There Is No "Cold Data" In Analytics

Splunk Performance at Enterprise Scale

Somu Rajarathinam | Solutions Architect, Pure Storage

October 2018 | Version 1.0

Forward-Looking Statements

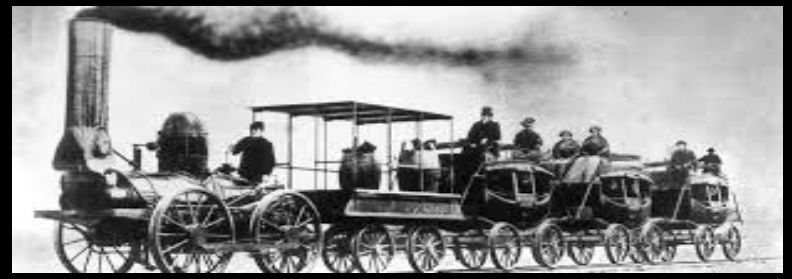
During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

splunk> .conf18

Dawn of 4th Industrial Revolution

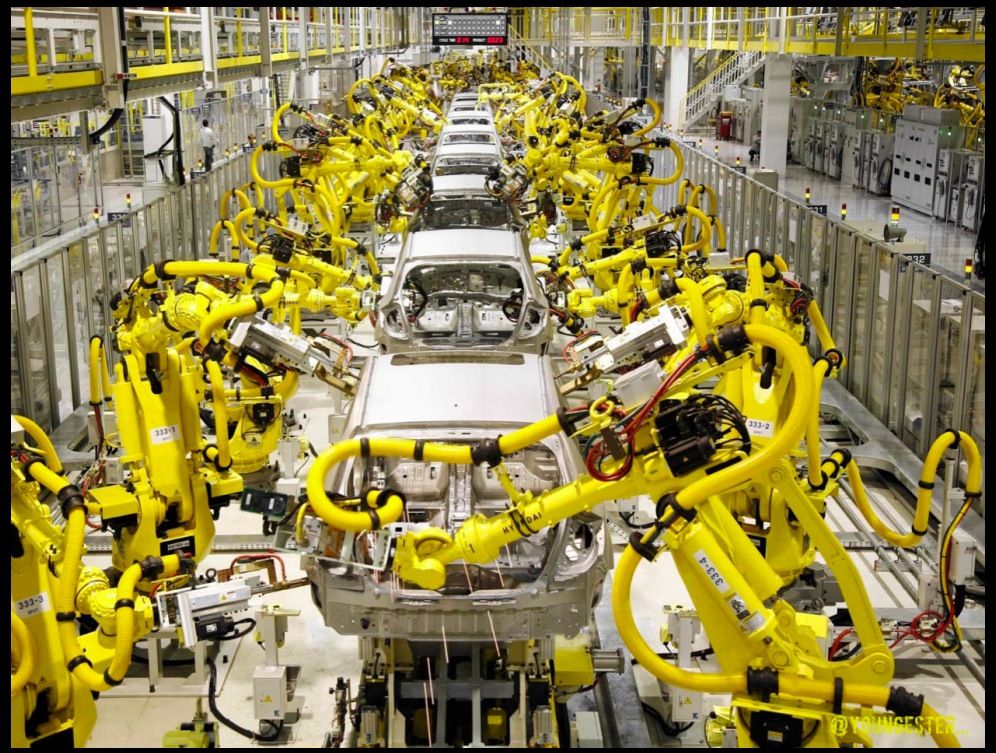
Big Data, Modern Analytics Driving Change In Every Industry



1st Revolution
1760-1820's
Steam Power
Rural to Urban

2nd Revolution
1870-1914
Mass Prod., Electricity
Urban to Analog

3rd Revolution
1980-2010
PC, Automation
Analog to Digital



4rd Revolution
2010-now
AI, Big Data, Cloud, IoT & Edge Computing
Digital to Intelligence

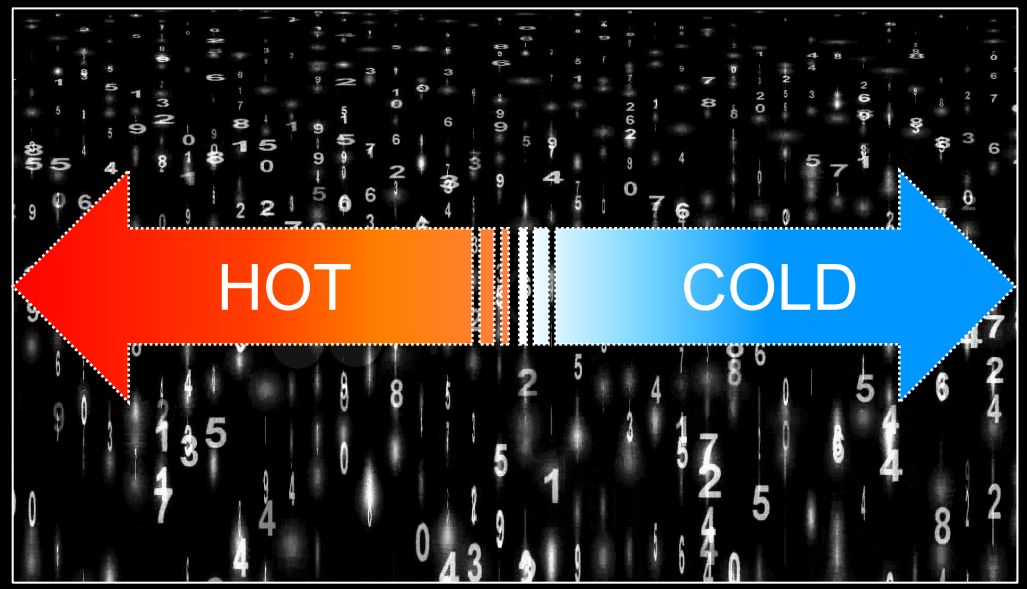
Data Growth

Source: IDC Data Age 2025 Study

Big Data

"Big data is no longer enough. It's now all about Fast Data "

"Big data is only as useful as its rate of analysis"

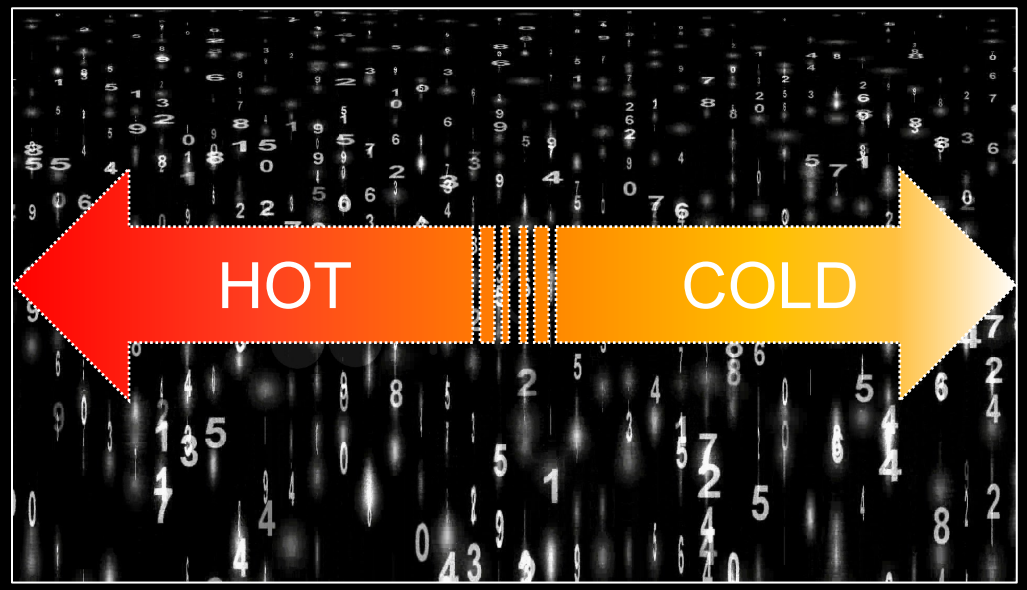


Source: entrepreneur.com

Big Data

"Big data is no longer enough. It's now all about Fast Data "

"Big data is only as useful as its rate of analysis"



Source: entrepreneur.com

Challenges With Data Growth

Storage Requirements

PERFORMANCE

Data Ingestion Volume

Search Performance

User Count

CAPACITY

Data Ingestion Volume

Data Retention (Hot/Cold)

Deployment Model

MANAGEABILITY

Data Protection

Non-Disruptive Upgrades

Scalability

Data Services

Challenges With Direct Attached Storage

STORAGE UPGRADE
DISRUPTIVE
SEAMLESS (NDU)

MANAGEMENT
LABOR-INTENSIVE, ERROR-PRONE
SIMPLE, CENTRAL MANAGEMENT

STORAGE AVAILABILITY
LIMITED WITHIN THE HOST
AVAILABLE TO ALL HOSTS

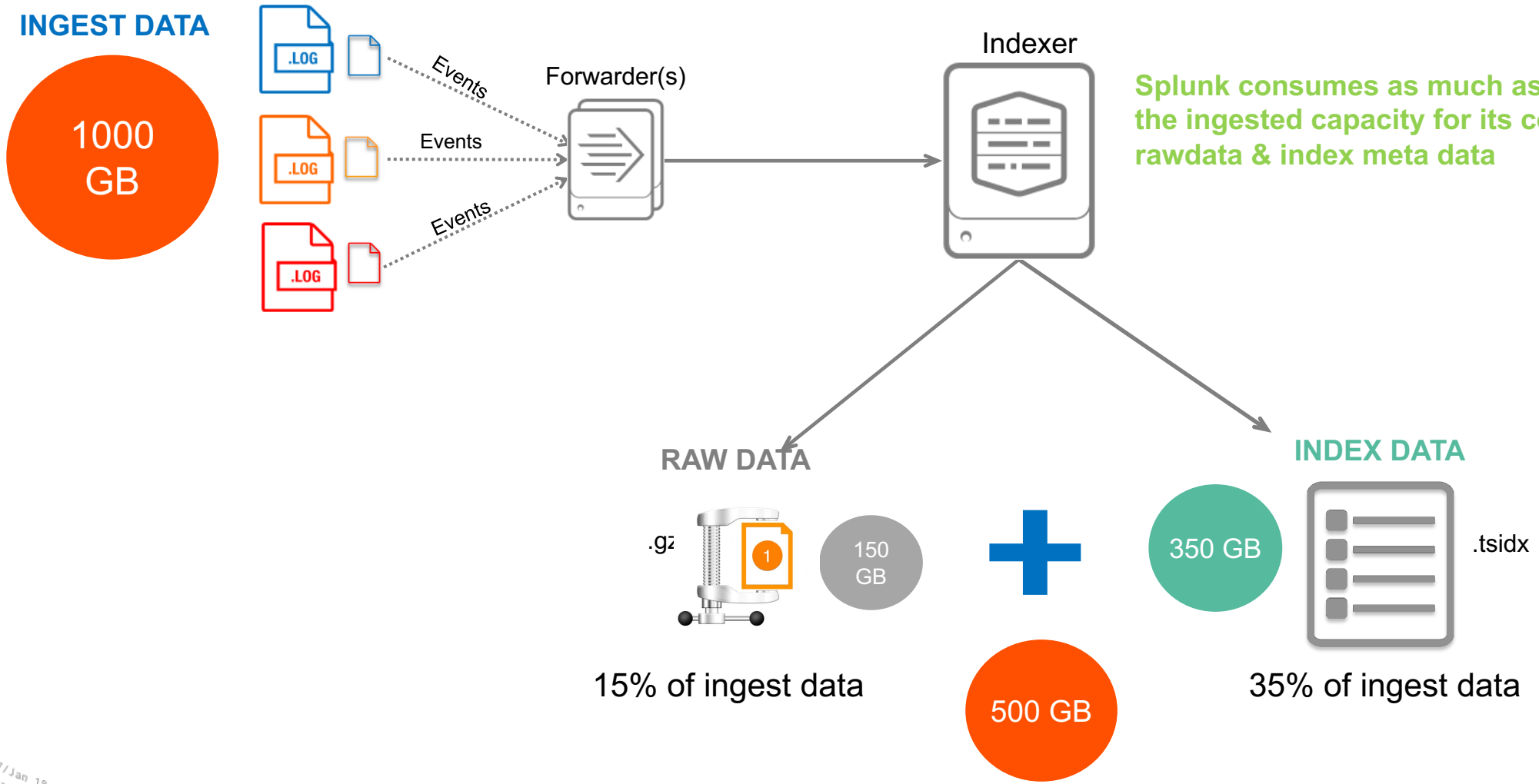
SCALABILITY
LIMITED BY EXPANSION SLOTS
HIGHLY SCALABLE

DATA PROTECTION
MANUAL RAID
RAID-HA

HOST USAGE
HOST CPU OVERHEAD
NO HOST CPU OVERHEAD

DATA SERVICES
NO DATA REDUCTION, ENCRYPTION
INCLUDED AND ALWAYS ON

How Does Splunk Indexer Store Data?



1 – For syslog type of data. Splunk reduction rate can vary based on the type of data

Pure's Value With Splunk Data!

On Pure FlashArray

Replicated rawdata (.gz) files are exact copy on peer nodes and are deduplicated on Pure FlashArray

Index files (.tsidx) are logical copy on peer nodes. While they don't benefit on deduplication they get an average of 3 to 1 compression on Pure

Data services like encryption, data reduction, replication and snapshots, for free on Pure FlashArray

Indexer Cluster Storage Usage

1 TB

Replication Factor (RF): 3

Search Factor (SF): 2

1150 GB

382 GB

Data reduction → 2.6 : 1

splunk> .conf18

Splunk Space Usage[1] on Pure

For 1 TB of INGEST DATA

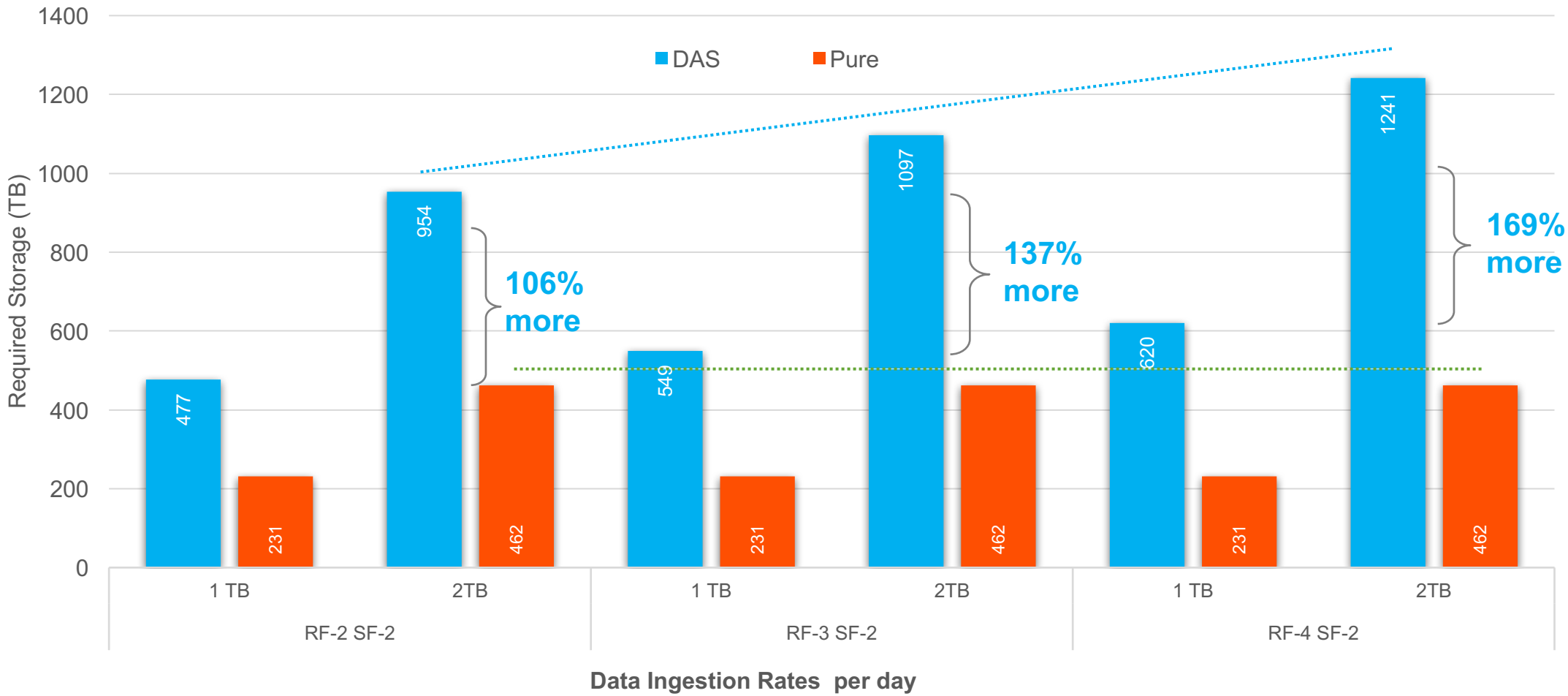
	No RF/SF	RF-2, SF-2	RF-3, SF-2*	RF-4, SF-2
RAW DATA	150	150 150	150 150 150	150 150 150 150
INDEX DATA	350	350 350	350 350	350 350
SAVED DATA	500 GB	1000 GB	1150 GB	1300 GB
	150	150	150	150
	116	116 116	116 116	116 116
	266	382	382	382

1 - Space usage for one day of ingest and not considering Hot/Warm or Cold tiers

2 - Splunk recommends RF-3, SF-2

Raw Storage Requirements Comparison

Pure FlashArray vs Direct Attached Storage (DAS) for 90 days of Hot/Warm and 270 days of Cold = 360 days total

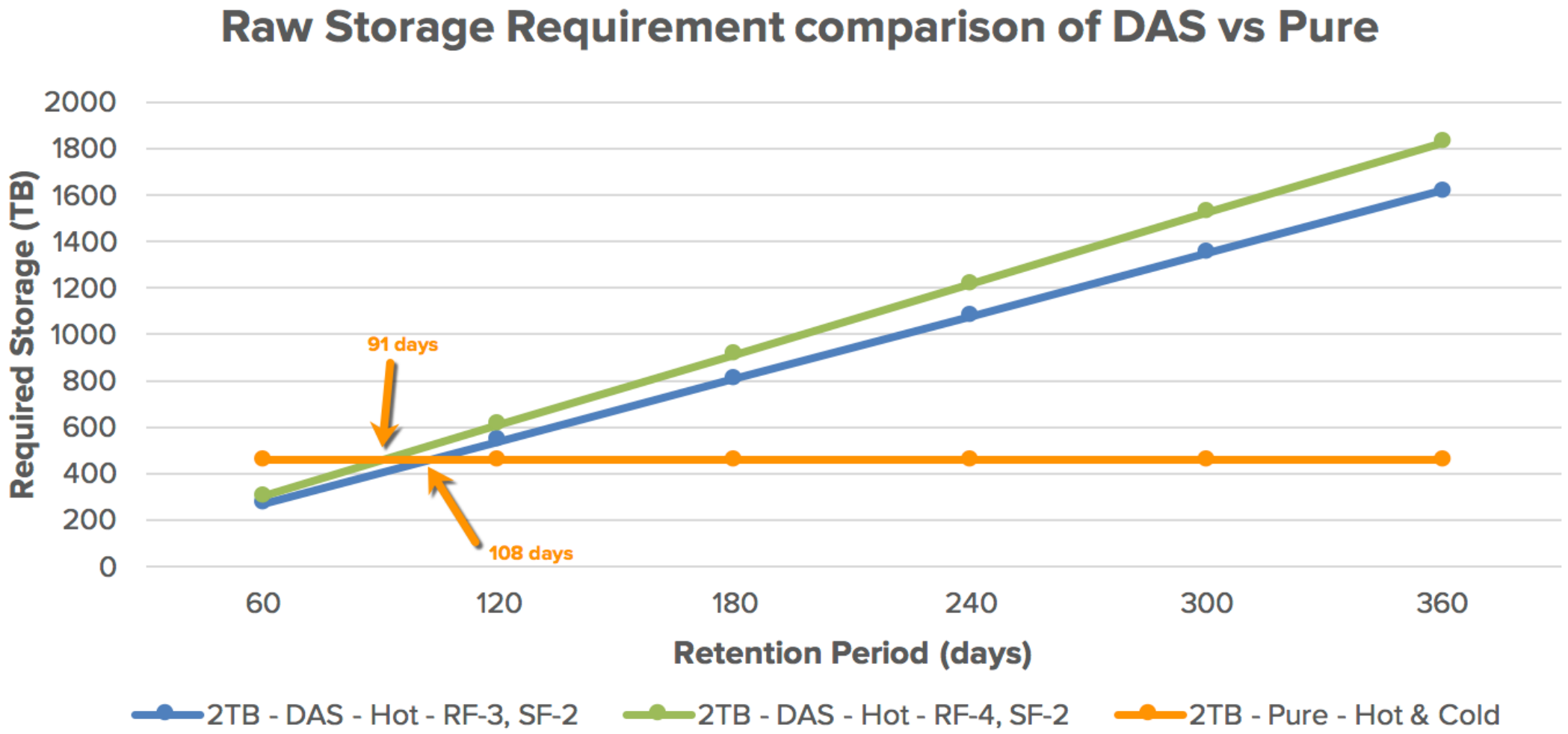


Direct Attached Storage space usage includes standard Splunk data reduction through compression

splunk> .conf18

Raw Storage Requirements Comparison

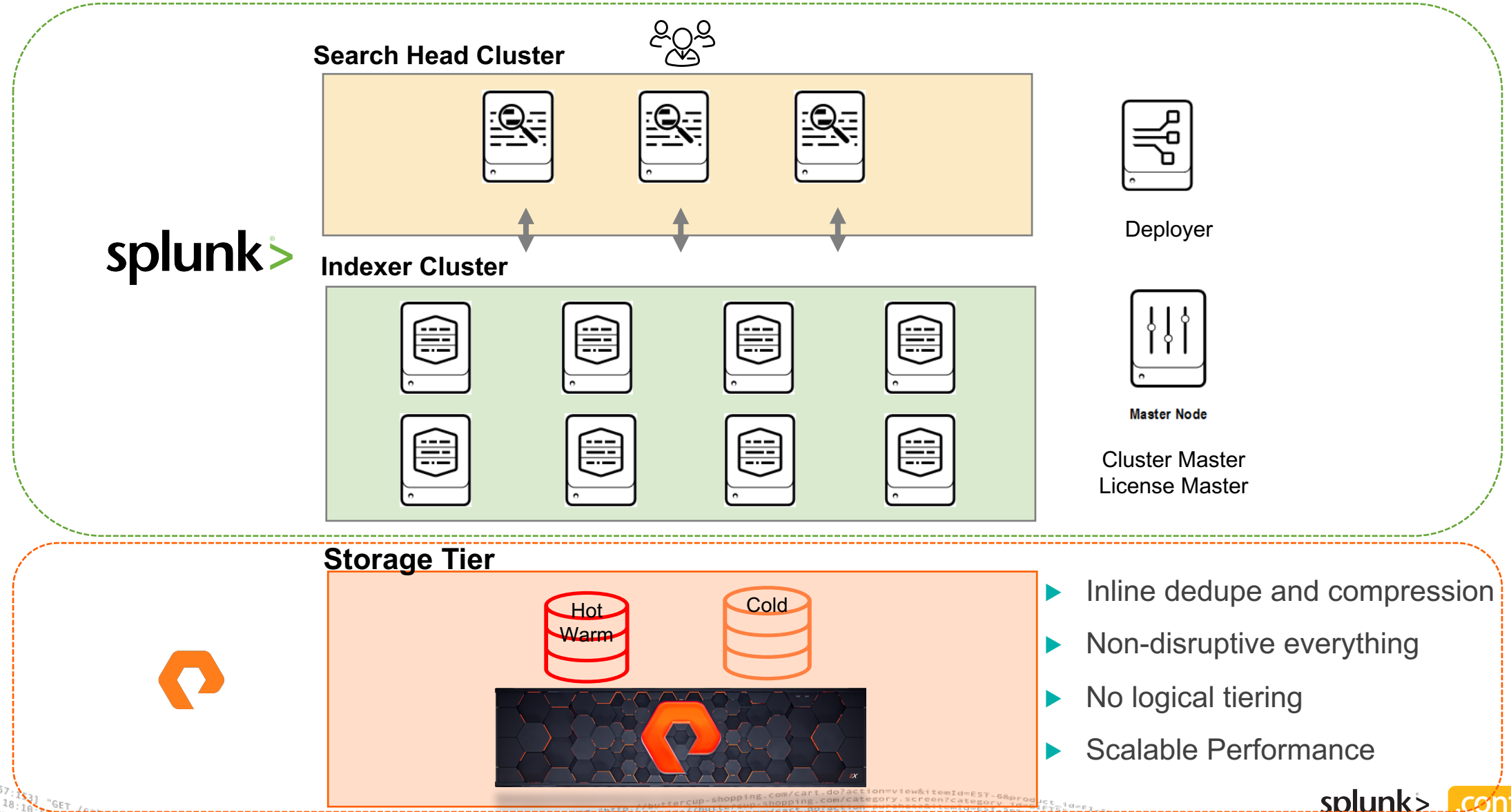
Pure FlashArray vs Direct Attached Storage (DAS) for 90 days of Hot/Warm and 270 days of Cold = 360 days total



splunk> .conf18

Flashstack for Splunk – Ref Arch

Logical Architecture using Pure FlashArray



- Inline dedupe and compression
- Non-disruptive everything
- No logical tiering
- Scalable Performance

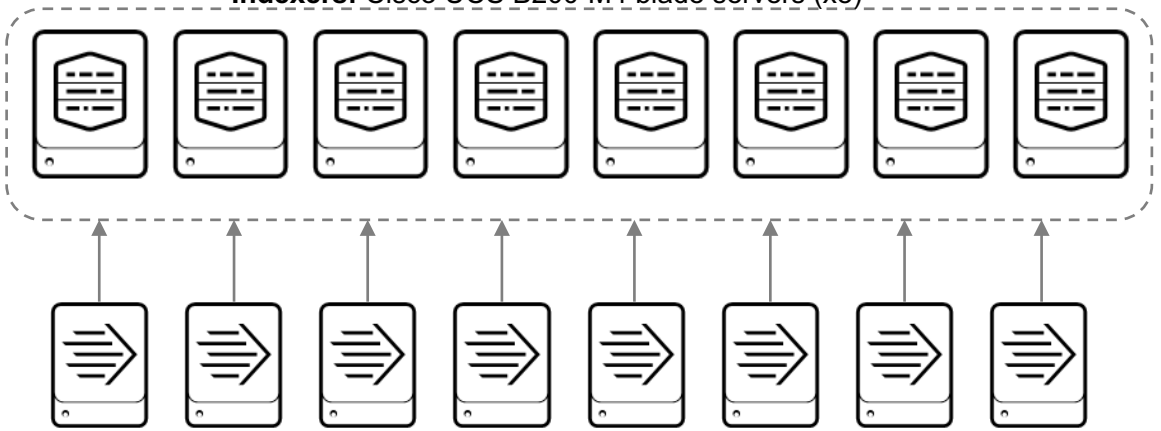
splunk> .conf18

Flashstack for Splunk – Ref Arch

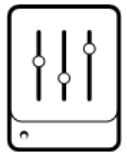
Physical Architecture using Pure FlashArray

Search Heads:
Cisco UCS B200-M4 blade servers (x3)

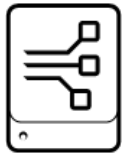
Indexers: Cisco UCS B200-M4 blade servers (x8)



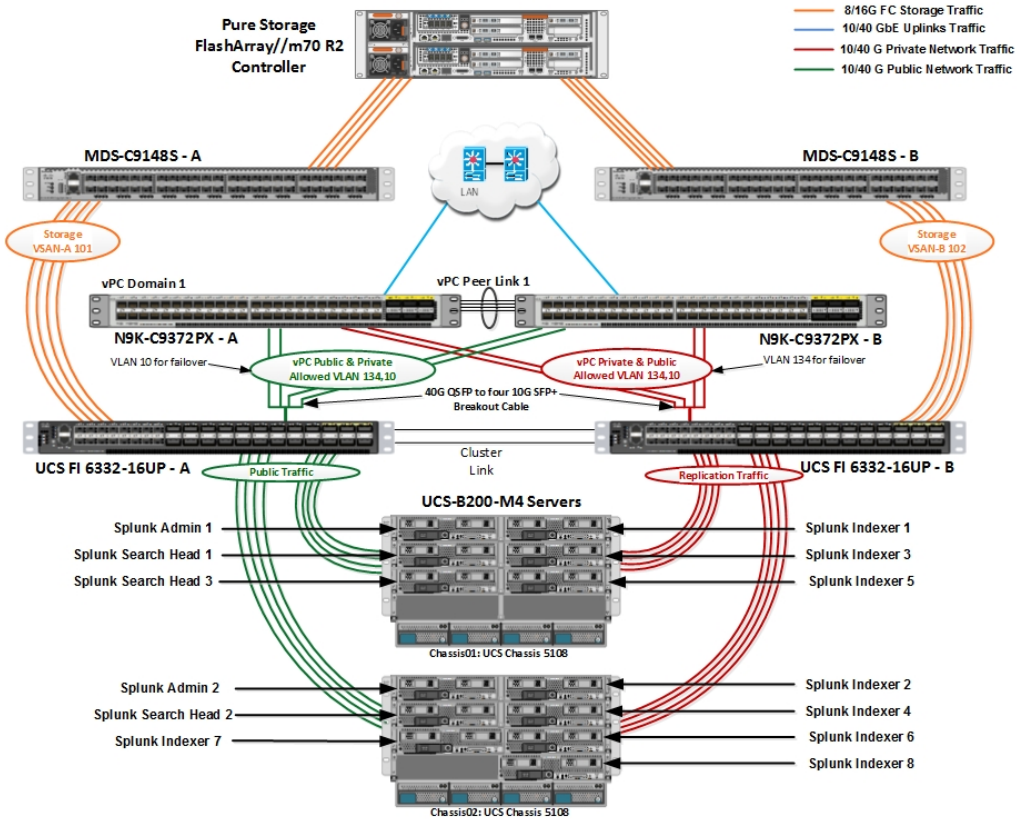
Forwarders: Cisco UCS C240-M4 Rack servers (x8)



Master Node:
Cisco UCS B200-M4 blade server



Deployment Server:
Cisco UCS B200-M4 blade server

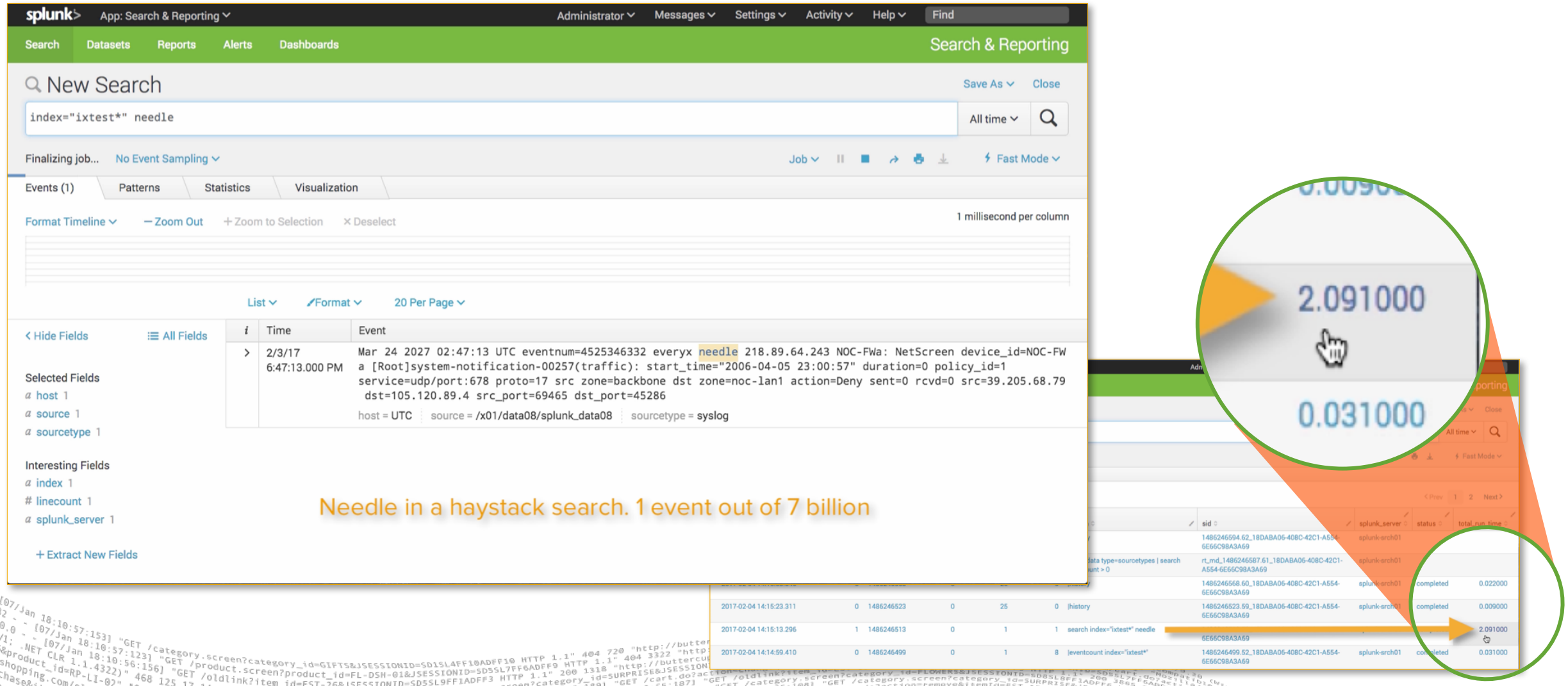


https://support.purestorage.com/Solutions/Splunk/Reference/FlashStack_CI_for_Splunk_Reference_Architecture

splunk> .conf18

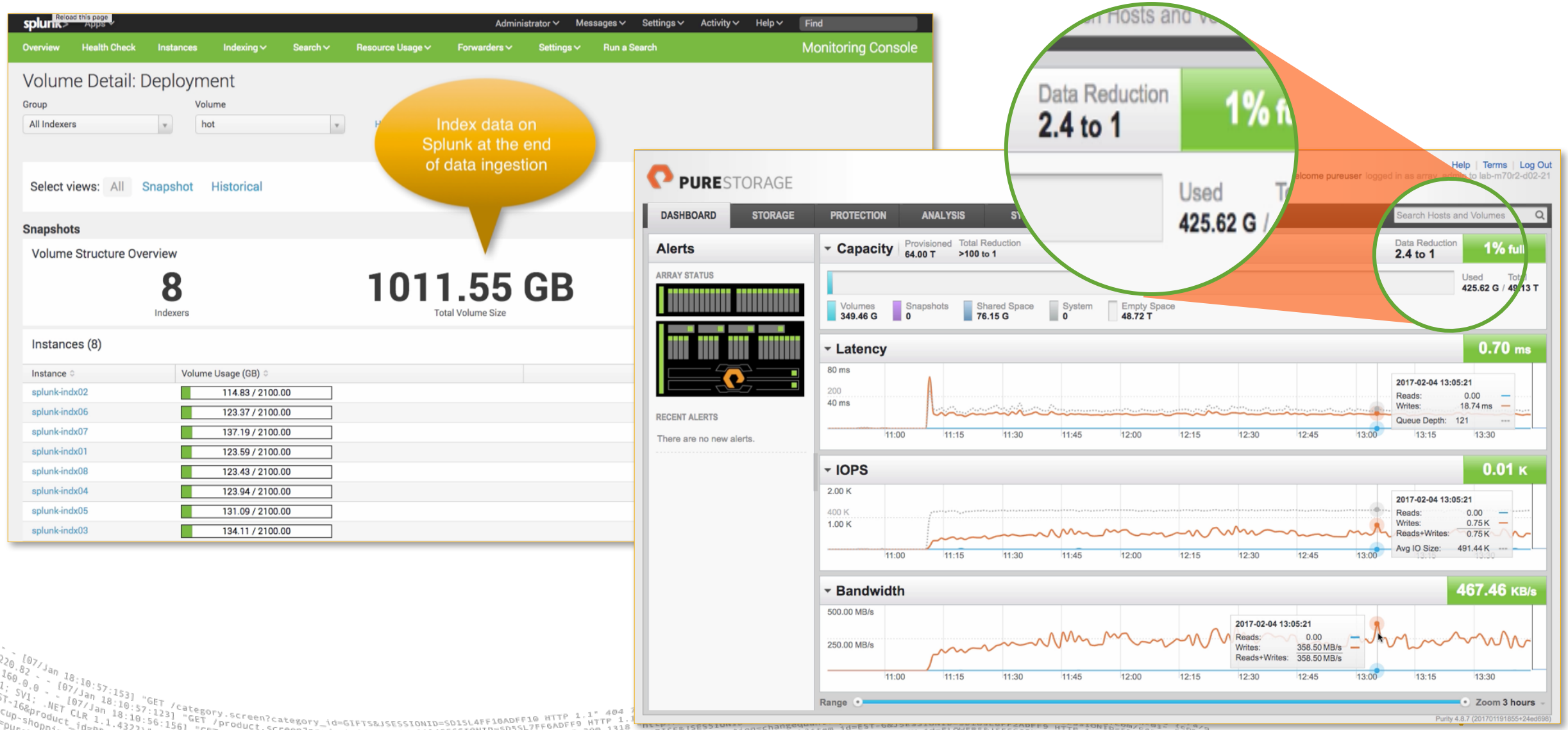
Searching 7 Billion Events in 2 Seconds

Finding A Needle In A Haystack:



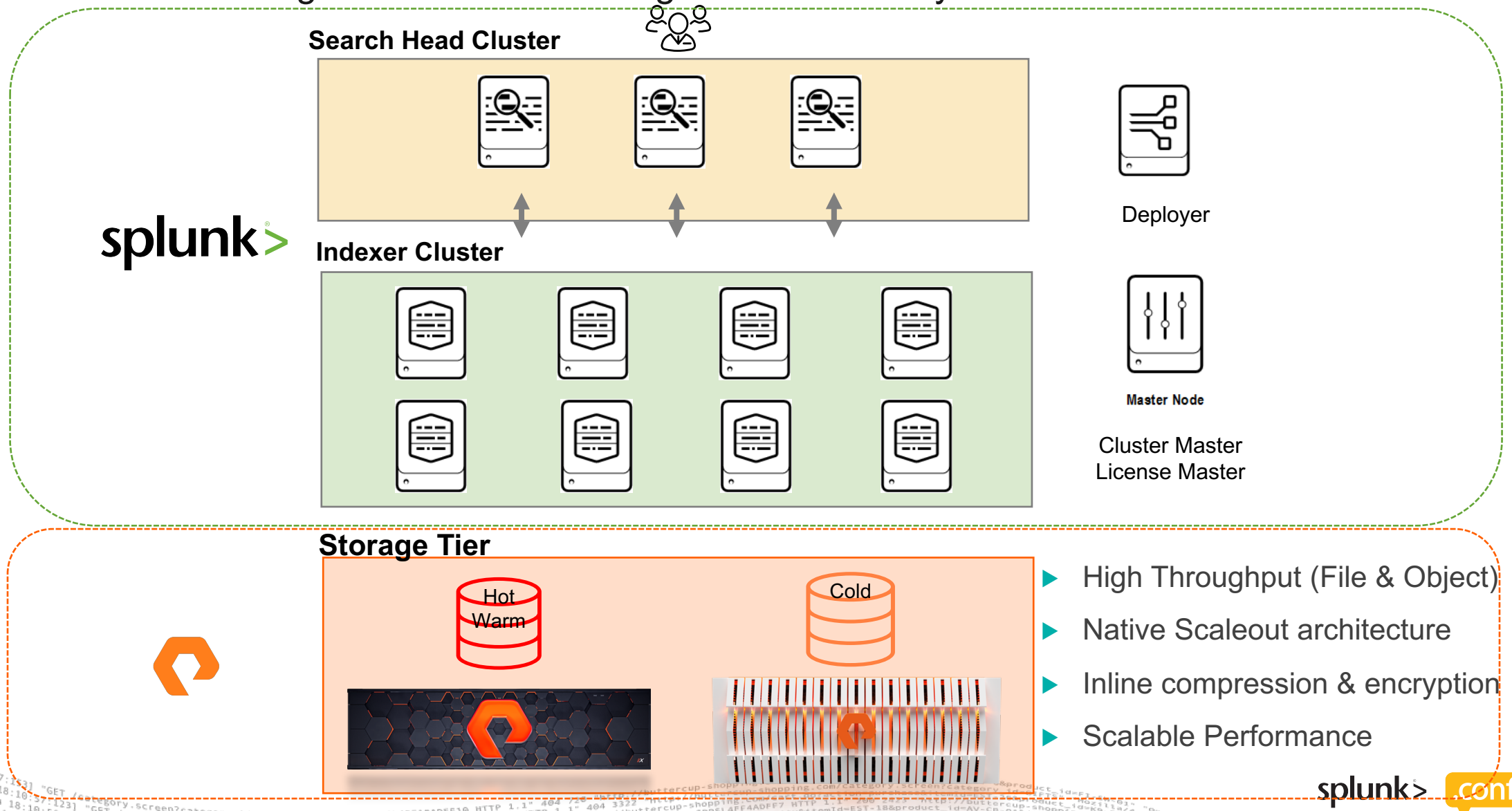
2-3x Data Reduction

Reducing Splunk's Replication Factor & Search Factor



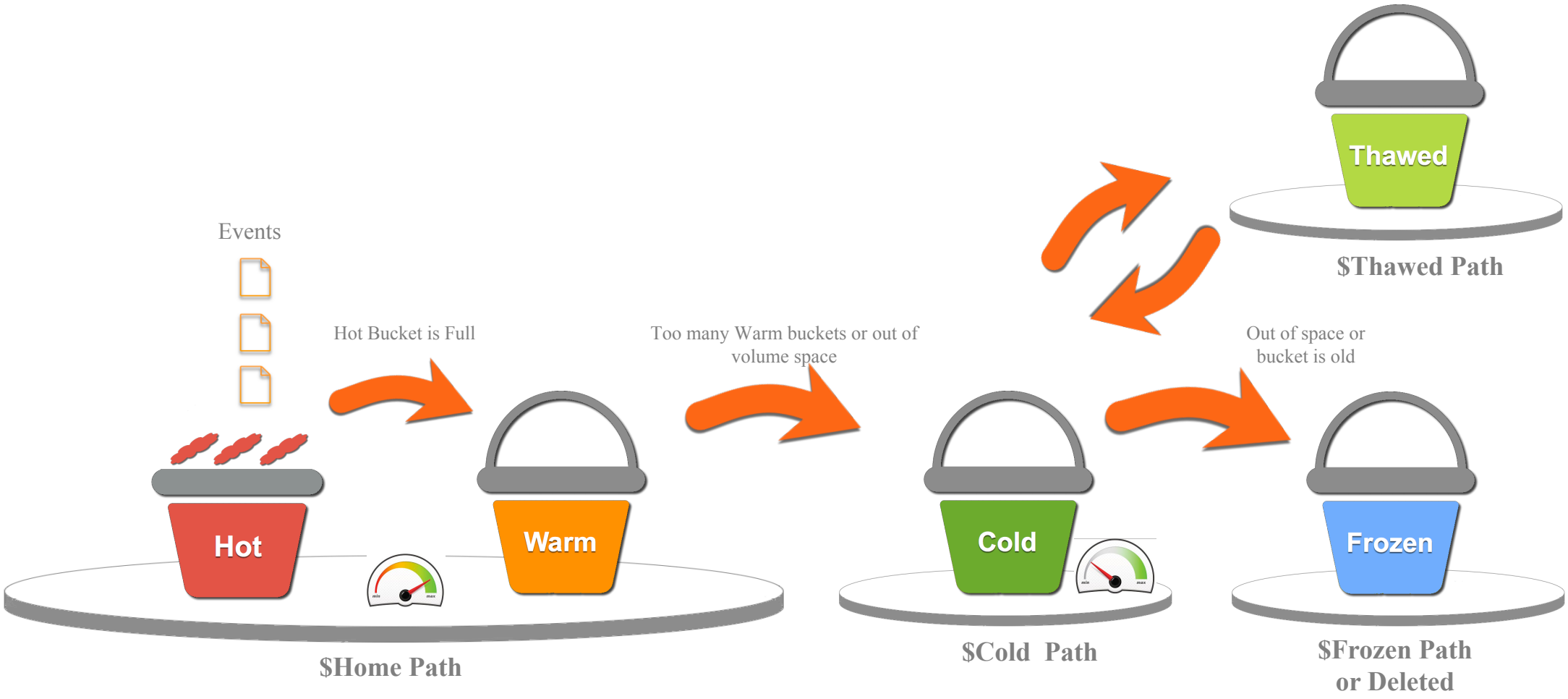
Upcoming Flashstack for Splunk – Ref Arch

Logical Architecture using Pure FlashArray & FlashBlade



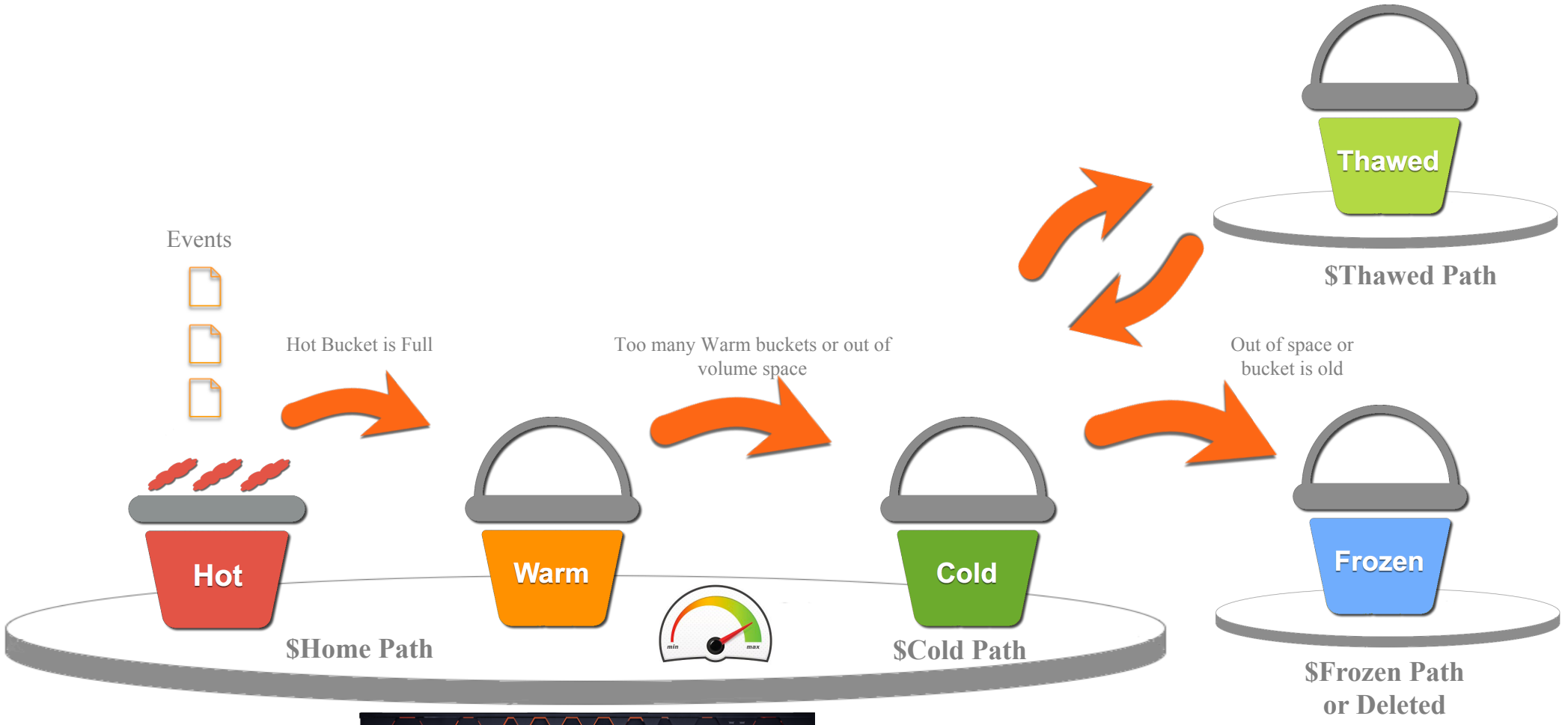
Splunk Buckets Across Tiers

How Splunk stores data?



Splunk Buckets on Pure Storage #1

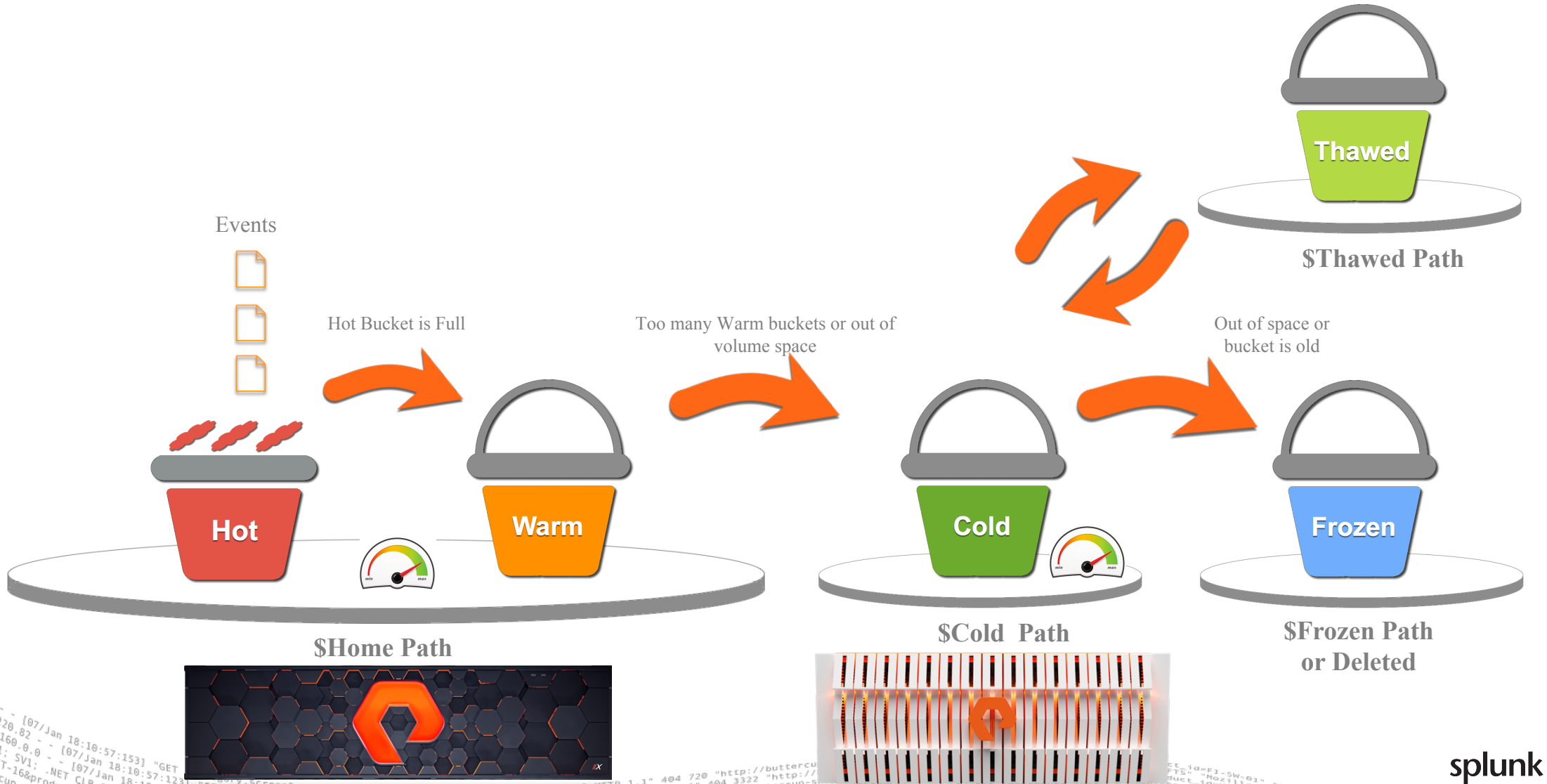
Eliminate logical tiering between Hot/Warm and Cold



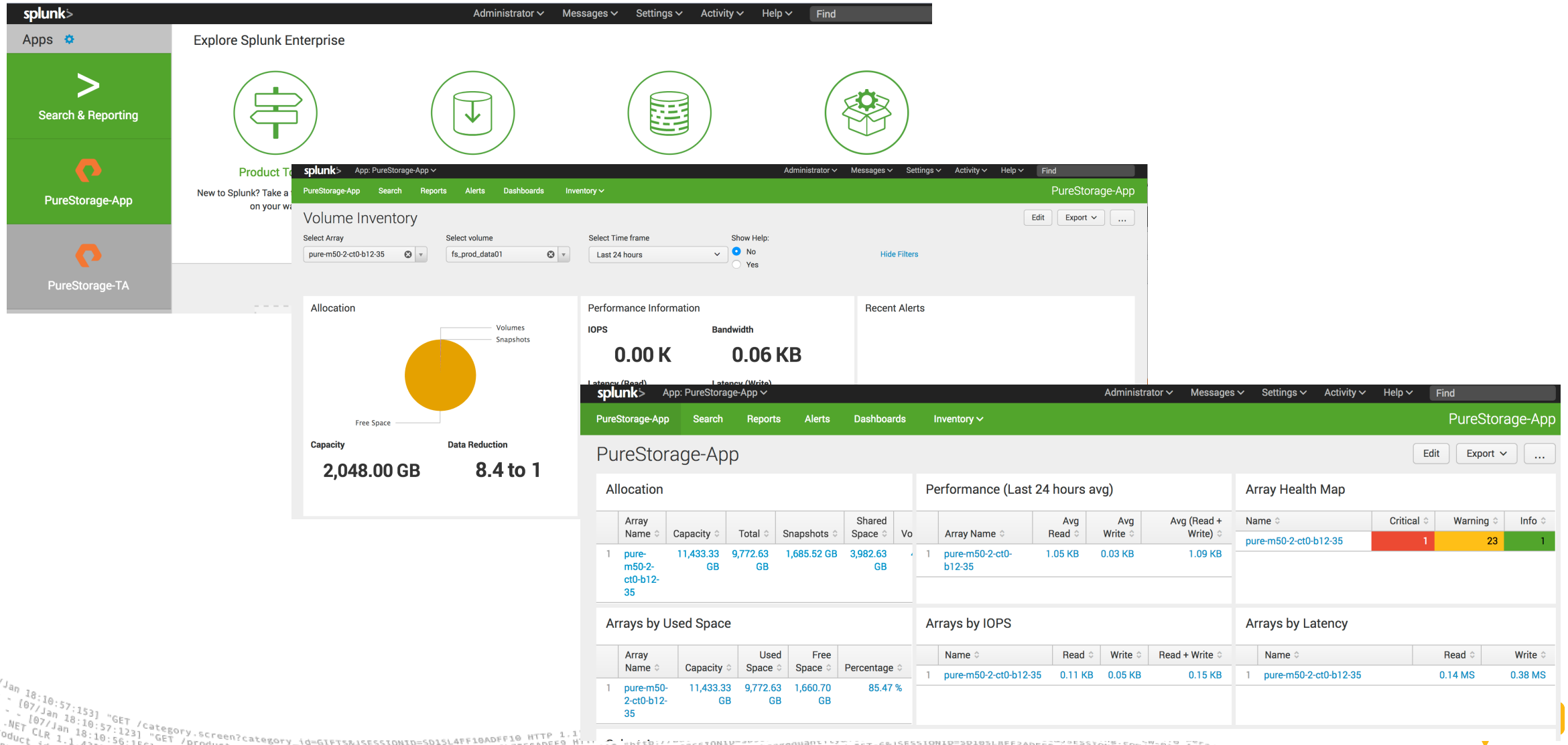
splunk> .conf18

Splunk Buckets on Pure Storage #2

Gain scalable capacity at better performance



Splunk App for Pure Flasharray



Splunk App for Pure Flasharray

PureStorage-App | Search | Reports | Alerts | Dashboards | Inventory | PureStorage-App

Array Inventory

Edit | Export | ...

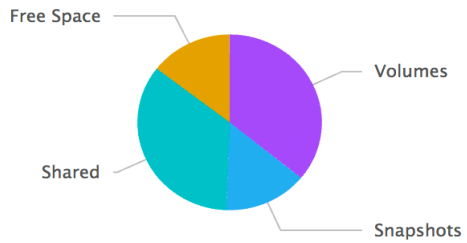
Select Array: pure-m50-2-ct0-b12-35

Select time frame: Last 24 hours

Show Help: (•) No ( ) Yes

Hide Filters

Allocation



Capacity
11,433.33 GB

Data Reduction
8.4 to 1

Performance Infomation

IOPS
1.09 K

Bandwidth
4.50 KB

Latency (Read)
0.14 ms

Latency (Write)
0.44 ms

Recent Alerts

No results found.

Volumes by Used Space

	Volume Name	Capacity	Used Space	Free Space	Percentage
1	ds-splunk-data04	3,072.00 GB	1,137.82 GB	1,934.18 GB	37.04 %
2	ds-splunk-data01	3,072.00 GB	945.83 GB	2,126.17 GB	30.79 %
3	ds-splunk-data02	3,072.00 GB	890.55 GB	2,181.45 GB	28.99 %

Volumes by IOPS

	Name	Read	Write	Read + Write
1	fs-ebs-2clone600-db01	47.79 k	0.00 k	47.79 k
2	ds-splunkidx-data02	41.56 k	0.00 k	41.56 k
3	ds-splunkidx-data02	19.59 k	0.00 k	19.59 k
4	ds-splunk-idx-data	5.60 k	0.00 k	5.60 k
5	ds-splunk-data03	5.11 k	0.00 k	5.11 k
6	fs-boot-lun	4.59 k	0.00 k	4.59 k

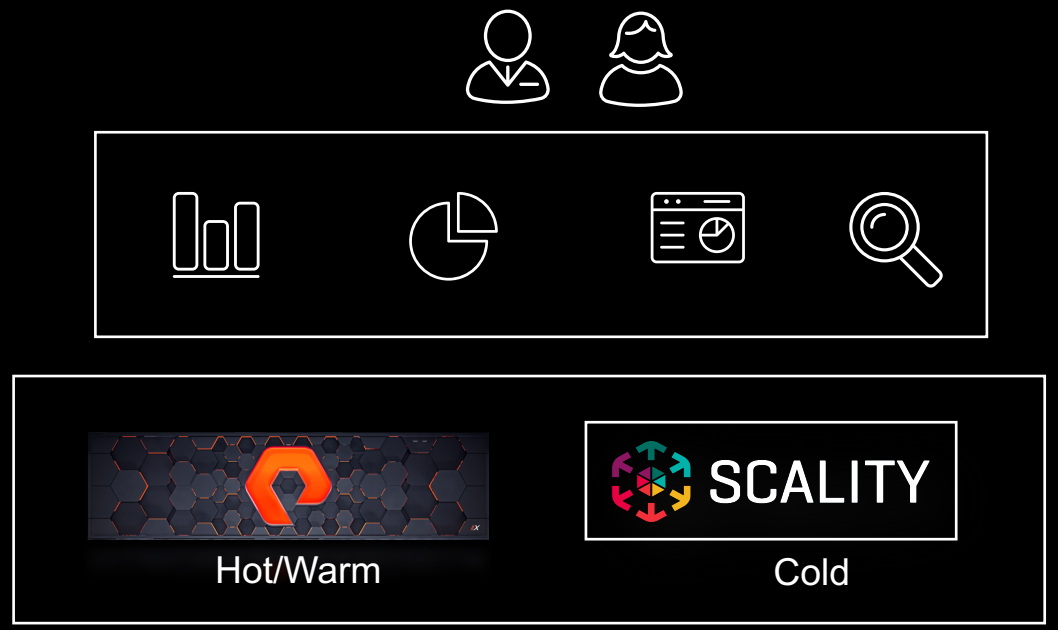
Volumes by Latency

	Name	Read	Write
1	ds-oel72asm	1.15 ms	0.00 ms
2	fs-ebs-clone200-db01	1.07 ms	0.43 ms
3	ds-splunk-data02	0.94 ms	0.44 ms
4	fb-dss-infra	0.90 ms	0.55 ms
5	ds-splunk-idx-data	0.73 ms	0.29 ms
6	fb-dss-infra	0.70 ms	0.55 ms

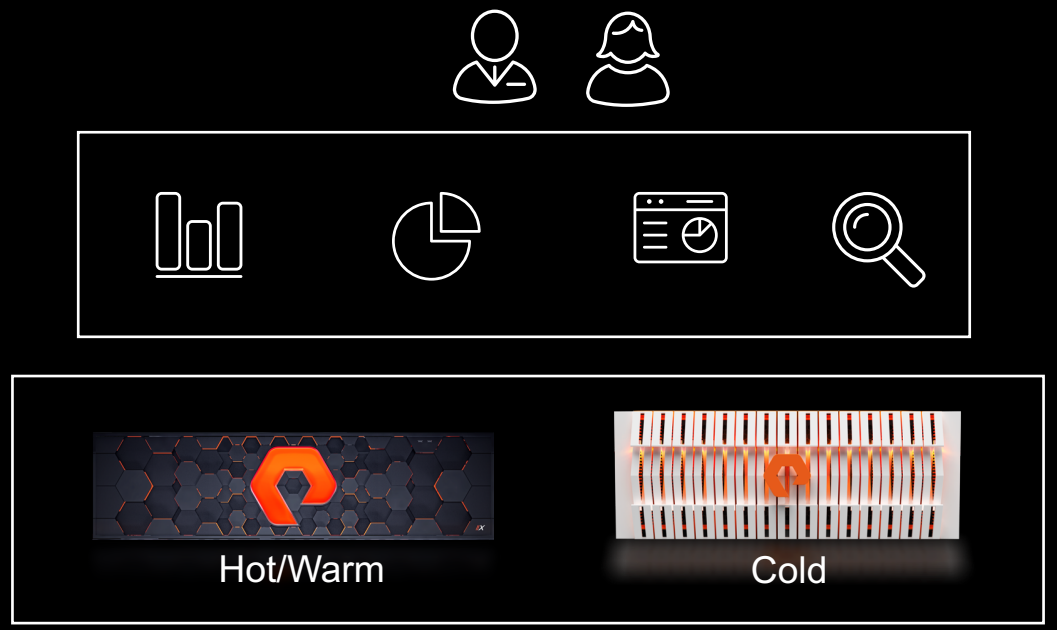
Use Cases

There is No Cold Data in Analytics

A Joint Pure/Splunk Customer



THEN **VS** NOW



- Requirement to query historical data frequently
- Searches across Cold data hung
- Significant Data Growth
- Inefficiencies & added complexity of scaling their Cold tier
- Resulted in poor Customer Satisfaction

- Searches across Cold data performed faster
- Cold is also effectively the Hot data
- Capacity can be added seamlessly
- Improved Customer Satisfaction

splunk> .conf18

Splunk on Pure: Healthcare

Use Case:

- Multi-site, ingest of 1TB/day
- Splunk was setup on DAS storage

Business Value

- Ability to standardize for Core IT and Security

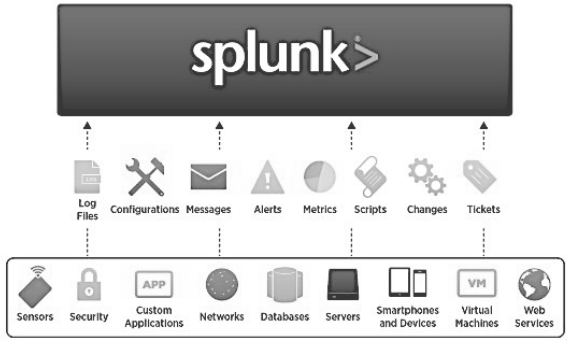
Technical Value

- Ease of scale and flexibility while providing best in class performance
- At-rest encryption for the Security team

Splunk on Pure: Credit Union

Use Case:

- Splunk, VDI and VMware all supported by FA//m20



vmware Horizon View

vmware

Business Value

- Simplicity and scalability

Technical Value

- Ability to consolidate workloads
- Data reduction keeps footprint small (3U)

splunk> .conf18

Splunk on Pure Flasharray

Better Performance

- Faster searches across longer period of time (No tiering)
- Scale compute and storage independently

Scalable Capacity

- Linear scalability of capacity and performance
- Better data reduction means more capacity

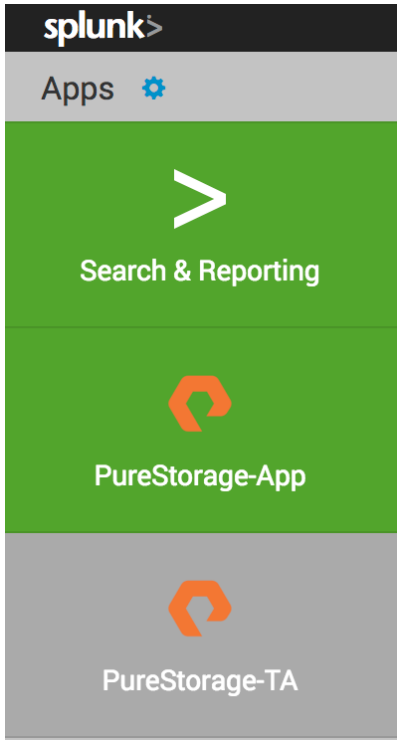
Data Services

- Always on Encryption
- Always on deduplication, compression
- Snapshots for Backups

Splunk Resources

FlashStack for Splunk Reference Architecture

https://support.purestorage.com/Solutions/Applications/Splunk



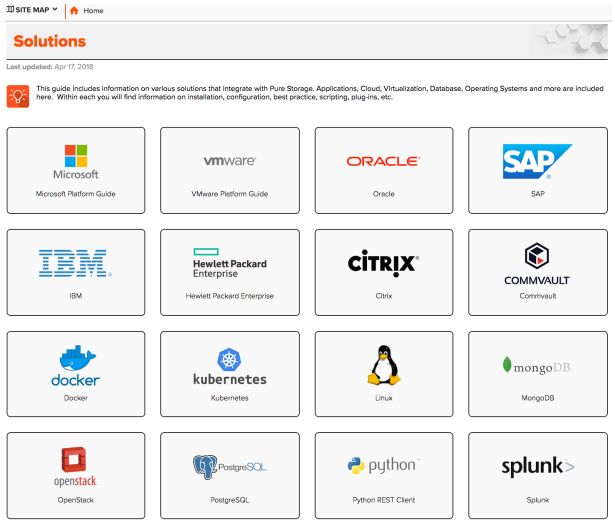
Splunk App for Pure FA

TA - https://splunkbase.splunk.com/app/3659/
Apps - https://splunkbase.splunk.com/app/3660/

splunk> .conf18

More Info

https://support.purestorage.com



Splunk Solutions Page on Pure

https://support.purestorage.com/Solutions/Splunk

Somu Rajarathinam

@purelydb

www.somu.us

https://github.com/rsomu/

somu@purestorage.com

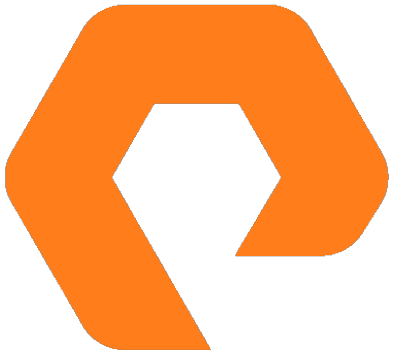
AUTODESK Inspirage logitech ORACLE

Splunk on Pure Storage

http://www.purestorage.com/splunk

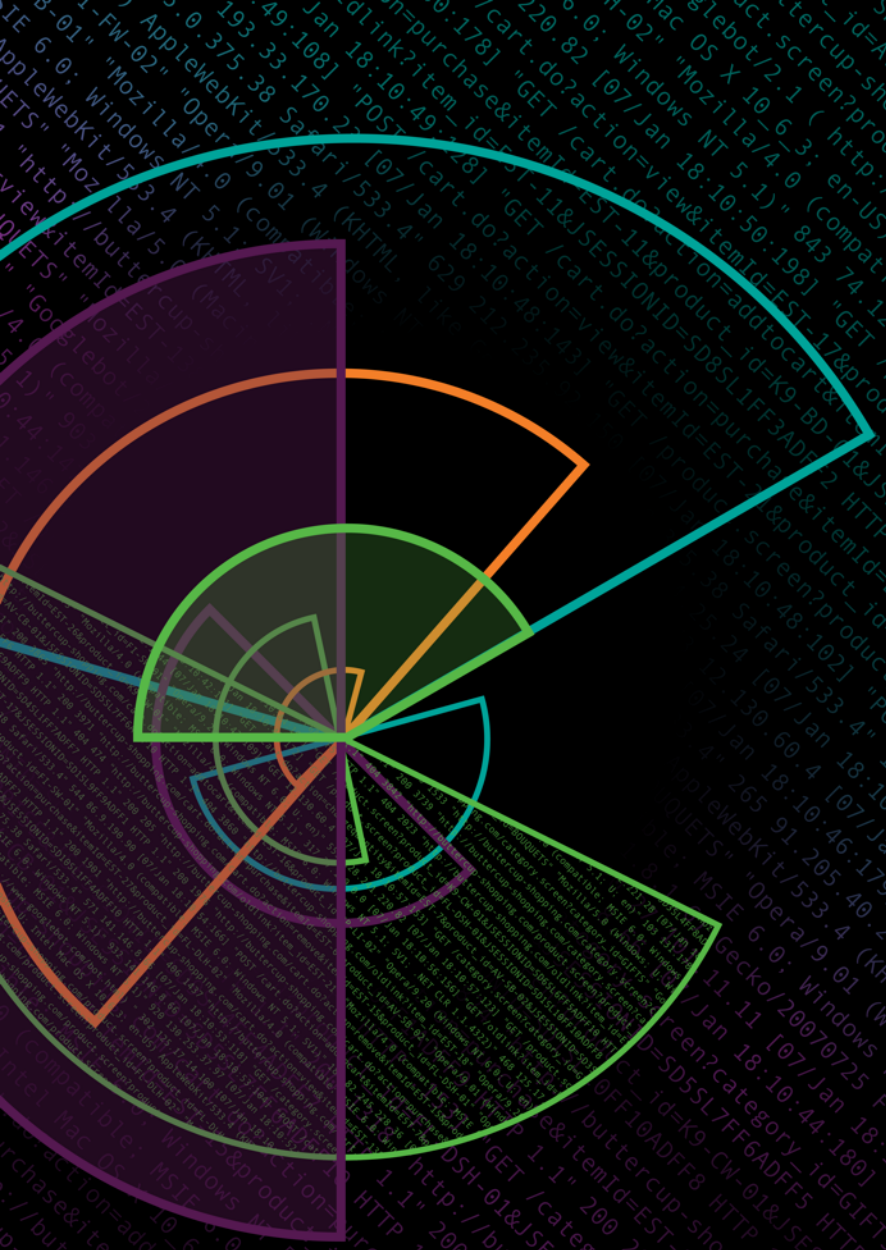
splunk> .conf18

For more information



Visit PURE Storage booth at

SOURCE = *PAVILION



Q&A

splunk> .conf18

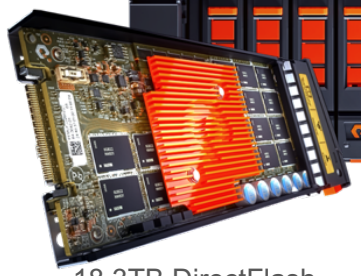
Thank You

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

FlashArray //X

The Market-defining All-flash Array: Now In 100% Nvme

512TB DirectFlash Shelf



18.3TB DirectFlash Module

100% NVMe DENSITY:
1PB
IN JUST 3U

100% NVMe PERFORMANCE:
12 GB/s
BANDWIDTH

>_FC
>_iSCSI
>_VVOL
>_NFS
>_SMB

MULTI-PROTOCOL
BUILT-IN

99.9999%
AVAILABILITY & NDU EVERYTHING

DIRECTFLASH SHELF NOT CURRENTLY GA

splunk> .conf18

FlashBlade™ FB

The All-flash Data Hub For Modern Analytics

DENSITY
10+ PBs / RACK

BIG + FAST
80 GB/S BW
5M+ NFS OPS

CONVERGED
FILE & OBJECT

SIMPLE ELASTIC SCALE
JUST ADD A BLADE!