## What's New in Splunk 7.3

Dave Shpritz, Aplura Splunk Practice Lead
June/July, 2019

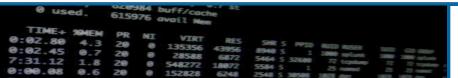
**APLUR** 

Many Solutions, One Goal.

## 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







## 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





## SmartStore (S2)

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

Many Solutions, One Goal.

#### 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 of indexers
  - Rolling upgrade of indexers
- There is a timeout, so longer running searches will still be subject to kill, as will indexed real-time searches (mostly in hot, so mostly should be ok, but edge cases)





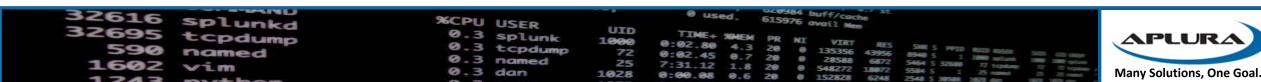
## Indexer Clustering Performance

- 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



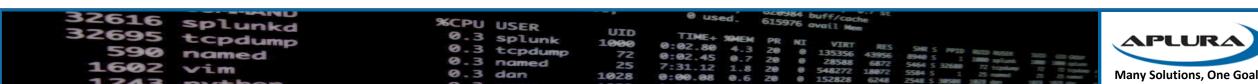
#### Search Performance - Stats to tstats

- 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 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



# 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





#### 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

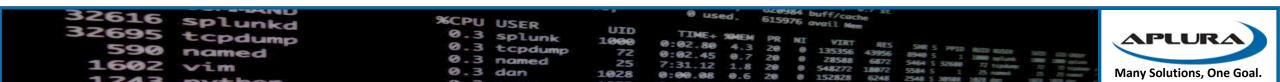
- Better interface for the index manager page
  - Makes SmartStore retention easier
- Relative Search Concurrency
  - Now in the UI
  - On prem too
  - Includes "max\_searches\_perc" and "auto\_summary\_perc"
  - Check yo' self

**APLUR** 

Many Solutions, One Goal.

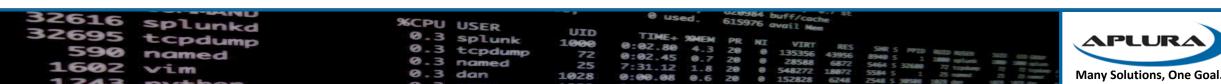
## SHC Deployer Changes

- 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



#### SHC Push Modes

- merge\_to\_default
  - Default
  - Like current behavior in < 7.2</li>
- 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



### Indexing Pipeline - Metrics

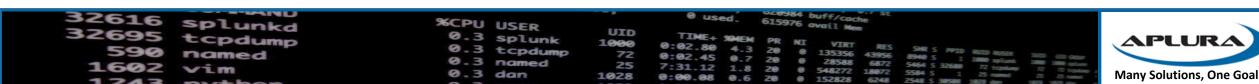
- Better metrics on pipeline usage
  - Better instrumentation
  - "metrics.log" "group=dutycycle"
  - Includes management, ingest, misc types
  - The "ratio" field is a measurement of busyness (via maths)
  - Will fluctuate at first before it stabilizes
- Why do we need this? (aside from better logging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





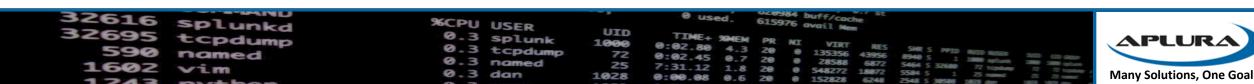
## Indexing Pipeline – Pipeline Set Selection

- 7.2 and prior 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



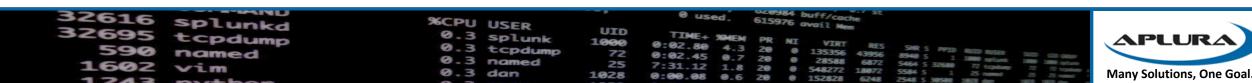
#### Metrics

- 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



## 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



#### 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



#### 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



