What's in my Data?

Field Analysis for the Advanced Engineer

Ryan Wood

Who am I?

- Daily Splunk User for 9 years.
- Currently working as a Splunk PS delivery engineer at Guidepoint Security.
- Background crossing healthcare, banking, financial industries before delivering PS.
- Love the front-end more than the architecture side; it's where I began.

#Content

How is this relevant to you?

If you've ever asked yourself...

"Which app did ['field'] come from? How is it defined?"

"What fields are available in my data, and what values do they contain?"

"How could I efficiently compare field contents of two datasets?"

"Do I want to learn some useful SPL?"

Fields Derivations

dataSource	endpoint_title	field_name	stanza	attribute	acl_app	value
props-calcfields	aix_secure : EVAL-action	action	[aix_secure]	EVAL-action	Splunk_TA_nix	if(app="su" AND isnull(action),"success",action)
props-calcfields	audittrail : EVAL-app	app	[audittrail]	EVAL-app	Splunk_SA_CIM	if(match(_raw,"action\=login\sattempt"),"splunk",app)
props-extractions	aix_secure : REPORT-dest_for_aix_secure	dest_for_aix_secure	[aix_secure]	REPORT-dest_for_aix_secure	Splunk_TA_nix	host_as_dest
props-extractions	aix_secure : REPORT-signature_for_aix_secure_timesync	signature_for_aix_secure_timesync	[aix_secure]	REPORT-signature_for_aix_secure_timesync	Splunk_TA_nix	signature_for_nix_timesync
props-fieldaliases	audittrail : FIELDALIAS-dest_for_splunk_access	dest	[audittrail]	FIELDALIAS-dest_for_splunk_access	Splunk_SA_CIM	host as dest
props-fieldaliases	audittrail : FIELDALIAS-file_acl_for_splunk_filesystem_change	file_acl	[audittrail]	FIELDALIAS-file_acl_for_splunk_filesystem_change	Splunk_SA_CIM	mode as file_acl
props-lookups (automatic lookups)	aix_secure : LOOKUP-action_for_osx_secure	action	[aix_secure]	LOOKUP-action_for_osx_secure	Splunk_TA_nix	nix_action_lookup vendor_action OUTPUTNEW action
props-lookups (automatic lookups)	audittrail : LOOKUP-dmc_add_instance_info	machine search_group	[audittrail]	LOOKUP-dmc_add_instance_info	splunk_monitoring_console	dmc_assets host OUTPUTNEW machine search_group
transforms-extractions	Account_Domain_as_dest_nt_domain		[Account_Domain_as_dest_nt_domain]		Splunk_TA_windows	(?:(?:[^\n]+)\n)?(.+)
transforms-extractions	Account_Domain_as_src_nt_domain		[Account_Domain_as_src_nt_domain]		Splunk_TA_windows	(?!^~$)([^\n]+)\n

Formatted Fields Content Summary

(no groupby)

field	Count of Events w/ Field	Perc of Total Events w/ Field	Distinct Values	Top values with count of each
dest_is_expected dest_requires_av dest_should_timesync dest_should_update	1428	100.0%	1 (Exact)	"false" :1428
index	1428	100.0%	3 (Exact)	"osnix" :1355 "oswin" :46 "oswinsec" :27
source	1428	100.0%	6 (Estimate)	"/var/log/messages" :1251 "/var/log/dnf.log" :92 "XmlWinEventLog:System" :30 "WinEventLog:Security" :27 "XmlWinEventLog:Application" :16
sourcetype	1428	100.0%	3 (Exact)	"syslog" :1355 "XmlWinEventLog" :46 "WinEventLog" :27
process	1350	94.5%	25 (Estimate)	"systemd" :732 "dbus-daemon" :480 "syslog-ng" :24 "dnf" :14 "reviving" :10

Formatted Fields Content Summary

(BY *wildcard index filter*)

index	field	Count of Events w/ Field	Perc of Total Events w/ Field	Distinct Values	Top values with count of each
osnix*	source	1355	100.0%	3 (Exact)	"/var/log/messages" :1251 "/var/log/dnf.log" :92 "/var/log/cron" :12
osnix*	sourcetype	1355	100.0%	1 (Exact)	"syslog" :1355
osnix*	process	1339	98.8%	25 (Estimate)	"systemd" :732 "dbus-daemon" :480 "syslog-ng" :24 "dnf" :14 "reviving" :10
oswin*	Error_Code	73	100.0%	1 (Exact)	"-" :73
oswin*	source	73	100.0%	3 (Exact)	"XmlWinEventLog:System" :30 "WinEventLog:Security" :27 "XmlWinEventLog:Application" :16
oswin*	tag	73	100.0%	18 (Estimate)	"endpoint" :73 "filesystem" :73 "os" :73 "process" :73 "report" :73

BSIDES splunk>

Formatted Fields Content Summary

(BY *index, sourcetype, source*)

index	sourcetype	source	field	count	diffPerc	distinctValues	values
osnix	syslog	/var/log/messages	process	1251	100.0%	5 (Estimate)	"systemd" :732 "dbus-daemon" :480 "syslog-ng" :24 "dnf" :14 "chronyd" :1
osnix	syslog	/var/log/dnf.log	process	76	82.6%	18 (Estimate)	"reviving" :10 "timer" :6 "Command" :4 "Installroot" :4 "Releasever" :4
oswin	XmlWinEventLog	XmlWinEventLog:System	ThreadID	30	100.0%	7 (Estimate)	"'1800'" :9 "'2492'" :6 "'4552'" :6 "'576'" :5 "'3344'" :2
oswin	XmlWinEventLog	XmlWinEventLog:Application	ProcessID ThreadID	16	100.0%	1 (Exact)	"'0'" :16
oswinsec	WinEventLog	WinEventLog:Security	Security_ID	27	100.0%	2 (Exact)	"S-1-5-18" :29 "S-1-0-0" :7

Disclaimer

“There’s a thousand ways you can build a SPL query.”

- The SPL queries made available through this presentation are not necessarily “best practice.”
- These are written with a goal of readability, version applicability (7.3+), and universal scalability (1gb – 100tb)
  - This can result in verbose syntax/odd ordering of commands, or “non optimized” segments of SPL.
- The views, information, or opinions expressed during this presentation are my own and do not necessarily represent those of Guidepoint Security.
- My Development Process of SPL:

  Understanding, Adoption, Innovation, Sharing.

Before we dive in deep…

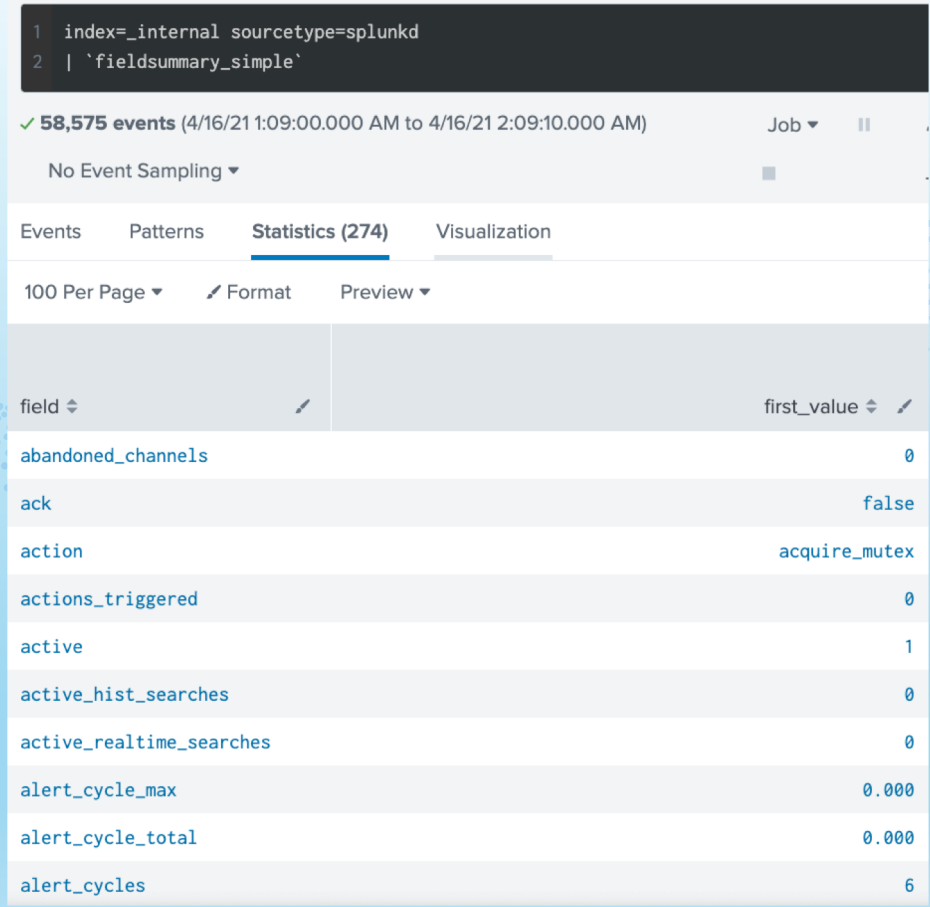
A simple data summarization!

```
| stats first(*) AS *
| transpose 0 column_name="field"
| rename "row 1" as first_value
```

Provides a very direct “show me the fields” view that can save a lot of time and be run on the fly when developing.

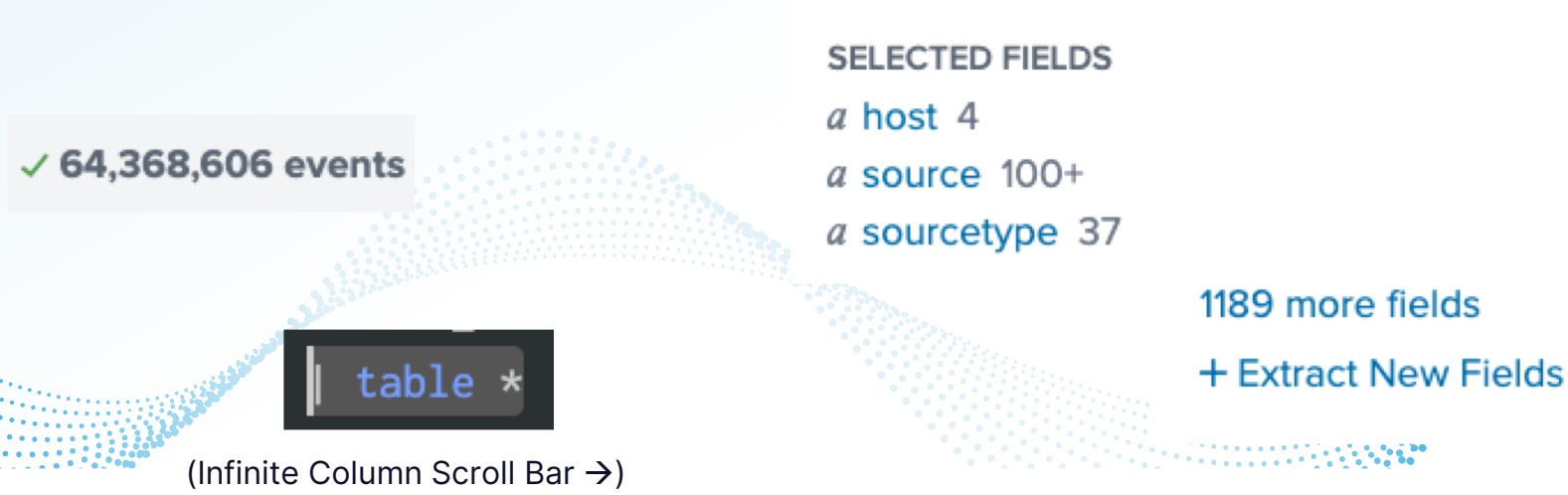
Extremely macro-able.

NOTE: This will return the **first** value found in any event, not just the first event’s values. Returns only the first value of multi-value fields.



Framing the Problem

- We've got data, but the basic methods of breaking it apart are too labor-intensive, and a basic summarization isn't enough.



| fieldsummary

Specify the list of fields (or don't to run it on all fields) to return a set of characteristics about the dataset available *at time of command call.*

Syntax

fieldsummary [maxvals=<unsigned_int>] [<wc-field-list>]

Optional arguments

maxvals

Syntax: maxvals=<unsigned_int>

Description: Specifies the maximum distinct values to return for each field. Cannot be negative. Set `maxvals = 0` to return all available distinct values for each field.

Default: 100

wc-field-list

Syntax: <field> ...

Description: A single field name or a space-delimited list of field names. You can use the asterisk (*) as a wildcard to specify a list of fields with similar names. For example, if you want to specify all fields that start with "value", you can use a wildcard such as `value*`.

https://docs.splunk.com/Documentation/Splunk/latest/SearchReference/Fieldsummary

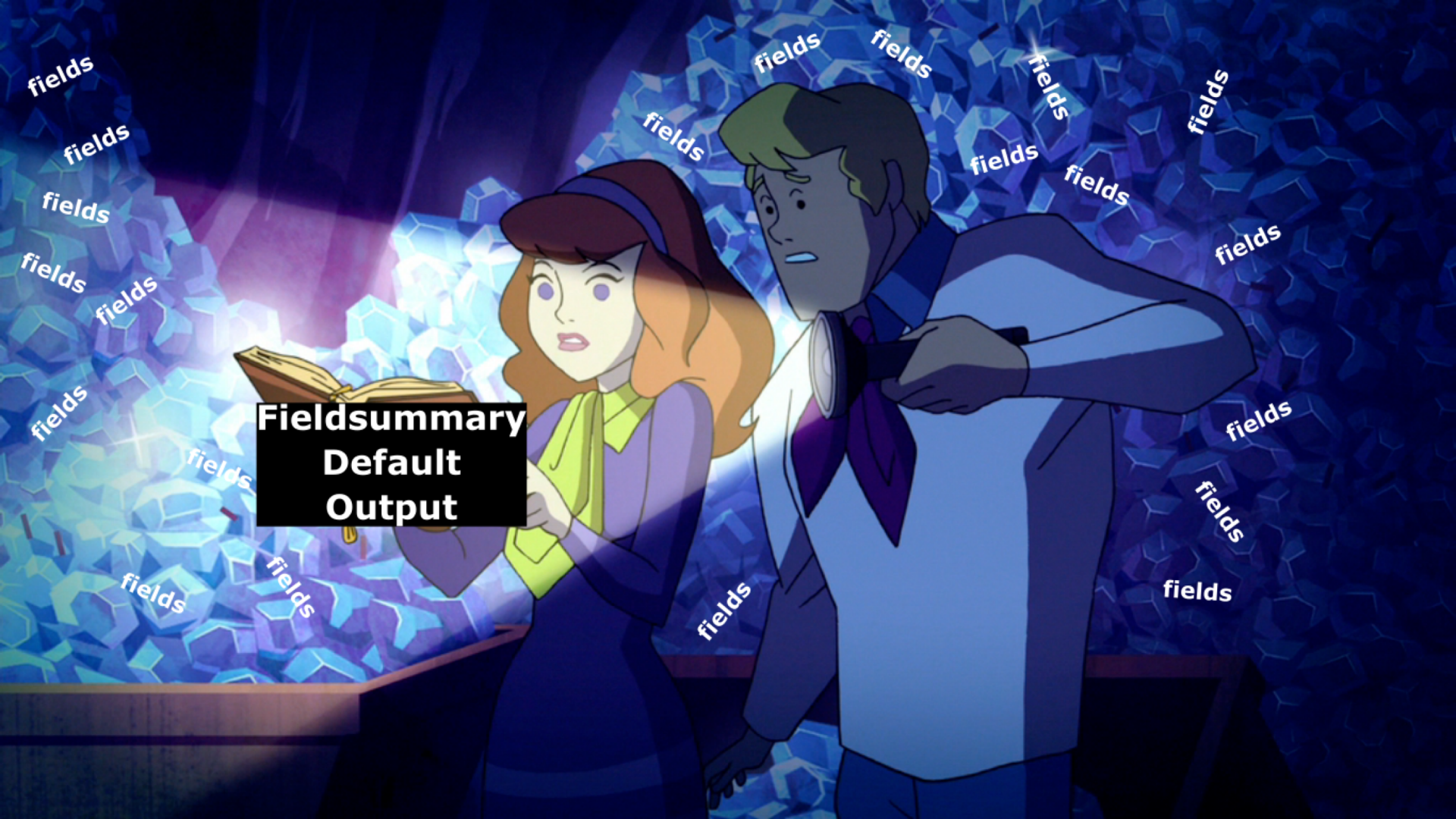
BSIDES splunk>

Default `fieldsummary` Output

field	count	distinct_count	is_exact	max	mean	min	numeric_count	stdev	values
abandoned_channels	1863	1	1	0	0	0	1863	0	[{"value":"0","count":1863}]
ack	1398	1	1				0		[{"value":"false","count":1398}]
action	73	1	1				0		[{"value":"acquire_mutex","count":73}]
actions_triggered	466	2	1	1	0.0193	0	466	0.1378	[{"value":"0","count":457},{"value":"1","count":9}]
active	1152	6	1	5	0.3793	0	1152	0.8559	[{"value":"0","count":912},{"value":"1","count":104},{"value":"2","count":97},{"value":"3","count":23},{"value":"4","count":10},{"value":"5","count":6}]
active_hist_searches	2385	21	0	25	1.1287	0	2385	2.2016	[{"value":"0","count":1492},{"value":"1","count":320},{"value":"2","count":162},{"value":"3","count":140},{"value":"4","count":97},{"value":"5","count":73},{"value":"6","count":41},{"value":"7","count":14}]
active_realtime_searches	2385	1	1	0	0	0	2385	0	[{"value":"0","count":2385}]
alert_cycle_max	466	3	1	0.002	0.0000	0.000	466	0.0001	[{"value":"0.000","count":464},{"value":"0.001","count":1},{"value":"0.002","count":1}]
alert_cycle_total	466	3	1	0.002	0.0000	0.000	466	0.0001	[{"value":"0.000","count":464},{"value":"0.001","count":1},{"value":"0.002","count":1}]
alert_cycles	466	17	0	23	5.1910	3	466	2.8864	[{"value":"4","count":134},{"value":"3","count":113},{"value":"5","count":98},{"value":"6","count":58},{"value":"11","count":12},{"value":"7","count":11},{"value":"12","count":9},{"value":"10","count":7}]

https://docs.splunk.com/Documentation/Splunk/latest/SearchReference/Fieldsummary

BSIDES splunk>


fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
fields
Fieldsummary
Default
Output

```
index=_internal sourcetype=splunkd
| fields - date_*, host, linecount, punct, splunk_server*, timestartpos, timeendpos, timestamp
| fieldsummary maxvals=5
| eval distinctValues = case(
    (is_exact == 1), (distinct_count . " (Exact)"),
    (is_exact == 0),(distinct_count . " (Estimate)")
    )
| eventstats max(count) AS eventCount
| eval diffPerc = (round(((count / eventCount) * 100),1) . "%")
| where (tonumber(rtrim(diffPerc,"%")) > 10)
| fields - distinct_count, eventCount, is_exact, mean, max, min, stdev
| spath input=values path="{}.value" output="value_strings"
| spath input=values path="{}.count" output="value_counts"
| eval value_strings = split("\"" . mvjoin(value_strings, "\":::\"") . "\"", ":::")
| eval values = mvzip(value_strings, value_counts, ": ")
| fields - value_counts, value_strings
| mvexpand values
| rex field=values "^(?<fieldValueString>\".+\"):\s(?<fieldValueCount>\d+)"
| eventstats max(eval(len(fieldValueString))) AS max_fieldValue_len BY field
| eval whitespaceInsertAmount = ((max_fieldValue_len + 4) - len(fieldValueString))
| eval values = fieldValueString . substr("
                                  ",1,whitespaceInsertAmount) . ":" . fieldValueCount
| fields - whitespaceInsertAmount, max_fieldValue_len, fieldValueString, fieldValueCount
| stats list(values) AS values, values(values) AS values_ThisMustBeSeparateToRetainOrdering, values(*) AS * BY field
| fields - values_ThisMustBeSeparateToRetainOrdering
| eval valuesHash = md5(mvjoin(values, ":::"))
| eval checkSum = md5( count . ":::" . diffPerc . ":::" . numeric_count . ":::" . distinctValues . ":::" . valuesHash)
| eventstats values(field) AS field BY checkSum
| sort 0 - checkSum
| streamstats count AS duplicateCount BY checkSum
| where duplicateCount < 2
| sort 0 - count
| table field, count, diffPerc, numeric_count, distinctValues, values
| rename count AS "Count of Events w/ Field", diffPerc AS "Perc of Total Events w/ Field", distinctValues AS "Distinct Values", numeric_count AS "Numeric Count",
    values AS "Top values with count of each"
```

Re: These Examples

Reference utility queries posted at (github) include comments segmenting them into purpose-driven components for easier understanding.

Each individual component of each usecase is marked with a “START/END” comment:

```
| eval comment = if(1==1, null(), "
END initial fieldsummary command execution via map.


START fieldsummary enrichment and base calculations/filtering.
")
| eval distinctValues=case(
    (is_exact == 1), (distinct_count . " (Exact)"),
    (is_exact == 0),(distinct_count . " (Estimate)")
    )
| eventstats max(count) as eventCount BY index, sourcetype, source
| eval diffPerc=(round(((count / eventCount) * 100),1) . "%")
| eval comment = if(1==1, null(), "
    Note: Modify | where command below based on desired minimum % of eve
| where (tonumber(rtrim(diffPerc,"%")) > 10)
| fields - distinct_count, eventCount, is_exact, mean, max, min, stdev
| eval comment = if(1==1, null(), "
END fieldsummary enrichment and base calculations/filtering.
```

The two notable lines for customization:

| where (tonumber(rtrim(diffPerc,"%")) > 10)

- Controls minimum coverage % for that specific field to display in final results

```
| where (tonumber(rtrim(diffPerc,"%")) > 10)
```

| fieldsummary maxvals=(number)

- maxvals determines how many unique values are returned in the values field.
- Caution, as more values needs higher compute usage.

```
| fieldsummary maxvals=5
```

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Fetch data, run fieldsummary

```
index=_internal sourcetype=splunkd
| fields - date_*, host, linecount, punct, splunk_server*, timestartpos, timeendpos, timestamp
| fieldsummary maxvals=5
| eval distinctValues = case((is_exact == 1),(distinct_count . " (Exact)"),(is_exact == 0),(distinct_count . " (Estimate)"))
| eventstats max(count) AS eventCount
| eval diffPerc = (round(((count / eventCount) * 100),1) . "%")
| where (tonumber(rtrim(diffPerc,"%")) > 10)
| fields - distinct_count, eventCount, is_exact, mean, max, min, stdev
| spath input=values path="{}.value" output="value_strings"
| eval value_strings = split("\"" . mvjoin(value_strings, "\":::\"") . "\"", ":::")
| spath input=values path="{}.count" output="value_counts"
| eval values = mvzip(value_strings, value_counts, ": ")
| fields - value_counts, value_strings
| mvexpand values
| rex field=values "^(?<fieldValueString>\".+\"):\s(?<fieldValueCount>\d+)"
| eventstats max(eval(len(fieldValueString))) AS max_fieldValue_len BY field
| eval whitespaceInsertAmount = ((max_fieldValue_len + 4) - len(fieldValueString))
| eval values = fieldValueString . substr("
                            ",1,whitespaceInsertAmount) . ":" . fieldValueCount
| fields - whitespaceInsertAmount, max_fieldValue_len, fieldValueString, fieldValueCount
| stats list(values) AS values, values(values) AS values_ThisMustBeSeparateToRetainOrdering, values(*) AS * BY field
| fields - values_ThisMustBeSeparateToRetainOrdering
| eval valuesHash = md5(mvjoin(values, ":::"))
| eval checkSum = md5( count . ":::" . diffPerc . ":::" . numeric_count . ":::" . distinctValues . ":::" . valuesHash)
| eventstats values(field) AS field BY checkSum
| sort 0 - checkSum
| streamstats count AS duplicateCount BY checkSum
| where duplicateCount < 2
| sort 0 - count
| table field, count, diffPerc, numeric_count, distinctValues, values
| rename count AS "Count of Events w/ Field", diffPerc AS "Perc of Total Events w/ Field", distinctValues AS "Distinct Values", numeric_count AS "Numeric Count", values AS
    "Top values with count of each"
```

```
index=_internal sourcetype=splunkd
| fields - date_*, host, linecount, punct, splunk_server*, timestartpos, timeendpos, timestamp
| fieldsummary maxvals=5
| eval distinctValues = case((is_exact == 1),(distinct_count . " (Exact)"),(is_exact == 0),(distinct_count . " (Estimate)"))
| eventstats max(count) AS eventCount
| eval diffPerc = (round(((count / eventCount) * 100),1) . "%")
| where (tonumber(rtrim(diffPerc,"%")) > 10)
| fields - distinct_count, eventCount, is_exact, mean, max, min, stdev
| spath input=values path="{}.value" output="value_strings"
| eval value_strings = split("\"" . mvjoin(value_strings, "\":::\"") . "\"", ":::")
| spath input=values path="{}.count" output="value_counts"
| eval values = mvzip(value_strings, value_counts, ": ")
| fields - value_counts, value_strings
| mvexpand values
| rex field=values "^(?<fieldValueString>\".+\"):\s(?<fieldValueCount>\d+)"
| eventstats max(eval(len(fieldValueString))) AS max_fieldValue_len BY field
| eval whitespaceInsertAmount = ((max_fieldValue_len + 4) - len(fieldValueString))
| eval values = fieldValueString . substr("
                            ",1,whitespaceInsertAmount) . ":" . fieldValueCount
| fields - whitespaceInsertAmount, max_fieldValue_len, fieldValueString, fieldValueCount
| stats list(values) AS values, values(values) AS values_ThisMustBeSeparateToRetainOrdering, values(*) AS * BY field
| fields - values_ThisMustBeSeparateToRetainOrdering
| eval valuesHash = md5(mvjoin(values, ":::"))
| eval checkSum = md5( count . ":::" . diffPerc . ":::" . numeric_count . ":::" . distinctValues . ":::" . valuesHash)
| eventstats values(field) AS field BY checkSum
| sort 0 - checkSum
| streamstats count AS duplicateCount BY checkSum
| where duplicateCount < 2
| sort 0 - count
| table field, count, diffPerc, numeric_count, distinctValues, values
| rename count AS "Count of Events w/ Field", diffPerc AS "Perc of Total Events w/ Field", distinctValues AS "Distinct Values", numeric_count AS "Numeric Count", values AS
    "Top values with count of each"
```

Fetch data, run fieldsummary

Fieldsummary Output
Enrichment & Filtering

```
index=_internal sourcetype=splunkd
| fields - date_*, host, linecount, punct, splunk_server*, timestartpos, timeendpos, timestamp
| fieldsummary maxvals=5
| eval distinctValues = case((is_exact == 1),(distinct_count . " (Exact)"),(is_exact == 0),(distinct_count . " (Estimate)"))
| eventstats max(count) AS eventCount
| eval diffPerc = (round(((count / eventCount) * 100),1) . "%")
| where (tonumber(rtrim(diffPerc,"%")) > 10)
| fields - distinct_count, eventCount, is_exact, mean, max, min, stdev
| spath input=values path="{}.value" output="value_strings"
| eval value_strings = split("\"" . mvjoin(value_strings, "\":::\"") . "\"", ":::")
| spath input=values path="{}.count" output="value_counts"
| eval values = mvzip(value_strings, value_counts, ": ")
| fields - value_counts, value_strings
| mvexpand values
| rex field=values "^(?<fieldValueString>\".+\"):\s(?<fieldValueCount>\d+)"
| eventstats max(eval(len(fieldValueString))) AS max_fieldValue_len BY field
| eval whitespaceInsertAmount = ((max_fieldValue_len + 4) - len(fieldValueString))
| eval values = fieldValueString . substr("
                              ",1,whitespaceInsertAmount) . ":" . fieldValueCount
| fields - whitespaceInsertAmount, max_fieldValue_len, fieldValueString, fieldValueCount
| stats list(values) AS values, values(values) AS values_ThisMustBeSeparateToRetainOrdering, values(*) AS * BY field
| fields - values_ThisMustBeSeparateToRetainOrdering
| eval valuesHash = md5(mvjoin(values, ":::"))
| eval checkSum = md5( count . ":::" . diffPerc . ":::" . numeric_count . ":::" . distinctValues . ":::" . valuesHash)
| eventstats values(field) AS field BY checkSum
| sort 0 - checkSum
| streamstats count AS duplicateCount BY checkSum
| where duplicateCount < 2
| sort 0 - count
| table field, count, diffPerc, numeric_count, distinctValues, values
| rename count AS "Count of Events w/ Field", diffPerc AS "Perc of Total Events w/ Field", distinctValues AS "Distinct Values", numeric_count AS "Numeric Count", values AS
    "Top values with count of each"
```

Fetch data, run fieldsummary

Fieldsummary Output
Enrichment & Filtering

‘values’ JSON Manipulation

```
index=_internal sourcetype=splunkd
| fields - date_*, host, linecount, punct, splunk_server*, timestartpos, timeendpos, timestamp
| fieldsummary maxvals=5
| eval distinctValues = case((is_exact == 1),(distinct_count . " (Exact)"),(is_exact == 0),(distinct_count . " (Estimate)"))
| eventstats max(count) AS eventCount
| eval diffPerc = (round(((count / eventCount) * 100),1) . "%")
| where (tonumber(rtrim(diffPerc,"%")) > 10)
| fields - distinct_count, eventCount, is_exact, mean, max, min, stdev
| spath input=values path="{}.value" output="value_strings"
| eval value_strings = split("\"" . mvjoin(value_strings, "\":::\"") . "\"", ":::")
| spath input=values path="{}.count" output="value_counts"
| eval values = mvzip(value_strings, value_counts, ": ")
| fields - value_counts, value_strings
| mvexpand values
| rex field=values "^(?<fieldValueString>\".+\"):\s(?<fieldValueCount>\d+)"
| eventstats max(eval(len(fieldValueString))) AS max_fieldValue_len BY field
| eval whitespaceInsertAmount = ((max_fieldValue_len + 4) - len(fieldValueString))
| eval values = fieldValueString . substr("
                            ",1,whitespaceInsertAmount) . ":" . fieldValueCount
| fields - whitespaceInsertAmount, max_fieldValue_len, fieldValueString, fieldValueCount
| stats list(values) AS values, values(values) AS values_ThisMustBeSeparateToRetainOrdering, values(*) AS * BY field
| fields - values_ThisMustBeSeparateToRetainOrdering
| eval valuesHash = md5(mvjoin(values, ":::"))
| eval checkSum = md5( count . ":::" . diffPerc . ":::" . numeric_count . ":::" . distinctValues . ":::" . valuesHash)
| eventstats values(field) AS field BY checkSum
| sort 0 - checkSum
| streamstats count AS duplicateCount BY checkSum
| where duplicateCount < 2
| sort 0 - count
| table field, count, diffPerc, numeric_count, distinctValues, values
| rename count AS "Count of Events w/ Field", diffPerc AS "Perc of Total Events w/ Field", distinctValues AS "Distinct Values", numeric_count AS "Numeric Count", values AS
    "Top values with count of each"
```

Fetch data, run fieldsummary

Fieldsummary Output
Enrichment & Filtering

‘values’ JSON Manipulation

Whitespace Column Alignment

```
index=_internal sourcetype=splunkd
| fields - date_*, host, linecount, punct, splunk_server*, timestartpos, timeendpos, timestamp
| fieldsummary maxvals=5
| eval distinctValues = case((is_exact == 1),(distinct_count . " (Exact)"),(is_exact == 0),(distinct_count . " (Estimate)"))
| eventstats max(count) AS eventCount
| eval diffPerc = (round(((count / eventCount) * 100),1) . "%")
| where (tonumber(rtrim(diffPerc,"%")) > 10)
| fields - distinct_count, eventCount, is_exact, mean, max, min, stdev
| spath input=values path="{}.value" output="value_strings"
| eval value_strings = split("\"" . mvjoin(value_strings, "\":::\"") . "\"", ":::")
| spath input=values path="{}.count" output="value_counts"
| eval values = mvzip(value_strings, value_counts, ": ")
| fields - value_counts, value_strings
| mvexpand values
| rex field=values "^(?<fieldValueString>\".+\"):\s(?<fieldValueCount>\d+)"
| eventstats max(eval(len(fieldValueString))) AS max_fieldValue_len BY field
| eval whitespaceInsertAmount = ((max_fieldValue_len + 4) - len(fieldValueString))
| eval values = fieldValueString . substr("
                            ",1,whitespaceInsertAmount) . ":" . fieldValueCount
| fields - whitespaceInsertAmount, max_fieldValue_len, fieldValueString, fieldValueCount
| stats list(values) AS values, values(values) AS values_ThisMustBeSeparateToRetainOrdering, values(*) AS * BY field
| fields - values_ThisMustBeSeparateToRetainOrdering
| eval valuesHash = md5(mvjoin(values, ":::"))
| eval checkSum = md5( count . ":::" . diffPerc . ":::" . numeric_count . ":::" . distinctValues . ":::" . valuesHash)
| eventstats values(field) AS field BY checkSum
| sort 0 - checkSum
| streamstats count AS duplicateCount BY checkSum
| where duplicateCount < 2
| sort 0 - count
| table field, count, diffPerc, numeric_count, distinctValues, values
| rename count AS "Count of Events w/ Field", diffPerc AS "Perc of Total Events w/ Field", distinctValues AS "Distinct Values", numeric_count AS "Numeric Count", values AS
    "Top values with count of each"
```

Fetch data, run fieldsummary

Fieldsummary Output
Enrichment & Filtering

'values' JSON Manipulation

Whitespace Column Alignment

Result Rows Aggregation

```
index=_internal sourcetype=splunkd
| fields - date_*, host, linecount, punct, splunk_server*, timestartpos, timeendpos, timestamp
| fieldsummary maxvals=5
| eval distinctValues = case((is_exact == 1),(distinct_count . " (Exact)"),(is_exact == 0),(distinct_count . " (Estimate)"))
| eventstats max(count) AS eventCount
| eval diffPerc = (round(((count / eventCount) * 100),1) . "%")
| where (tonumber(rtrim(diffPerc,"%")) > 10)
| fields - distinct_count, eventCount, is_exact, mean, max, min, stdev
| spath input=values path="{}.value" output="value_strings"
| eval value_strings = split("\"" . mvjoin(value_strings, "\":::\"") . "\"", ":::")
| spath input=values path="{}.count" output="value_counts"
| eval values = mvzip(value_strings, value_counts, ": ")
| fields - value_counts, value_strings
| mvexpand values
| rex field=values "^(?<fieldValueString>\".+\"):\s(?<fieldValueCount>\d+)"
| eventstats max(eval(len(fieldValueString))) AS max_fieldValue_len BY field
| eval whitespaceInsertAmount = ((max_fieldValue_len + 4) - len(fieldValueString))
| eval values = fieldValueString . substr("
                              ",1,whitespaceInsertAmount) . ":" . fieldValueCount
| fields - whitespaceInsertAmount, max_fieldValue_len, fieldValueString, fieldValueCount
| stats list(values) AS values, values(values) AS values_ThisMustBeSeparateToRetainOrdering, values(*) AS * BY field
| fields - values_ThisMustBeSeparateToRetainOrdering
| eval valuesHash = md5(mvjoin(values, ":::"))
| eval checkSum = md5( count . ":::" . diffPerc . ":::" . numeric_count . ":::" . distinctValues . ":::" . valuesHash)
| eventstats values(field) AS field BY checkSum
| sort 0 - checkSum
| streamstats count AS duplicateCount BY checkSum
| where duplicateCount < 2
| sort 0 - count
| table field, count, diffPerc, numeric_count, distinctValues, values
| rename count AS "Count of Events w/ Field", diffPerc AS "Perc of Total Events w/ Field", distinctValues AS "Distinct Values", numeric_count AS "Numeric Count", values AS
    "Top values with count of each"
```

Fetch data, run fieldsummary

Fieldsummary Output Enrichment & Filtering

'values' JSON Manipulation

Whitespace Column Alignment

Result Rows Aggregation

```
| makeresults | fields - _time
| eval list = "oswin*
osnix*"
| makemv list tokenizer="([^\n]+)"
| mvexpand list
| rename list AS index
| map maxsearches=10 search="search (index=\"$index$\")
    | fields - \"date_*\", eventtype, host, linecount, punct, \"splunk_server*\", timestartpos,
        timeendpos, timestamp
    | fieldsummary maxvals=5
    | where (count > 0)
    | eval index = \"$index$\" "
```

Passing a single field of wildcarded values

Passing multiple fields, dynamically generated

```
| tstats count WHERE index=* BY index, sourcetype, source
| map maxsearches=100 search="search (index=\"$index$\" sourcetype=\"$sourcetype$\" source=\"$source$\")
    | fields - \"date_*\", host, linecount, punct, \"splunk_server*\", timestartpos, timeendpos, timestamp
    | fieldsummary maxvals=5
    | where (count > 0)
    | eval index = \"$index$\"
    | eval sourcetype = \"$sourcetype$\"
    | eval source = \"$source$\"
    "
```

```
| makeresults | fields - _time
| eval list = "oswin*
osnix*"
| makemv list tokenizer="([^\n]+)"
| mvexpand list
| rename list AS index
| map maxsearches=10 search="search (index=\"$index$\")
    | fields - \"date_*\", eventtype, host, linecount, punct, \"splunk_server*\", timestartpos,
        timeendpos, timestamp
    | fieldsummary maxvals=5
    | where (count > 0)
    | eval index = \"$index$\" "
```

Passing a single field of wildcarded values

Passing multiple fields, dynamically generated

```
| tstats count WHERE index=* BY index, sourcetype, source
| map maxsearches=100 search="search (index=\"$index$\" sourcetype=\"$sourcetype$\" source=\"$source$\")
    | fields - \"date_*\", host, linecount, punct, \"splunk_server*\", timestartpos, timeendpos, timestamp
    | fieldsummary maxvals=5
    | where (count > 0)
    | eval index = \"$index$\"
    | eval sourcetype = \"$sourcetype$\"
    | eval source = \"$source$\"
    "
```

First Things First: Fetch Your Input Data

Input is Input, No Matter the Method

```
index=_internal sourcetype=splunkd
```

```
| inputlookup reference_lookup.csv
| table host, tag, index
```

```
| makeresults
| eval input_list = "oswin*
osnix*"
| makemv input_list tokenizer="([^\n]+)"
| mvexpand input_list
| rename input_list AS index
```

```
| tstats count WHERE index IN ("oswin*", "osnix*") BY index, sourcetype, source
```

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Run Your Fieldsummary

(specify *maxvals*)

On specified fields

```
| fieldsummary *dest*, *host*, *server* maxvals=25
```

```
| fields - date_*, host, linecount, punct, splunk_server*, timestartpos, timeendpos, timestamp
| fieldsummary maxvals=5
```

On all fields after dropping unwanted

Iteratively, passing a filter input value

```
| map maxsearches=10 search="search (index=\"$index$\")
    | fields - \"date_*\", eventtype, host, linecount, punct, \"splunk_server*\", timestartpos, timeendpos, timestamp
    | fieldsummary maxvals=5
    | where (count > 0)
    | eval index = \"$index$\" "
```

```
| map maxsearches=100 search="search (index=\"$index$\" sourcetype=\"$sourcetype$\" source=\"$source$\")
    | fields - \"date_*\", host, linecount, punct, \"splunk_server*\", timestartpos, timeendpos, timestamp
    | fieldsummary maxvals=5
    | where (count > 0)
    | eval index = \"$index$\"
    | eval sourcetype = \"$sourcetype$\"
    | eval source = \"$source$\"
    "
```

Iteratively, passing multiple filter values

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Default `fieldsummary` Output

(as a reminder)

field	count	distinct_count	is_exact	max	mean	min	numeric_count	stdev	values
abandoned_channels	1863	1	1	0	0	0	1863	0	[{"value":"0","count":1863}]
ack	1398	1	1				0		[{"value":"false","count":1398}]
action	73	1	1				0		[{"value":"acquire_mutex","count":73}]
actions_triggered	466	2	1	1	0.0193	0	466	0.1378	[{"value":"0","count":457},{"value":"1","count":9}]
active	1152	6	1	5	0.3793	0	1152	0.8559	[{"value":"0","count":912},{"value":"1","count":104},{"value":"2","count":97},{"value":"3","count":23},{"value":"4","count":10},{"value":"5","count":6}]
active_hist_searches	2385	21	0	25	1.1287	0	2385	2.2016	[{"value":"0","count":1492},{"value":"1","count":320},{"value":"2","count":162},{"value":"3","count":140},{"value":"4","count":97},{"value":"5","count":73},{"value":"6","count":41},{"value":"7","count":14}]
active_realtime_searches	2385	1	1	0	0	0	2385	0	[{"value":"0","count":2385}]
alert_cycle_max	466	3	1	0.002	0.0000	0.000	466	0.0001	[{"value":"0.000","count":464},{"value":"0.001","count":1},{"value":"0.002","count":1}]
alert_cycle_total	466	3	1	0.002	0.0000	0.000	466	0.0001	[{"value":"0.000","count":464},{"value":"0.001","count":1},{"value":"0.002","count":1}]
alert_cycles	466	17	0	23	5.1910	3	466	2.8864	[{"value":"4","count":134},{"value":"3","count":113},{"value":"5","count":98},{"value":"6","count":58},{"value":"11","count":12},{"value":"7","count":11},{"value":"12","count":9},{"value":"10","count":7}]

https://docs.splunk.com/Documentation/Splunk/latest/SearchReference/Fieldsummary

BSIDES splunk>

Transform Default Output, Add Percentiles, Filter.

(concat, create coverage %)

```
START fieldsummary transformations and base calculations/filtering.
")
| eval distinctValues = case(
    (is_exact == 1), (distinct_count . " (Exact)"),
    (is_exact == 0),(distinct_count . " (Estimate)")
    )
| eventstats max(count) as eventCount
| eval diffPerc = (round(((count / eventCount) * 100),1) . "%")
| where (tonumber(rtrim(diffPerc,"%")) > 10)
| fields - distinct_count, eventCount, is_exact, mean, max, min, stdev
| eval comment = if(1==1, null(), "
END fieldsummary transformations and base calculations/filtering.
```

field	count	eventCount	diffPerc	distinctValues	distinct_count	is_exact
index	1355	1355	100.0%	1 (Exact)	1	1
comm	240	1355	17.7%	1 (Exact)	1	1
uid	240	1355	17.7%	1 (Exact)	1	1
processed	24	1355	1.8%	24 (Estimate)	24	0
source	1355	1355	100.0%	3 (Exact)	3	1
process	1339	1355	98.8%	25 (Estimate)	25	0
pid	1263	1355	93.2%	20 (Estimate)	20	0
name	240	1355	17.7%	1 (Exact)	1	1

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Transform Default Output, Add Percentiles, Filter.

(filter to minimum coverage using *diffPerc*)

```
13 START fieldsummary transformations and base calculations/filtering.
14 ")
15 | eval distinctValues = case(
16     (is_exact == 1), (distinct_count . " (Exact)"),
17     (is_exact == 0),(distinct_count . " (Estimate)")
18     )
19 | eventstats max(count) as eventCount
20 | eval diffPerc = (round(((count / eventCount) * 100),1) . "%")
21 | where (tonumber(rtrim(diffPerc,"%")) > 10)
22 | fields - distinct_count, eventCount, is_exact, mean, max, min, stdev
23 | eval comment = if(1==1, null(), "
24 END fieldsummary transformations and base calculations/filtering.
```

field	count	eventCount	diffPerc	distinctValues	distinct_count	is_exact
index	1355	1355	100.0%	1 (Exact)	1	1
comm	240	1355	17.7%	1 (Exact)	1	1
uid	240	1355	17.7%	1 (Exact)	1	1
processed	24	1355	1.8%	24 (Estimate)	24	0
source	1355	1355	100.0%	3 (Exact)	3	1
process	1339	1355	98.8%	25 (Estimate)	25	0
pid	1263	1355	93.2%	20 (Estimate)	20	0
name	240	1355	17.7%	1 (Exact)	1	1

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Transform Default Output, Add Percentiles, Filter.

(drop unwanted fields)

```
13 START fieldsummary transformations and base calculations/filtering.
14 ")
15 | eval distinctValues = case(
16     (is_exact == 1), (distinct_count . " (Exact)"),
17     (is_exact == 0),(distinct_count . " (Estimate)")
18     )
19 | eventstats max(count) as eventCount
20 | eval diffPerc = (round(((count / eventCount) * 100),1) . "%")
21 | where (tonumber(rtrim(diffPerc,"%")) > 10)
22 | fields - distinct_count, eventCount, is_exact, mean, max, min, stdev
23 | eval comment = if(1==1, null(), "
24 END fieldsummary transformations and base calculations/filtering.
```

Modify the *fields* line if there's anything you want to keep!

field	count	diffPerc	distinctValues	values
index	1355	100.0%	1 (Exact)	[{"value":"osnix","count":1355}]
comm	240	17.7%	1 (Exact)	[{"value":"timedatectl","count":240}]
uid	240	17.7%	1 (Exact)	[{"value":"1002","count":240}]
source	1355	100.0%	3 (Exact)	[{"value":"/var/log/messages","count":1251}, {"value":"/var/log/dnf.log","count":92}, {"value":"/var/log/cron","count":12}]
process	1339	98.8%	25 (Estimate)	[{"value":"systemd","count":732},{"value":"dbus-daemon","count":480},{"value":"syslog-ng","count":24},{"value":"dnf","count":14}, {"value":"reviving","count":10}]
pid	1263	93.2%	20 (Estimate)	[{"value":"1","count":732}, {"value":"1149","count":480}, {"value":"1144","count":24}, {"value":"1946073","count":6}, {"value":"2159122","count":6}]
name	240	17.7%	1 (Exact)	[{"value":"'org.freedesktop.timedate1'","count":240}]

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Readable *values* Field Transformation

(extract JSON elements using *spath*)

```
START reformatting of 'values' JSON data into readable Multivalue field.
")
| spath input=values path="{}.value" output="value_strings"
| spath input=values path="{}.count" output="value_counts"
| eval value_strings = split("\"" . mvjoin(value_strings, "\":::\"") . "\"", ":::")
| eval values = mvzip(value_strings, value_counts, ": ")
| fields - value_counts, value_strings
| eval comment = if(1==1, null(), "
END reformatting of 'values' JSON data into readable Multivalue field.
```

field	count	diffPerc	distinctValues	value_counts	value_strings	values
source	1355	100.0%	3 (Exact)	1251 92 12	/var/log/messages /var/log/dnf.log /var/log/cron	[{"value":"/var/log/messages","count":1251}, {"value":"/var/log/dnf.log","count":92}, {"value":"/var/log/cron","count":12}]
process	1339	98.8%	25 (Estimate)	732 480 24 14 10	systemd dbus-daemon syslog-ng dnf reviving	[{"value":"systemd","count":732},{"value":"dbus-daemon","count":480}, {"value":"syslog-ng","count":24},{"value":"dnf","count":14}, {"value":"reviving","count":10}]
pid	1263	93.2%	20 (Estimate)	732 480 24 6 6	1 1149 1144 1946073 2159122	[{"value":"1","count":732},{"value":"1149","count":480}, {"value":"1144","count":24},{"value":"1946073","count":6}, {"value":"2159122","count":6}]
name	240	17.7%	1 (Exact)	240	'org.freedesktop.timedate1'	[{"value":"'org.freedesktop.timedate1'","count":240}]

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

Readable *values* Field Transformation

(wrap the MV *value_strings* field with "doublequotes")

```
27 START reformatting of 'values' JSON data into readable Multivalue field.
28 ")
29 | spath input=values path="{}.value" output="value_strings"
30 | spath input=values path="{}.count" output="value_counts"
31 | eval value_strings = split("\"" . mvjoin(value_strings, "\":::\"") . "\"", ":::")
32 | eval values = mvzip(value_strings, value_counts, ": ")
33 | fields - value_counts, value_strings
34 | eval comment = if(1==1, null(), "
35 END reformatting of 'values' JSON data into readable Multivalue field.
```

field	count	diffPerc	distinctValues	value_counts	value_strings	values
source	1355	100.0%	3 (Exact)	1251 92 12	"/var/log/messages" "/var/log/dnf.log" "/var/log/cron"	[{"value":"/var/log/messages","count":1251}, {"value":"/var/log/dnf.log","count":92}, {"value":"/var/log/cron","count":12}]
process	1339	98.8%	25 (Estimate)	732 480 24 14 10	"systemd" "dbus-daemon" "syslog-ng" "dnf" "reviving"	[{"value":"systemd","count":732},{"value":"dbus- daemon","count":480},{"value":"syslog-ng","count":24}, {"value":"dnf","count":14},{"value":"reviving","count":10}]
pid	1263	93.2%	20 (Estimate)	732 480 24 6 6	"1" "1149" "1144" "1946073" "2159122"	[{"value":"1","count":732},{"value":"1149","count":480}, {"value":"1144","count":24},{"value":"1946073","count":6}, {"value":"2159122","count":6}]
name	240	17.7%	1 (Exact)	240	"'org.freedesktop.timedate1'"	[{"value":"'org.freedesktop.timedate1'","count":240}]

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

Readable *values* Field Transformation

(wrap the MV *value_strings* field with "doublequotes")

For Splunk 8+, you could also use *mvmap()*

https://docs.splunk.com/Documentation/Splunk/latest/SearchReference/MultivalueEvalFunctions#mvmap.28X.2CY.29

```
| eval value_strings = mvmap(value_strings, "\"" . value_strings . "\"")
```

field	count	diffPerc	distinctValues	value_counts	value_strings	values
source	1355	100.0%	3 (Exact)	1251 92 12	"/var/log/messages" "/var/log/dnf.log" "/var/log/cron"	[{"value":"/var/log/messages","count":1251}, {"value":"/var/log/dnf.log","count":92}, {"value":"/var/log/cron","count":12}]
process	1339	98.8%	25 (Estimate)	732 480 24 14 10	"systemd" "dbus-daemon" "syslog-ng" "dnf" "reviving"	[{"value":"systemd","count":732},{"value":"dbus- daemon","count":480},{"value":"syslog-ng","count":24}, {"value":"dnf","count":14},{"value":"reviving","count":10}]
pid	1263	93.2%	20 (Estimate)	732 480 24 6 6	"1" "1149" "1144" "1946073" "2159122"	[{"value":"1","count":732},{"value":"1149","count":480}, {"value":"1144","count":24},{"value":"1946073","count":6}, {"value":"2159122","count":6}]
name	240	17.7%	1 (Exact)	240	"'org.freedesktop.timedate1'"	[{"value":"'org.freedesktop.timedate1'","count":240}]

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

Readable *values* Field Transformation

(combine the two MV fields, drop each individual field)

```
START reformatting of 'values' JSON data into readable Multivalue field.
")
| spath input=values path="{}.value" output="value_strings"
| spath input=values path="{}.count" output="value_counts"
| eval value_strings = split("\"" . mvjoin(value_strings, "\":::\"") . "\"", ":::")
| eval values = mvzip(value_strings, value_counts, ": ")
| fields - value_counts, value_strings
| eval comment = if(1==1, null(), "
END reformatting of 'values' JSON data into readable Multivalue field.
```

field	count	diffPerc	distinctValues	values
source	1355	100.0%	3 (Exact)	"/var/log/messages": 1251 "/var/log/dnf.log": 92 "/var/log/cron": 12
process	1339	98.8%	25 (Estimate)	"systemd": 732 "dbus-daemon": 480 "syslog-ng": 24 "dnf": 14 "reviving": 10
pid	1263	93.2%	20 (Estimate)	"1": 732 "1149": 480 "1144": 24 "1946073": 6 "2159122": 6
name	240	17.7%	1 (Exact)	"'org.freedesktop.timedate1'": 240

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Column Alignment of *values* Field

(expand MV *values* field, extract string and count into new fields)

```
38 START column alignment formatting for values field.
39     Note: column alignment, if calculated on the aggregate dataset without GROUPBY, can result in wrapped lines if the dataset contains any
           fields with long values.")
40 | mvexpand values
41 | rex field=values "^(?<fieldValueString>\".+\"):\s(?<fieldValueCount>\d+)"
42 | eventstats max(eval(len(fieldValueString))) AS max_fieldValue_len BY field
43 | eval whitespaceInsertAmount=((max_fieldValue_len + 4) - len(fieldValueString))
44 | eval values = fieldValueString . substr("
                                                                    ",1,whitespaceInsertAmount) . ":" . fieldValueCount
45 | fields - whitespaceInsertAmount, max_fieldValue_len, fieldValueString, fieldValueCount
46 | stats list(values) AS values, values(values) AS values_ThisMustBeSeparateToRetainOrdering, values(*) AS * BY field
47 | fields - values_ThisMustBeSeparateToRetainOrdering
48 | eval comment = if(1==1, null(), "
49 END column alignment formatting for values field.
```

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Column Alignment of *values* Field

(expand MV *values* field, extract string and count into new fields)

```
| mvexpand values
| rex field=values "^(?<fieldValueString>\".+\"):\s(?<fieldValueCount>\d+)"
```

field	values
index	"osnix": 1355
comm	"timedatectl": 240
uid	"1002": 240
source	"/var/log/messages": 1251 "/var/log/dnf.log": 92 "/var/log/cron": 12
process	"systemd": 732 "dbus-daemon": 480 "syslog-ng": 24 "dnf": 14 "reviving": 10
pid	"1": 732 "1149": 480 "1144": 24 "1946073": 6 "2159122": 6
name	"'org.freedesktop.timedate1'": 240

field	fieldValueString	fieldValueCount	values
index	"osnix"	1355	"osnix": 1355
comm	"timedatectl"	240	"timedatectl": 240
uid	"1002"	240	"1002": 240
source	"/var/log/messages"	1251	"/var/log/messages": 1251
source	"/var/log/dnf.log"	92	"/var/log/dnf.log": 92
source	"/var/log/cron"	12	"/var/log/cron": 12
process	"systemd"	732	"systemd": 732
process	"dbus-daemon"	480	"dbus-daemon": 480
process	"syslog-ng"	24	"syslog-ng": 24
process	"dnf"	14	"dnf": 14
process	"reviving"	10	"reviving": 10
pid	"1"	732	"1": 732
pid	"1149"	480	"1149": 480
pid	"1144"	24	"1144": 24
pid	"1946073"	6	"1946073": 6
pid	"2159122"	6	"2159122": 6
name	"'org.freedesktop.timedate1'"	240	"'org.freedesktop.timedate1'": 240

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

Column Alignment of *values* Field

(calculate max field length, whitespace insert size)

```
38 START column alignment formatting for values field.
39     Note: column alignment, if calculated on the aggregate dataset without GROUPBY, can result in wrapped lines if the dataset contains any
           fields with long values.")
40 | mvexpand values
41 | rex field=values "^(?<fieldValueString>\".+\"):\s(?<fieldValueCount>\d+)"
42 | eventstats max(eval(len(fieldValueString))) AS max_fieldValue_len BY field
43 | eval whitespaceInsertAmount=((max_fieldValue_len + 4) - len(fieldValueString))
44 | eval values = fieldValueString . substr("
                                                    ",1,whitespaceInsertAmount) . ":" . fieldValueCount
45 | fields - whitespaceInsertAmount, max_fieldValue_len, fieldValueString, fieldValueCount
46 | stats list(values) AS values, values(values) AS values_ThisMustBeSeparateToRetainOrdering, values(*) AS * BY field
47 | fields - values_ThisMustBeSeparateToRetainOrdering
48 | eval comment = if(1==1, null(), "
49 END column alignment formatting for values field.
```

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Column Alignment of *values* Field

(calculate max field length, whitespace insert size)

```
| eventstats max(eval(len(fieldValueString))) AS max_fieldValue_len BY field
| eval whitespaceInsertAmount = ((max_fieldValue_len + 4) - len(fieldValueString))
```

field	fieldValueString	fieldValueCount
index	"osnix"	1355
comm	"timedatectl"	240
uid	"1002"	240
source	"/var/log/messages"	1251
source	"/var/log/dnf.log"	92
source	"/var/log/cron"	12
process	"systemd"	732
process	"dbus-daemon"	480
process	"syslog-ng"	24
process	"dnf"	14
process	"reviving"	10
pid	"1"	732
pid	"1149"	480
pid	"1144"	24
pid	"1946073"	6
pid	"2159122"	6
name	"'org.freedesktop.timedate1'"	240

field	fieldValueString	fieldValueCount	max_fieldValue_len	whitespaceInsertAmount
index	"osnix"	1355	7	4
comm	"timedatectl"	240	13	4
uid	"1002"	240	6	4
source	"/var/log/messages"	1251	19	4
source	"/var/log/dnf.log"	92	19	5
source	"/var/log/cron"	12	19	8
process	"systemd"	732	13	8
process	"dbus-daemon"	480	13	4
process	"syslog-ng"	24	13	6
process	"dnf"	14	13	12
process	"reviving"	10	13	7
pid	"1"	732	9	10
pid	"1149"	480	9	7
pid	"1144"	24	9	7
pid	"1946073"	6	9	4
pid	"2159122"	6	9	4
name	"'org.freedesktop.timedate1'"	240	29	4

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Column Alignment of *values* Field

(append whitespace onto *value* to align rows)

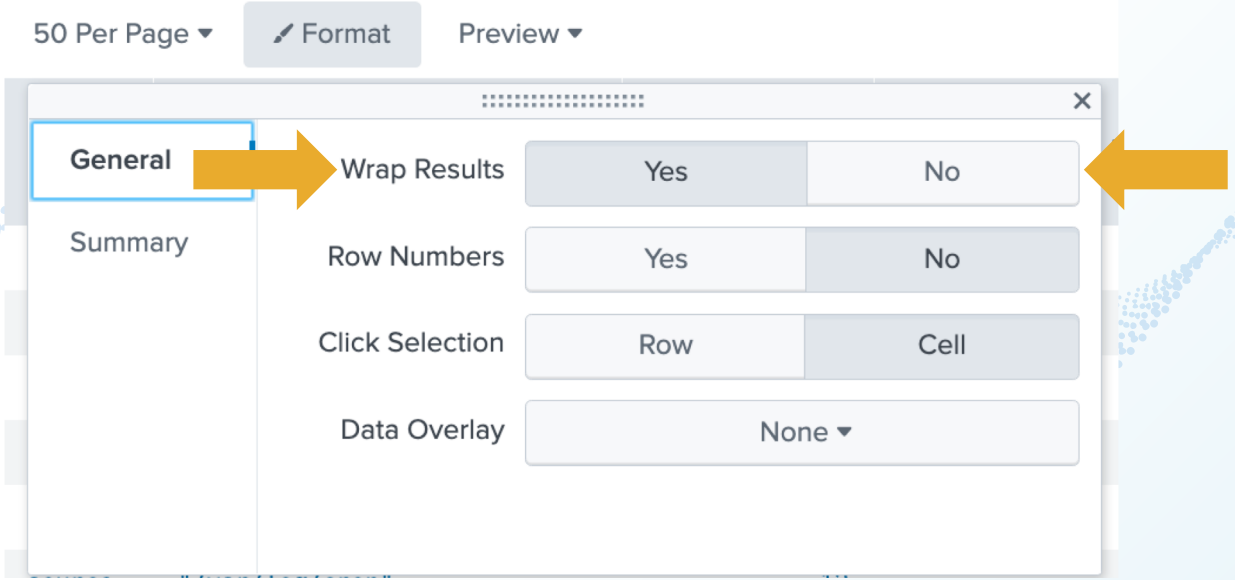
```
START column alignment formatting for values field.
    Note: column alignment, if calculated on the aggregate dataset without GROUPBY, can result in wrapped lines if the dataset contains any
        fields with long values.")
| mvexpand values
| rex field=values "^(?<fieldValueString>\".+\"):\s(?<fieldValueCount>\d+)"
| eventstats max(eval(len(fieldValueString))) AS max_fieldValue_len BY field
| eval whitespaceInsertAmount=((max_fieldValue_len + 4) - len(fieldValueString))
| eval values = fieldValueString . substr("
                                                    ",1,whitespaceInsertAmount) . ":" . fieldValueCount
| fields - whitespaceInsertAmount, max_fieldValue_len, fieldValueString, fieldValueCount
| stats list(values) AS values, values(values) AS values_ThisMustBeSeparateToRetainOrdering, values(*) AS * BY field
| fields - values_ThisMustBeSeparateToRetainOrdering
| eval comment = if(1==1, null(), "
END column alignment formatting for values field.
```

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Column Alignment of *values* Field

NOTE: “Wrap Results” MUST be set to Yes to *see* the alignment.



Column Alignment of *values* Field

(expand MV *values* field, extract string and count into new fields)

```
| eval whitespaceInsertAmount = ((max_fieldValue_len + 4) - len(fieldValueString))
| eval values = fieldValueString . substr("
                                        ",1,whitespaceInsertAmount) . ":" . fieldValueCount
```

field	fieldValueString	fieldValueCount	max_fieldValue_len	whitespaceInsertAmount
index	"osnix"	1355	7	4
comm	"timedatectl"	240	13	4
uid	"1002"	240	6	4
source	"/var/log/messages"	1251	19	4
source	"/var/log/dnf.log"	92	19	5
source	"/var/log/cron"	12	19	8
process	"systemd"	732	13	8
process	"dbus-daemon"	480	13	4
process	"syslog-ng"	24	13	6
process	"dnf"	14	13	12
process	"reviving"	10	13	7
pid	"1"	732	9	10
pid	"1149"	480	9	7
pid	"1144"	24	9	7
pid	"1946073"	6	9	4
pid	"2159122"	6	9	4
name	"'org.freedesktop.timedate1'"	240	29	4

field	values
process	"systemd" :732
process	"dbus-daemon" :480
process	"syslog-ng" :24
process	"dnf" :14
process	"reviving" :10
pid	"1" :732
pid	"1149" :480
pid	"1144" :24
pid	"1946073" :6
pid	"2159122" :6

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

Column Alignment of *values* Field

(aggregate rows together, retaining order of values)

```
38 START column alignment formatting for values field.
39     Note: column alignment, if calculated on the aggregate dataset without GROUPBY, can result in wrapped lines if the dataset contains any
           fields with long values.")
40 | mvexpand values
41 | rex field=values "^(?<fieldValueString>\".+\"):\s(?<fieldValueCount>\d+)"
42 | eventstats max(eval(len(fieldValueString))) AS max_fieldValue_len BY field
43 | eval whitespaceInsertAmount=((max_fieldValue_len + 4) - len(fieldValueString))
44 | eval values = fieldValueString . substr("
                                                                    ",1,whitespaceInsertAmount) . ":" . fieldValueCount
45 | fields - whitespaceInsertAmount, max_fieldValue_len, fieldValueString, fieldValueCount
46 | stats list(values) AS values, values(values) AS values_ThisMustBeSeparateToRetainOrdering, values(*) AS * BY field
47 | fields - values_ThisMustBeSeparateToRetainOrdering
48 | eval comment = if(1==1, null(), "
49 END column alignment formatting for values field.
```

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Column Alignment of *values* Field

(aggregate rows together, retaining order of values)

```
| fields - whitespaceInsertAmount, max_fieldValue_len, fieldValueString, fieldValueCount
| stats list(values) AS values, values(values) AS values_ThisMustBeSeparateToRetainOrdering, values(*) AS * BY field
| fields - values_ThisMustBeSeparateToRetainOrdering
```

field	values
process	"systemd" :732
process	"dbus-daemon" :480
process	"syslog-ng" :24
process	"dnf" :14
process	"reviving" :10
pid	"1" :732
pid	"1149" :480
pid	"1144" :24
pid	"1946073" :6
pid	"2159122" :6

field	count	diffPerc	distinctValues	values
index	1355	100.0%	1 (Exact)	"osnix" :1355
comm	240	17.7%	1 (Exact)	"timedatectl" :240
uid	240	17.7%	1 (Exact)	"1002" :240
source	1355	100.0%	3 (Exact)	"/var/log/messages" :1251 "/var/log/dnf.log" :92 "/var/log/cron" :12
process	1339	98.8%	25 (Estimate)	"systemd" :732 "dbus-daemon" :480 "syslog-ng" :24 "dnf" :14 "reviving" :10
pid	1263	93.2%	20 (Estimate)	"1" :732 "1149" :480 "1144" :24 "1946073" :6 "2159122" :6
name	240	17.7%	1 (Exact)	"'org.freedesktop.timedate1'" :240

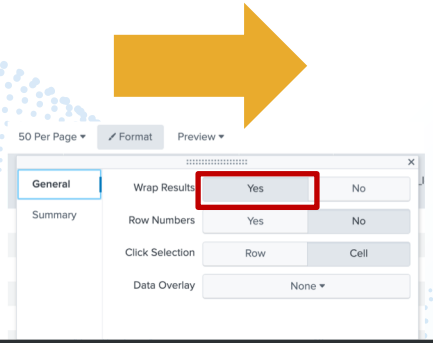
https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Column Alignment values Field

(adding whitespace to align multi-value items, recap)

values
"Metrics": 1300088
"PerProcess": 85400
"PeriodicHealthReporter": 65229
"SavedSplunker": 22543
"KVStoreCollectionStats": 17364
"Hostwide": 8758
"Indexes": 7446
"TcpOutputProc": 5862
"HttpPubSubConnection": 5447
"KVStoreServerStats": 3245
"LocalAppsAdminHandler": 3100



values
"Metrics" :1295761
"PerProcess" :85123
"PeriodicHealthReporter" :65024
"SavedSplunker" :22465
"KVStoreCollectionStats" :17245
"Hostwide" :8730
"Indexes" :7395
"TcpOutputProc" :5843
"HttpPubSubConnection" :5428
"KVStoreServerStats" :3235
"LocalAppsAdminHandler" :3089

```
START column alignment formatting for values field.
    Note: column alignment, if calculated on the aggregate dataset without GROUPBY, can result in wrapped lines if the dataset contains any
        fields with long values.")
| mvexpand values
| rex field=values "^(?<fieldValueString>\".+\"):\s(?<fieldValueCount>\d+)"
| eventstats max(eval(len(fieldValueString))) AS max_fieldValue_len BY field
| eval whitespaceInsertAmount = ((max_fieldValue_len + 4) - len(fieldValueString))
| eval values = fieldValueString . substr("
                                                    ",1,whitespaceInsertAmount) . ":" . fieldValueCount
| fields - whitespaceInsertAmount, max_fieldValue_len, fieldValueString, fieldValueCount
| stats list(values) AS values, values(values) AS values_ThisMustBeSeparateToRetainOrdering, values(*) AS * BY field
| fields - values_ThisMustBeSeparateToRetainOrdering
```

Aggregate Common Field Values

(create *checkSum* md5 hash for each row)

```
START row deduplication/aggregation of duplicate information across multiple fields into a single row through using a md5 hash.
    If all the columns EXCEPT field name are equivalent, aggregate.
")
| eval valuesHash = md5(mvjoin(values, ":::"))
| eval checkSum = md5(count . ":::" . diffPerc . ":::" . numeric_count . ":::" . distinctValues . ":::" . valuesHash)
| eventstats values(field) AS field BY checkSum
| sort 0 - checkSum
| streamstats count AS duplicateCount BY checkSum
| where duplicateCount < 2
```

field	count	diffPerc	distinctValues	values	valuesHash	checkSum
dest_pci_domain	1355	100.0%	1 (Exact)	"untrust" :1355	bb10778bded86abef03219886e9482f2	93271e9826567ba2b772980511f7f544
dest_is_expected dest_should_timesync dest_should_update	1355	100.0%	1 (Exact)	"false" :1355	d9d407bcc9c1b82146baba04bfeeb116	086d4d71ff171f096efbaab037114916
dest_is_expected dest_should_timesync dest_should_update	1355	100.0%	1 (Exact)	"false" :1355	d9d407bcc9c1b82146baba04bfeeb116	086d4d71ff171f096efbaab037114916
dest_is_expected dest_should_timesync dest_should_update	1355	100.0%	1 (Exact)	"false" :1355	d9d407bcc9c1b82146baba04bfeeb116	086d4d71ff171f096efbaab037114916

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

Aggregate Common Field Values

(identify “duplicates” using *eventstats* + *streamstats*)

```
START row deduplication/aggregation of duplicate information across multiple fields into a single row through using a md5 hash.
    If all the columns EXCEPT field name are equivalent, aggregate.
")
| eval valuesHash = md5(mvjoin(values, ":::"))
| eval checkSum = md5(count . ":::" . diffPerc . ":::" . numeric_count . ":::" . distinctValues . ":::" . valuesHash)
| eventstats values(field) AS field BY checkSum
| sort 0 - checkSum
| streamstats count AS duplicateCount BY checkSum
| where duplicateCount < 2
```

field	count	diffPerc	distinctValues	values	checkSum	duplicateCount
dest_pci_domain	1355	100.0%	1 (Exact)	"untrust" :1355	93271e9826567ba2b772980511f7f544	1
dest_is_expected dest_should_timesync dest_should_update	1355	100.0%	1 (Exact)	"false" :1355	086d4d71ff171f096efbaab037114916	1
dest_is_expected dest_should_timesync dest_should_update	1355	100.0%	1 (Exact)	"false" :1355	086d4d71ff171f096efbaab037114916	2
dest_is_expected dest_should_timesync dest_should_update	1355	100.0%	1 (Exact)	"false" :1355	086d4d71ff171f096efbaab037114916	3

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Aggregate Common Field Values

(consolidate)

```
START row deduplication/aggregation of duplicate information across multiple fields into a single row through using a md5 hash.
    If all the columns EXCEPT field name are equivalent, aggregate.
")
| eval valuesHash = md5(mvjoin(values, ":::"))
| eval checkSum = md5(count . ":::" . diffPerc . ":::" . numeric_count . ":::" . distinctValues . ":::" . valuesHash)
| eventstats values(field) AS field BY checkSum
| sort 0 - checkSum
| streamstats count AS duplicateCount BY checkSum
| where duplicateCount < 2
```

field	count	diffPerc	distinctValues	values	checkSum	duplicateCount
dest_pci_domain	1355	100.0%	1 (Exact)	"untrust" :1355	93271e9826567ba2b772980511f7f544	1
dest_is_expected dest_should_timesync dest_should_update	1355	100.0%	1 (Exact)	"false" :1355	086d4d71ff171f096efbaab037114916	1

Drops Multi-value

~~`| stats first(*) AS * BY checkSum`~~

Loses Sort Order

~~`| stats values(*) AS * BY checkSum`~~

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

```
index=_internal sourcetype=splunkd
| fields - date_*, host, linecount, punct, splunk_server*, timestartpos, timeendpos, timestamp
| fieldsummary maxvals=5
| eval distinctValues = case((is_exact == 1),(distinct_count . " (Exact)"),(is_exact == 0),(distinct_count . " (Estimate)"))
| eventstats max(count) AS eventCount
| eval diffPerc = (round(((count / eventCount) * 100),1) . "%")
| where (tonumber(rtrim(diffPerc,"%")) > 10)
| fields - distinct_count, eventCount, is_exact, mean, max, min, stdev
| spath input=values path="{}.value" output="value_strings"
| eval value_strings = split("\"" . mvjoin(value_strings, "\":::\"") . "\"", ":::")
| spath input=values path="{}.count" output="value_counts"
| eval values = mvzip(value_strings, value_counts, ": ")
| fields - value_counts, value_strings
| mvexpand values
| rex field=values "^(?<fieldValueString>\".+\"):\s(?<fieldValueCount>\d+)"
| eventstats max(eval(len(fieldValueString))) AS max_fieldValue_len BY field
| eval whitespaceInsertAmount = ((max_fieldValue_len + 4) - len(fieldValueString))
| eval values = fieldValueString . substr("
                            ",1,whitespaceInsertAmount) . ":" . fieldValueCount
| fields - whitespaceInsertAmount, max_fieldValue_len, fieldValueString, fieldValueCount
| stats list(values) AS values, values(values) AS values_ThisMustBeSeparateToRetainOrdering, values(*) AS * BY field
| fields - values_ThisMustBeSeparateToRetainOrdering
| eval valuesHash = md5(mvjoin(values, ":::"))
| eval checkSum = md5( count . ":::" . diffPerc . ":::" . numeric_count . ":::" . distinctValues . ":::" . valuesHash)
| eventstats values(field) AS field BY checkSum
| sort 0 - checkSum
| streamstats count AS duplicateCount BY checkSum
| where duplicateCount < 2
| sort 0 - count
| table field, count, diffPerc, numeric_count, distinctValues, values
| rename count AS "Count of Events w/ Field", diffPerc AS "Perc of Total Events w/ Field", distinctValues AS "Distinct Values", numeric_count AS "Numeric Count", values AS
    "Top values with count of each"
```

Fetch data, run fieldsummary

Fieldsummary Output
Enrichment & Filtering

'values' JSON Manipulation

Whitespace Column Alignment

Result Rows Aggregation

Quick Review

Fieldsummary Usecase Components:

1. Initial fieldsummary calculation on input data.
   (including | map for GROUPBY examples)

2. Transform fieldsummary default output, add percentiles, filter minimum coverage threshold.

3. Reformat values field JSON into multi-value field, retaining value and count.

4. Format column view of *values* field to align values across groupby's by using whitespace.

5. Aggregate *field* where 100% equal to other rows in the dataset (BY groupby) using checkSum hash.

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Fieldsummary

(base, no groupby)

field	count	diffPerc	numeric_count	distinctValues	values
status	22512	98.7%	0	2 (Exact)	"success" :12796 "skipped" :9716
savedsearch_name	22512	98.7%	0	35 (Estimate)	"Threat - Correlation Searches - Lookup Gen" :1447 "_ACCELERATE_DM_SA-ThreatIntelligence_Incident_Management_ACCELERATE_" :964 "_ACCELERATE_DM_DA-ESS-ThreatIntelligence_Threat_Intelligence_ACCELERATE_" :961 "_ACCELERATE_DM_SA-ThreatIntelligence_Risk_ACCELERATE_" :955 "_ACCELERATE_DM_SA-NetworkProtection_Domain_Analysis_ACCELERATE_" :872 "_ACCELERATE_3E64B466-3A4F-475E-BC61-6C5DD2B2F161_InfoSec_App_for_Splunk_nobody_0ec400bee57dd68a_ACCELERATE_" :660 "_ACCELERATE_3E64B466-3A4F-475E-BC61-6C5DD2B2F161_InfoSec_App_for_Splunk_nobody_ff4c80c65c5958bd_ACCELERATE_" :645
priority	22512	98.7%	0	2 (Exact)	"highest" :12474 "default" :10038
search_type	22512	98.7%	0	3 (Exact)	"datamodel_acceleration" :16226 "scheduled" :4981 "report_acceleration" :1305
app	22512	98.7%	0	19 (Estimate)	"Splunk_SA_CIM" :12474 "SA-ThreatIntelligence" :3664 "DA-ESS-ThreatIntelligence" :1839 "InfoSec_App_for_Splunk" :1646 "SA-NetworkProtection" :1064 "TA-linux_auditd" :513 "SA-Utils" :371
user_watchlist	22512	98.7%	0	1 (Exact)	"false" :22512
user	22512	98.7%	0	2 (Exact)	"nobody" :18605 "admin" :3907

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Fieldsummary

(BY *index* wildcard)

index	field	count	diffPerc	numeric_count	distinctValues	values
osnix*	index	6929	100.0%	0	2 (Exact)	"osnix" :4782 "osnixsec" :2147
osnix*	sourcetype	6929	100.0%	0	2 (Exact)	"syslog" :4984 "auditd" :1945
osnix*	source	6929	100.0%	0	5 (Exact)	"/var/log/messages" :4066 "auditd" :1945 "/var/log/dnf.log" :507 "/var/log/cron" :209 "secure" :202
oswin*	Keywords	453	100.0%	0	7 (Estimate)	"Audit Success" :209 "0x8080000000000000" :202 "0x80000000000000" :33 "0x8000000000002004" :4 "0x8000000000000018" :2 "0x8000000000002008" :2 "0x4000000000000000" :1
oswin*	Error_Code	453	100.0%	0	1 (Exact)	"-" :453
oswin*	EventCode	453	100.0%	453	22 (Estimate)	"7036" :188 "4799" :60 "4624" :46 "4672" :46 "4634" :35

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

Fieldsummary

(BY *index, sourcetype, source*)

Performance warning at scale!

index	sourcetype	source	field	count	diffPerc	distinctValues	values
osnix	syslog	/var/log/messages	process	1251	100.0%	5 (Estimate)	"systemd" :732 "dbus-daemon" :480 "syslog-ng" :24 "dnf" :14 "chronyd" :1
osnix	syslog	/var/log/dnf.log	process	76	82.6%	18 (Estimate)	"reviving" :10 "timer" :6 "Command" :4 "Installroot" :4 "Releasever" :4
oswin	XmlWinEventLog	XmlWinEventLog:System	ThreadID	30	100.0%	7 (Estimate)	"'1800'" :9 "'2492'" :6 "'4552'" :6 "'576'" :5 "'3344'" :2
oswin	XmlWinEventLog	XmlWinEventLog:Application	ProcessID ThreadID	16	100.0%	1 (Exact)	"'0'" :16
oswinsec	WinEventLog	WinEventLog:Security	Security_ID	27	100.0%	2 (Exact)	"S-1-5-18" :29 "S-1-0-0" :7

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Fieldsummary

(BY *index, sourcetype, source* with custom event filters)

```
| tstats count WHERE index IN ("oswin*", "osnix*") BY index, sourcetype, source
| eval specificFilterString = case(
    match(source, "^WinEventLog:Security"), "(TargetUserName=* AND (SubjectUserName=* OR SubjectUserSid=*))",
    match(source, "^XmlWinEventLog:System"), "(SAMAccountName=* AND signature_id=4755)",
    match(sourcetype, "^XmlWinEventLog"), "earliest=1618349400 latest=1618363811",
    match(index, "^osnix"), "(uid IN (16255, 20755, 38169)",
    1==1, "")
| eval comment = if( 1==1, null(), "
specificFilterString is a way to pass in conditional filter strings to the first pipe of the searches executed by the map command, allowing filtering of the data to avoid heavier impact. ")
| map maxsearches=100 search="search (index=\"$index$\" sourcetype=\"$sourcetype$\" source=\"$source$\") [| makeresults | eval search = $specificFilterString$ | return $search]
    | fields - \"date_*\", host, linecount, punct, \"splunk_server*\", timestartpos, timeendpos, timestamp
    | fieldsummary maxvals=10
    | where (count > 0)
    | eval index = \"$index$\"
    | eval sourcetype = \"$sourcetype$\"
    | eval source = \"$source$\"
    "
```

index	sourcetype	source	count	specificFilterString
osnix	syslog	/var/log/cron	212	(uid IN (16255, 20755, 38169)
osnix	syslog	/var/log/dnf.log	546	(uid IN (16255, 20755, 38169)
osnix	syslog	/var/log/messages	11433	(uid IN (16255, 20755, 38169)
osnixsec	syslog	secure	188	(uid IN (16255, 20755, 38169)
oswin	XmlWinEventLog	XmlWinEventLog:Application	32	earliest=1618349400 latest=1618363811
oswin	XmlWinEventLog	XmlWinEventLog:System	187	(SAMAccountName=* AND signature_id=4755)
oswinsec	WinEventLog	WinEventLog:Security	202	(TargetUserName=* AND (SubjectUserName=* OR SubjectUserSid=*))

Fieldsummary

(BY *index, sourcetype, source* with custom event filters)

```
| tstats count WHERE index IN ("oswin*", "osnix*") BY index, sourcetype, source
| eval specificFilterString = case(
    match(source, "^WinEventLog:Security"), "(TargetUserName=* AND (SubjectUserName=* OR SubjectUserSid=*))",
    match(source, "^XmlWinEventLog:System"), "(SAMAccountName=* AND signature_id=4755)",
    match(sourcetype, "^XmlWinEventLog"), "earliest=1618349400 latest=1618363811",
    match(index, "^osnix"), "(uid IN (16255, 20755, 38169)",
    1==1, "")
| eval comment = if( 1==1, null(), "
specificFilterString is a way to pass in conditional filter strings to the first pipe of the searches executed by the map command, allowing filtering of the data to avoid heavier impact. ")
| map maxsearches=100 search="search (index=\"$index$\" sourcetype=\"$sourcetype$\" source=\"$source$\") [| makeresults | eval search = $specificFilterString$ | return $search]
    | fields - \"date_*\", host, linecount, punct, \"splunk_server*\", timestartpos, timeendpos, timestamp
    | fieldsummary maxvals=10
    | where (count > 0)
    | eval index = \"$index$\"
    | eval sourcetype = \"$sourcetype$\"
    | eval source = \"$source$\"
    "
```

index	sourcetype	source	count	specificFilterString
osnix	syslog	/var/log/cron	212	(uid IN (16255, 20755, 38169)
osnix	syslog	/var/log/dnf.log	546	(uid IN (16255, 20755, 38169)
osnix	syslog	/var/log/messages	11433	(uid IN (16255, 20755, 38169)
osnixsec	syslog	secure	188	(uid IN (16255, 20755, 38169)
oswin	XmlWinEventLog	XmlWinEventLog:Application	32	earliest=1618349400 latest=1618363811
oswin	XmlWinEventLog	XmlWinEventLog:System	187	(SAMAccountName=* AND signature_id=4755)
oswinsec	WinEventLog	WinEventLog:Security	202	(TargetUserName=* AND (SubjectUserName=* OR SubjectUserSid=*))

Fieldsummary

(BY *index, sourcetype, source* with event count limit)

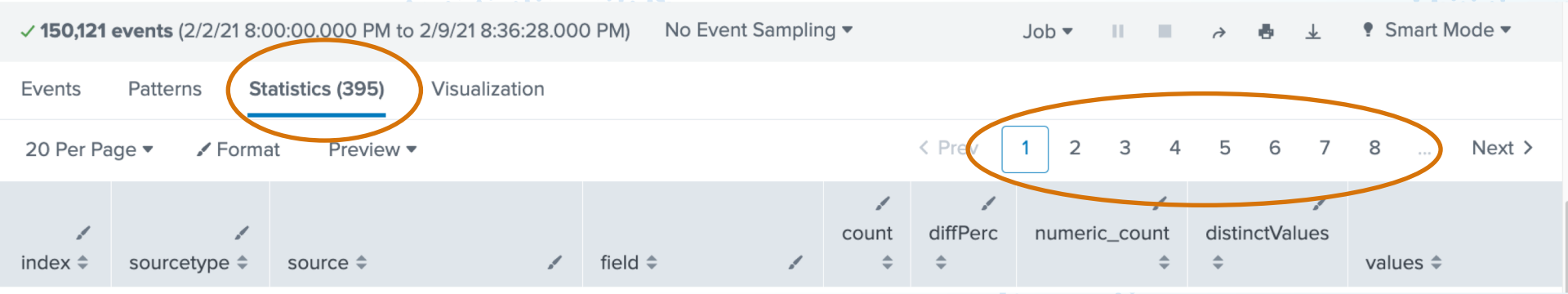
```
| tstats count WHERE index IN ("oswin*", "osnix*") BY index, sourcetype, source
| eval head_limit = case(
    match(source, "^WinEventLog:Security"), "1000000",
    match(source, "^XmlWinEventLog:System"), "50000",
    match(index, "^osnix"), "5000000",
    1==1, "100000000")
| fields - count
| map maxsearches=100 search="search (index=\"$index$\" sourcetype=\"$sourcetype$\" source=\"$source$\")
    | head $head_limit$
    | fields - \"date_*\", host, linecount, punct, \"splunk_server*\", timestartpos, timeendpos, timestamp
    | fieldsummary maxvals=10
    | where (count > 0)
    | eval index = \"$index$\"
    | eval sourcetype = \"$sourcetype$\"
    | eval source = \"$source$\"
    "
```

index	sourcetype	source	head_limit
osnix	syslog	/var/log/cron	5,000,000
osnix	syslog	/var/log/dnf.log	5,000,000
osnix	syslog	/var/log/messages	5,000,000
osnixsec	syslog	secure	5,000,000
oswin	XmlWinEventLog	XmlWinEventLog:Application	100,000,000
oswin	XmlWinEventLog	XmlWinEventLog:System	50,000
oswinsec	WinEventLog	WinEventLog:Security	1,000,000

splunk>

Too many results!
Running it over and over is expensive!

(fieldsummary by *index, sourcetype, source*)



If only there was a way to re-use the results…

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

| loadjob

| loadjob is a command that accesses search artifacts that are available on the Splunk search head you run it on by querying for a specific Search ID (sid).

(no replication of ad-hoc searches across SHC)

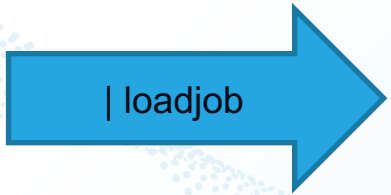
```
1 index=_internal sourcetype=splunkd
2 | stats count BY date_hour, color
```

✓ **1,461,911 events** (4/13/21 12:00:00.000 AM to 4/14/21 12:53:45.0

Events Patterns **Statistics (48)** Visualization

100 Per Page ▾ Format Preview ▾

date_hour ▲	color	count
0	green	3076
0	red	480
1	green	2320
1	red	480
2	green	2320
2	red	480
3	green	2320
3	red	480




```
1 | loadjob 1618361126.44930
2 | chart sum(count) AS count OVER date_hour BY color
```

✓ 24 results (4/13/21 12:00:00.000 AM to 4/14/21 12:56:31.000 AM) No

Events Patterns **Statistics (24)** Visualization

100 Per Page ▾ Format Preview ▾

date_hour	green	red
0	2957	480
1	2320	480
2	2320	480
3	2320	480
4	2320	480
5	2332	484
6	2320	480
7	2320	480

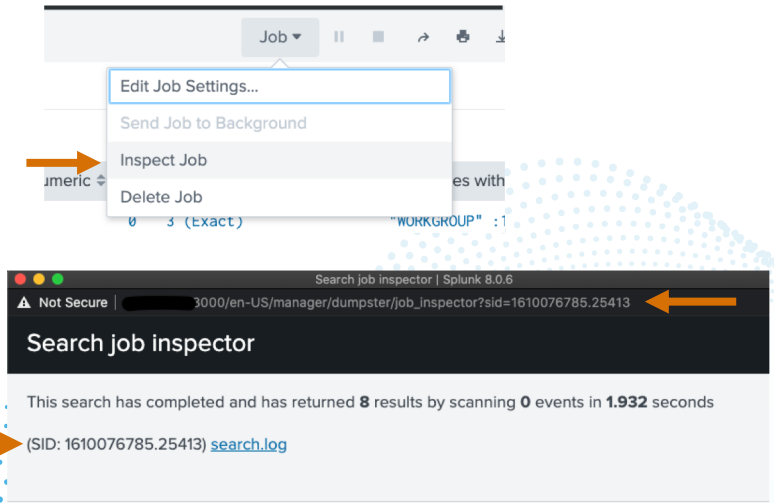
https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

| loadjob

Collecting the Search ID

You can find the SID of the search in the Job Inspector:



Or use *addinfo* to add the metadata to the search results: (macro-friendly!)

```
| makeresults
| fields - _time
| addinfo
| rename info_min_time AS search_earliest_time, info_max_time AS search_latest_time, info_search_time
    AS search_execution_time, info_sid AS sid
| foreach search_*
    [| eval <<FIELD>> = round('<<FIELD>>', 0)]
```

✓ 1 result (4/13/21 1:00:00.000 AM to 4/14/21 1:05:56.000 AM) Job

No Event Sampling

Events Patterns Statistics (1) Visualization

100 Per Page Format Preview

search_earliest_time	search_execution_time	search_latest_time	sid
1618275600	1618362357	1618362356	1618362356.45023

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

| loadjob

field	values	count	diffPerc	distinctValues	numeric_count	search_id
dest_is_expected	"false" :9224	9224	100.0%	1 (Exact)	0	1612886410.5255
dest_requires_av	"false" :9224	9224	100.0%	1 (Exact)	0	1612886410.5255
dest_should_timesync	"false" :9224	9224	100.0%	1 (Exact)	0	1612886410.5255
dest_should_update	"false" :9224	9224	100.0%	1 (Exact)	0	1612886410.5255

```
1 | loadjob 1612886410.5255
2 | eval valuesHash = md5(mvjoin(values, ":::"))
3 | eval checkSum = md5( count . ":::" . diffPerc . ":::" . numeric_count . ":::" . distinctValues . "
    :::" . valuesHash)
4 | eventstats values(field) AS field BY checkSum
5 | sort 0 - checkSum
6 | streamstats count AS duplicateCount BY checkSum
7 | where duplicateCount < 2
8 | table field, count, diffPerc, numeric_count, distinctValues, values
```

field	count	diffPerc	numeric_count	distinctValues	values
dest_is_expected dest_requires_av dest_should_timesync dest_should_update	9224	100.0%	0	1 (Exact)	"false" :9224

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

| loadjob

(Getting the most out of your results)

```
1 | loadjob 1618452906.50781
2 | search field="*name*"
```

✓ 13 results (4/14/21 2:00:00.000 AM to 4/15/21 2:16:09.000 AM) No Event Sampling

Events Patterns **Statistics (13)** Visualization

100 Per Page ▾ Format Preview ▾

index	sourcetype	source	field
osnix	syslog	/var/log/messages	name
oswin	XmlWinEventLog	XmlWinEventLog:System	Name
oswinsec	WinEventLog	WinEventLog:Security	ComputerName dvc
oswinsec	WinEventLog	WinEventLog:Security	LogName
oswinsec	WinEventLog	WinEventLog:Security	SourceName
oswinsec	WinEventLog	WinEventLog:Security	Account_Name

```
1 | loadjob 1618452906.50781
2 | search values="*Windows*"
```

✓ 17 results (4/14/21 2:00:00.000 AM to 4/15/21 2:18:01.000 AM) No Event Sampling

Events Patterns **Statistics (17)** Visualization

100 Per Page ▾ Format Preview ▾

index	sourcetype	source	field	count	diffPerc	distinctValues	values
oswin	XmlWinEventLog	XmlWinEventLog:System	System_Props_Xml	207	100.0%	25 (Estimate)	"<Provider Name='Event <Channel>System</Chan :1 "<Provider Name='Event <Channel>System</Chan :1 "<Provider Name='Micr <TimeCreated SystemTi UserID='S-1-5-20'/>" "<Provider Name='Micr <TimeCreated SystemTi UserID='S-1-5-18'/>" "<Provider Name='Micr <TimeCreated SystemTi UserID='S-1-5-18'/>"
oswin	XmlWinEventLog	XmlWinEventLog:System	EventData_Xml	207	100.0%	25 (Estimate)	"<Data Name='param1'>V "<Data Name='param1'>V "<Data Name='param1'>V "<Data Name='param1'>V "<Data Name='param1'>V

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Field Derivations

dataSource	endpoint_title	field_name	stanza	attribute	acl_app	value	conf_specific_properties_mv
transforms-extractions	WorkstationName_as_Source_Workstation		[WorkstationName_as_Source_Workstation]		Splunk_TA_windows	(.+)	FORMAT::::Source_Workstation::$1 CAN_OPTIMIZE::::1 CLEAN_KEYS::::1 SOURCE_KEY::::WorkstationName WRITE_META::::False KEEP_EMPTY_VALS::::0 LOOKAHEAD::::4096 MATCH_LIMIT::::100000 MV_ADD::::0
transforms-extractions	Workstation_Name_as_src_nt_host		[Workstation_Name_as_src_nt_host]		Splunk_TA_windows	(?:[\\]+)?([^~].*)	FORMAT::::src_nt_host::"$1" CAN_OPTIMIZE::::1 CLEAN_KEYS::::1 SOURCE_KEY::::Workstation_Name WRITE_META::::False KEEP_EMPTY_VALS::::0 LOOKAHEAD::::4096 MATCH_LIMIT::::100000 MV_ADD::::0
props-lookups (automatic lookups)	xmlwineventlog : LOOKUP-1severity_for_windows	severity	[xmlwineventlog]	LOOKUP-1severity_for_windows	Splunk_TA_windows	windows_severity_lookup Type OUTPUTNEW severity	overwrite::::0 transforms_stanza::::windows_severity_lookup lookup_input::::Type lookup_output::::severity
props-lookups (automatic lookups)	XmlWinEventLog : LOOKUP-1severity_for_windows	severity	[XmlWinEventLog]	LOOKUP-1severity_for_windows	Splunk_TA_windows	windows_severity_lookup Type OUTPUTNEW severity	overwrite::::0 transforms_stanza::::windows_severity_lookup lookup_input::::Type lookup_output::::severity
props-fieldaliases	xmlwineventlog : FIELDALIAS-user_id_for_windows	user_id	[xmlwineventlog]	FIELDALIAS-user_id_for_windows	Splunk_TA_windows	UserID AS user_id	alias.UserID::::user_id
props-fieldaliases	XmlWinEventLog : FIELDALIAS-user_id_for_windows	user_id	[XmlWinEventLog]	FIELDALIAS-user_id_for_windows	Splunk_TA_windows	UserID AS user_id	alias.UserID::::user_id
props-extractions	xmlwineventlog : REPORT-Sub_Status_from_xml	Sub_Status_from_xml	[xmlwineventlog]	REPORT-Sub_Status_from_xml	Splunk_TA_windows	SubStatus_as_Sub_Status	
props-extractions	XmlWinEventLog : REPORT-Sub_Status_from_xml	Sub_Status_from_xml	[XmlWinEventLog]	REPORT-Sub_Status_from_xml	Splunk_TA_windows	SubStatus_as_Sub_Status	

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

https://conf.splunk.com/files/2019/slides/FN1061.pdf
"Lesser Known Search Commands"

rest

The rest command reads a Splunk REST API endpoint and returns the resource data as a search result.[1]

- MUST be the first search command in a search block
- Is "time agnostic" - It only queries - so time is not a factor in execution
- Limits results to what the requesting user is allowed to access

```
| rest /services/data/indexes splunk_server=local count=0
| dedup title
| fields title
```

title
_audit
_internal
_introspection
_thefishbucket
apps
firedalerts
history
httpd
main
minecraft
minecraft_madscience

[1] https://docs.splunk.com/Documentation/Splunk/latest/SearchReference/Rest

Field Derivations

“Where does this field come from?”

- Accounting for props, transforms, calculated fields, lookups, it can all get very complex when tracing where a field comes from.
- Let’s leverage some REST endpoints to make this easier!

- Gathers field object information and combines it together into a single view
  - Calculated Fields / props.conf / transforms.conf / Aliases / Lookups
  - Outputs final field set of common elements like title, field name, stanza, app, etc, then stores all conf-specific configurations/properties in the ***conf_specific_properties_mv*** field

```
| rest /services/data/props/calcfields
| dedup id
| rename field.name AS field_name, eai:acl.* AS acl_*
| table dataSource, title, field_name, stanza, type, attribute, acl_app, acl_owner, acl_sharing, value
| eval dataSource = "props-calcfields"
| append
    [| rest /services/data/props/extractions
    | dedup id
    | rename eai:acl.* AS acl_*
    | rex field=attribute "^[^-]+-(?<field_name>.+)"
    | table dataSource, title, field_name, stanza, type, attribute, acl_app, acl_owner, acl_sharing, value
    | eval dataSource = "props-extractions"]
| append
    [| rest /services/data/props/fieldaliases
    | dedup id
    | foreach alias.*
        [| eval conf_specific_properties_mv = case(
            isnull(conf_specific_properties_mv) AND isnotnull('<<FIELD>>'), "<<FIELD>>:::" . '<<FIELD>>',
            isnotnull(conf_specific_properties_mv) AND isnotnull('<<FIELD>>'), mvappend(conf_specific_properties_mv, "<<FIELD>>:::" . '<<FIELD>>'),
            isnotnull(conf_specific_properties_mv) AND isnull('<<FIELD>>'), conf_specific_properties_mv
            )]
    | rename eai:acl.* AS acl_*
    | rex field=attribute "^[^-]+-(?<field_name>.+)"
    | table dataSource, title, field_name, stanza, type, attribute, acl_app, acl_owner, acl_sharing, value, conf_specific_properties_mv
    | eval dataSource = "props-fieldaliases"]
| append
    [| rest /services/data/transforms/extractions
```

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Field Derivations Query

(A note on the methodology)

```
| makeresults | where 1==2
| append
    [| rest splunk_server=* /servicesNS/-/-/data/props/fieldaliases
    | dedup id
    | foreach alias.*
       [| eval conf_specific_properties_mv = case(
           isnull(conf_specific_properties_mv) AND isnotnull('<<FIELD>>'), "<<FIELD>>:::" . '<<FIELD>>',
           isnotnull(conf_specific_properties_mv) AND isnotnull('<<FIELD>>'), mvappend(conf_specific_properties_mv, "<<FIELD>>:::" . '<<FIELD>>'),
           isnotnull(conf_specific_properties_mv) AND isnull('<<FIELD>>'), conf_specific_properties_mv
           )]
    | rename eai:acl.* AS acl_*
    | rex field=attribute "^[^-]+-(?<field_name>.+)"
    | table dataSource, title, field_name, stanza, type, attribute, acl_app, acl_owner, acl_sharing, value, conf_specific_properties_mv
    | eval dataSource = "props-fieldaliases"]
| append
    [| rest splunk_server=* /servicesNS/-/-/data/transforms/extractions
    | dedup id
    | rename eai:acl.* AS acl_*, REGEX AS value
    | foreach *
       [| eval <<FIELD>> = if( match('<<FIELD>>', "."), '<<FIELD>>', null() )]
    | foreach FORMAT, CAN_OPTIMIZE, CLEAN_KEYS, DEFAULT_VALUE, SOURCE_KEY, DEST_KEY, WRITE_META, DELIMS, FIELDS, KEEP_EMPTY_VALS, LOOKAHEAD,
      MATCH_LIMIT, MV_ADD, REPEAT_MATCH
       [| eval conf_specific_properties_mv = case(isnull(conf_specific_properties_mv) AND isnotnull('<<FIELD>>'), "<<FIELD>>:::" . '<<FIELD>>',
           isnotnull(conf_specific_properties_mv) AND isnotnull('<<FIELD>>'), mvappend(conf_specific_properties_mv, "<<FIELD>>:::" . '<<FIELD>>'),
           isnotnull(conf_specific_properties_mv) AND isnull('<<FIELD>>'), conf_specific_properties_mv)]
    | eval stanza = title
    | table dataSource, title, stanza, acl_app, acl_owner, acl_sharing, value, conf_specific_properties_mv
    | eval dataSource = "transforms-extractions"]
| table dataSource, field_name, stanza, type, attribute, acl_app, acl_owner, acl_sharing, title, value, conf_specific_properties_mv
```

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Field Derivations Query

(Using | *union* in Splunk 8+)

```
| union
    [| rest /services/data/props/calcfields
    | dedup id
    | rename field.name AS field_name, eai:acl.* AS acl_*
    | table dataSource, title, field_name, stanza, type, attribute, acl_app, acl_owner, acl_sharing, value
    | eval dataSource = "props-calcfields"]

    [| rest /services/data/props/extractions
    | dedup id
    | rename eai:acl.* AS acl_*
    | rex field=attribute "^[^-]+-(?<field_name>.+)"
    | table dataSource, title, field_name, stanza, type, attribute, acl_app, acl_owner, acl_sharing, value
    | eval dataSource = "props-extractions"]

    [| rest /services/data/props/fieldaliases
    | dedup id
    | foreach alias.*
        [| eval conf_specific_properties_mv = case(
            isnull(conf_specific_properties_mv) AND isnotnull('<<FIELD>>'), "<<FIELD>>:::" . '<<FIELD>>',
            isnotnull(conf_specific_properties_mv) AND isnotnull('<<FIELD>>'), mvappend(conf_specific_properties_mv, "<<FIELD>>:::" . '<<FIELD>>'),
            isnotnull(conf_specific_properties_mv) AND isnull('<<FIELD>>'), conf_specific_properties_mv
            )]
    | rename eai:acl.* AS acl_*
    | rex field=attribute "^[^-]+-(?<field_name>.+)"
    | table dataSource, title, field_name, stanza, type, attribute, acl_app, acl_owner, acl_sharing, value, conf_specific_properties_mv
    | eval dataSource = "props-fieldaliases"]

    [| rest /services/data/transforms/extractions
    | dedup id
    | rename eai:acl.* AS acl_*, REGEX AS value
    | foreach *
        [| eval <<FIELD>> = if( match('<<FIELD>>', "."), '<<FIELD>>', null() )]
    | foreach FORMAT, CAN_OPTIMIZE, CLEAN_KEYS, DEFAULT_VALUE, SOURCE_KEY, DEST_KEY, WRITE_META, DELIMS, FIELDS, KEEP_EMPTY_VALS, LOOKAHEAD,
        MATCH_LIMIT, MV_ADD, REPEAT_MATCH
        [| eval conf_specific_properties_mv = case(isnull(conf_specific_properties_mv) AND isnotnull('<<FIELD>>'), "<<FIELD>>:::" . '<<FIELD>>',
            isnotnull(conf_specific_properties_mv) AND isnotnull('<<FIELD>>'), mvappend(conf_specific_properties_mv, "<<FIELD>>:::" . '<<FIELD>>'),
            isnotnull(conf_specific_properties_mv) AND isnull('<<FIELD>>'), conf_specific_properties_mv)
            ]
    | eval stanza = title
    | table dataSource, title, stanza, acl_app, acl_owner, acl_sharing, value, conf_specific_properties_mv
    | eval dataSource = "transforms-extractions"]
```

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Configured Datamodel Fields

(bonus)

field	inheritance	datamodel	object
user	All_Changes.Endpoint_Changes	Change	Endpoint_Changes
vendor_product	All_Changes.Endpoint_Changes	Change	Endpoint_Changes
action	All_Changes.Endpoint_Changes	Change	Endpoint_Changes
_time	All_Changes.Endpoint_Changes.Endpoint_Restarts	Change	Endpoint_Restarts
_raw	All_Changes.Endpoint_Changes.Endpoint_Restarts	Change	Endpoint_Restarts
source	All_Changes.Endpoint_Changes.Endpoint_Restarts	Change	Endpoint_Restarts
sourcetype	All_Changes.Endpoint_Changes.Endpoint_Restarts	Change	Endpoint_Restarts
host	All_Changes.Endpoint_Changes.Endpoint_Restarts	Change	Endpoint_Restarts
dest_bunit	All_Changes.Endpoint_Changes.Endpoint_Restarts	Change	Endpoint_Restarts
dest_category	All_Changes.Endpoint_Changes.Endpoint_Restarts	Change	Endpoint_Restarts
dest_priority	All_Changes.Endpoint_Changes.Endpoint_Restarts	Change	Endpoint_Restarts
src_bunit	All_Changes.Endpoint_Changes.Endpoint_Restarts	Change	Endpoint_Restarts
src_category	All_Changes.Endpoint_Changes.Endpoint_Restarts	Change	Endpoint_Restarts
src_priority	All_Changes.Endpoint_Changes.Endpoint_Restarts	Change	Endpoint_Restarts

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Configured Datamodel Fields

(bonus)

```
| datamodelsimple type="models"
| map maxsearches=1000 search="| datamodelsimple type=objects datamodel=$datamodel$
    | eval datamodel=\"$datamodel$\""
| map maxsearches=1000 search="| datamodelsimple type=attributes datamodel=$datamodel$ object=$object$ nodename=$lineage$
    | eval datamodel=\"$datamodel$\"
    | eval object=\"$object$\"
    | eval lineage=\"$lineage$\""
| eval comment = if(1==1, null(), "
These map commands have a seemingly extreme maxsearches setting, but be assured that the datamodelsimple command reads the
    internal configuration of the SH host and doesn't require much compute.

This search is running to compile a breakout of all datamodels, objects, and fields configured in the environment.")
| rename attribute AS field, lineage AS inheritance
```

https://github.com/TheWoodRanger/splunk_fields_analysis_presentation

BSIDES splunk>

Additional Reading

"I stand on the shoulders of giants."

There are some incredible resources, examples, and usecases available in these presentations.

- The "Gotchas" of Splunk (Users Beware!)
  (https://splunkcommunity.com/wp-content/uploads/2019/11/Splunk.conf19-Gotchas.pdf)
- Security Ninjutsu Part 4
  (http://conf.splunk.com/files/2017/slides/security-ninjutsu-part-four-attackers-be-gone-in-45-minutes-of-epic-spl.pdf)
- Security Ninjutsu Part 6
  (https://www.davidveuve.com/splunk.html#ninjutsupartsix)
- Tricks for better SPL (SPLendid uses for SPL in SPLunk)
  (https://conf.splunk.com/files/2019/slides/FN1300.pdf)
- Lesser Known Search Commands
  (https://conf.splunk.com/files/2019/slides/FN1061.pdf)
- Master Joining Datasets without Using Join
  (https://conf.splunk.com/files/2020/slides/TRU1761C.pdf)
- Turning Security Use Cases into SPL
  (https://static.rainfocus.com/splunk/splunkconf18/sess/1523489574149001lr6z/finalPDF/SEC1583_TurningSecurityUseCases_Final_1538510573435001VmSq.pdf)

Thanks for attending!

Any questions?

You can find me at

- **Splunk UserGroups Slack: TheWoodRanger**
- **Email: ryan.wood@guidepointsecurity.com**
- **#BSides21-FieldAnalysis**