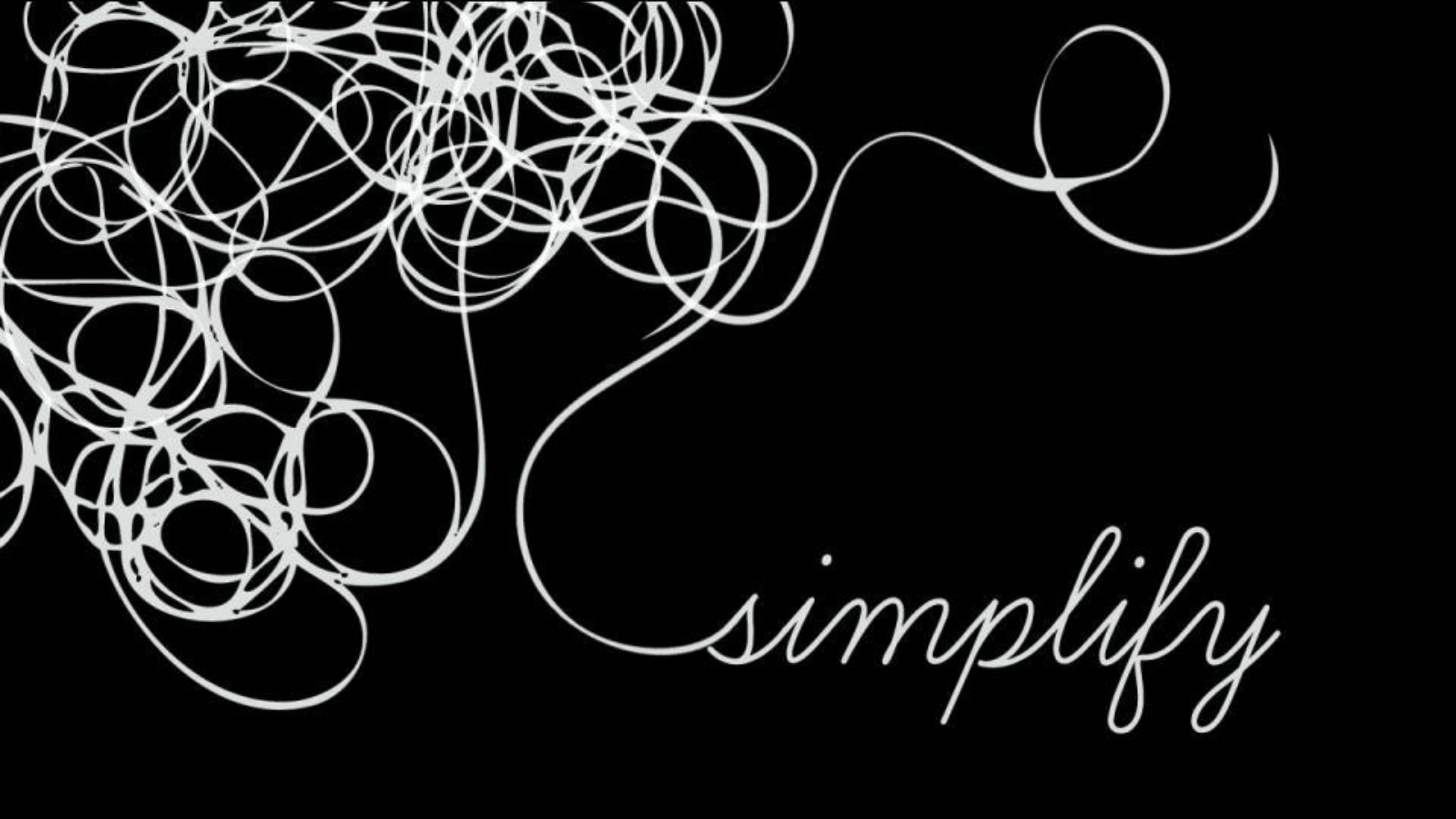
Unofficial, but really helpful tool

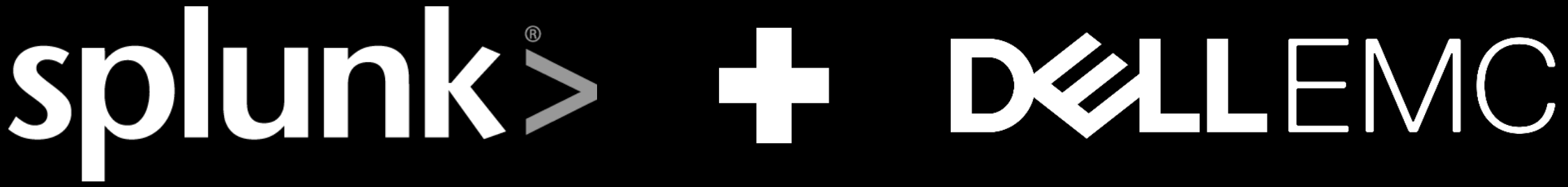


http://splunk-sizing.appspot.com/

Splunk Sizing Questionnaire

- What is the licensed daily ingest rate for Splunk (expressed in some amount of GB/Day or TB/day)?
- What is the retention period for Hot/Warm and Cold (days kept in each tier)?
- Any data being sent to frozen? If so, what is the retention period and requirement for doing so?
- Is indexer clustering being leveraged? If so, what are the settings for Replication and Search Factor?
- How many indexer and search servers are deployed? Do you have a visualization you can share of the deployment?
- Is Splunk being run as a single site or multiple sites? If multiple, is multi-site clustering being leveraged?
- Is the Enterprise Security App or ITSI for Splunk deployed?


simplify



The right solutions to optimize your Splunk deployment

Dell EMC Ready Solutions for Splunk

Ready System

VxRack + Isilon

VxRail + Isilon

Ready Bundle

PowerEdge + Isilon

*"Meets or **EXCEEDS** minimum hardware requirements"*

Dell EMC Ready Solutions for Splunk

Ready Bundle for Splunk

PowerEdge

DAS
SSD/HHD/Hybrid
Bare Metal or Virtual

Ready System for Splunk

VxRack Flex

ScaleIO SDS
SSD/HHD/Hybrid
Bare Metal or Virtual

VxRail

VMware VSAN
SSD
Virtual

PLACEHOLDER for S2 Solutions

We are in beta now and will be adding solution guidance for our S2 architecture platform in mid-September as the beta and performance testing completes.

Choosing the right Ready Solution for Splunk

VMware Capable

Bare Metal

One or more

- ☑ Prefer DAS or commodity deployment
- ☑ Ability to start small & scale (<100GB)
- ☑ Prefer to scale compute & storage together linearly
- ☑ Dedicated HW for Splunk

One or more

- ☑ VxRack already installed
- ☑ Prefer ScaleIO for SDS
- ☑ Hypervisor choice | option for no hypervisor for indexer
- ☑ Scale compute and storage independently
- ☑ Want networking included at factory

One or more

- ☑ VxRail already installed
- ☑ Prefer VSAN for SDS
- ☑ Standardized or prefer VMware for hypervisor
- ☑ Scale compute and storage independently
- ☑ Looking for ability for snapshots for data protection

Cold Storage Option

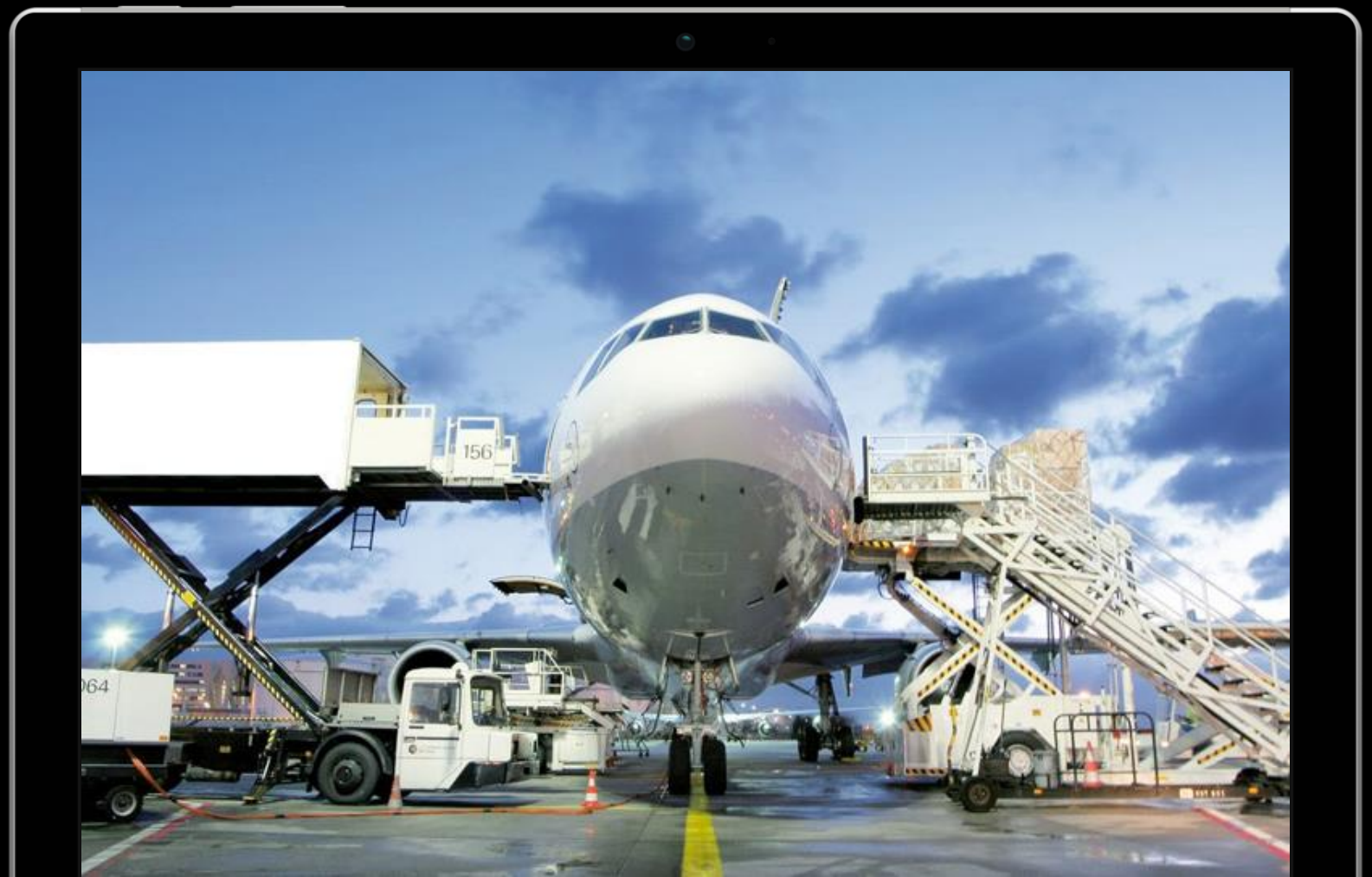
One or more

- ☑ Isilon already installed
- ☑ Cold bucket size to expected to be > 100TB

splunk> .conf18

Logistics Leader

- Doug called them out on Q1FY18 earnings call…



- Simplified acquisition
- Leveraged Ninjas
- Deployed apps for all Dell EMC platforms
- Adopted Ready System for Splunk with Isilon

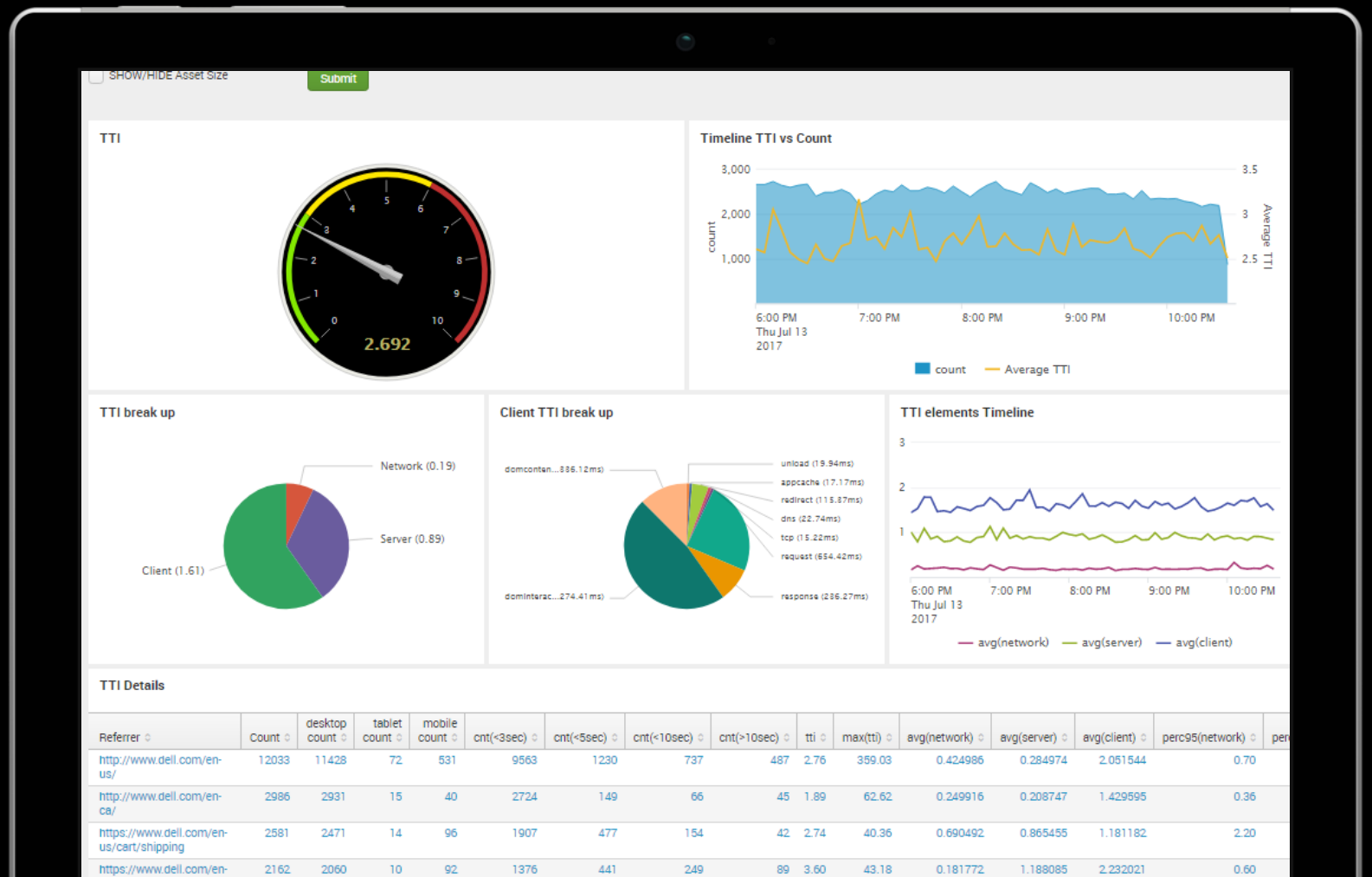
Wholesale Club Retailer

- Flashed Splunk
- Bottomless cold with Isilon…over 1PB!
- Decreased floor space by 30%
- Growing to +3TB/day

Winter is coming…

Splunk at Dell EMC

- Our defense against Black Friday…



- eCommerce IT services
- Marketing effectiveness
- Security and threats
- Replatformed to Ready Bundle for Splunk

Splunk applications from Dell EMC

Extend the power of Splunk to Dell EMC Platforms

What are Splunk Apps?

Splunk applications and add-ons allow user to import data into Splunk from specific sources

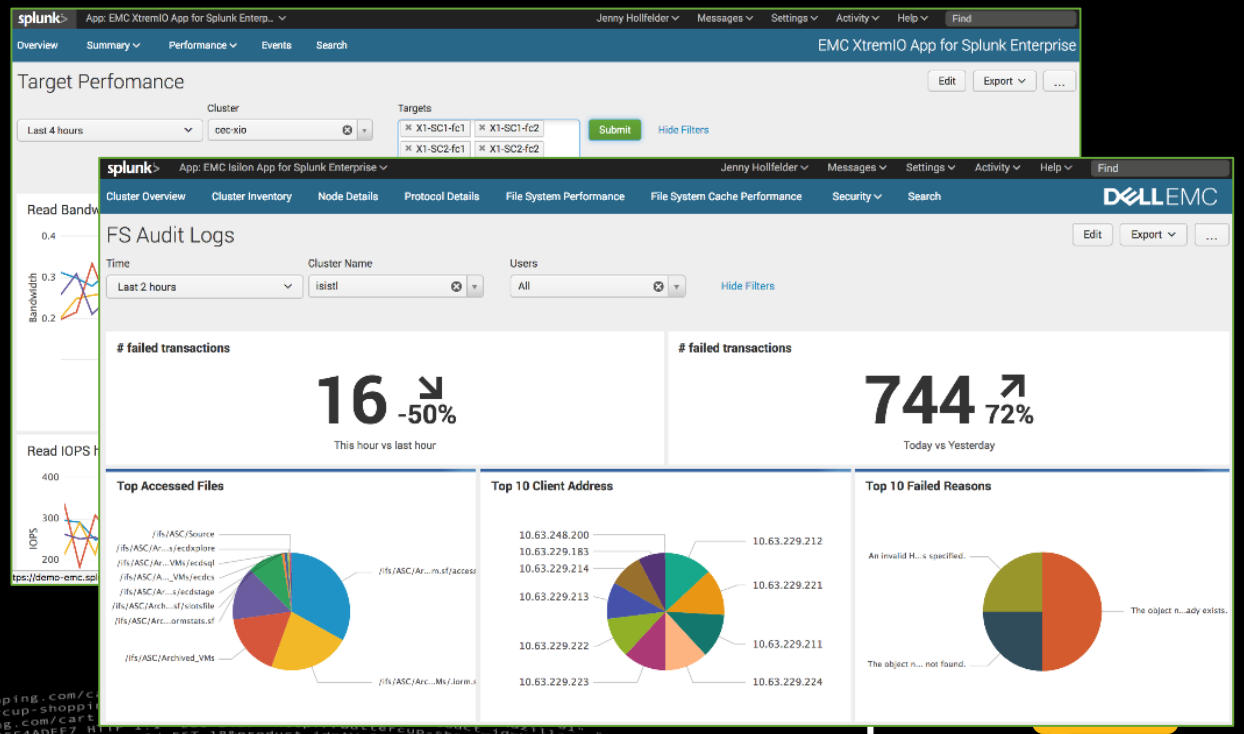
Splunk & its partners have created a rich community called SplunkBase that has 1000s+ applications

Why are Splunk Apps important?

Splunk apps and add-ons allow customers to incorporate new use cases and extend their Splunk environment. This leads to increased Splunk License needs as well as additional Hardware

Dell EMC has apps for the following:

- *VMAX*
- *XtremIO*
- *Isilon*
- *VNX*



Global Solution Centers

Validate. Evaluate. Collaborate. Innovate

Solution centers
Staffed with engineers and Blueprint solution experts

Engagements begin with your challenges

- Briefings with a team of experts
- Architectural design sessions
- Proofs of concept

DELL EMC

Let our Splunk Ninjas help you!

- Trained by Splunk
- Splunk Architecture Experts
- Dell EMC Portfolio Experts
- Religious about Best Practices
- Available across the GLOBE!!!

Email Splunk.Ninjas@emc.com

Don't forget to rate this session in the .conf18 mobile app