

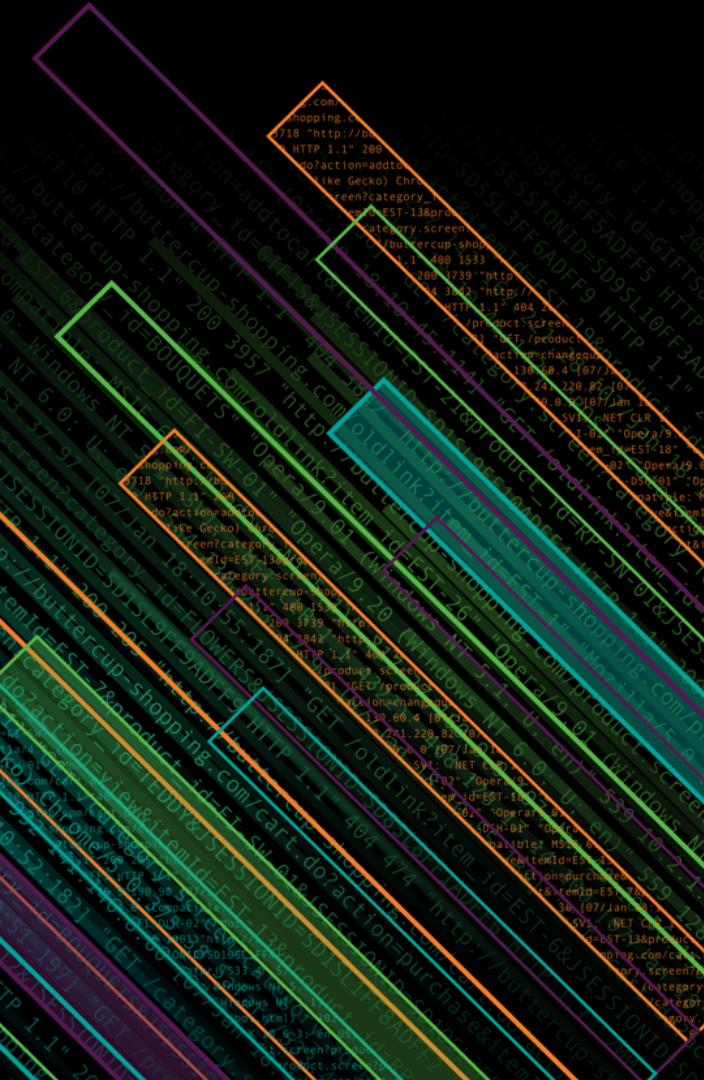


splunk> There Is No “Cold Data” In Analytics

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, Splunk>, Listen to Your Data, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2018 Splunk Inc. All rights reserved.

Dawn of 4th Industrial Revolution

Big Data, Modern Analytics Driving Change In Every Industry



1st Revolution

1760-1820's Steam Power Rural to Urban



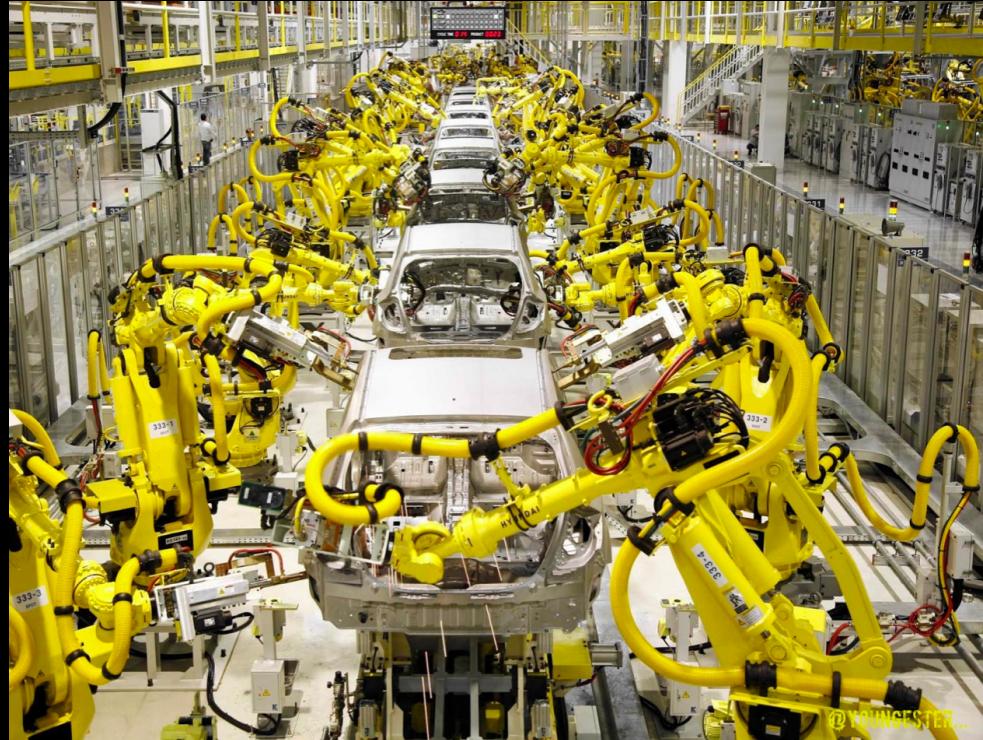
2nd Revolution

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

2020

50 ZB

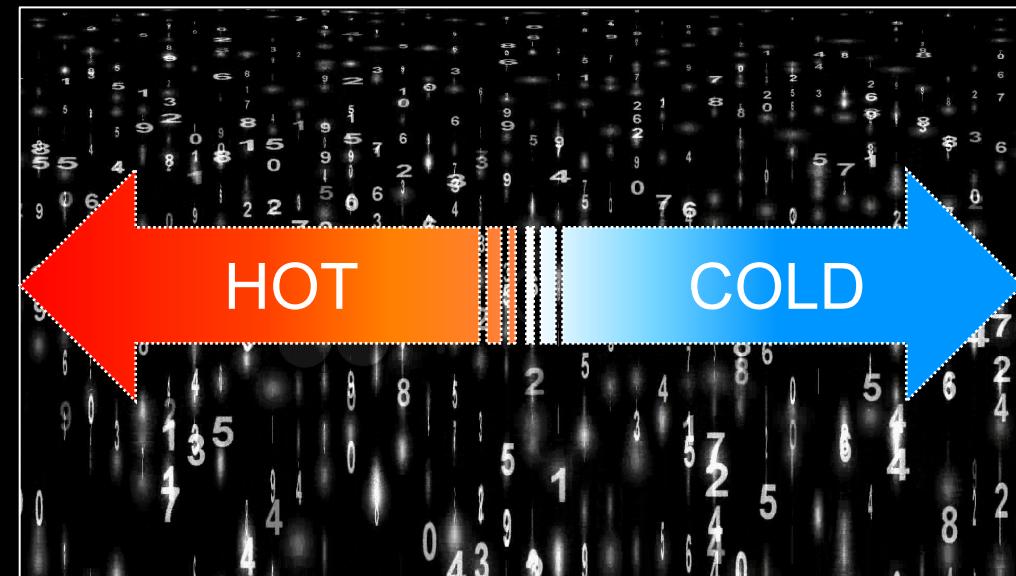
2025

163 ZB

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”

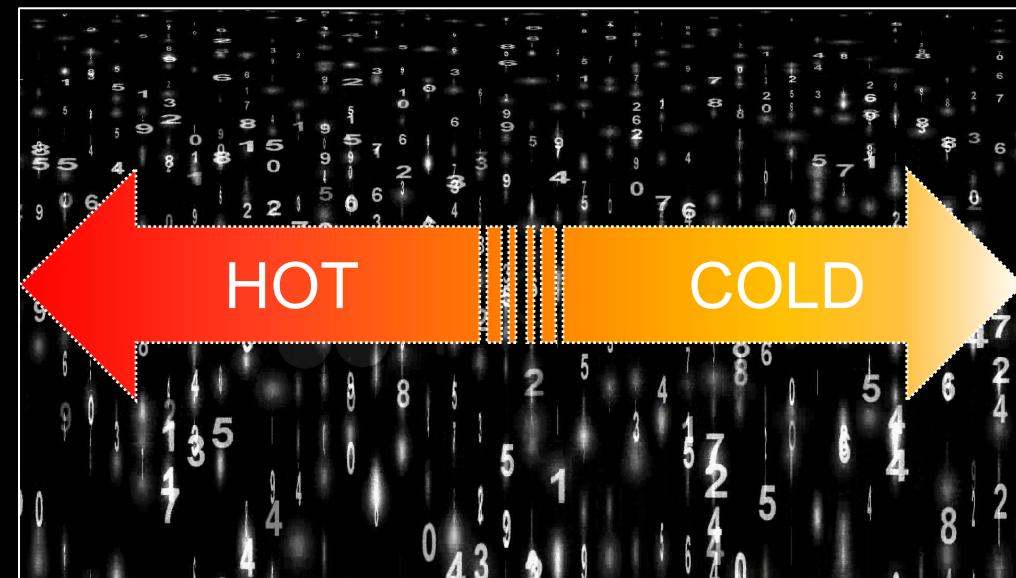


JET CLR 1.1.4200.1 Jan 18 10:56:23 2013 "GET /PERFORM/HS/

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



DATA SERVICES

NO DATA REDUCTION, ENCRYPTION

INCLUDED AND ALWAYS ON



SCALABILITY

LIMITED BY EXPANSION SLOTS

HIGHLY SCALABLE



DATA PROTECTION

MANUAL RAID

RAID-HA

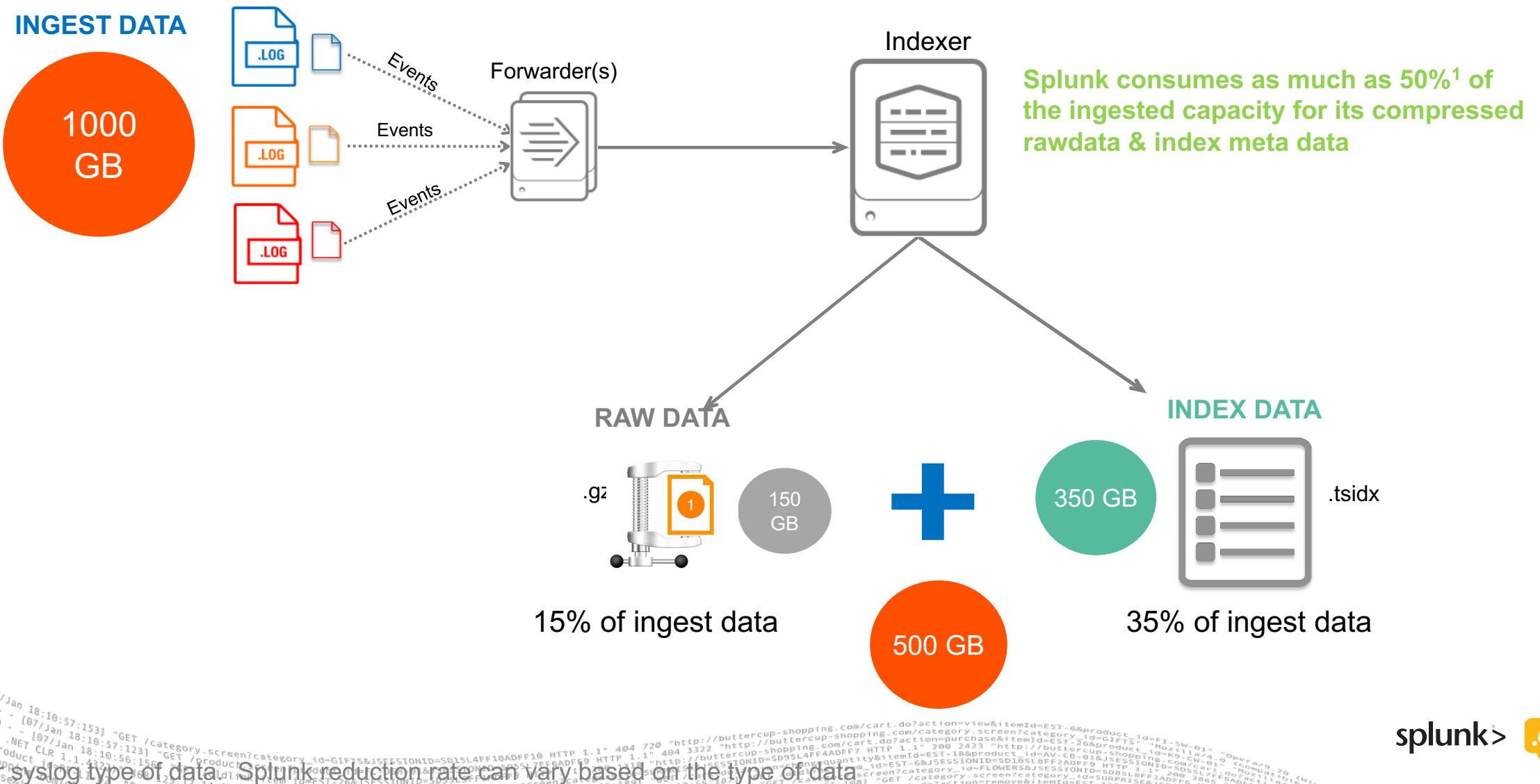


HOST USAGE

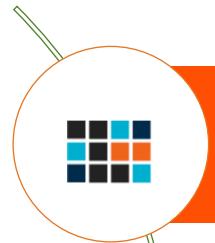
HOST CPU OVERHEAD

NO HOST CPU OVERHEAD

How Does Splunk Indexer Store 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



150G



150G



150G



350G



350G



1150 GB

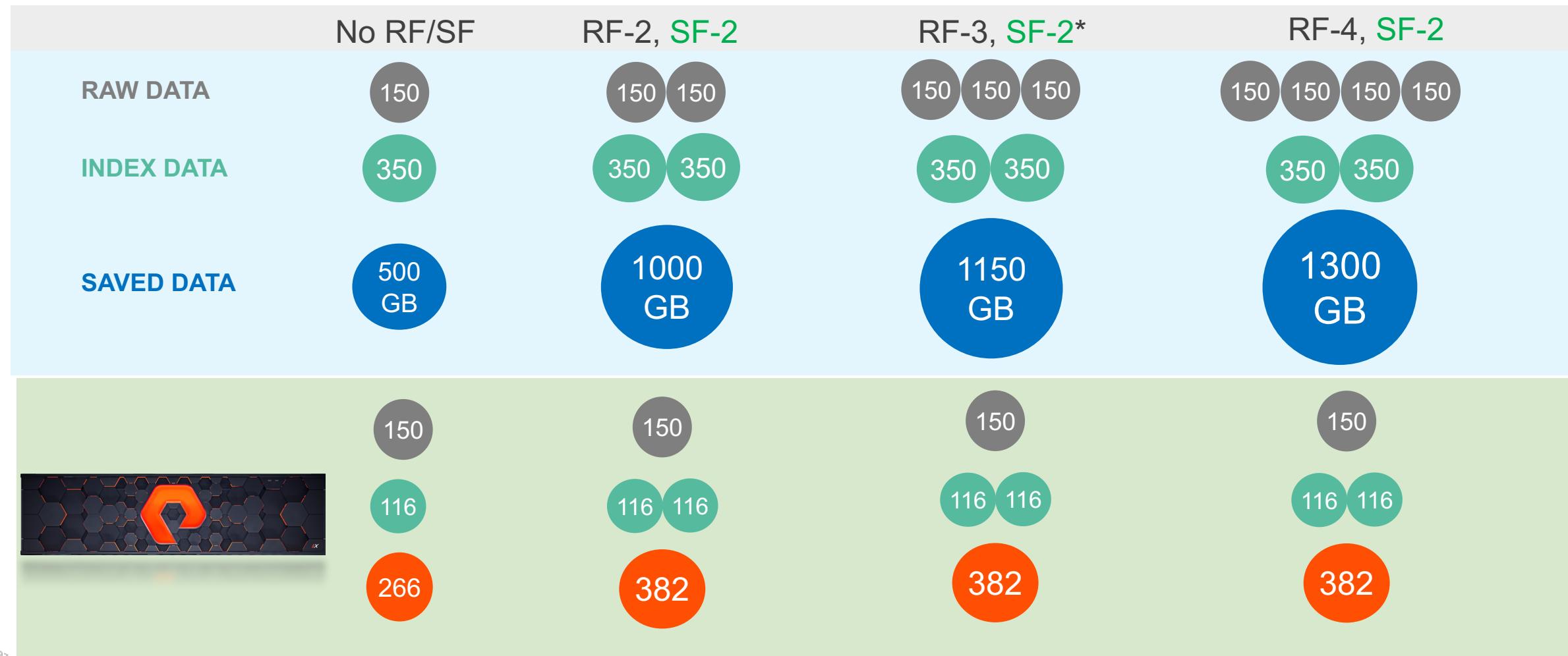


382 GB

Data reduction → 2.6 : 1

Splunk Space Usage¹ on Pure

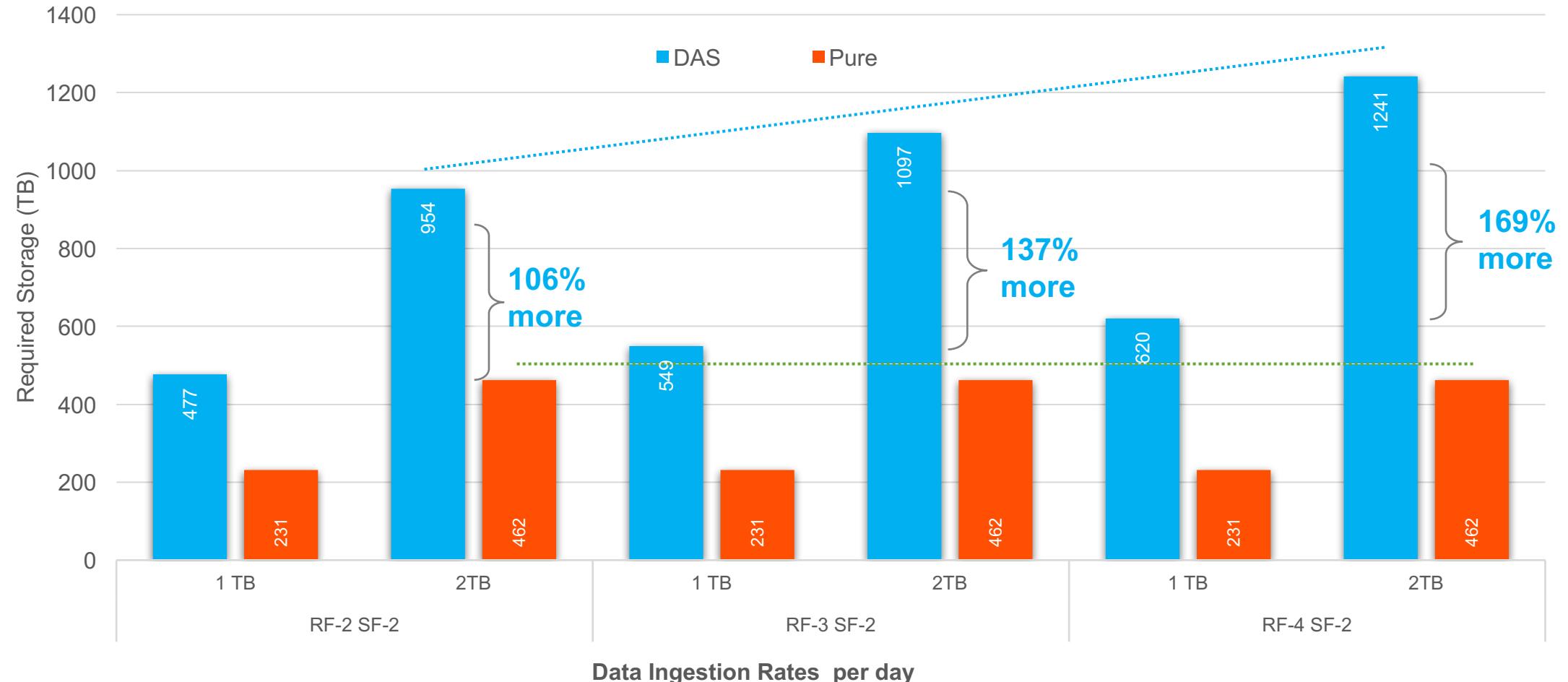
For 1 TB of INGEST DATA



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

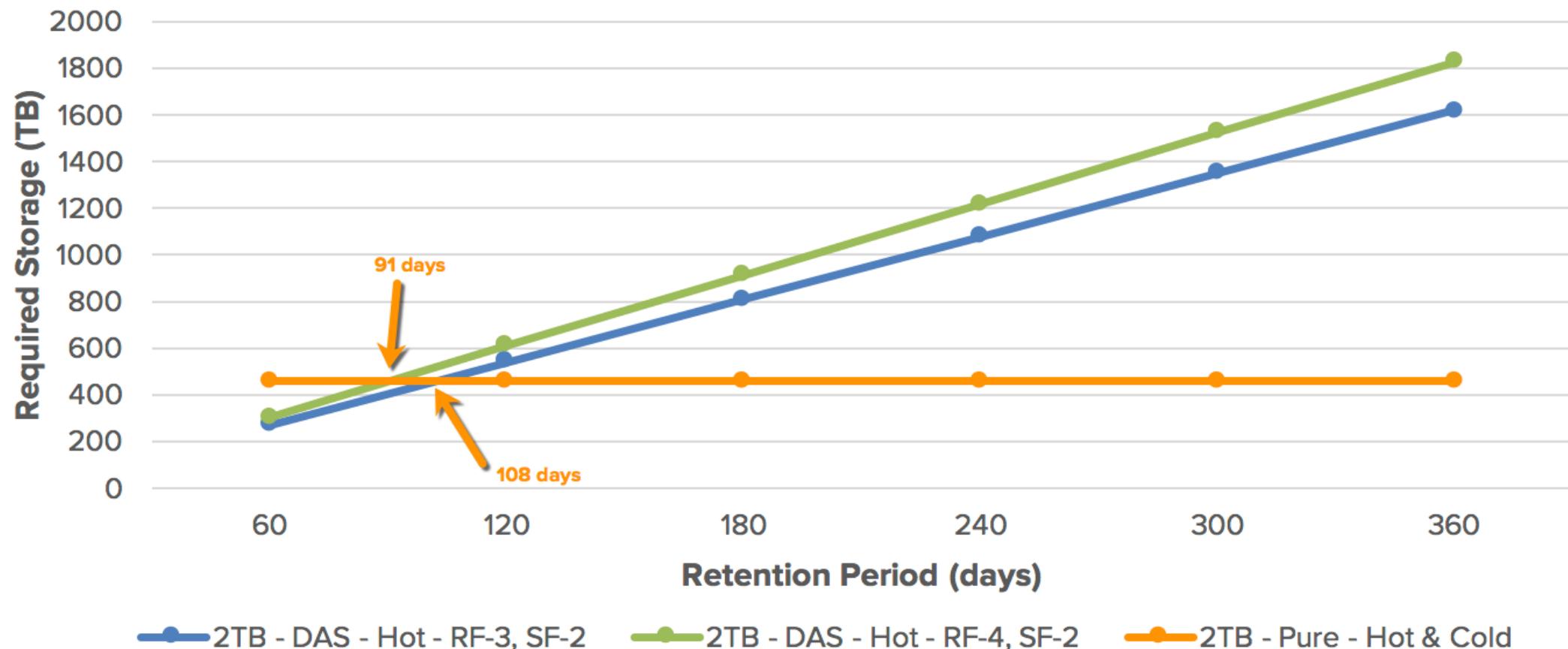


Direct Attached Storage space usage includes standard Splunk data reduction through compression.

Raw Storage Requirements Comparison

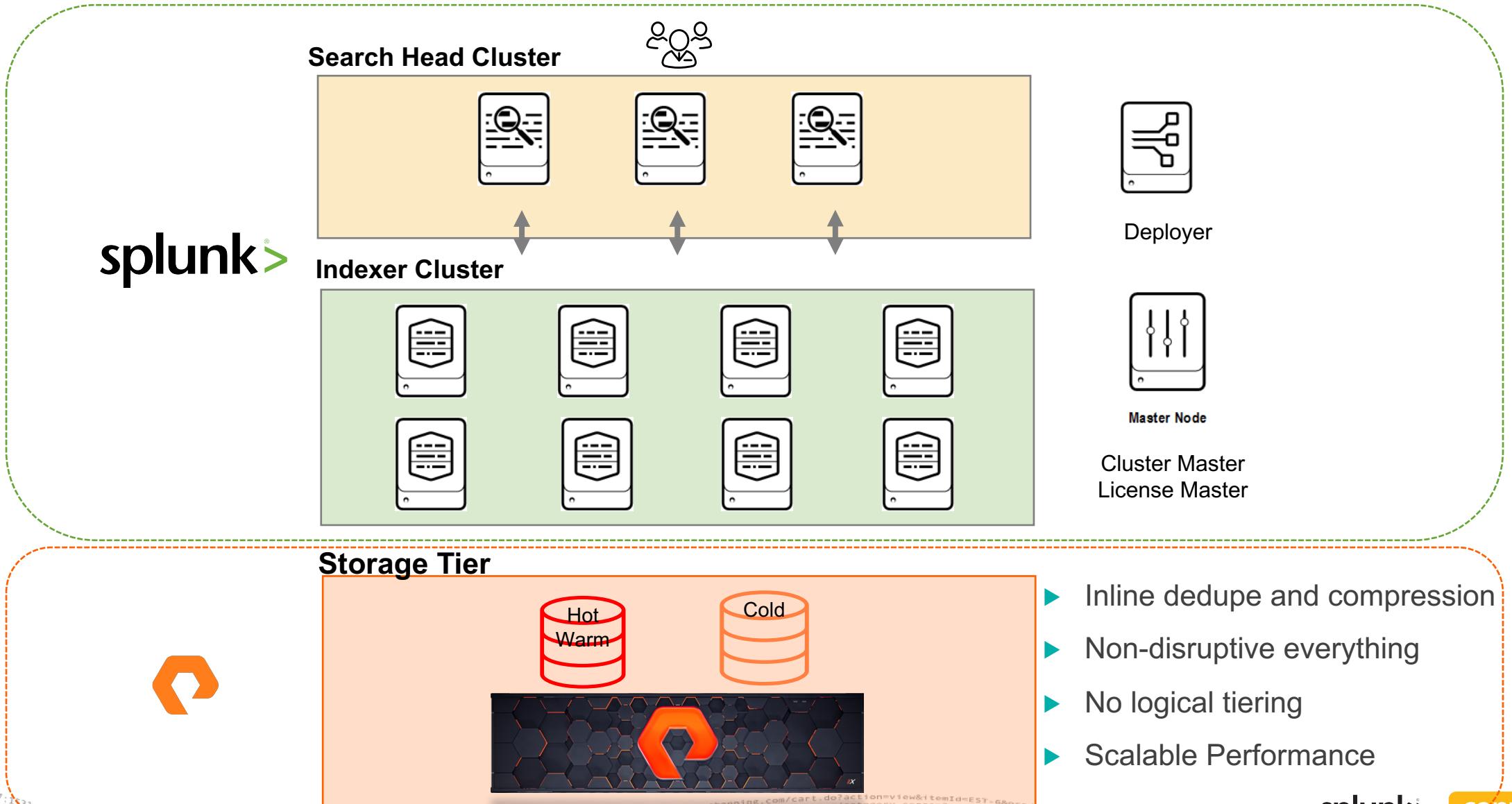
Pure FlashArray vs Direct Attached Storage (DAS) for 90 days of Hot/Warm and 270 days of Cold = 360 days total

Raw Storage Requirement comparison of DAS vs Pure



Flashstack for Splunk – Ref Arch

Logical Architecture using Pure FlashArray

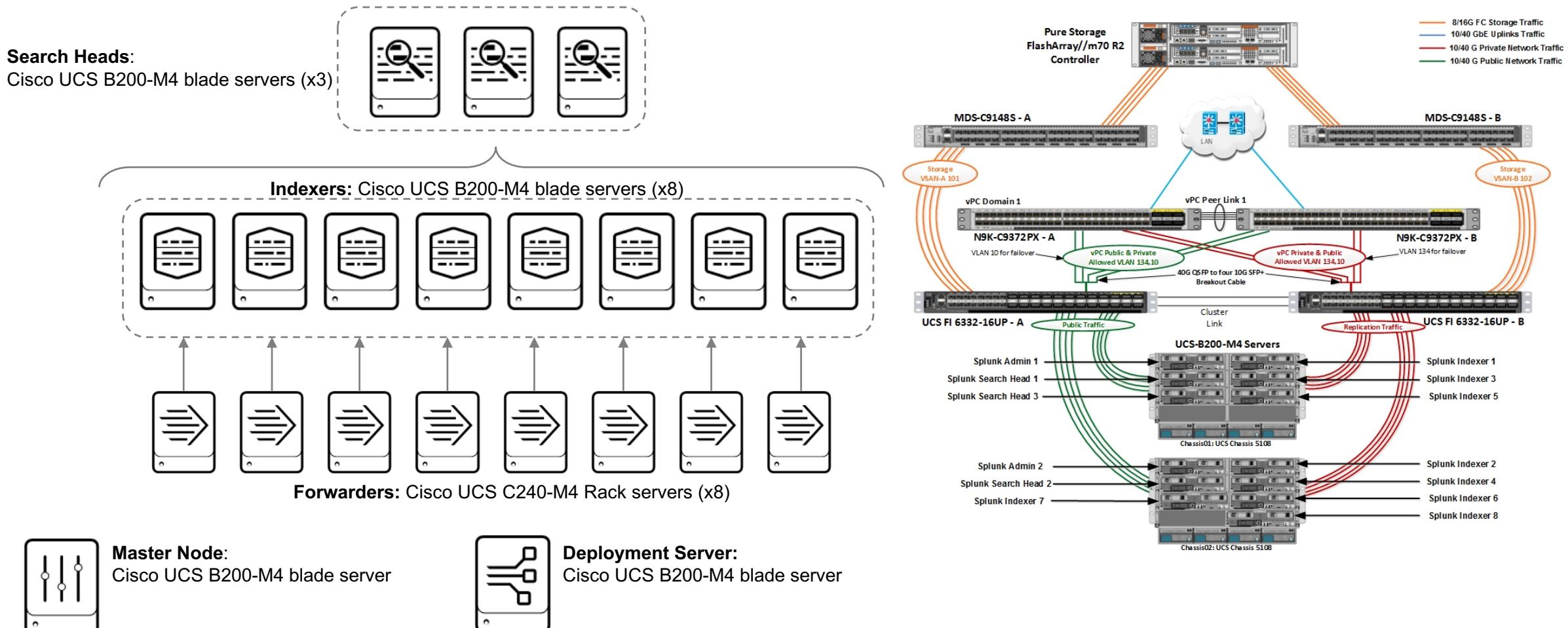


- ▶ Inline dedupe and compression
 - ▶ Non-disruptive everything
 - ▶ No logical tiering
 - ▶ Scalable Performance

splunk> conf18

Flashstack for Splunk – Ref Arch

Physical Architecture using Pure FlashArray



Searching 7 Billion Events in 2 Seconds

Finding A Needle In A Haystack:

splunk> App: Search & Reporting

Administrator Messages Settings Activity Help Find

Search Datasets Reports Alerts Dashboards Search & Reporting

New Search index="ixtest*" needle Save As Close

Finalizing job... No Event Sampling

Events (1) Patterns Statistics Visualization

Format Timeline - Zoom Out + Zoom to Selection × Deselect 1 millisecond per column

List Format 20 Per Page

< Hide Fields All Fields i Time Event

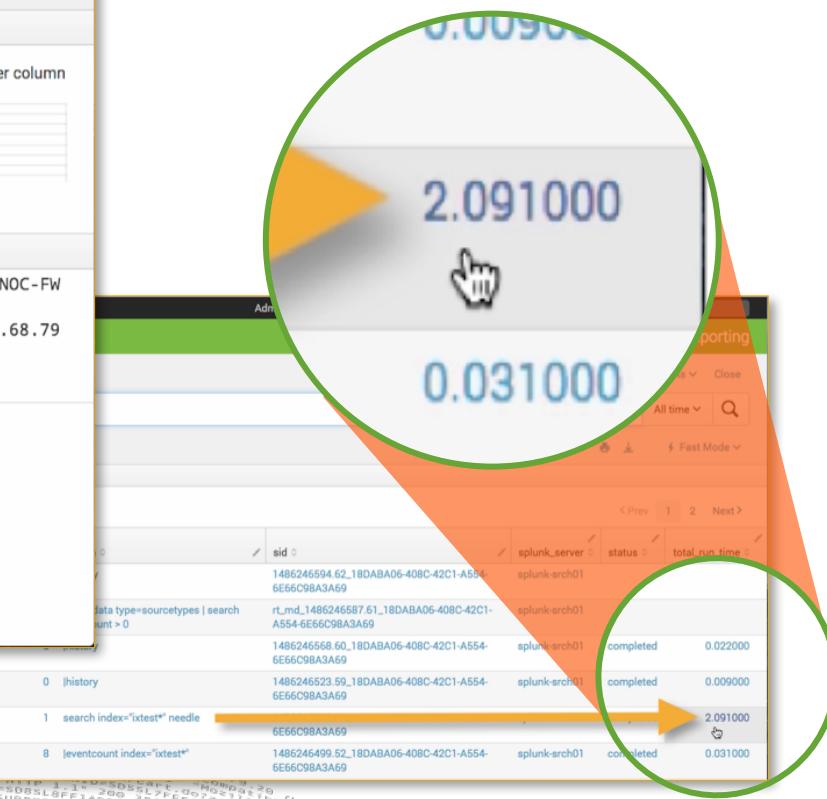
> 2/3/17 Mar 24 2027 02:47:13 UTC eventnum=4525346332 everyx needle 218.89.64.243 NOC-FWa: NetScreen device_id=NOC-FW a [Root]system-notification-00257(traffic): start_time="2006-04-05 23:00:57" duration=0 policy_id=1 service=udp/port:678 proto=17 src=backbone dst zone=noc-lan1 action=Deny sent=0 rcvd=0 src=39.205.68.79 dst=105.120.89.4 src_port=69465 dst_port=45286 host = UTC | source = /x01/data08/splunk_data08 | sourcetype = syslog

Selected Fields a host 1 a source 1 a sourcetype 1

Interesting Fields a index 1 # linecount 1 a splunk_server 1

+ Extract New Fields

Needle in a haystack search. 1 event out of 7 billion



2-3x Data Reduction

Reducing Splunk's Replication Factor & Search Factor

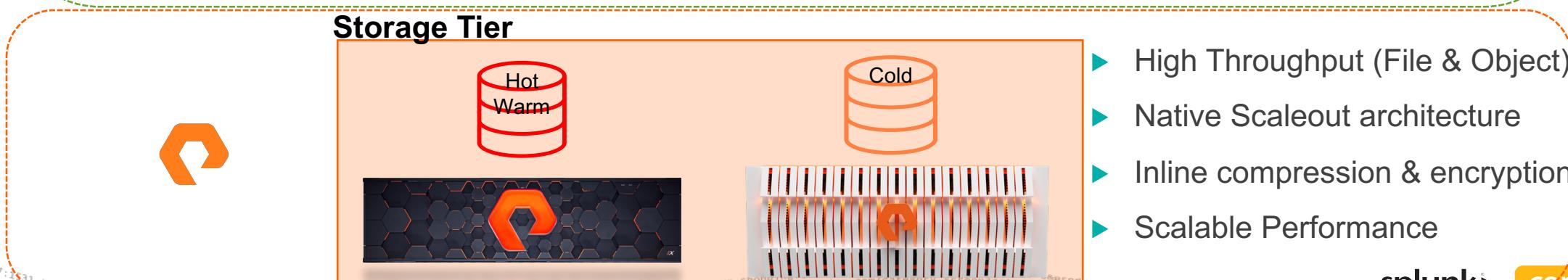
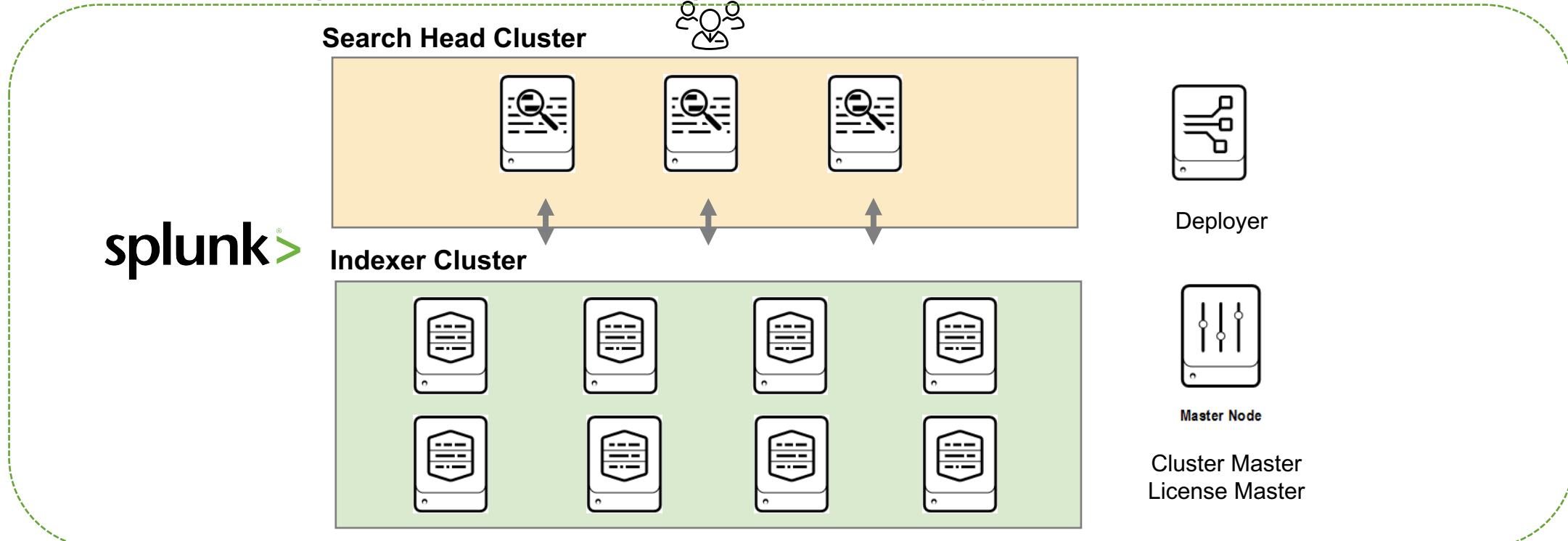
The image illustrates the process of reducing data replication and search factors by utilizing a hybrid storage solution. It shows three main components:

- Splunk Monitoring Console:** Displays 'Volume Detail: Deployment' for a 'hot' volume. It shows 8 Indexers with a total volume size of 1011.55 GB. A yellow callout bubble points to the Pure Storage dashboard, containing the text: "Index data on Splunk at the end of data ingestion".
- Pure Storage Dashboard:** Monitors storage performance metrics including Capacity, Latency, IOPS, and Bandwidth. It highlights a 'Data Reduction 2.4 to 1' and '1% full' status.
- Search Hosts and Volumes:** A search interface showing a summary of data reduction and usage across hosts and volumes.

At the bottom of the image, there is a large amount of raw log data from a Splunk search, which is partially visible and truncated.

Upcoming Flashstack for Splunk – Ref Arch

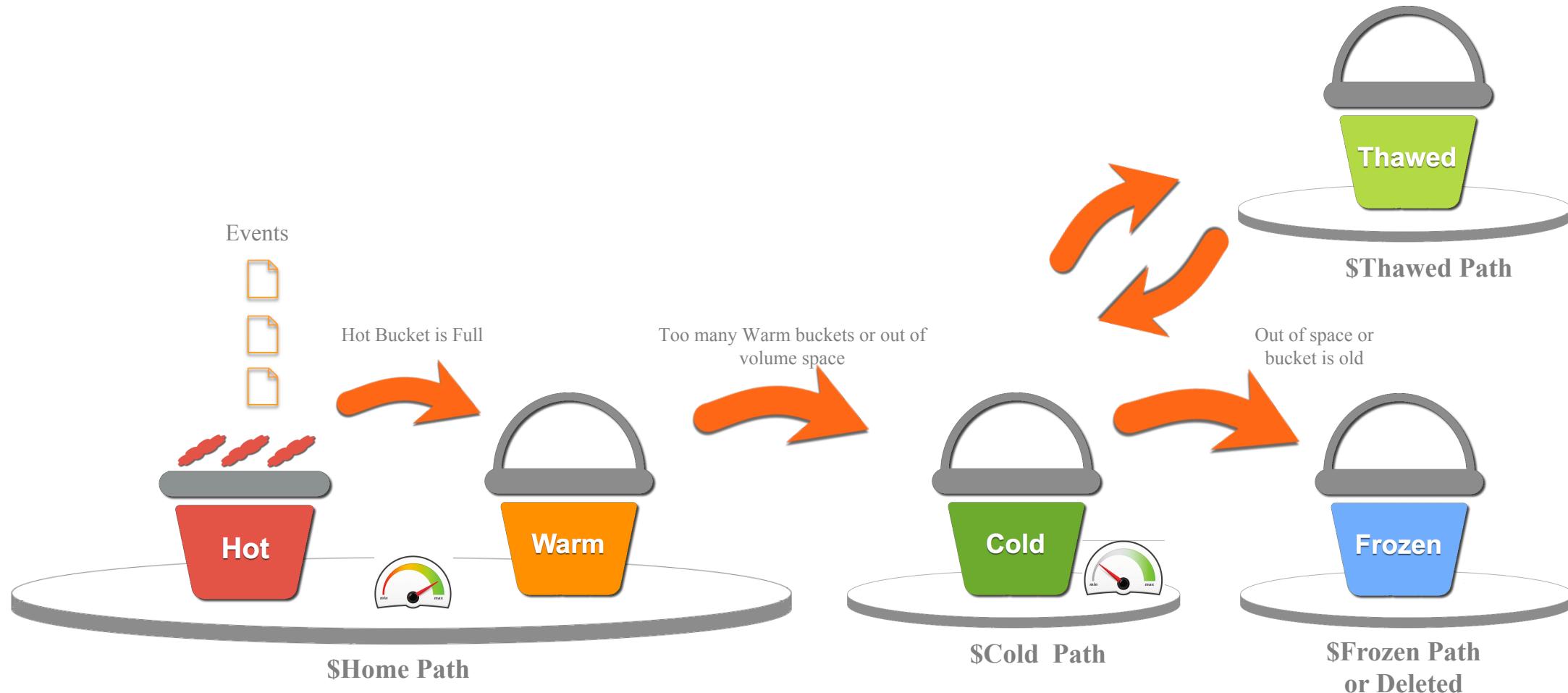
Logical Architecture using Pure FlashArray & FlashBlade



- ▶ High Throughput (File & Object)
 - ▶ Native Scaleout architecture
 - ▶ Inline compression & encryption
 - ▶ Scalable Performance

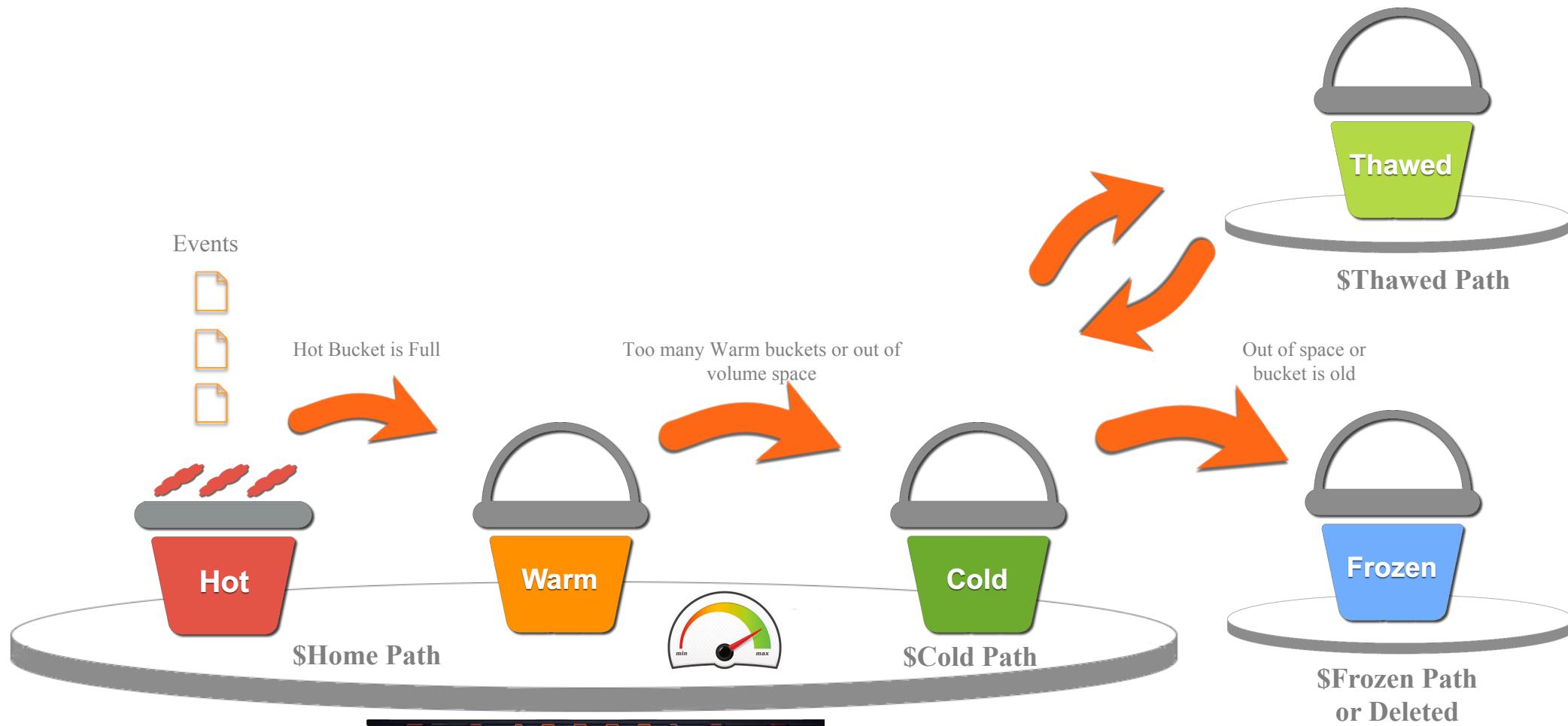
Splunk Buckets Across Tiers

How Splunk stores data?



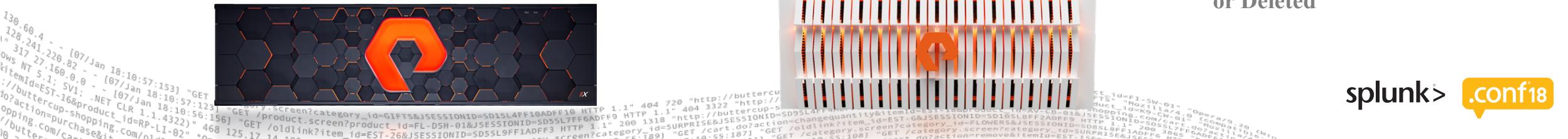
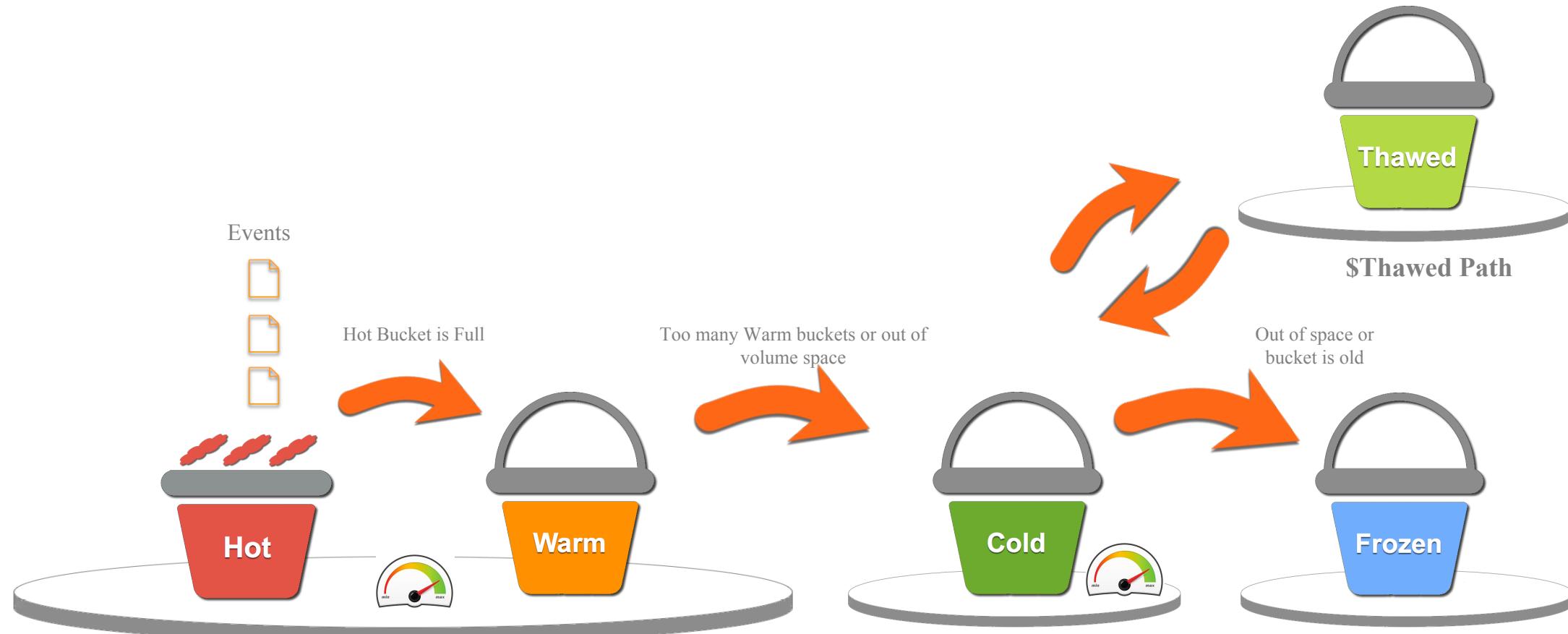
Splunk Buckets on Pure Storage #1

Eliminate logical tiering between Hot/Warm and Cold



Splunk Buckets on Pure Storage #2

Gain scalable capacity at better performance



Splunk App for Pure Flasharray

The screenshot displays the Splunk Enterprise interface with the PureStorage App installed. The top navigation bar includes links for Administrator, Messages, Settings, Activity, Help, and Find. The left sidebar features a 'Search & Reporting' icon and three app cards: 'PureStorage-App' (selected), 'PureStorage-TA', and 'PureStorage-DB'. The main content area is titled 'Explore Splunk Enterprise' and shows four circular icons representing different data types. Below this, a modal window for 'Volume Inventory' is open, showing details for array 'pure-m50-2-ct0-b12-35' and volume 'fs_prod_data01' over the last 24 hours. The interface then transitions into a detailed dashboard for the 'PureStorage-App', which includes sections for Allocation (Capacity: 2,048.00 GB, Data Reduction: 8.4 to 1), Performance Information (IOPS: 0.00 K, Bandwidth: 0.06 KB), and Recent Alerts. The bottom half of the dashboard provides tables for Array Health Map, Arrays by Used Space, Arrays by IOPS, and Arrays by Latency, all showing data for the array 'pure-m50-2-ct0-b12-35'.

Splunk App for Pure Flasharray

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

Array Inventory

Select Array: pure-m50-2-ct0-b12-35 | Select time frame: Last 24 hours | Show Help: No (Yes)

Allocation

Performance Information

IOPS: 1.09 K	Bandwidth: 4.50 KB
Latency (Read): 0.14 ms	Latency (Write): 0.44 ms

Recent Alerts

No results found.

Capacity: **11,433.33 GB**

Data Reduction: **8.4 to 1**

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

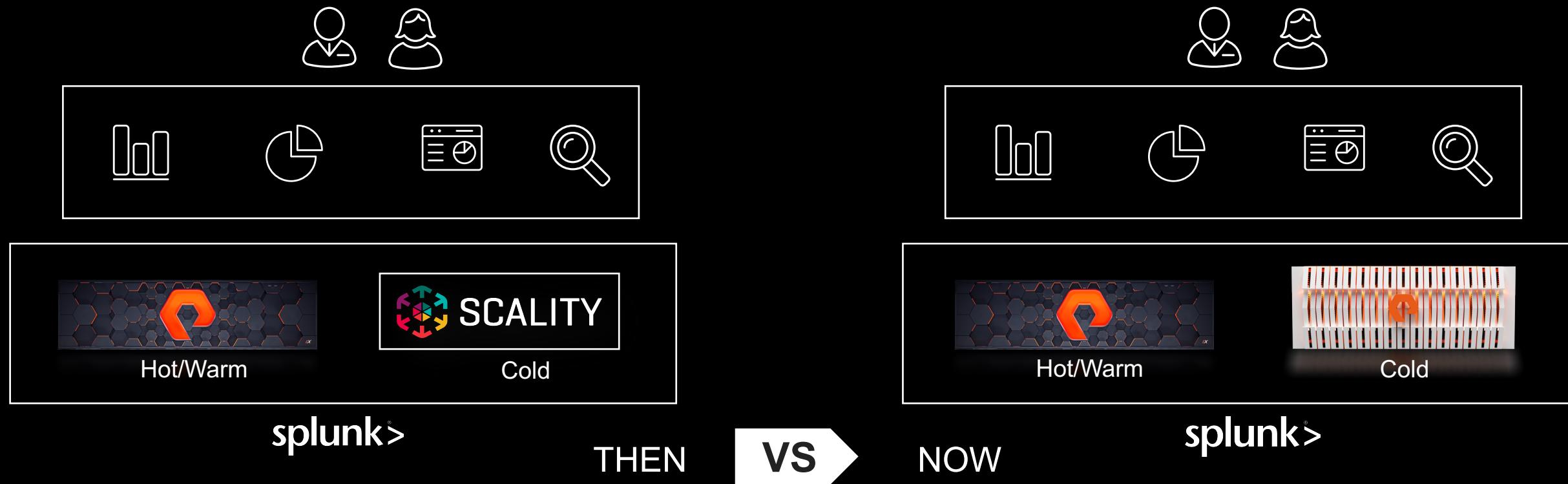
Name	Read	Write
1 ds-oel72asm	1.15 ms	0.00 ms
2 fs-ebs-clone200-db01	1.07 ms	0.43 ms
3 ds-splunk-data02	0.94 ms	0.44 ms
4 fb-dss-infra	0.90 ms	0.55 ms
5 ds-splunk-idx-data	0.73 ms	0.29 ms
6 fb-dss-infra	0.70 ms	0.55 ms

nf18

Use Cases

There is No Cold Data in Analytics

A Joint Pure/Splunk Customer



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

Splunk on Pure: Healthcare

Use Case:

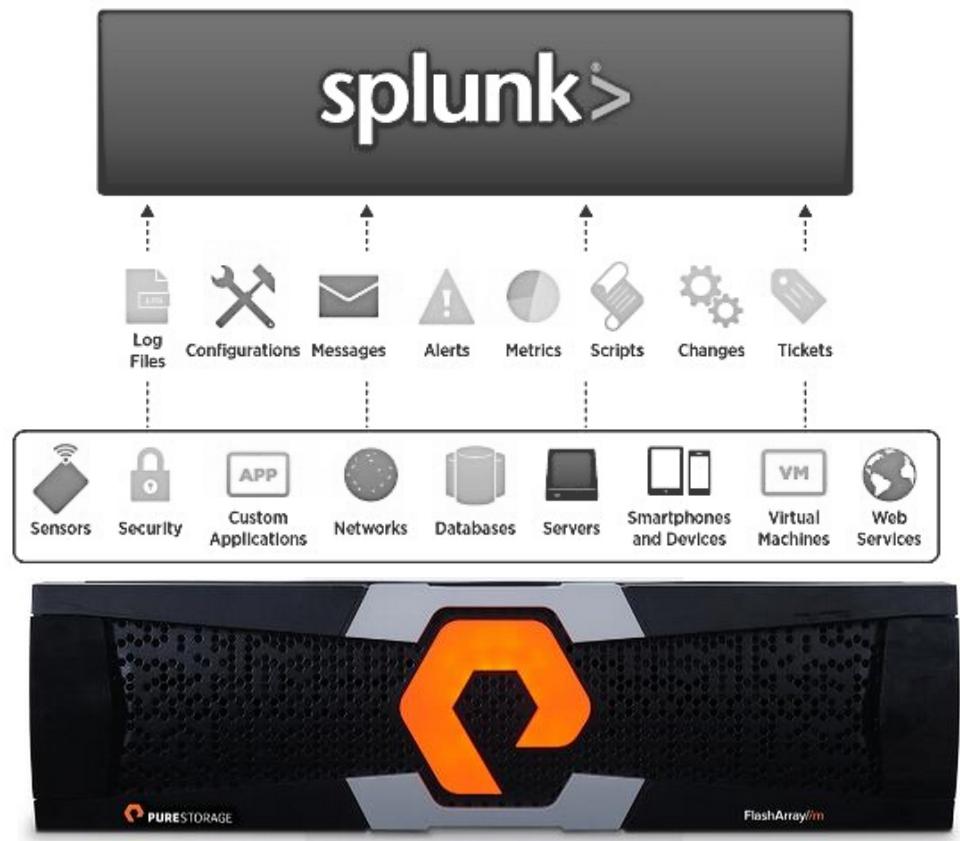
- ▶ Multi-site, ingest of 1TB/day
- ▶ Splunk was setup on DAS storage

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:

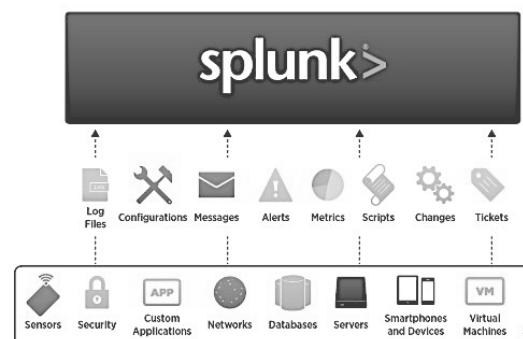
- ▶ Splunk, VDI and VMware all supported by FA//m20

Business Value

- ## ► Simplicity and scalability

Technical Value

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



VMware



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

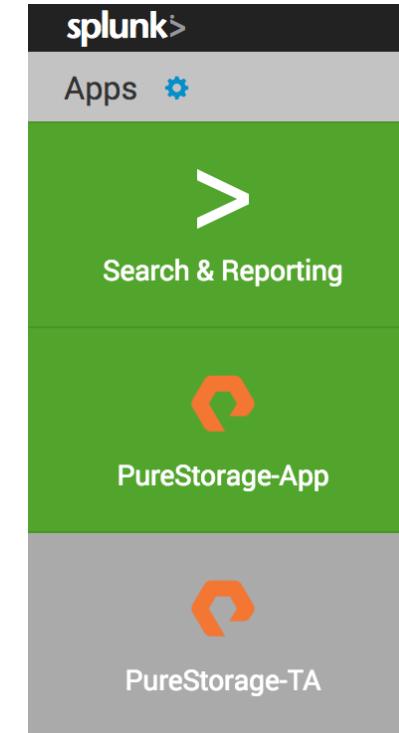


Splunk Resources



FlashStack for Splunk Reference Architecture

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

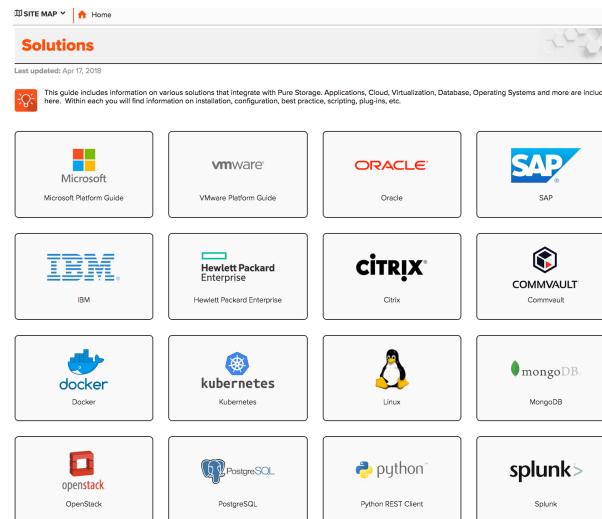


Splunk App for Pure FA

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

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



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

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

100% NVMe
DENSITY:
1PB
IN JUST 3U

MULTI- PROTOCOL BUILT-IN

100% NVMe
PERFORMANCE:
12 GB/s
BANDWIDTH

99.9999%
AVAILABILITY &
NDU EVERYTHING

DIRECTEL ASH SHELF NOT CURRENTLY GA

FlashBlade™

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!

