What's New in Splunk 7.3

Dave Shpritz, Aplura Splunk Practice Lead

Baltimore Splunk User Group

June 17th, 2019



Splunk 7.3

- Codename: PinkyPie
- "Dark Data"
- More getting data in (and getting in more data)
- Other market-y stuff (AR, Mobile)
- Not really interested in this stuff, so we aren't going to cover it



32616	Splunkd	%CPU USER		0 used. 615976 avail Men	
32695	tcpdump	0.3 splunk	1000	TIME+ SOMEM PR NI VIRT PC	APLURA
590	named	0.3 tcpdump 0.3 named	72	0:02.45 0.7 20 0 135356 43956 8940 5 1 1000 ptint and the	
1602	vim	0.3 dan	1028	7:31.12 1.8 20 0 548272 18072 5464 5 3268 72 toping 72 t	Any Solutions, One Goal.

What are we going to cover?

- SmartStore (S2)
- Searchable Data Rebalance
- Indexer Clustering Performance
- Search Performance
- Some Cloud
- SHC Deployer changes
- Indexing Pipeline
- Metrics
- Workload Management (WLM)
- Token Authentication (finally)
- LDAP
- Time Fields

32616	splunkd	%CPU USER		0 used. 615976 avail Mem	
32695 590	tcpdump named	0.3 splunk 0.3 tcpdump	1 000 72	TIME+ SHEM PR NI VIRT RES SHE S PPID NUTD NUSSER SHE S	APLURA
1602	Vim	0.3 dan	25 1028	7:31.12 1.8 20 0 548272 18072 5464 5 32680 72 taplan 72 72 72 74 74 74 75 75 75 75 75 75 75 75 75 75 75 75 75	Many Solutions, One Goal.

SmartStore (S2)

- Now supports Report and Data Model acclerations
- This means ES is now supported on SmartStore
- New retention settings (size)
- Support for non-clustered indexers and indexes
- Better resiliency (?)
- Better scalability (?)



Searchable Data Rebalance

- Everyone has to be on 7.3
- Previously, data rebalance is not search safe
- The removal of buckets could cause differences/inaccuracy in results
- 7.3 added workflow to make removal of excess buckets search safe
- Available as a checkbox when performing rebalance
- Other things aren't available when doing this
 - Excess bucker removal
 - Rolling restart
 - Rolling upgrade
- There is a timeout, so longer running searches will still be subject, as will indexed real-time searches (mostly in hot, so mostly should be ok)

32616	Splunkd	%CPU USEP		0 used. 615976 avail He	/
32695	tcpdump	0.3 splunk	UID	TIME+ SMEM PR NI VIRT	APLURA
590	named	0.3 tcpdump	72	0:02.45 0.7 20 0 135356 43956 8940 5 1 1860 mint and	
1602	Vim	0.3 dan	25 1028	7:31.12 1.8 20 0 548272 18072 5564 5 32680 72 trading 72 2 trading 72 12	Many Solutions, One Goal.

Indexer Clustering Performance

- Changes to how the UI displays data
- Used to show intermediate changes, now caches
- More logging (event=rfMet, event=sfMet, event=allSearchable)
- Could mean that on a rolling restart, SF/RF appears to "flap" more
- tsidxWritingLevel = 3



Search Performance

- Lots of smaller improvements
- Some larger ones too
 - Stats to tstats
 - tsidxWritingLevel and Data Model Acceleration
 - Data Model UI index constraints
 - CIDR matching
 - Compression
 - Datamodel command
 - Post Process

32616	Splunka	%CPU USER		0 used. 615976 avail the	
32695	tcpdump	0.3 splunk	1000	TIME+ SHEN PR NI VIRT	APLURA
590	named	0.3 tcpdump	72	0:02.45 0.7 20 0 135356 43956 8940 5 1 1000 mills sim ci anu	
1602	Vim	0.3 dan	1028	7:31.12 1.8 20 0 548272 18072 5584 5 1 25 tobar 72 topany 72 topan	Many Solutions, One Goal.

Search Performance - Stats to tstats

- Optimizes searches that use stats command
- Converts them to tstats under-the-hood
- On by default, but can be disabled using the "noop" command
- Will work with any indexed field
 - As long as they are in fields.conf
 - Remember, fields.conf is not sourcetype scoped, so, careful with that ax



Search Performance - tsidxWritingLevel and Data Model Acceleration

- Writing level in 7.2 (level 2 introduced in 7.2, 7.3 adds a level 3)
- All indexers and search heads have to have this set
- A collection of enhancements to how tsidx files are written/structured
- Large space and search performance gains
- Previously the DMA tsidx files were only using level 1 (even if level 2 was set for the index)
- DMA now will use the same enhancements

32616	Splunkd	%CPU USER		0 used. 615976 avail Mem	
32695	tcpdump	0.3 splunk	1000	TIME+ SMEM PR NI VIRT BE	APLURA
590	named	0.3 tcpdump 0.3 nomed	72	0:02.45 0.7 20 0 135356 43956 8940 5 1 1000 splust 3mm cm day	
1602	Vim	0.3 dan	1028	7:31.12 1.8 20 0 548272 18072 5464 5 32680 72 toplan 72 72 72 72 72 72 72 72 72 72 72 72 72	Many Solutions, One Goal.

Search Performance - Data Model UI index constraints

- Best practice is to have index constraints in place
- CIM app uses macros to implement this (please check yo' self)
- 7.3 enforces that a DM must have a constraint in place
- Should be macro aware (like the macros in CIM)
- Can still have non-constrained searches in JSON
- Index=* is a valid constraint ☺



Search Performance - CIDR matching

- General performance improvements
- tstats would not perform negated CIDR, now it does
- Search is now IPv6 CIDR aware (no love for tstats)



Search Performance - Compression

- Zstandard compression (look, Facebook did something good!)
- Less space, less CPU usage
- 7.2 introduced this for journals
- Search results still used gzip
- Now defaults to zstd
- You can use a splunkd command to decompress
- Alert actions still get gzip
- Note that there is no zstandard decompression module in the bundled python

32616	Splupkd	%CPU USED		0 used. 615976 gygil to	/
32695	tcpdump	0.3 splunk	UID	TIME+ SHEN PR NT	
590	named	0.3 tcpdump	72	0:02.80 4.3 20 0 135356 43956 8948 5 10 MID MISER SHI CT CAN	
160 2	Vim	0.3 named	25	7:31.12 1.8 20 0 548272 18072 5584 5 1 72 tuphane 72 tu	Many Solutions, One Cool
1242		0.3 dan	1028	0:00.08 0.6 20 0 152828 6248 2548 5 38588 1878 day	ivially solutions, One Goal.

Search Performance - Datamodel command

- New flag for the "datamodel" command to allow the use of "summariesonly=true"
- Should allow for faster drilldowns
- Not available for the "from" command



Search Performance - Post Process

- Post process searches used to be run by the same splunkd process
- Could cause memory issues
- Makes the execution of them smarter, moves them to search pipelines
- Config options can disable this if there are problems



Some Cloud

- Better interface for the index manager page
 - Makes SmartStore retention easier
- Relative Search Concurrency
 - Now in the UI
 - On prem too
 - Includes "max_searches_perc" and "auto_summary_perc"
 - Check yo self



SHC Deployer Changes

- Review: config merging
- Now we get some control on this via app.conf and the [shclustering] stanza
- deployer_lookups_push_mode
 - preserve_lookups (honors CLI)
 - always_preserve (ignores CLI)
 - always_overwrite(ignores CLI)
- deployer_push_mode
 - merge_to_default
 - local_only
 - default_only
 - full

32616	Splunkd	%CPU USER		0 used. 615976 avail Me	
32695	tcpdump	0.3 splunk	1000	TIME+ SHEM PR NI VIRT REF	RA)
590	named	0.3 tcpdump 0.3 named	72	0:02.45 0.7 20 0 135356 43956 8940 5 1 1000 plust 300 ch day	
1242	VIM	0.3 dan	1028	0:00.08 0.6 20 0 152828 5248 2548 5 1 2 Many Solutions	, One Goal.

SHC Push Modes

- merge_to_default
 - Default
 - Like what we currently have
- local_only
 - Only pushes /local configs
 - Could be used for something like built-in apps ("search")
 - Only delivered to the captain
- default_only
 - Only pushes /default configs
 - Gets delivered to all nodes/members
- full
 - No merging
 - default to default, local to local

32616	Splupkd	%CPU USER		0 used	615976	buff/cache			
32695	tcpdump	0.3 splunk	UID	TIME+ SMEM	PR NT				APUURA
590	named	0.3 tcpdump	72	0:02.80 4.3	20 0	135356 43956	SHR S PPID RUID RUSER	300 GD 000r	
1602	VIM	0.3 named 0.3 dan	25	7:31.12 1.8	20 0	28588 6872 548272 18072	5464 \$ 32688 72 topic		Many Solutions One Goal
			1020	0:00.08 0.6	20 0	152828 6248	2548 5 38588 1878 day	All in the second se	wany Johutions, one doan

Indexing Pipeline - Metrics

- Better metrics on pipeline usage
 - Better instrumentation
 - "metrics.log" "group=dutycycle"
 - Includes management, ingest, misc types
 - The "ratio" field is a measurement of busyness (via maths)
 - Will fluctuate at first before it stabilizes
- Why do we need this? (aside from better logging of a logging product)



Indexing Pipeline – Pipeline Set Selection

- Currently just uses round-robin
- Could lead to stuffed and starved pipelines
- New server.conf config for "pipelineSetSelectionPolicy"
- "round_robin" or "weighted_random"
- "weighted_random"
 - Uses more maths
 - Should improve throughput
 - There are settings to change some of the variables on this selection process

32616	Splunkd	%CPU USER		0 used. 615976 avail the	
32695	tcpdump	0.3 splunk	1000	TIME+ SMEM PR NI VIRT PT	LURA
590	named	0.3 tcpdump 0.3 named	72	0:02.45 0.7 20 0 135356 43956 8340 5 1 300 study at an at a	
1602	Vim	0.3 dan	1028	7:31.12 1.8 20 0 548272 18072 5564 5 32560 72 toping 72	lutions, One Goal.

Metrics

- Reduced storage footprint and increased search performance on metrics indexes
- Metrics Workspace now included with Splunk Enterprise
- Also available via Splunkbase
- Added multi-series charting
- Better aggregation of common fields across indexes
- Better accessibility and localization
- Metrics rollup
 - Think "summary indexing, but for metrics"
 - Take very fine measurements, roll them up into aggregates for faster searching

32616	Splupkd	%CPU LISER		0 used. 615976 avail the	/
32695	tcpdump	0.3 splunk	UID	TIME+ SHEN PR NT HTTP	APLURA
590	named	0.3 tcpdump	72	0:02.80 4.3 20 0 135356 43956 8946 5 1 100 NUSER SUD COLORD	
1602	Vim	0.3 dan	25 1028	7:31.12 1.8 20 0 548272 18872 5564 5 32660 72 toping 72 7	Many Solutions, One Goal.

Workload Management (WLM)

- Uses Pools
 - Get assigned CPU and memory resources
- Linux only (uses Linux cgroups under the hood)
- Prioritize searches (by app, user, type)
- Resource reservation
- System protection
- Splunkd processes run under pools
- Assignment by manual addition or by rules
- Now with a better UI

32695 tcpdump 0.3 splunk 1000 0:02.80 4.3 Street ALL APLURA	
200 0:02.80 4.2 TO URT RES SHEET	
590 named	
1602 vim 0.3 named 25 7:31.12 1.8 20 0 548272 18072 5464 5 32600 72 taken 7 taken 7 Many Solutions One (Goal.

Token Authentication

- On prem only
- Previously Splunk didn't have a great way for API usage (REST)
- This lead to people doing some pretty gross things
- JWT (JSON Web Tokens)
- Token gets put in the `Authorization` headers for requests
- New settings in authorize.conf for [tokens_auth]
- New role capabilities for token viewing and management

32616	Splupkd	%CPU LISER		0 used. 615976 gygil km	/
32695	tcpdump	0.3 splunk	UID	TIME+ SMEN PR NT HTT	
590	named	0.3 tcpdump	72	0:02.80 4.3 20 0 135356 43956 8940 5 1 MUD RUSER SUD CD GROW	
160 2	Vim	0.3 named 0.3 dan	25	7:31.12 1.8 20 0 548272 18072 5584 S 1 25 training 72	Many Solutions, One Goal
			1020	0:00.08 0.6 20 0 152828 5248 2548 5 2658 4 10	wany Johutions, One doan

LDAP

- Now has caching
- Caching has paging (lots of tweaking available)
- On prem and Cloud
- Should allow for very large LDAP queries (thousands of users/groups)



Time fields

- ADD_EXTRA_TIME_FIELDS in props.conf
- Used to be "true" and "false"
- Includes "date_hour, date_mday, date_minute, date_month, date_second, date_wday, date_year, date_zone, timestartpos, timeendpos, timestamp"
- Now has options
 - "none" or "false"
 - Um. None. Including sub-second info.
 - "all" or "true" (default)
 - Buddhist at a hot-dog stand
 - "subseconds"
 - None of the extra fields, but still the sub-second info

32616	Splunkd	%CPU USER		0 used. 615976 avail Men	
32695	tcpdump	0.3 splunk	UID 1000	TIME+ 304EM PR NI VIRT RES COL	APLURA
590	named	0.3 tcpdump 0.3 named	72	0:02.45 0.7 20 0 135356 43956 8940 5 1 1000 splast 300 300 and 5	
1002	Vim	0.3 dan	1028	0:00.08 0.6 20 0 152828 6748 2548 5 1 25 tuber 72 tuber 7	lany Solutions, One Goal.