.conf2015

Keeping Splunk in Check: Tools to Better Manage Your Investment

Aaron Kornhauser

Sr. Professional Services Consultant, Splunk, Inc.

Vladimir Skoryk

Sr. Professional Services Consultant, Splunk, Inc.

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.

Agenda

- Introduction
- Reference Hardware
- Available Tools
- Common Questions
- Scenarios/Troubleshooting
- Resources
- Q&A

Who Are We?

Hello, I'm:

Aaron Kornhauser

Sr. PS Consultant

akorn@splunk.com

Vladimir Skoryk

Sr. PS Consultant

vs@splunk.com

Reference Hardware

Role	Core Splunk*	Enterprise Security (ES) †
Indexer	12 CPU cores 12GB of RAM 800 IOPS/indexer RAID 1+0 **data ingest: 150-200GB/day**	12 CPU cores 12GB of RAM 800 IOPS/indexer RAID 1+0 **data ingest: 100GB/day**
Search Head	16 CPU cores 12GB of RAM 2x 300GB 10k rpm SAS in RAID1	16 CPU cores 16GB of RAM 2x 300GB 10k rpm SAS in RAID1

All instances x64, CPU > 2Ghz per core

* http://docs.splunk.com/Documentation/Splunk/latest/Capacity/Referencehardware

† http://docs.splunk.com/Documentation/ES/latest/Install/DeploymentPlanning

Available Tools

So what's out there and what's the difference?

Distributed Management Console (DMC) – Built in and only available on v6.2+

- http://docs.splunk.com/Documentation/Splunk/latest/Admin/ConfigureTheMonitoringConsole
- Splunk supported and focuses on all facets of the deployment
- New feature preso with Patrick/Octavio – make sure you go see it!

FireBrigade

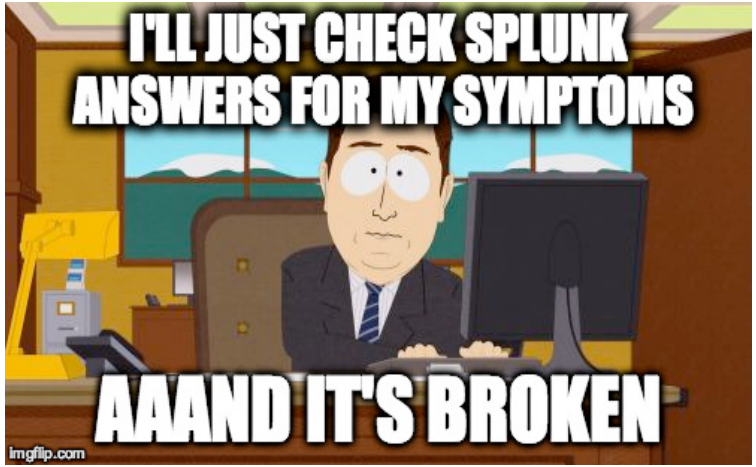
- https://splunkbase.splunk.com/app/1632/
- Detailed look at index/bucket activity and capacity

SoS (Splunk on Splunk)

- https://splunkbase.splunk.com/app/748/
- Legacy Splunk troubleshooting tool

Our health app – Splunk Health Overview

- https://splunkbase.splunk.com/app/1919/
- Combination of views found to be helpful in the field



Note:

Deployment monitor app is deprecated – try to stay away from it

Many of these app functionalities are being rolled in the DMC

How Are Things, Overall?

High level environment status – quick view of what's up/down/not reporting:

- Forwarder health - finding forwarders that we haven't seen for awhile
- Data source health - how are our data feeds doing?
- REST endpoints (| rest /services/server/info) - looking at system information, possibly under provisioned ones

Spotting warnings and errors within Splunk _internal:

- index=_internal sourcetype=splunkd (log_level=ERROR OR log_level=WARN) | cluster showcount=t | table cluster_count host log_level message | sort – cluster_count | rename cluster_count AS count, log_level AS level
- index=_internal sourceype=splunkd log_level!=INFO | timechart count by component

Track resource usage:

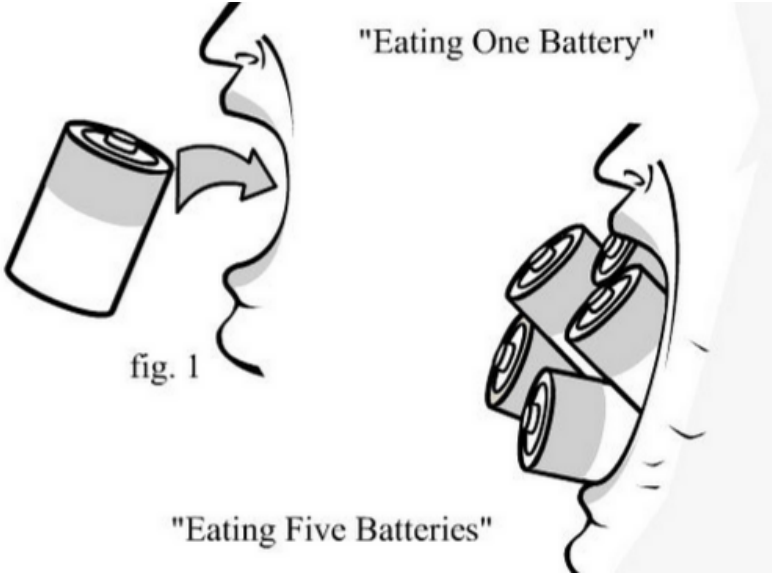
- Say hello to _introspection (Splunk 6.1+)
- Captures disk and other resource metrics (by default on full installs)
- http://docs.splunk.com/Documentation/Splunk/latest/Troubleshooting/Aboutthepl atforminstrumentationframework

Dashboards to help save the day:

- Health Status - Splunk Health Overview
- Instance - Distributed Management Console
- Indexing Performance - Distributed Management Console
- Resource Usage - Splunk Health Overview
- License Usage - Splunk Health Overview

Coming up

- Scenario based discussions around health topics
  - Environment overview
  - Data health
  - Configuration
  - Usage
  - Search insights



Scenario 1: Environment Overview

How to use the tools available to check overall health…

What are we reporting on?

- _internal
- _introspection
- metadata and using tstats
  http://docs.splunk.com/Documentation/Splunk/latest/SearchReference/Tstats
- REST endpoints
  - | rest /services/server/info
  - | rest /services/server/roles
  - | rest /services/server/status/resource-usage
- No need for additional addons

Scenario 1: Environment Overview

Splunk Health Overview – Health Status

Health Status

Edit | More Info

Today

This dashboard is intended to show the overall status and health of your Splunk deployment.
The dashboard provides server availability via REST calls, utilization for CPU and memory, as well as Splunk internal messages.
To drilldown onto a particular host in question, click on the server name and additional panels will be populated. Also, the timechart has a pan and zoom feature which allows you to specify a timerange on the chart to filter the panel below.

Historical Unreachable Splunk Server Instances

2m ago

Host	Roles	Message	Sparkline	Count	First Occurrence	Last Occurrence
		Unable to distribute to peer named [redacted] at uri https://[redacted] because peer has status = "Down".		2	08/31/2015 15:39:32.251	08/31/2015 15:39:34.902

Doesn't meet reference hardware

Current Splunk Server Status

2m ago

Status	Server	Role	OS	Cores	Avail Mem (MB)	CPU System Utilized	CPU User Utilized	Mem Utilized (MB)	Version
✓	[redacted] idx1	indexer cluster_slave	Linux	6	15949	2.73	2.00	13254.38	6.2.2
✓	[redacted] -idx4	indexer cluster_slave	Linux	6	11909	2.33	3.72	11751.90	6.2.2
✓	[redacted] idx3	indexer cluster_slave	Linux	6	15949	1.85	9.27	15483.36	6.2.2

Forwarder Status

Status	Forwarder	Current Time	Latest Event	Communicated Minutes Ago
✓	[redacted]	08/31/2015 15:41:25.000	08/31/2015 15:40:26.119	0.98
✓		08/31/2015 15:41:25.000	08/31/2015 15:40:39.527	0.76
✓		08/31/2015 15:41:25.000	08/31/2015 15:40:50.254	0.58
✓		08/31/2015 15:41:25.000	08/31/2015 15:41:01.987	0.38
✓		08/31/2015 15:41:25.000	08/31/2015 15:41:08.017	0.28
✓		08/31/2015 15:41:25.000	08/31/2015 15:41:14.620	0.17

Scenario 1: Environment Overview

Splunk Health Overview – Heath Status

Data Feed Status

4m ago

Status	sourcetype	Last Indexed	Events
✓	itseclog	08/31/2015 15:40:42	12,032,511
✓	postfix_syslog	08/31/2015 15:12:05	5,717,049
✓	Snare:Application	08/31/2015 15:30:55	1,225,604
✓	msql_agent_heartbeat	08/31/2015 15:39:50	427,238
✓	syslog	08/31/2015 15:41:09	96,508
✓	Snare:System	08/31/2015 15:35:01	62,855
✓	rhsm-too_small	08/31/2015	
✓	juniper:sslvpn	08/31/2015	
✓	rhsmcertd-too_small	08/31/2015	
!	bluecoat	08/31/2015 08:39:52	293,925,293

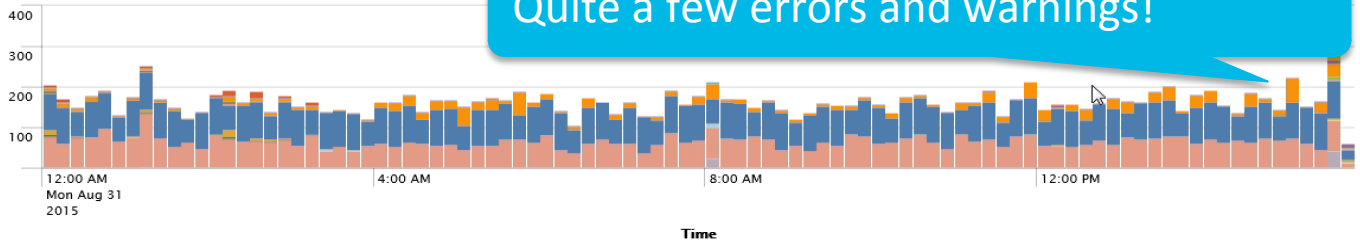
« prev 1 2 3 4 5 6 next »

Looks like source stopped sending data!

Splunk Internal Warning and Error Messages

4m ago

Quite a few errors and warnings!



Scenario 1: Environment Overview

DMC - Instances

Overview | Instances | Indexing Performance | Search Activity | Resource Usage | KV Store | Licensing | Setup | Search | Distributed Management Console

Instances

Splunk Instances are listed here. Learn More

Group:

All

17 instances

Issues accessing instance

Instance	Machine	Role	OS	Cores	RAM		Status	Action
-idx5		Indexer					Unreachable	Views
-idx6		Indexer	Linux	8	31.36 GB	6.2.2	Up	Views
-idx1		Indexer	Linux	6	15.58 GB	6.2.2	Up	Views
-idx2		Indexer	Linux	6	15.58 GB	6.2.2	Up	Views
-idx3		Indexer	Linux	6	15.58 GB	6.2.2	Up	Views
-idx4		Indexer	Linux	8	31.36 GB	6.2.2	Up	Views
-idx7		Indexer	Linux	6	15.58 GB	6.2.2	Up	Views
-idx8		Indexer	Linux	6	15.58 GB	6.2.2	Up	Views

Scenario 1: Environment Overview

DMC – Indexing Performance

Overview | Instances | Indexing Performance | Search Activity | Resource Usage | KV Store | Licensing | Setup | Search | Distributed Management Console

Indexing Performance: Deployment

Group:
Indexer

Select views: All Snapshot Historical

Snapshots

Overview of Indexing Performa

Slight ingestion imbalance

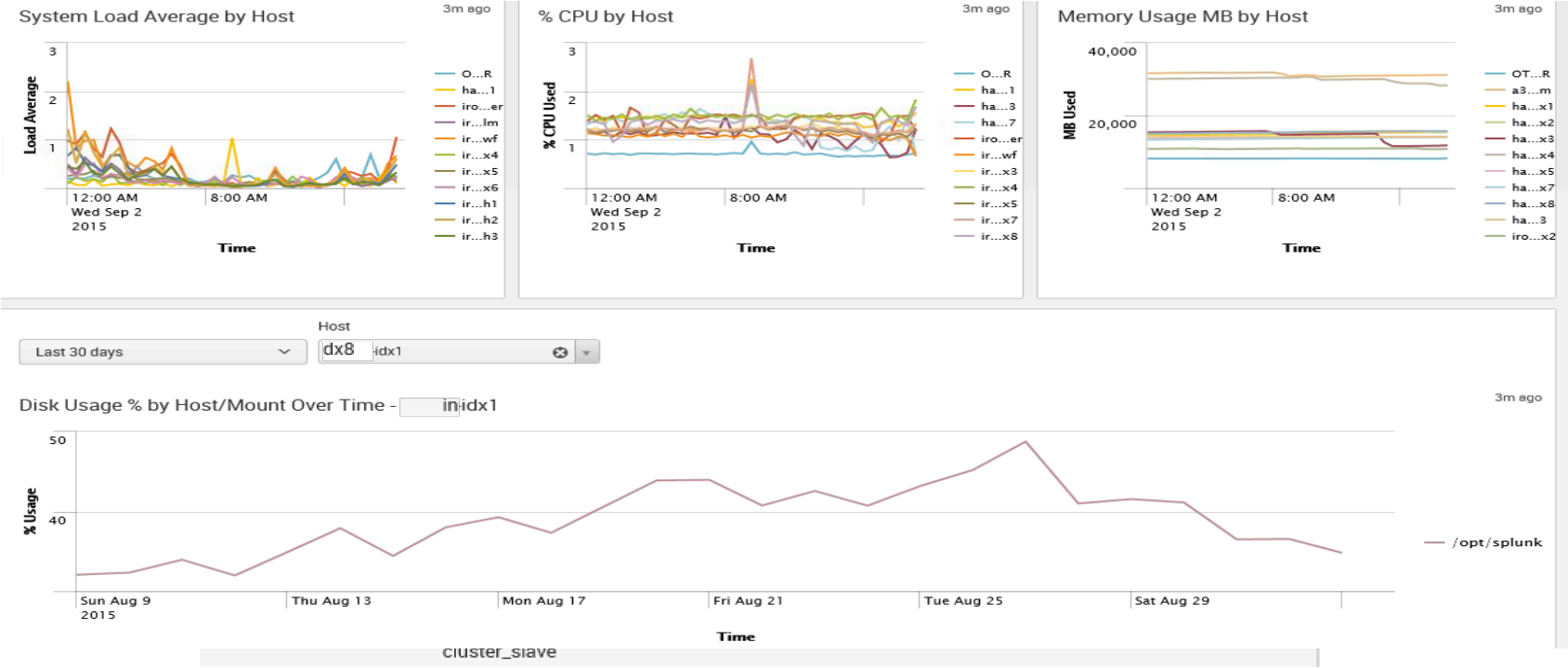
10	14 KB/s	9 KB/s
INDEXERS	TOTAL INDEXING RATE	AVERAGE INDEXING RATE	MEDIAN INDEXING RATE

Indexing Performance by Instance

Instance	Indexing Rate (KB/s)	Status	Parsing Queue Fill Ratio (%)	Aggregation Queue Fill Ratio (%)	Typing Queue Fill Ratio (%)	Indexing Queue Fill Ratio (%)
idx6	59	normal	0.00	0.00	0.00	0.00
idx8	20	normal	0.00	0.00	0.00	0.00
idx3	15	normal	0.00	0.00	0.00	0.00
idx6	12	normal	0.00	0.00	0.00	0.00
idx4	9	normal	0.00	0.00	0.00	0.00
idx3	9	normal	0.00	0.00	0.00	0.00

Scenario 1: Environment Overview

Splunk Health Overview – Resource Usage



Scenario 2: Data Imbalance

Splunk Health Overview – License Usage

License Overview

Edit | More Info

Split-by

Sourcetype

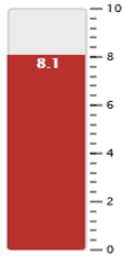
This dashboard is intended to show the overall status of license utilization across the different Splunk servers and data types.

The dashboard provides a high level overview of overall license used today, by index, and trending over time by index/environm

Splunk on Splunk.

Week over week license usage helps detect anomalies and growth

Todays Overall License Usage (GB) — 2m ago

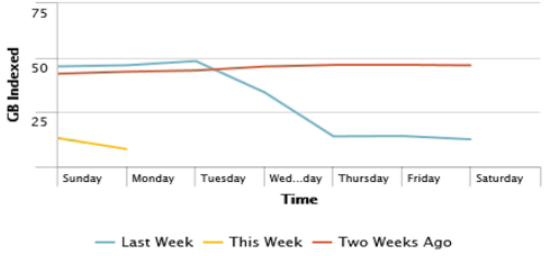


Total GB Ingested Today — 1m ago

Source Type	Total GB
oracle:audit:text	2.62
SOC_DB_AUD_ORACLE	1.73
ps	1.41
GoldenGate	1.22
top	0.49
bluecoat	0.21
netstat	0.21
msql_log	0.09
unix	0.06
cpu	0.02

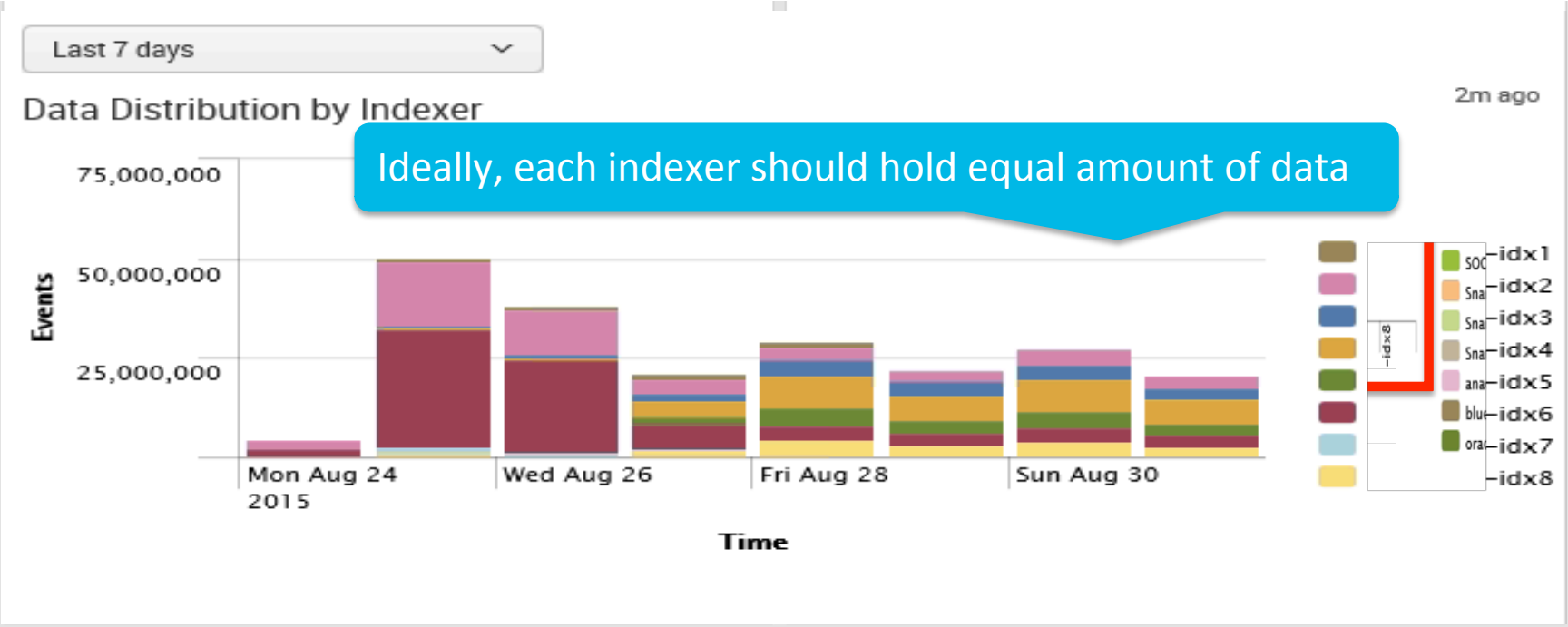
« prev 1 2 3 4 next »

Week on Week Usage - Current Day ... Partial — 2m ago



Scenario 2: Data Imbalance Continued

Splunk Health Overview – License Usage



Scenario 2: Data Imbalance - Troubleshoot/Wrapup

Troubleshooting:

- Validate firewall rules are in place
- Check that all forwarders have the correct outputs
- Ensure indexers all all listening on proper port
- Does splunkd.log have anything to say?
- Use the Indexing Overview and Configuration Overview (btool saves the day)

Possible Causes:

- Simple misconfiguration
- Data processing queues filling up and forwarders timing out and jumping to next indexer
  - Check Distributed Indexing Performance in the DMC for queue filling - typical sign of disk performance issues
- Indexer affinity - the forwarders get stuck to one indexer because EOF never met
  - forceTimebasedAutoLB can help! http://blogs.splunk.com/2014/03/18/time-based-load-balancing/
  - Updating syslog files - each file <1GB, host in the path, broken out by sourcetype, cron job/ logrotate to remove stale files.

Scenario 3: Data Health Checkup

How's your data feeling?

Feed still working
- Seeing recent data
- Gaps in data

Ingest issues
- Line breaking, time parsing, truncation

Indexing latency
- _time - _indextime

Predictive analytics…events in the future!
- incorrect time zone
- timestamp parsing issues
- time drift (NTP not set)

Make sure to see the "Onboarding Data Into Splunk" presentation!

Scenario 3: Data Quality

Greetings from the future!

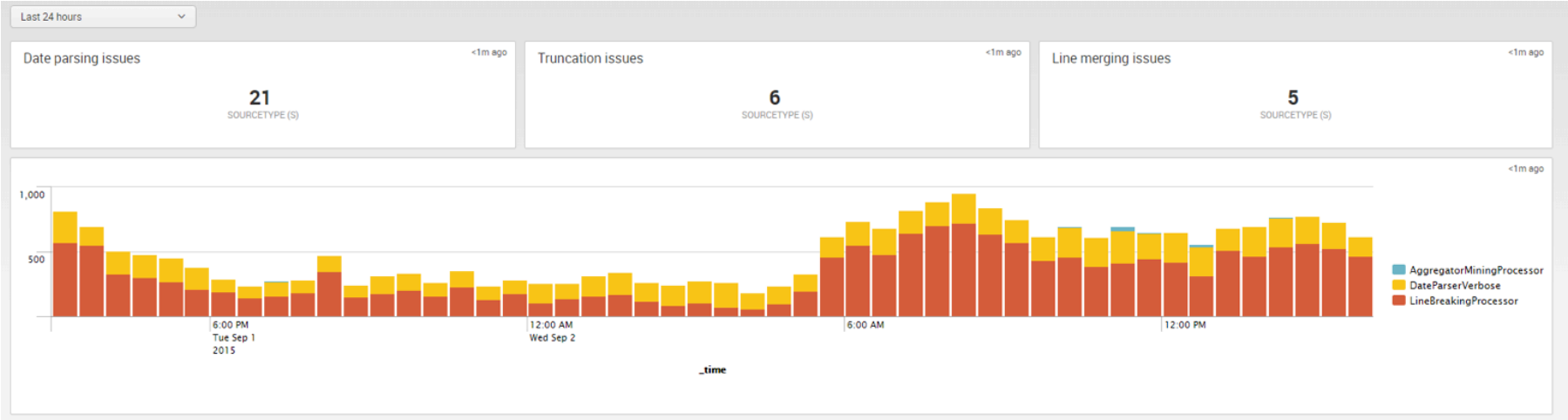
Your data is hours ahead of system time!

host	sourcetype	index	avg(ahead)
127.0.0.1	oracle:alert:xml	main	03:31:52.402025
127.0.0.1	oracle:listener:xml	main	03:31:53.755763

```
index=* earliest=+5m latest=+20y
        | eval ahead=abs(now() - _time)
        | stats avg(ahead) by host, sourcetype, index
        | eval avg(ahead)=tostring('avg(ahead)', "duration")
```

Scenario 3: Data Quality

Linebreaking and Timestamping




```
index=_internal sourcetype=splunkd
        component=LineBreakingProcessor OR component=DateParserVerbose OR
          component=Aggregator*
        | timechart count by component
```

Scenario 3: Data Quality

Indexing Delay

host	sourcetype	index	avg(delay)
10.250.140.48	NTSyslog:Security	main	131.750000
10.252.110.49	NTSyslog:Security	main	111.375000
127.0.0.1	oracle:alert:xml	main	26325.480000
127.0.0.1	oracle:audit:text		64.913043
127.0.0.1	oracle:audit:xml		72.913043
127.0.0.1	oracle:listener:xml	main	25731.600000
167.235.13.205	NTSyslog:Security	main	128.916667
ACME-001	Linux:SELinuxConfig	main	515.480000
ACME-001	Linux:Service	main	497.958333
ACME-001	Linux:Update	main	467.320000

We have latency!

```
| tstats earliest(_time) AS t earliest(_indextime) AS i WHERE index=* AND (earliest=-1d) BY host, sourcetype,
    index, _time span=1h
    | eval delay=abs(t - i)
    | where delay>5
    | stats avg(delay) BY host, sourcetype, index
```

Scenario 4: Consistency Is Key

- File order precedence
  - http://docs.splunk.com/Documentation/Splunk/latest/AdminWheretofindtheconfigurationfiles
  - Don't put configs in /etc/system/local
- Are like instances of Splunk uniformly configured?
  - Indexer A knows about more than Indexer B
  - Forwarder A knows about Indexer A & B
- Use configuration management tools
  - Deployment server, Chef, Puppet, SCCM, etc.
- Meaningful Splunk app naming conventions
  - org_group_application_configuration (acme_all_search_base)

Scenario 4: Consistency Is Key

Why do we have an extra app?

<1m ago

2
SPLUNK SERVERS

1
POSSIBLY MISCONFIGURED SERVERS

<1m ago

	sos_server	app_count	stanza_count	setting_count	app_deviation	setting_deviation	stanza_deviation	status
1	idx2	4	252	8435	1	1312	39	deviation found
2	idx3	3	213	7123	0	0	0	ok

Configuration Overview - Splunk Health Overview

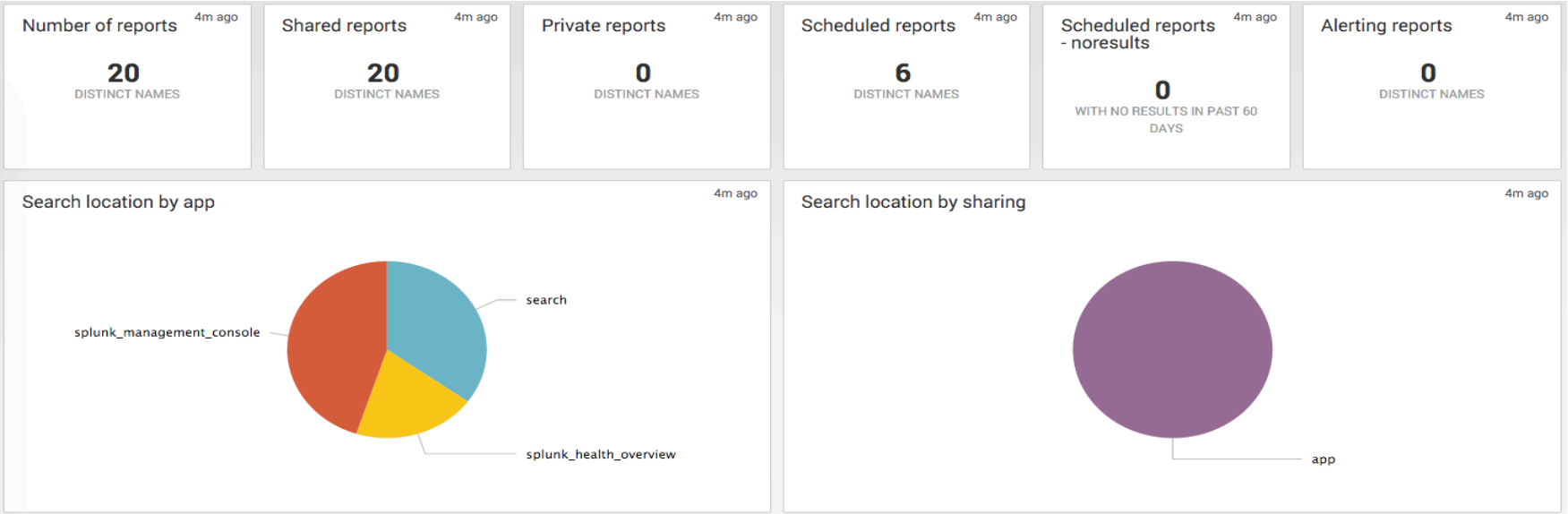
- Comparing btool output across like instances shows configuration inconsistencies

Scenario 5: Splunk Usage

- Inventories
  - Reports, dashboards, apps
- Search Activity
  - Are users running efficient searches?
  - http://docs.splunk.com/Documentation/Splunk/latest/Search/Writebettersearches
  - How are the scheduled jobs doing? Differed/Skipped?
- User activity monitoring
  - What views are being accessed
- Who has access to data
  - Roles and permissions
- Useful dashboards
  - Search Activity – Splunk Health Overview
  - Scheduler Activity – Splunk Health Overview
  - Search Activity: Instance – DMC

Useful app: Data Governance on apps.splunk.com

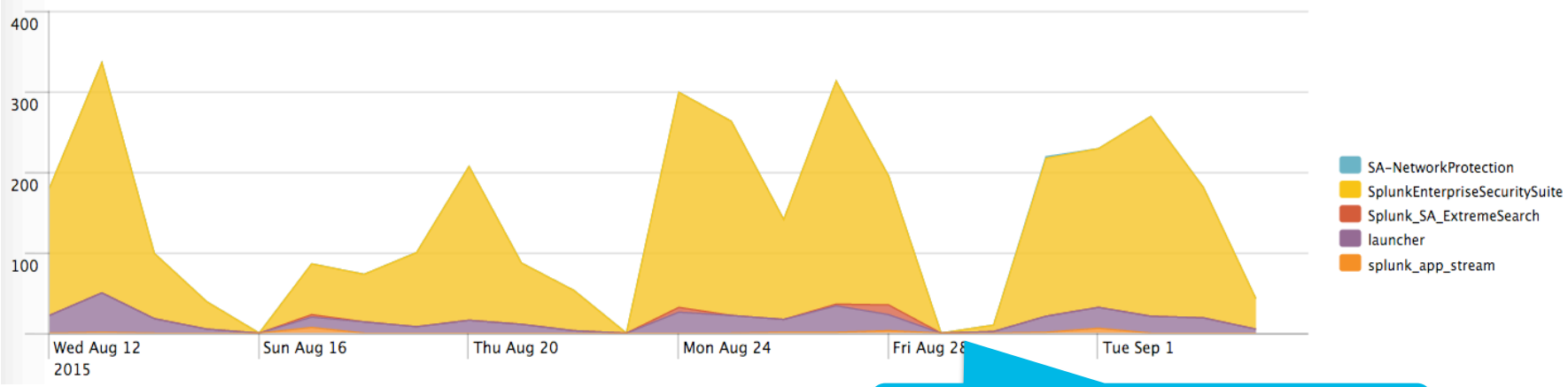
Scenario 5: Inventory Check



Saved Search Inventory - Splunk Health Overview

- | rest splunk_server_group=dmc_group_search_head /servicesNS/-/-/saved/searches

Scenario 5: User Activity



Can always spot weekends!

View and Dashboard Audit – Splunk Health Overview

- index="_internal" sourcetype=splunk_web_access GET app
  | rex "GET /[^/]+/app/(?<app>[^/ ?]+)/"
  | search app!="search" app=* AND user=* AND user!="-"
  | timechart limit=100 count by app

Scenario 6: Search Performance

Review search activity to ensure system and users are happy

Tools

- Search Activity – Splunk Health Overview
- Scheduler Activity – Splunk Health Overview
- http://docs.splunk.com/Documentation/Splunk/latest/Search/Writebettersearches
- Search Activity Instance – DMC

What to look for

- Long running searches
- Real time searches
- Concurrency
- Inefficient inline regular expressions
- Streaming commands before searching commands
- Scheduling - Frequently executed searches for long periods of time. ie running a search for the last day every minute

Scenario 6: Knowing What Is Being Searched

Search range by index

Split By:

Index

Values:

- Sum of Duration (Minutes)
- Average Duration (Minutes)
- Count of Searches

Looks like bulk of the searches cover 45 days

Search Window Statistics by Index

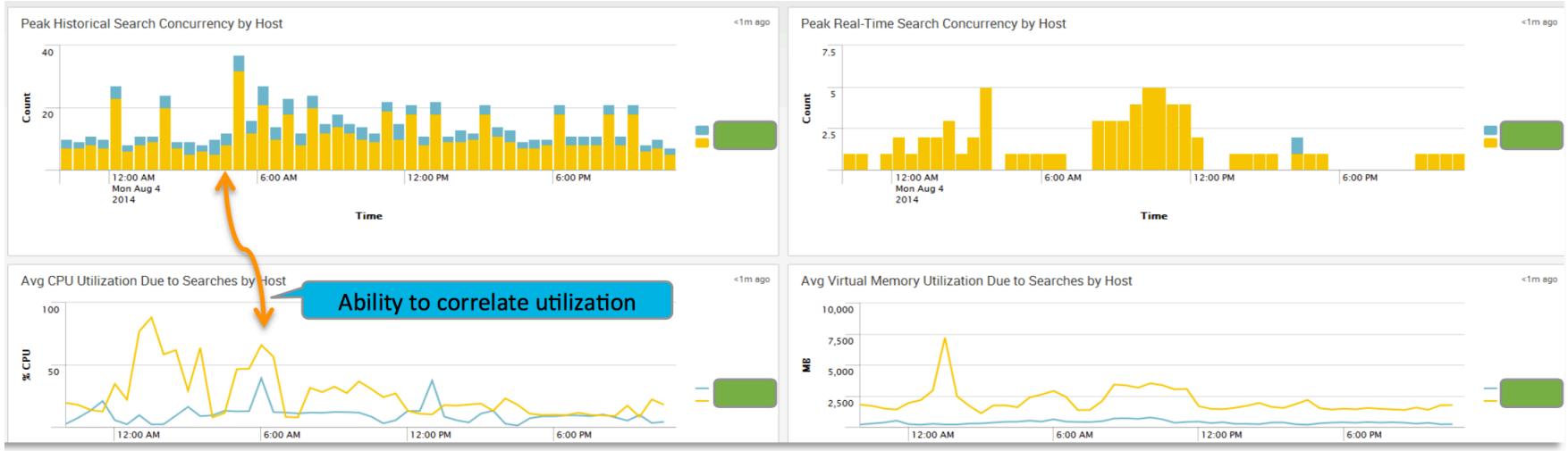
<1m ago

Index	5 Minutes	15 Minutes	1 Hour	4 Hours	1 Day	7 Days	30 Days	45 Days	45 Days +	All Time	Total
_audit	0	0	0	0	0	0	0	0	0	1	1
_internal	0	0	0	0	11	0	0	15	15	0	41
main	0	0	0	0	0	0	0	16	0	0	16
sos	0	0	0	0	3	0	0	0	0	0	3
summary	0	0	0	0	0	1	0	0	0	1	2
Total	0	0	0	0	14	1	0	31	15	2	63

Search Activity – Splunk Health Overview

Scenario 6: Search Performance

Understanding concurrency



Search Activity – Splunk Health Overview

- The total number of concurrent searches is base_max_searches + #cpus*max_searches_per_cpu
- max real-time searches = max_rt_search_multiplier x max historical searches
  - Set in limits.conf

Scenario 6: Search Performance

Inspecting Searches

Start Time	End Time	Search Earliest	Search Latest	count	range	Search	User	Run Time (Min)
09/04/2015 08:57:33.824148	09/04/2015 08:58:02.453392	ZERO_TIME	ZERO_TIME	2	All Time	'search index=internal | rex field=_raw "(?<AAA>.*)'"	admin	0.07
09/04/2015 08:57:43.087450	09/04/2015 08:58:02.461713	ZERO_TIME	ZERO_TIME	2	All Time	'search index=_internal | rex field=_raw "(?<AAA>.*)'"	admin	0.16
09/04/2015 08:58:01.287952	09/04/2015 08:58:32.458172	ZERO_TIME	ZERO_TIME	2	All Time	'search index=_internal Error* Fail*'	admin	0.08
09/04/2015 08:58:09.270373	09/04/2015 08:58:32.478625	Fri Sep 4 04:58:00 2015	Fri Sep 4 08:58:09 2015	2	04:00:09.000000	'search index=_internal Error* Fail*'	admin	0.07
09/04/2015 09:51:53.791042	09/04/2015 09:52:23.104457	Fri Sep 4 07:30:00 2015	Fri Sep 4 09:00:00 2015	2	01:30:00.000000	'search index=_audit host=akornhauser-mbp15 action=search user!=splunk-system-user search_id=* (info=granted OR info=completed) |rex field=apiStartTime '"	admin	0.24

Search Activity – Splunk Health Overview

Other helpful views:

Job inspector - http://docs.splunk.com/Documentation/Splunk/latest/Knowledge/ViewsearchjobpropertieswiththeJobInspector

Job Viewer

Search Activity Instance - DMC

Wrap up: Other Sanity Checks

Validate ulimit settings:

- -n open files (>2048)
- -f file size (unlimited?)
- -d data seg size (>1GB)

Ensure THP is disabled on Linux distros:

- http://docs.splunk.com/Documentation/Splunk/latest/ReleaseNotes/SplunkandTHP

Index sizing:

- Ensure that higher volume indexes (>10GB/day) are tuned with maxDataSize = auto_high_volume and have the appropriate number of maxHotBuckets
- Using Fire Brigade can help determine index bucket sizing.
- More buckets = more scanning = slower searches

Scaling Splunk – Knowing What To Look For

Key things to look for…

Meeting the reference hardware specs

- Indexing volume
- 150-200GB/day/indexer non-ES / ~100 GB ES
- Talk to your friendly sales rep!

Data retention

- Can you meet your retention SLA?

System load

- Is your system load > # cores?

Number of users/searches

- Check search concurrency - real time and historical

.conf2015

Q&A

splunk>

THANK YOU