Splunk Deletion Detector

When one search won’t do

Ellie Baum

Lead Software Engineer | Salesforce
During the course of this presentation, we may make forward-looking statements regarding future events or plans 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 may differ materially. 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, it may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements made herein.

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 functionalities described or to include any such feature or functionality in a future release.
Ellie Baum

Lead Software Engineer (PO and recovering SM) | Salesforce
Marketing Cloud CI Team
Everything from Developer desktops to CI builds

• Developer desktop productivity tooling
• Local and CI builds
• All products get telemetry built in

Without fail, adding data always finds bugs.
More data = More feedback. Less bugs.
Problem:
My company was deleting data out of my team’s Splunk index and nobody knew why
Be Agile. Get feedback. Iterate.
Data = Feedback

- Contains all our Continuous Integration (CI) data.
- Be a good scientist and hold some variables constant.

Within days of first having this data, we found a bug in our CI pipeline that cut 25% of time off of local dev builds
Losing This Data Is Not an Option.
Why Can’t I Use a “Regular” Search?

This alert only mattered in context.

Example: If the number of days in the index was 30 today, I only care if the number of days in the index yesterday was 40!
End Goal

Slack Notification

Visualization
...armed with “the data”, I get to bug MC Splunk team...
The Deletion Detector.
Big Picture.

Three Components:
1. Report
2. Lookup Table
3. Alert
1. Report.

This report will only produce one row of data when run. It summarizes how much data is in the index.
SPL for REPORT

```
index=BASE SEARCH
| stats min(_time) as EarliestTime max(_time) as LatestTime
count as NumberOfEvents
...
```
SPL for REPORT

index=BASE SEARCH
| stats min(_time) as EarliestTime max(_time) as LatestTime count as NumberOfEvents
| eval EarliestTime_Human = strftime(EarliestTime,"%Y-%m-%d %H:%M:%S.%Q"), LatestTime_Human = strftime(LatestTime,"%Y-%m-%d %H:%M:%S.%Q")
...
SPL for REPORT

```
index=BASE SEARCH
| stats min(_time) as EarliestTime max(_time) as LatestTime
count as NumberOfEvents
| eval EarliestTime_Human = strftime(EarliestTime,"%Y-%m-%d
%H:%M:%S.%Q"), LatestTime_Human = strftime(LatestTime,"%Y-
%m-%d %H:%M:%S.%Q")
| eval NumberOfDaysInIndex = round((LatestTime -
EarliestTime) / (60 * 60 * 24), 1)
```
SPL for REPORT

index=BASE SEARCH
| stats min(_time) as EarliestTime max(_time) as LatestTime count as NumberOfEvents
| eval EarliestTime_Human = strftime(EarliestTime,"%Y-%m-%d %H:%M:%S.%Q"), LatestTime_Human = strftime(LatestTime,"%Y-%m-%d %H:%M:%S.%Q")
| eval NumberOfDaysInIndex = round((LatestTime - EarliestTime) / (60 * 60 * 24), 1)
| eval LastRunTime = now(), LastRunTime_Human = strftime(LastRunTime,"%Y-%m-%d %H:%M:%S.%Q")
...

splunk> .conf20
SPL for REPORT

```splunk
index=BASE SEARCH
| stats min(_time) as EarliestTime max(_time) as LatestTime count as NumberOfEvents
| eval EarliestTime_Human = strftime(EarliestTime,"%Y-%m-%d %H:%M:%S.%Q"), LatestTime_Human = strftime(LatestTime,"%Y-%m-%d %H:%M:%S.%Q")
| eval NumberOfDaysInIndex = round((LatestTime - EarliestTime) / (60 * 60 * 24), 1)
| eval LastRunTime = now(), LastRunTime_Human = strftime(LastRunTime,"%Y-%m-%d %H:%M:%S.%Q")
| table EarliestTime_Human LatestTime_Human LastRunTime_Human NumberOfEvents NumberOfDaysInIndex EarliestTime LatestTime LastRunTime
```
2. Lookup Table.

The Lookup table will store the results of the Report over time.

This is how we get that sweet, sweet context.
SPL for REPORT to LOOKUP table

| inputlookup indexHistory.csv |
| where LastRunTime > relative_time(now(), "-180d") |
| append [ ] |
| outputlookup indexHistory.csv |
SPL for REPORT to LOOKUP table

| inputlookup indexHistory.csv |
| where LastRunTime > relative_time(now(), "-180d") |
| append |

  search index=BASE SEARCH |
| stats min(_time) as EarliestTime max(_time) as LatestTime |
| eval EarliestTime_Human = strftime(EarliestTime, "%Y-%m-%d %H:%M:%S.%Q"), LatestTime_Human = strftime(LatestTime, "%Y-%m-%d %H:%M:%S.%Q") |
| eval NumberOfDaysInIndex = round((LatestTime - EarliestTime) / (60 * 60 * 24), 1) |
| eval LastRunTime = now(), LastRunTime_Human = strftime(LastRunTime, "%Y-%m-%d %H:%M:%S.%Q") |
| table EarliestTime_Human LatestTime_Human LastRunTime_Human NumberOfDaysInIndex EarliestTime LatestTime LastRunTime |

| outputlookup indexHistory.csv |
### Lookup Table Results

This will be the basis for the alert we create

<table>
<thead>
<tr>
<th>EarliestTime</th>
<th>EarliestTime_Human</th>
<th>LastRunTime</th>
<th>LastRunTime_Human</th>
<th>LatestTime</th>
<th>LatestTime_Human</th>
<th>NumberOfDaysIndex</th>
<th>NumberOfEvents</th>
</tr>
</thead>
<tbody>
<tr>
<td>159454276</td>
<td>2020-07-13 09:51:16.000</td>
<td>1596432600</td>
<td>2020-08-03 01:30:00.000</td>
<td>1596432599</td>
<td>2020-08-03 01:29:59.000</td>
<td>20.7</td>
<td>41066</td>
</tr>
<tr>
<td>159454276</td>
<td>2020-07-13 09:51:16.000</td>
<td>1596448800</td>
<td>2020-08-03 06:00:00.000</td>
<td>1596448657</td>
<td>2020-08-03 05:57:37.000</td>
<td>20.8</td>
<td>61386</td>
</tr>
<tr>
<td>159454276</td>
<td>2020-07-13 09:51:16.000</td>
<td>1596449700</td>
<td>2020-08-03 06:15:00.000</td>
<td>1596449652</td>
<td>2020-08-03 06:14:12.000</td>
<td>20.8</td>
<td>61327</td>
</tr>
<tr>
<td>159454276</td>
<td>2020-07-13 09:51:16.000</td>
<td>1596450600</td>
<td>2020-08-03 06:30:00.000</td>
<td>1596450599</td>
<td>2020-08-03 06:29:59.000</td>
<td>20.9</td>
<td>61392</td>
</tr>
<tr>
<td>159454276</td>
<td>2020-07-13 09:51:16.000</td>
<td>1596451500</td>
<td>2020-08-03 06:45:00.000</td>
<td>1596451238</td>
<td>2020-08-03 06:40:38.000</td>
<td>20.9</td>
<td>61399</td>
</tr>
<tr>
<td>159454276</td>
<td>2020-07-13 09:51:16.000</td>
<td>1596452400</td>
<td>2020-08-03 07:00:00.000</td>
<td>1596452257</td>
<td>2020-08-03 06:57:37.000</td>
<td>20.9</td>
<td>61405</td>
</tr>
<tr>
<td>159454276</td>
<td>2020-07-13 09:51:16.000</td>
<td>1596453300</td>
<td>2020-08-03 07:15:00.000</td>
<td>1596453247</td>
<td>2020-08-03 07:14:07.000</td>
<td>20.9</td>
<td>61438</td>
</tr>
</tbody>
</table>
3. Alert.

The Alert is how we get notified.

Based off the data in the lookup table, did a deletion just occur?
SPL for Alert

| inputlookup indexHistory.csv
| sort 3 -LastRunTime
| sort 0 LastRunTime

...
SPL for Alert

| inputlookup indexHistory.csv
| sort 3 -LastRunTime
| sort 0 LastRunTime
| streamstats current=f last(NumberOfDaysInIndex) as prevDayCount
...
SPL for Alert

| inputlookup indexHistory.csv |
| sort 3 -LastRunTime |
| sort 0 LastRunTime |
| streamstats current=f last(NumberOfDaysInIndex) as prevDayCount |
| where prevDayCount > NumberOfDaysInIndex AND NumberOfDaysInIndex < 180 |

...
SPL for Alert

| inputlookup indexHistory.csv
| sort 3 -LastRunTime
| sort 0 LastRunTime
| streamstats current=f last(NumberOfDaysInIndex) as prevDayCount
| where prevDayCount > NumberOfDaysInIndex AND NumberOfDaysInIndex < 180
| eval DaysRemoved = prevDayCount - NumberOfDaysInIndex
...
SPL for Alert

| inputlookup indexHistory.csv |
| sort 3 -LastRunTime |
| sort 0 LastRunTime |
| streamstats current=f last(NumberOfDaysInIndex) as prevDayCount |
| where prevDayCount > NumberOfDaysInIndex AND NumberOfDaysInIndex < 180 |
| eval DaysRemoved = prevDayCount - NumberOfDaysInIndex |
| rename DaysRemoved AS "Number of Days just removed from our index", NumberOfDaysInIndex AS "Number of Days currently in our index" |
| table "Number of Days just removed from our index" "Number of Days currently in our index" |
The Deletion Detector: Putting it all together

1. REPORT
   That does
   Summaries data on a search

   Every 15
   min, append To...

2. LOOKUP TABLE

<table>
<thead>
<tr>
<th>NUMBER_OF_DAYS</th>
<th>LAST_RUN_TIME</th>
<th>EARLIEST_TIME</th>
<th>LATEST_TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>01/01/2020 1:00 PM</td>
<td>12/01/2019 12:00 PM</td>
<td>01/01/2020 12:00 PM</td>
</tr>
<tr>
<td>30.1</td>
<td>01/01/2020 1:15 PM</td>
<td>12/01/2020 12:00 PM</td>
<td>01/01/2020 12:15 PM</td>
</tr>
</tbody>
</table>

3. ALERT
   Is the last
   NUMBER_OF_DAYS
   value less than the 2nd
to last value?

   15 Min
   (Chron
   Expression)

   FALSE
   Cool. Do
   nothing.

   TRUE
   TELL A HUMAN!
Slack Notification

Data loss is sad! :( 

The alert condition for 'McciDataDeletionDetector' was triggered.

Check out the index's history using this report:

Number of Days just removed from our index | Number of Days currently in our index
---|---
17.4 | 27.9

If you believe you've received this email in error, please see your Splunk administrator.

`splunk > the engine for machine data`
Turns Out Our Splunk Infrastructure Couldn’t Handle the Data Load.

Moving to a new environment we got our retention back up to 180 days.
The Deletion Detector…
or Is It More Than That?

Alerts that require context!
Lessons Learned

• When I have an alert that requires **context**, I have a formula for solving the problem

• It’s good solution for short term problems but consider when to use summary indexes
Thank You

Please provide feedback via the SESSION SURVEY