splunk> .conf2017

Using Splunk Enterprise To Optimize Tailored Long-term Data Retention

Tomasz Bania | Incident Response Lead, Dolby

Eric Krieser | Splunk Professional Services

September 2017 | Washington, DC

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.

Overview

Problems To Address

- Limited Retention Capabilities
- Search Completion Time for Extended Searches

Goals

- Maintain key event fields for a specific retention period
- Maintain the smallest footprint possible for archiving this retained content

Challenges

- Storage Budget Limitations
- Significant Bloat in Current Data Set

Considerations

- Standard summaries using stats, and sistats adds key=value to raw event causing the summary events to use more space than the original raw event in many cases
- Standard summaries using stats, and sistats create very heavy tsidx indexes causing additional overhead
- Standard summaries add additional field content (metrics etc.) to the summary event causing additional bloat
- Multiple sourcetypes are many times stored across different indexes defined reduced and disparate retention settings

Bloat In Standard Summary Event

- Original Source Event:

```
Jun 22 12:07:04 1,2017/06/22 12:07:03,001606001116,THREAT,url,1,2017/06/22
12:07:03,192.168.0.2,184.106.31.170,0.0.0.0,0.0.0.0,rule1,crusher,,web-
browsing,vsys1,trust,untrust,ethernet1/2,ethernet1/1,blocked,2017/06/22
12:07:03,37148,1,52586,80,0,0,0x208000,tcp,alert,"modern-design.cn/rex/config.bin",(9999),not-
resolved,informational,client-to-server,0,0x0,192.168.0.0-192.168.255.255,United States,0,text/html
```

Reduction Using stats/sistats And Fields

- Summary search using stats/sistats command:

```
index=main sourcetype=pan:threat | fillnull value "nan" action dest_ip src_ip | eval orig_sourcetype=sourcetype | sistats
count by _time, host, orig_sourcetype, action, dest_ip, src_ip | fields - count | collect index=summary addtime=0
```

- Summary search using fields command:

```
index=main sourcetype=pan:threat | fillnull value "nan" action dest_ip src_ip | eval orig_sourcetype=sourcetype | fields
_time, host, orig_sourcetype, action, dest_ip, src_ip | collect index=summary addtime=0
```

Desired Results

- Output using sistats/table statement results:

```
search_name=sum_panthreat_sistats, search_now=1498155300.000, info_min_time=1498154880.000,
info_max_time=1498155180.000, info_search_time=1498155301.125, action=blocked, dest_ip=”194.106.31.170",
orig_host="127.0.0.1", orig_sourcetype="pan:threat", psrsvd_gc=1, psrsvd_v=1, src_ip="192.168.0.2"
```

- Output using fields statement results:

```
Jun 22 11:57:59 1,2017/06/22 11:57:59,001606001116,THREAT,url,1,2017/06/22
11:57:59,192.168.0.2,184.106.31.170,0.0.0.0,0.0.0.0,rule1,crusher,,web-
browsing,vsys1,trust,untrust,ethernet1/2,ethernet1/1,forwardAll,2017/06/22
11:57:59,53995,1,56068,80,0,0,0x208000,tcp,alert,"www.st-resources.net/config.bin",(9999),not-
resolved,informational,client-to-server,0,0x0,192.168.0.0-192.168.255.255,United States,0,text/html
```

- Desired output:

```
1498154279,127.0.0.1,pan:threat,blocked,192.168.0.2,184.106.31.170
```

Implementing The Desired Result

- Constructing a summary search using a CSV output:

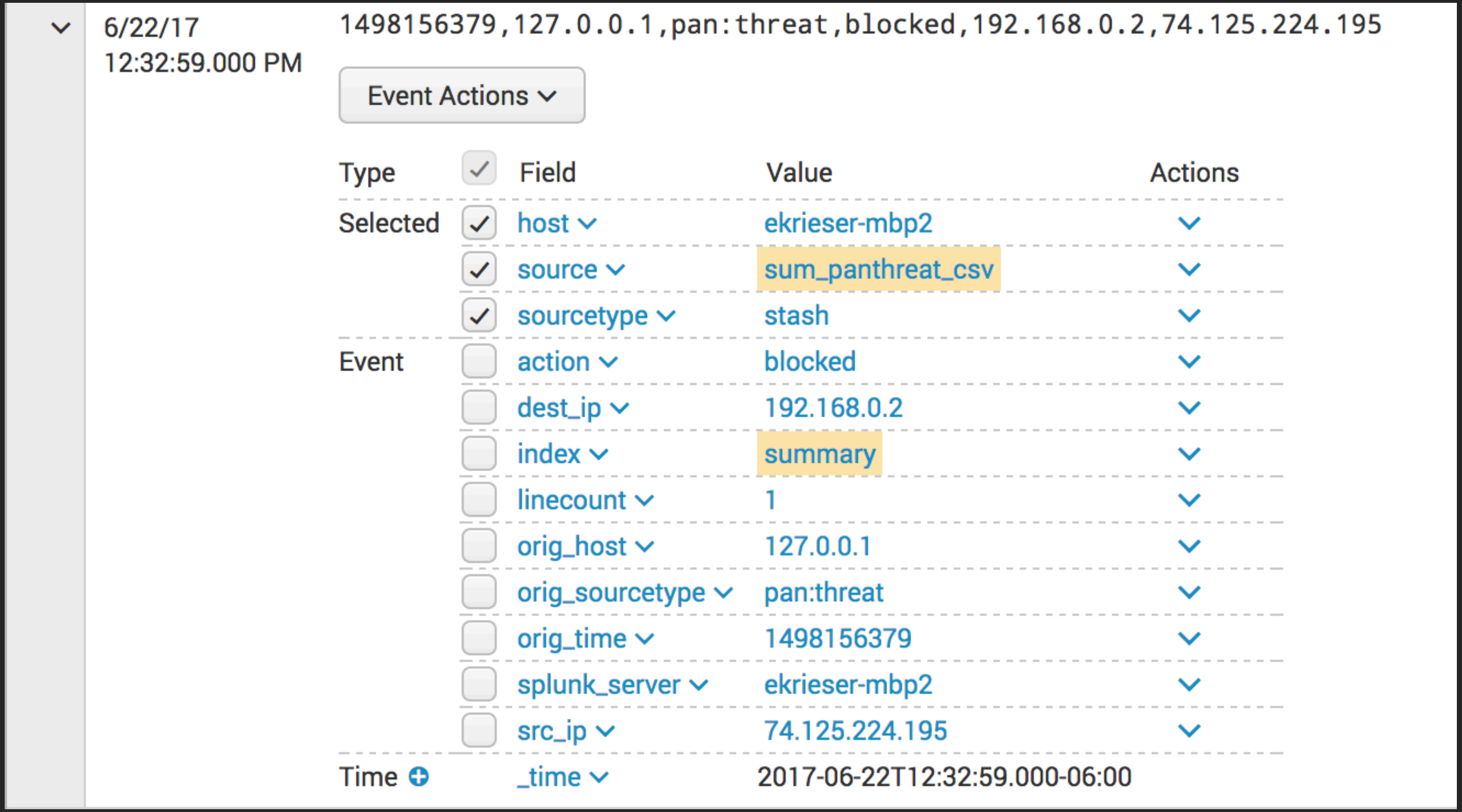
```
index=main sourcetype=pan:threat | fillnull value "nan" action dest_ip src_ip | eval orig_sourcetype=sourcetype | eval
_raw= _time.",".host.",".orig_sourcetype.",".action.",".dest_ip.",".src_ip | fields _time _raw | collect index=summary
addtime=0
```

- Build the summary Field Definition to Define Your Fields

```
props.conf

 [source::sum_panthreat_csv]
KV_MODE=none
REPORT-parse_sum_panthreat_csv = parse_sum_panthreat_csv

transforms.conf

 [parse_sum_panthreat_csv]
DELIMS = ","
FIELDS = orig_time, orig_host, orig_sourcetype, action, dest_ip, src_ip
```

Final Result Of Bloat Reduction

Extracted Fields View



6/22/17
12:32:59.000 PM

```
1498156379,127.0.0.1,pan:threat,blocked,192.168.0.2,74.125.224.195
```

Event Actions ∨

Type	☑	Field	Value	Actions
Selected	☑	host ∨	ekrieser-mbp2	∨
	☑	source ∨	sum_panthreat_csv	∨
	☑	sourcetype ∨	stash	∨
Event	□	action ∨	blocked	∨
	□	dest_ip ∨	192.168.0.2	∨
	□	index ∨	summary	∨
	□	linecount ∨	1	∨
	□	orig_host ∨	127.0.0.1	∨
	□	orig_sourcetype ∨	pan:threat	∨
	□	orig_time ∨	1498156379	∨
	□	splunk_server ∨	ekrieser-mbp2	∨
	□	src_ip ∨	74.125.224.195	∨
Time ⊕		_time ∨	2017-06-22T12:32:59.000-06:00	

Review Of Key Points To The Solution

- Use summary searching to to consolidate critical data stored across disparate sourcetypes and indexes
- Construct a single field called _raw that contains the desired summary content in some type of a CSV or character delimited form
- Pass _time, and _raw to the summary index
- Define the field format of the summary event using DELIM, and FIELDS transforms definitions

Things To Take Into Consideration

- Parsing fields with equal (=) signs can be a pain
- Take particular care in defining delimiter to use (CSV may not be the best in your use case)

Average Savings Stats

72% Reduction In Storage Volume

26% Faster Searches

Over 3x Longer Retention at Zero Additional Cost

Q&A

Tomasz Bania | Incident Response Lead, Dolby

Eric Krieser | Splunk Professional Services

Thank You

Don't forget to rate this session in the .conf2017 mobile app

splunk> .conf2017