Splunk…so Big and Flashy

Building Massive and Efficient Indexer Storage Environments for Splunk

Cory Minton

Principal SE and Data Fabrics Leader, Emerging Technologies @ EMC

splunk>

Disclaimer

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 the 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.

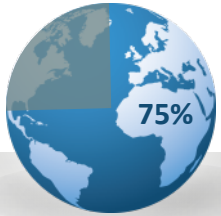
In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not, be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

These ARE The Droids You Are Looking For…

Agenda

- Data and Storage Tech Trends
- Splunk Architecture
- Why Flash Your Home Path?
- The Big, Cold Data Lake
- Converged Solutions
- Resources – Sweet Apps and Ninjas

Data Growth



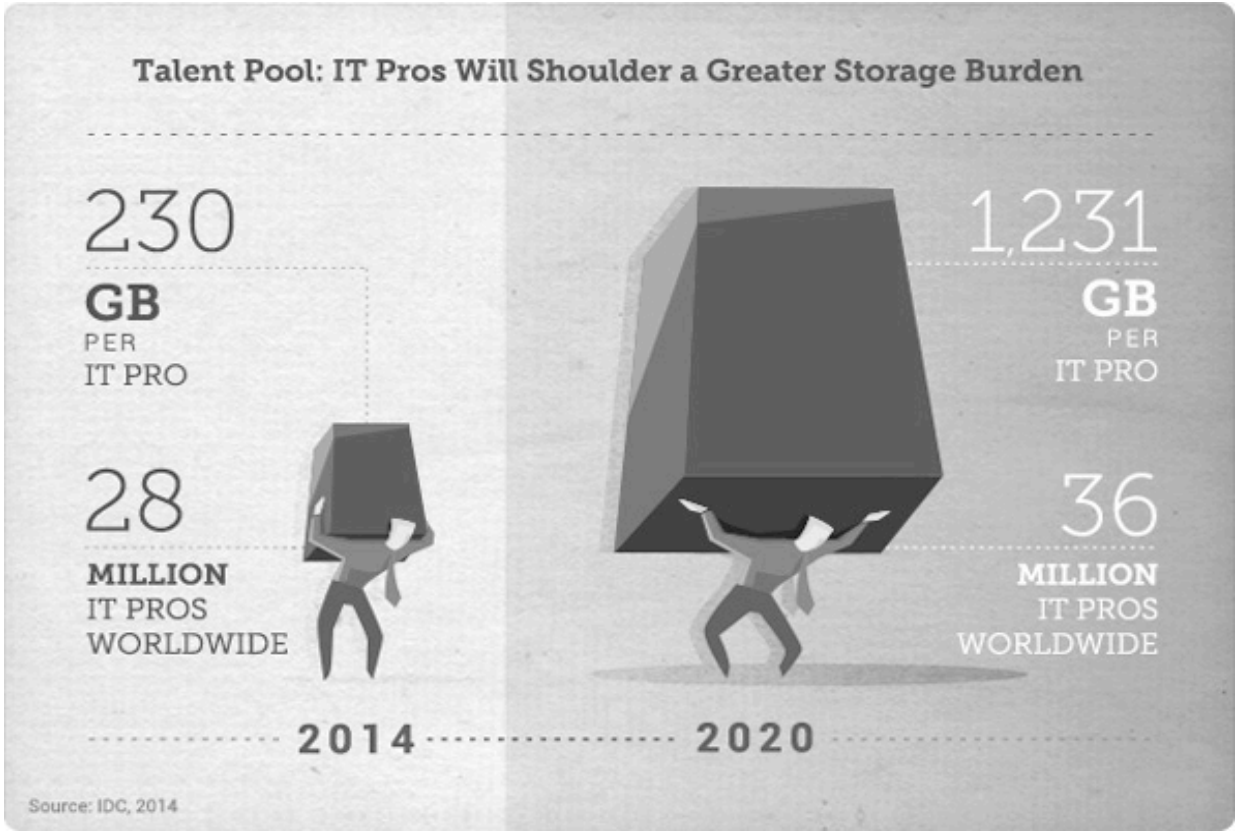
2015	2016	2017
71 EB	106 EB	133 EB

Source: IDC

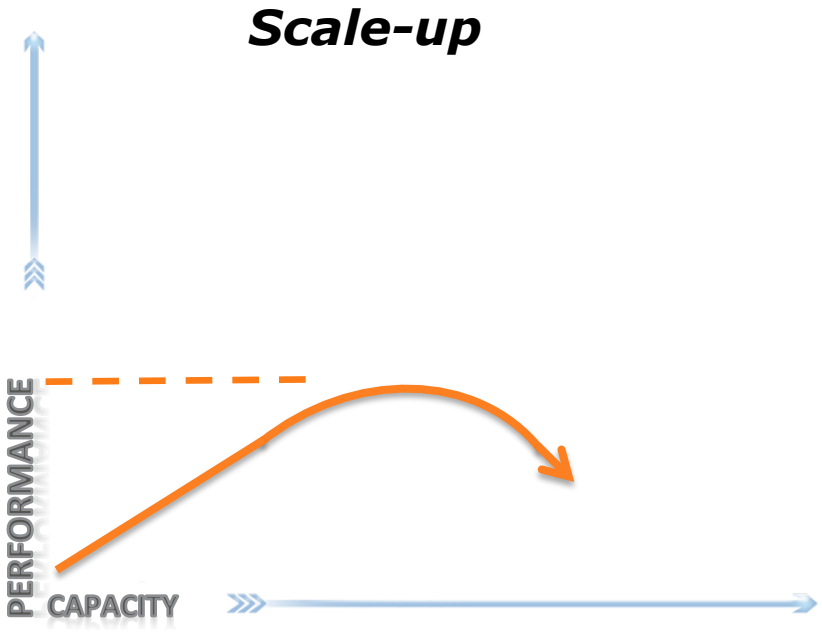
Total Capacity Shipped, Worldwide

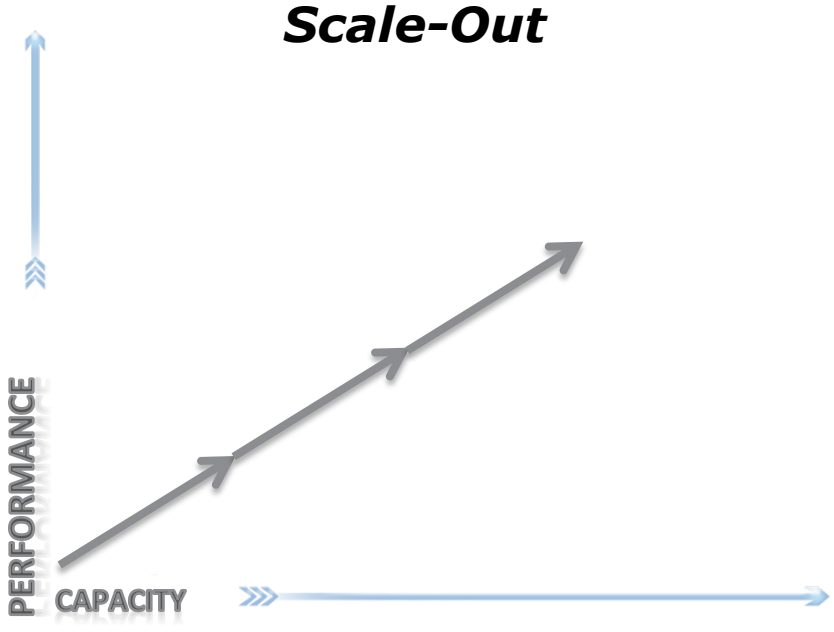
% of Unstructured Data

Do More With Less…



Architecture Matters…

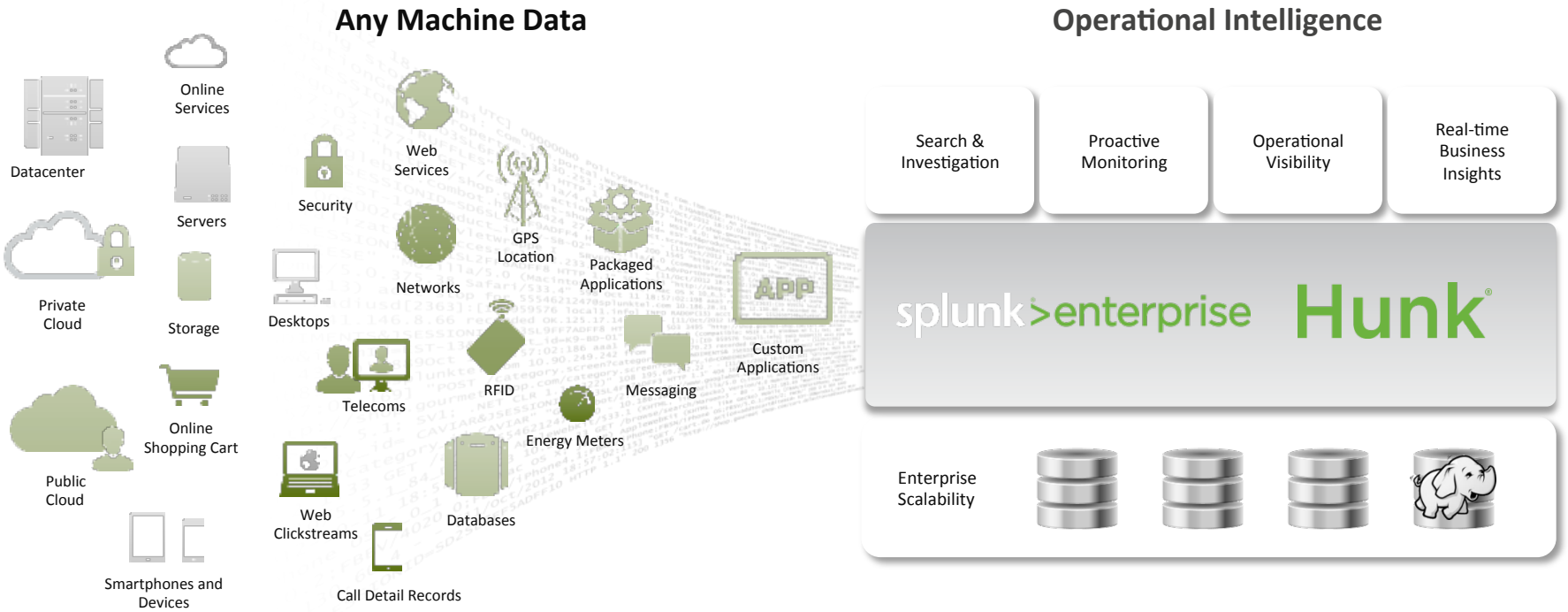




Enter SPLUNK ENTERPRISE

EMC²

Industry-leading Platform For Machine Data



Splunk Architecture

Search Heads

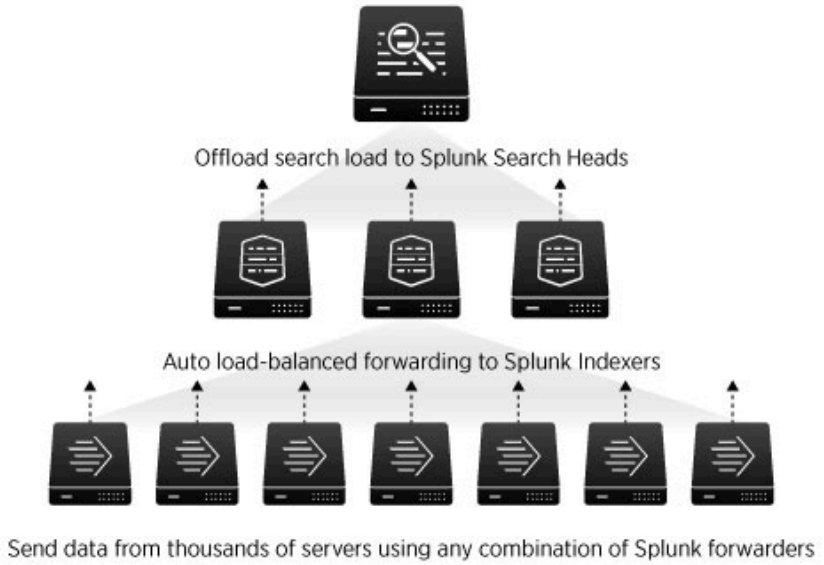
Query information across indexers and are usually CPU and memory intensive.

Indexers

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

Forwarders

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

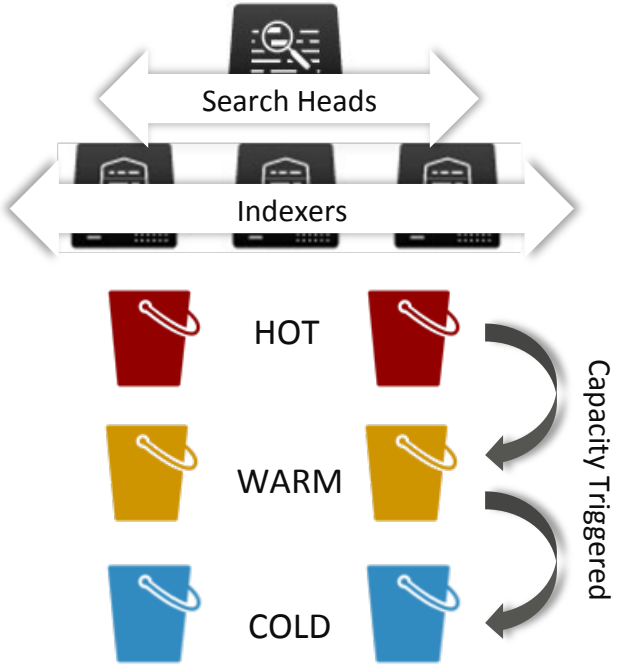


http://docs.splunk.com/Documentation/Splunk/latest/Overview/AboutSplunkEnterprisedeployments

Splunk Storage Requirements

Enterprise Performance And Data Services

- High-Performance Storage
  - Rare & Sparse Searches
- High-Capacity Storage
  - Long-Term Retention
- Scale-Out Infrastructure
  - Indexer & Search Heads
- De-dupe & Compression
  - Clustered Indexer Deployments
- Backup & Security
  - Data Protection & Compliance



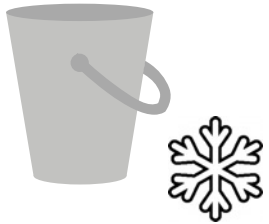
Splunk Indexer Buckets

HOT / WARM

- Recent searches/dashboards
- Usually block LUN
- High Random reads
- Sequential reads / writes

COLD

- Rare searches
- Usually NAS share
- Light Random Reads
- Sequential reads / writes



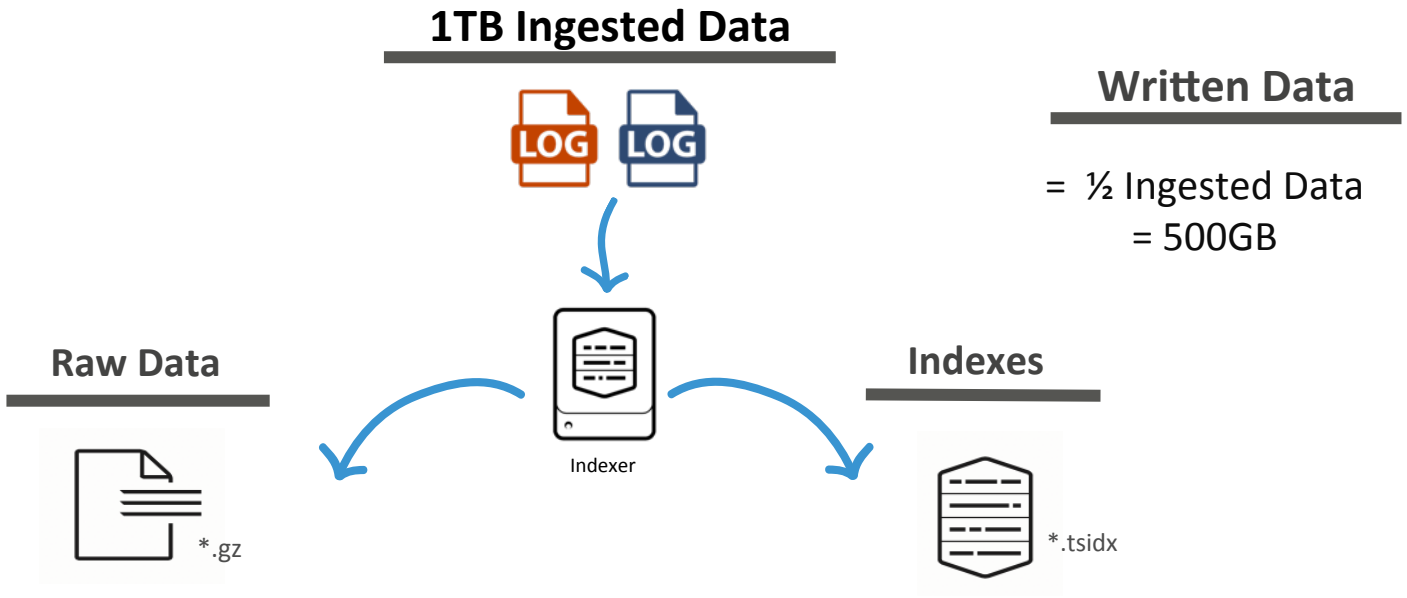
FROZEN

- Not searchable
- Usually offline media
- Only sequential write

Splunk moves indexes from between from hot/warm to cold to frozen based on user configuration

Size of the buckets impacts performance

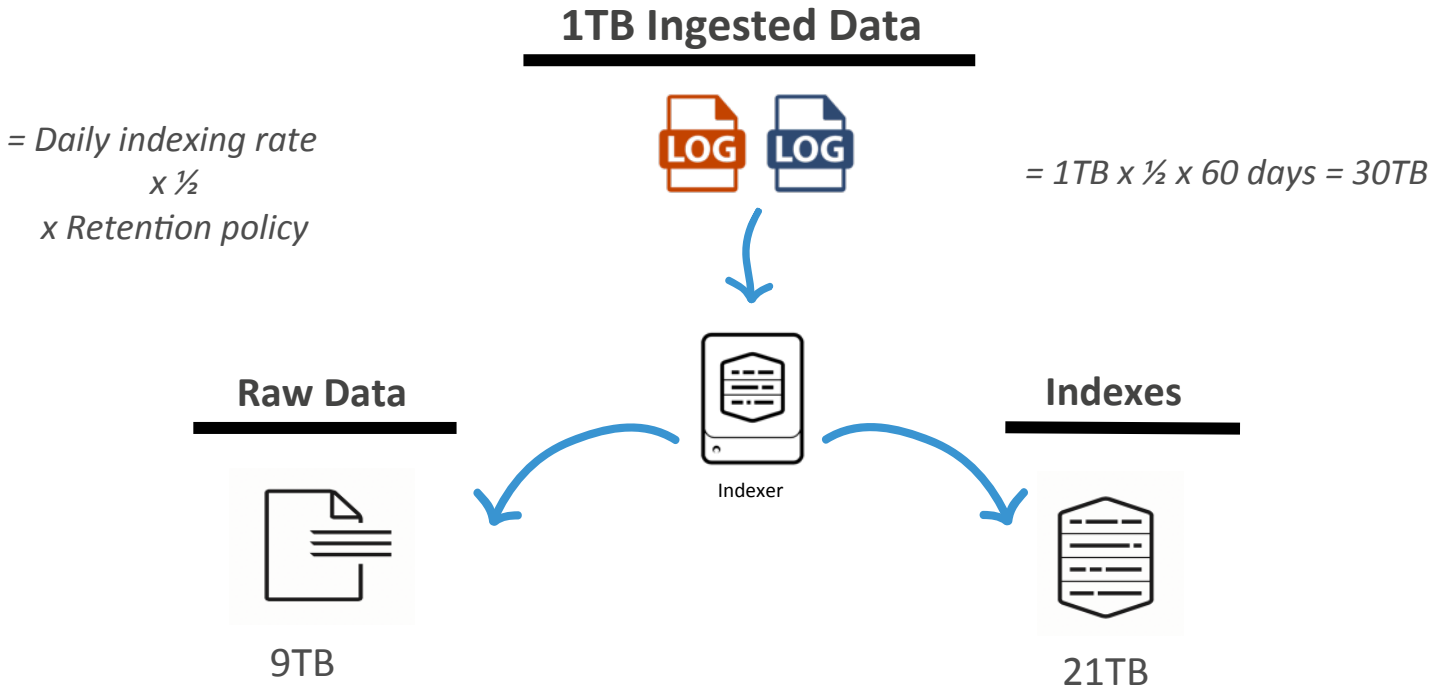
Indexer Storage Capacity



Compressed Raw data
30% of written data
→ 150GB

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

How Much Storage Do You Need?

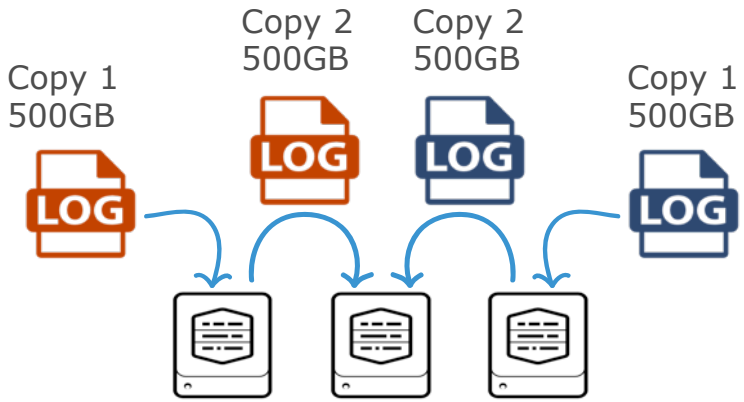


Splunk Indexer Availability

EMC²

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 **

STORAGE CAPACITY

MULTIPLIES!

Just How Much Storage?

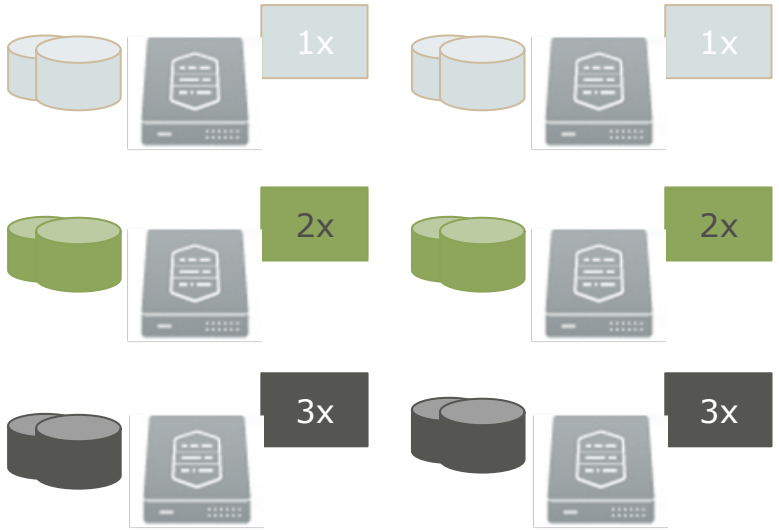
Storage Requirements in TB												
								1 Year	2 Years	3 Years	4 Years	5 Years
Splunk License (GB/DAY)	Retention (Days)	1	7	14	30	90	180	365	730	1095	1460	1825
	25	0.025	0.175	0.35	0.75	2.25	4.5	9.125	18.25	27.375	36.5	45.625
	50	0.05	0.35	0.7	1.5	4.5	9	18.25	36.5	54.75	73	91.25
	100	0.1	0.7	1.4	3	9	18	36.5	73	109.5	146	182.5
	250	0.25	1.75	3.5	7.5	22.5	45	91.25	182.5	273.75	365	456.25
	500	0.5	3.5	7	15	45	90	182.5	365	547.5	730	912.5
	1000	1	7	14	30	90	180	365	730	1095	1460	1825
	2000	2	14	28	60	180	360	730	1460	2190	2920	3650
	3000	3	21	42	90	270	540	1095	2190	3285	4380	5475
	4000	4	28	56	120	360	720	1460	2920	4380	5840	7300
	5000	5	35	70	150	450	900	1825	3650	5475	7300	9125
	6000	6	42	84	180	540	1080	2190	4380	6570	8760	10950
	7000	7	49	98	210	630	1260	2555	5110	7665	10220	12775
	10000	10	70	140	300	900	1800	3650	7300	10950	14600	18250

*Assumes RF/SF = 2

DAS Presents Challenges

1	Dedicated Storage Infrastructure • Silo that only runs Splunk
2	Compromised Availability • SSDs & servers fail • Index rebuilds can take hours to days
3	Lack of Enterprise Data Protection • No Snapshots or Compliance • DR limited to Multisite Clustering
4	Poor Storage Efficiency • Multiple copies of data • Multisite Clustering Increases Overhead
5	Non-Optimized Growth • Fixed compute to storage ratio • Servers must maintain storage symmetry
6	Management complexity • Multiple management points

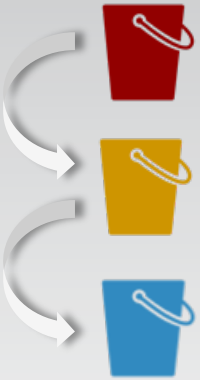
SPLUNK DAS ENVIRONMENT



Why EMC For Splunk

Optimized Infrastructure For Big & Fast Data

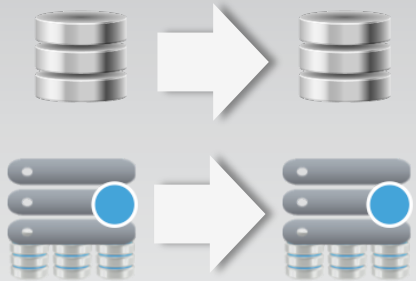
Optimized Shared Storage & Tiering



Hot & Warm Data Deployed On XtremIO or ScaleIO

Cold & Frozen Data Deployed On Isilon

Cost-Effective & Flexible Scale-Out



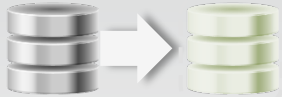
Scale-Out Capacity & Compute Independently Or As Converged Platform

Powerful Data Services

Encyption & Security

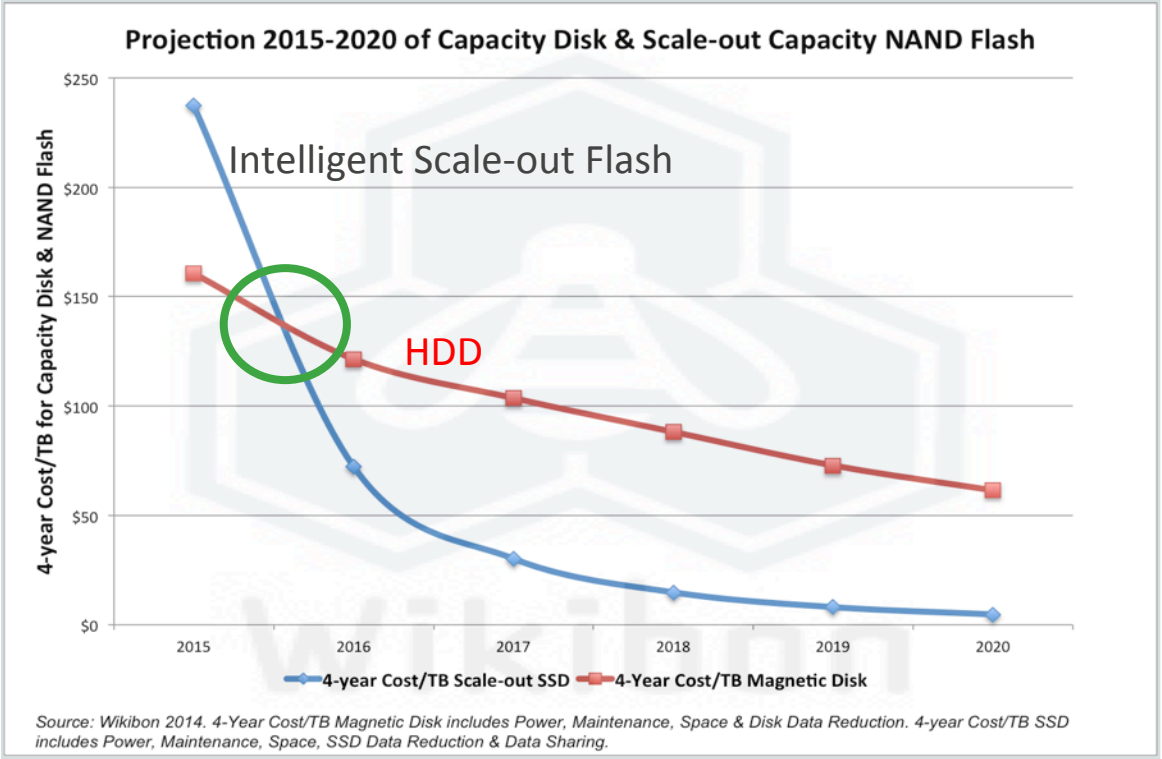
Index File Compression

Snapshots For Backups



Deduplication Of Clustered Indexes

Why Flash?!?

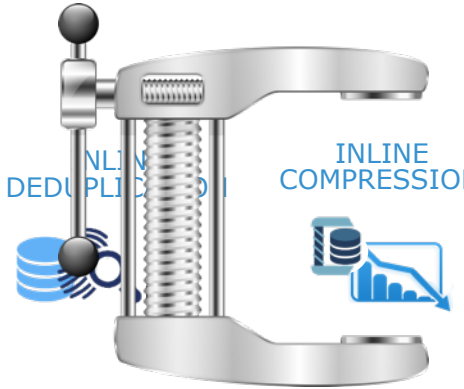


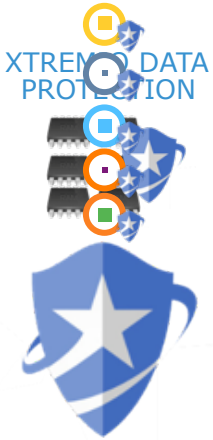
Economic Influences

- ✓ Consumer Demand
- ✓ Data Services Allowing free Copies of Application Data
- ✓ Flash technology has improved at a faster rate than Moore's Law

Xtremio Data Services

Always-on, Inline, Zero Penalty, Free



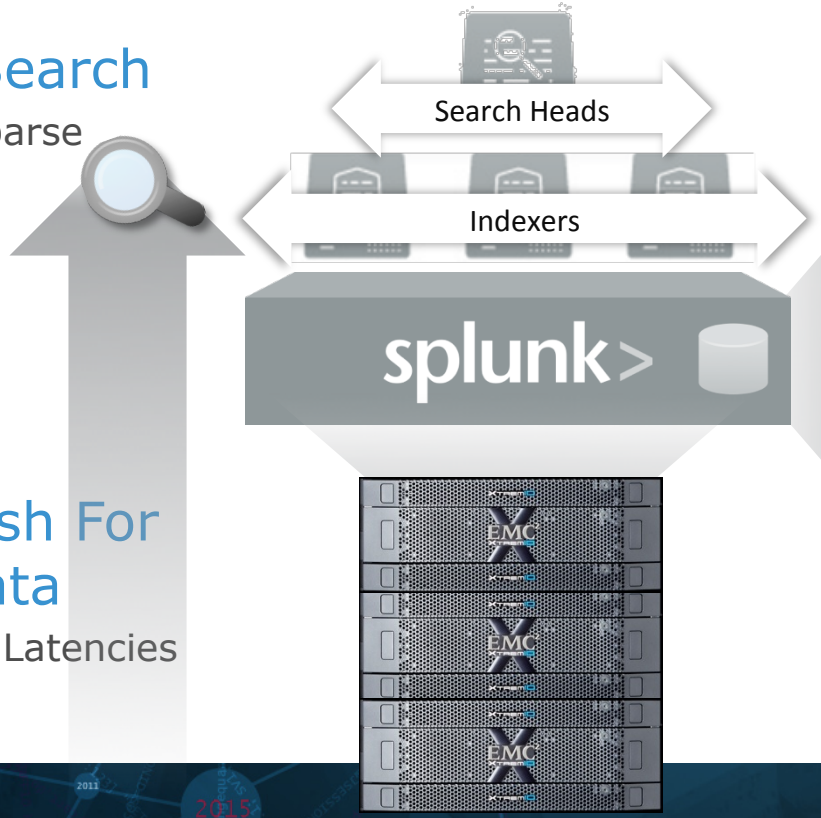


EMC Xtremio & Splunk

All-flash Infrastructure For Hot & Warm Data

High-Speed Search

Accelerate SuperSparse & Rare Searches



Scale-Out Flash For I/O-Bound Data

>1M IOPS & <1ms Latencies

Data Services For Hot & Warm Data

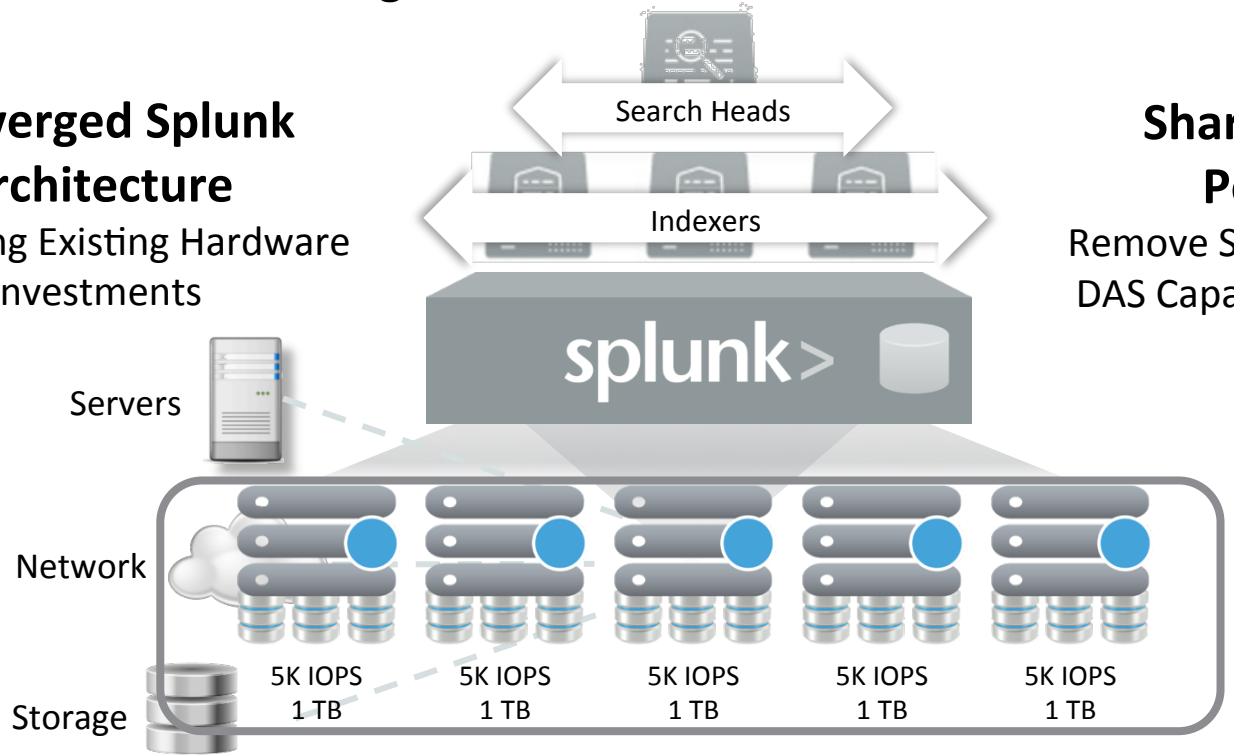
- Self-Encrypting Flash Drives
- Index File Compression
- In-Memory Data Copy Services
- Dedupe Clustered Index Copies

EMC Scaleio & Splunk

Converged Architecture For Hot & Warm Data

Converged Splunk Architecture

Leveraging Existing Hardware Investments

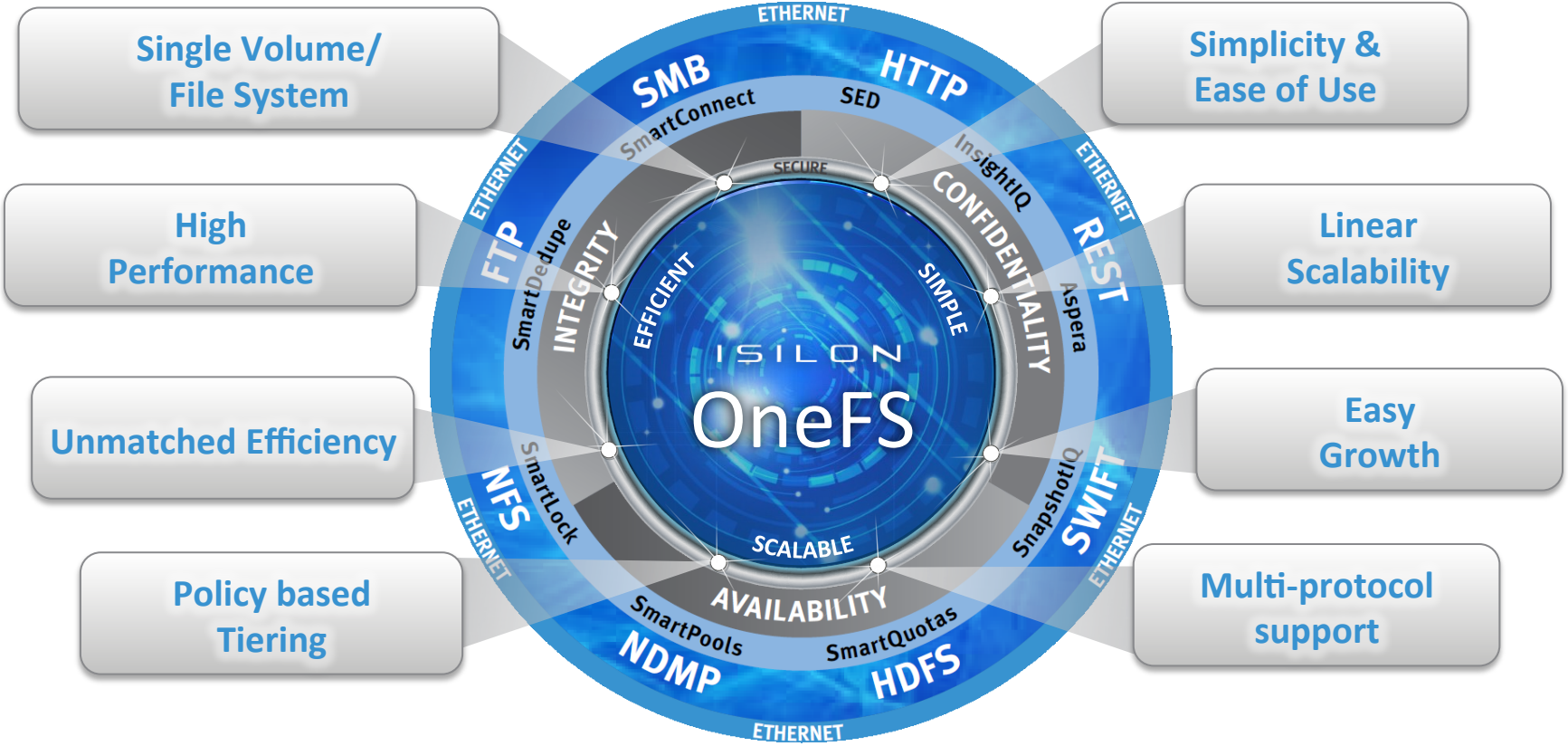


Shared Capacity & Performance

Remove Silos & Increase ROI On DAS Capacity & No Single Point Of Failure

25K IOPS & 5TB

EMC Isilon – Deep and WIDE Storage

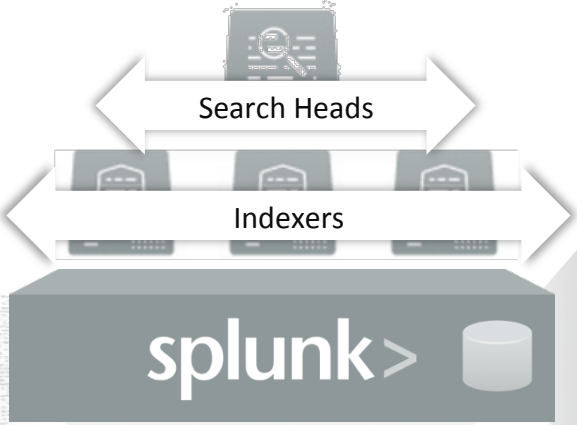


EMC Isilon & Splunk

Low-cost & Secure Scale-out For Cold Data

High-Speed Ingest

& Long-Term Retention With
Native HDFS Integration



Scale-Out Capacity

Up To 50PB Of Highly
Available Capacity

Consolidate, Protect & Secure Cold Data

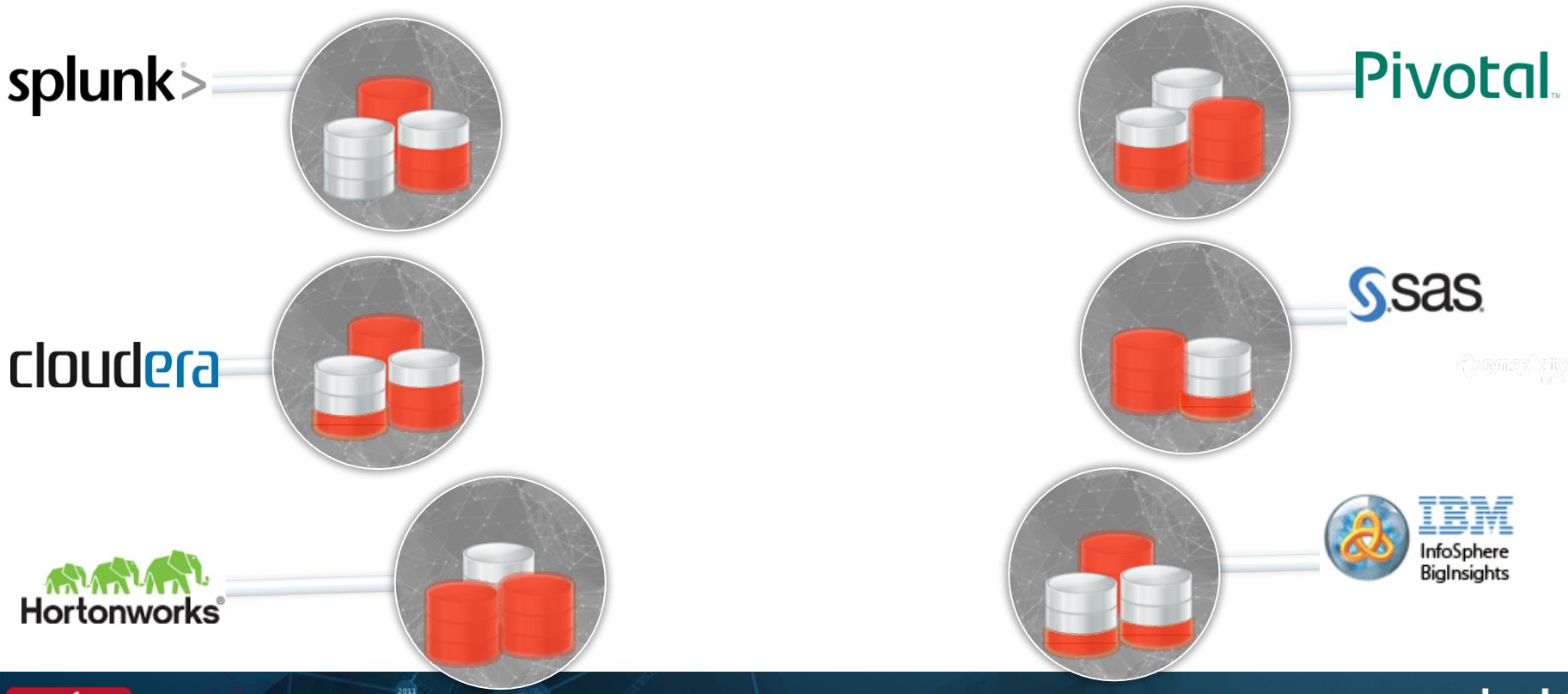
Self-Encrypting Drives

SmartLock Protects Cold & Frozen Data

Snapshots IQ For Backups

SmartDedupe For Clustered Indexes

Data Silos vs Consolidated Data Lake



Data Silos vs Consolidated Data Lake

EMC²

splunk>

cloudera

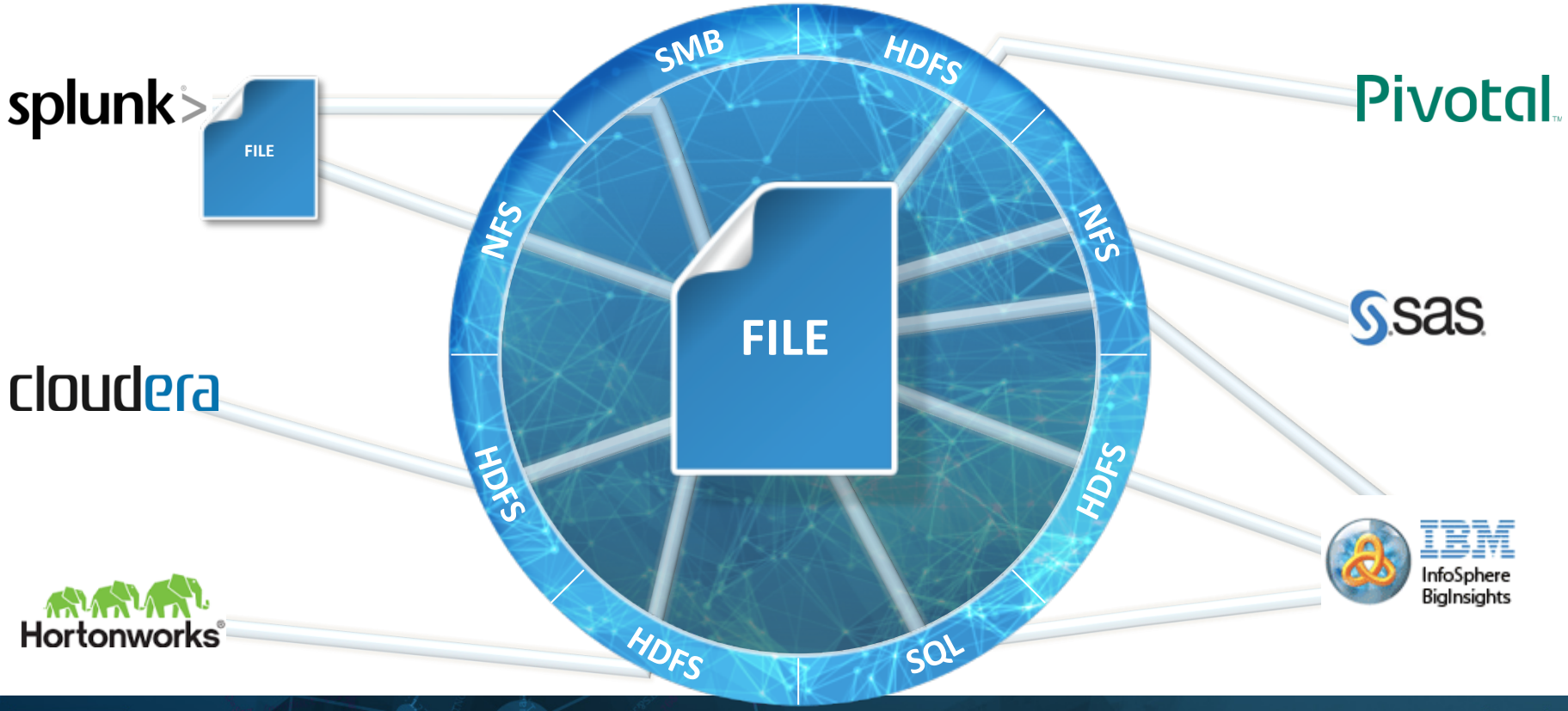
Hortonworks

Pivotal

sas

IBM InfoSphere BigInsights

EMC Isilon Next-Gen Access Methods



EMC Reference Architectures For Splunk Enterprise

XtremIO and Isilon Reference Architecture
ScaleIO and Isilon Reference Architecture

EMC Reference Architectures

EMC²

Single-Instance

Indexers

Distributed

Indexers Search

HOT & WARM

Mostly searches
Heavy Random reads
Sequential writes

XtremIO

Scale-Out 160 TB Flash & No Tuning

No RAID Configuration Needed

Many Copies & No Overhead

Deduplication & Compression

COLD

Adhoc searches
Light Random Reads
Sequential Writes

Isilon

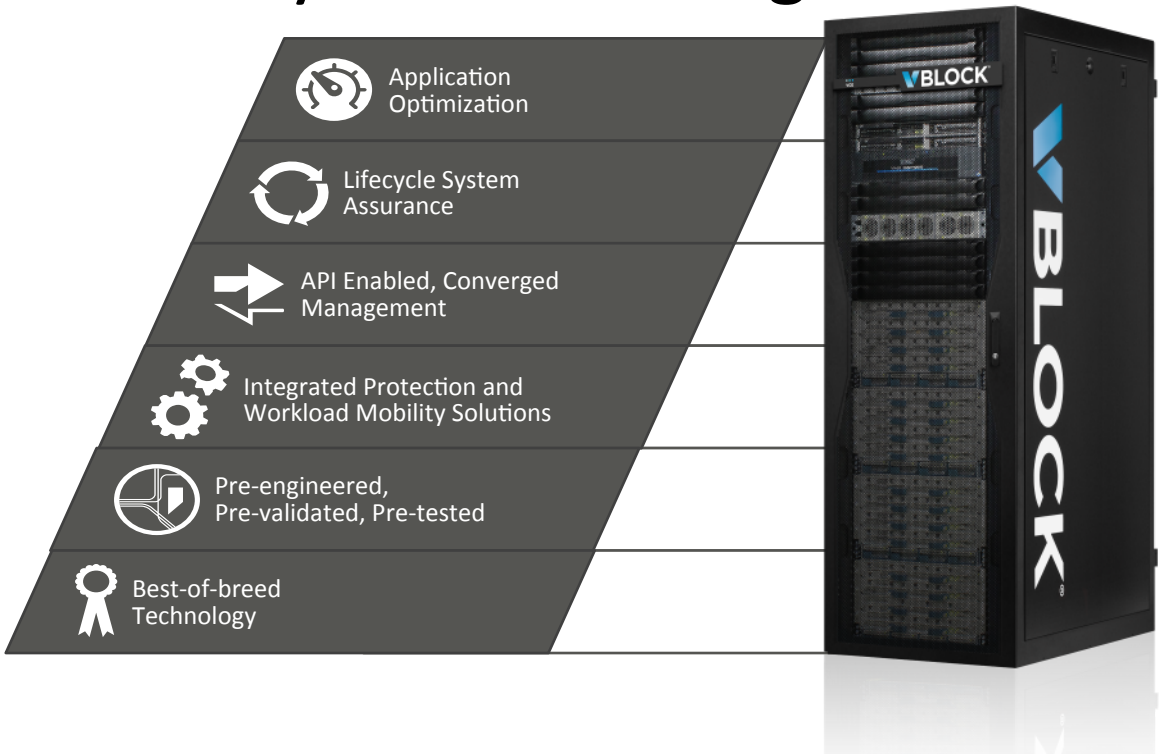
Scale-Out 50 PB Of Cold & Frozen Data

Multi-Protocol = Always Searchable

Tier “Frozen” Data Without Migration

SmartDedupe & SmartLock

Vblock® Systems –
The Only True Converged Infrastructure



Customer Experience

- Fastest Time-to-Business
- Highest Performance
- Highest Availability
- Converged Management
- Lowest Risk
- Lowest TCO

WHY VCE FOR SPLUNK?

FOCUS ON BUSINESS OPERATIONS, NOT MAINTAINING INFRASTRUCTURE

SPEED TIME-TO-DEPLOYMENT

- Factory Physical and Logical Build
- Compliance-Ready
- Performance and Availability

SIMPLIFY ONGOING OPERATIONS

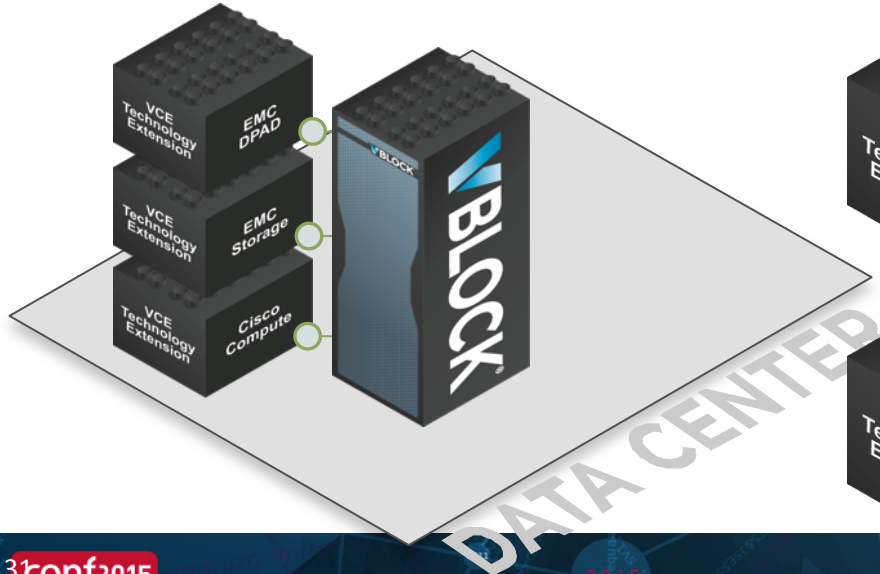
- Roadmap and New Feature Planning
- Configuration and Patch Management
- Single Support Through VCE

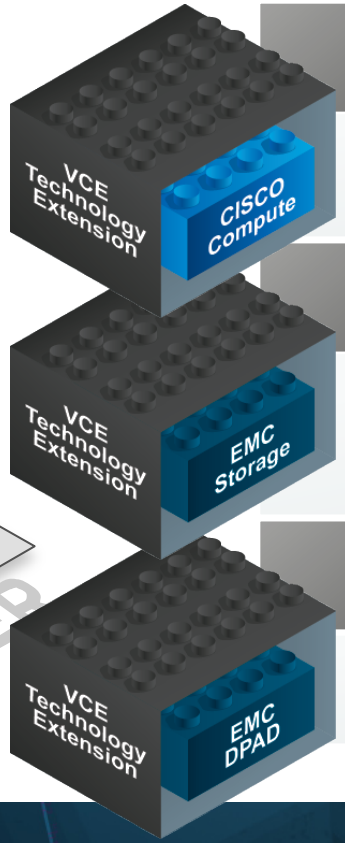
VCE Vscale Architecture

Modular Grow-as-needed Architectural Design

EMC²

Flexible combinations of EMC Isilon and XtremIO





VCE™ TECHNOLOGY EXTENSION FOR COMPUTE

- General Purpose Compute
- GPU Cluster
- DAS Based HDFS Cluster(s)

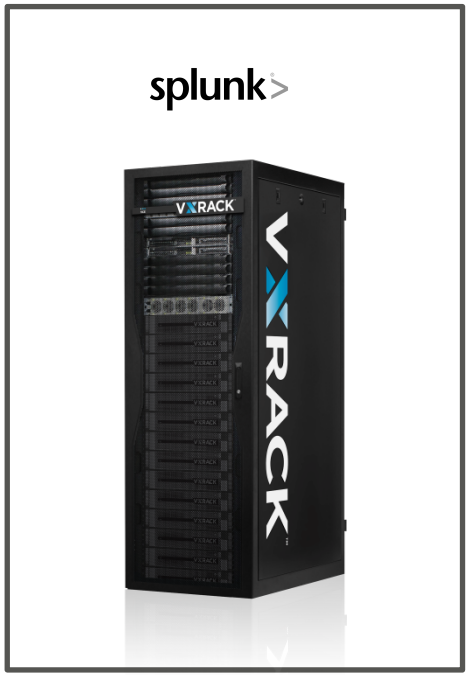
VCE™ TECHNOLOGY EXTENSION FOR STORAGE

- General Purpose Block Storage
- Object or Elastic Cloud Storage
- Data Lakes
- Solid State Flash

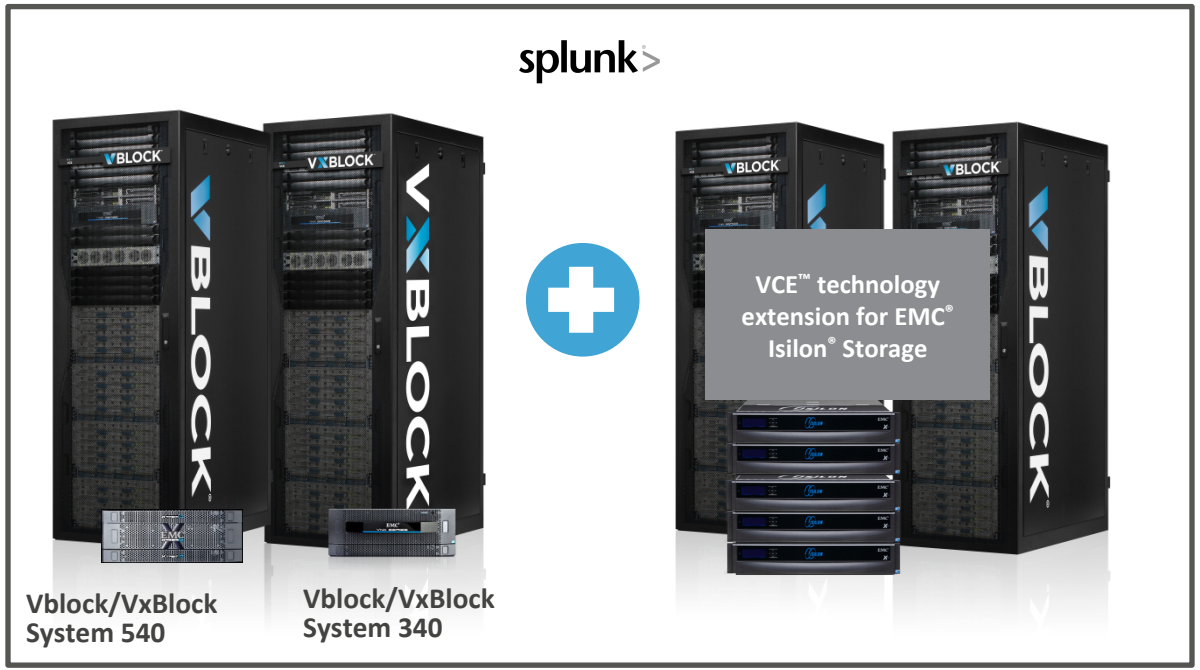
VCE™ TECHNOLOGY EXTENSION FOR DATA PROTECTION

- Data Protection
- Data Replication

VCE to Address Three Opportunities Splunk



Rack-Scale Bundle



Vblock/VxBlock System 540

Vblock/VxBlock System 340

HOT/WARM

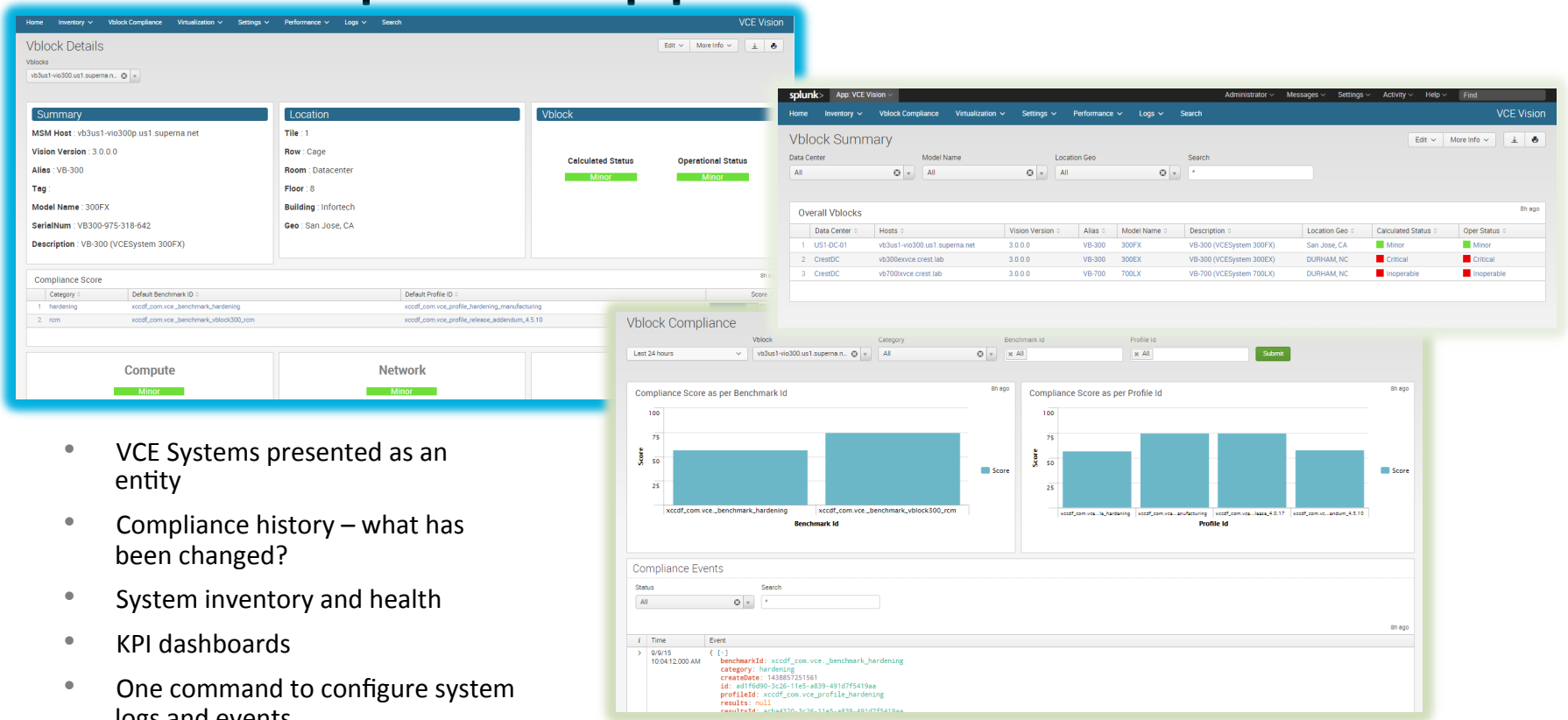
COLD

Block/Scale-Out Bundle

Sweet Apps and Ninjas

splunk>

Splunk App For VCE Vision

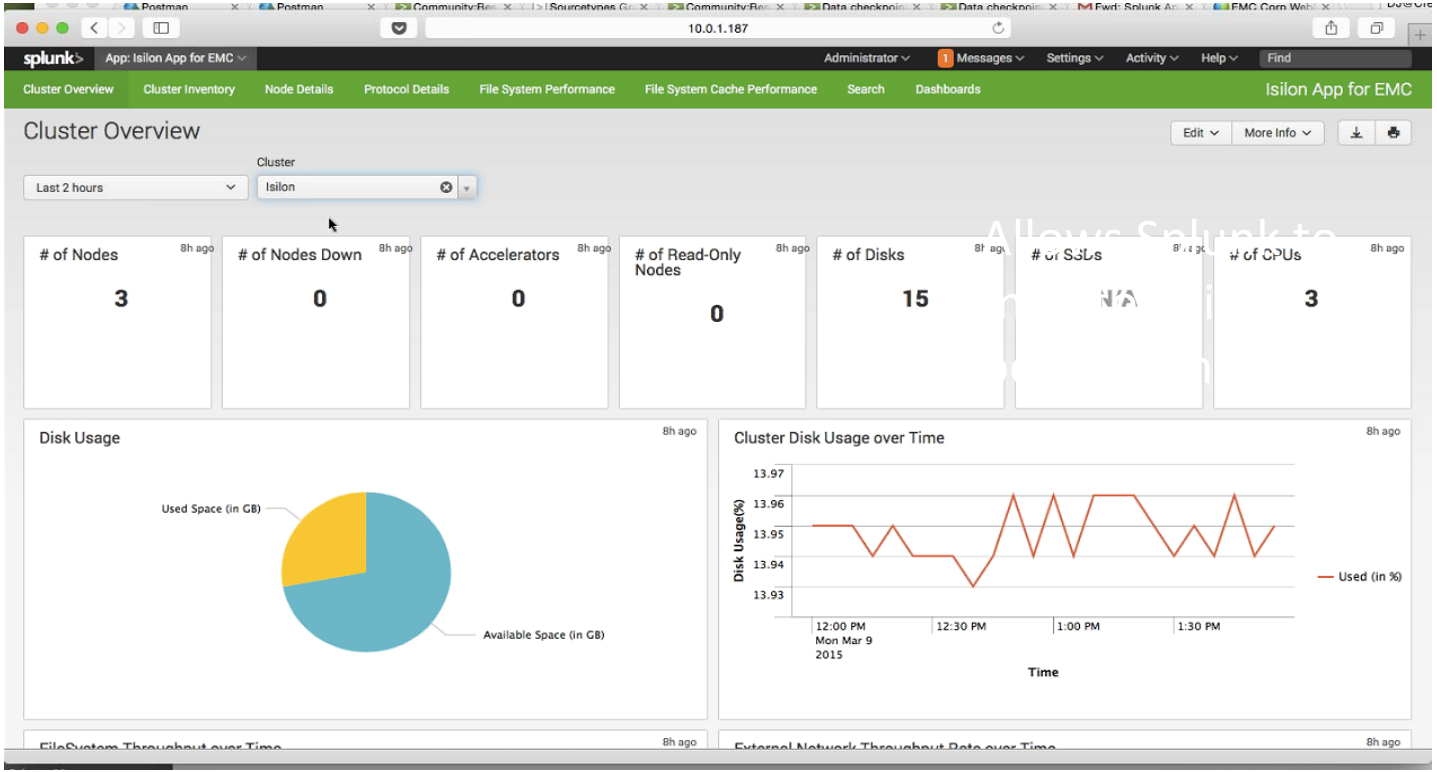


- VCE Systems presented as an entity
- Compliance history – what has been changed?
- System inventory and health
- KPI dashboards
- One command to configure system logs and events

Splunk Apps

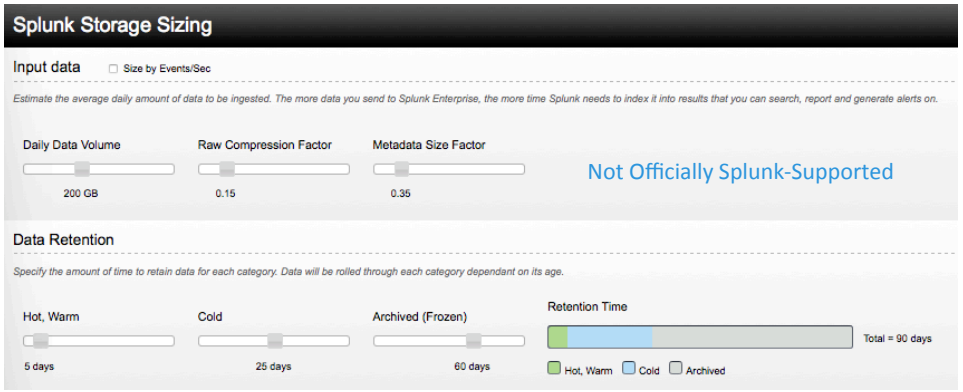
XtremIO App

Isilon App



How We Size Splunk Infrastructure

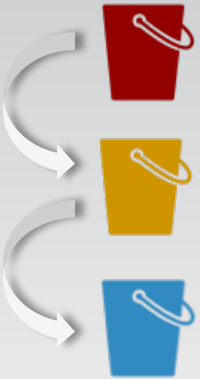
- We Use Splunk Best Practices “Religiously”
- We build “Converged Systems” first
- We have our own Ninjas to help!
- http://splunk-sizing.appspot.com/



Why EMC For Splunk

Optimized Infrastructure For Big & Fast Data

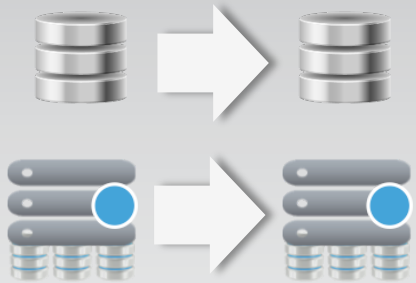
Optimized Shared Storage & Tiering



Hot & Warm Data Deployed On XtremIO or ScaleIO

Cold & Frozen Data Deployed On Isilon

Cost-Effective & Flexible Scale-Out



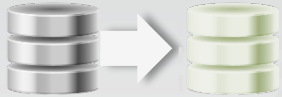
Scale-Out Capacity & Compute Independently Or As Converged Platform

Powerful Data Services

Encryption & Security

Index File Compression

Snapshots For Backups



Deduplication Of Clustered Indexes

These ARE The Droids You Are Looking For…


EMC²

THANK YOU

splunk>