Best Practices for Developing Splunk Apps and Add-ons



Disclaimer

During the course of this presentation, we may make forward looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC. The forwardlooking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not, be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

whoami

- jason.conger@splunk.com
- @JasonConger
- in http://www.linkedin.com/in/JasonConger
- blogs.splunk.com/author/jconger/ www.JasonConger.com

4+ years at Splunk

Created or consulted on numerous Splunkbase applications



Staff Solutions Architect Global Strategic Alliances

Agenda

- 1. Creating a Splunk Application
- 2. Getting data into Splunk
- 3. Asking questions of your data with Splunk

"I wish I knew these things before I ever built my first Splunk Application"

- Jason Conger





1. Good of the state of the sta





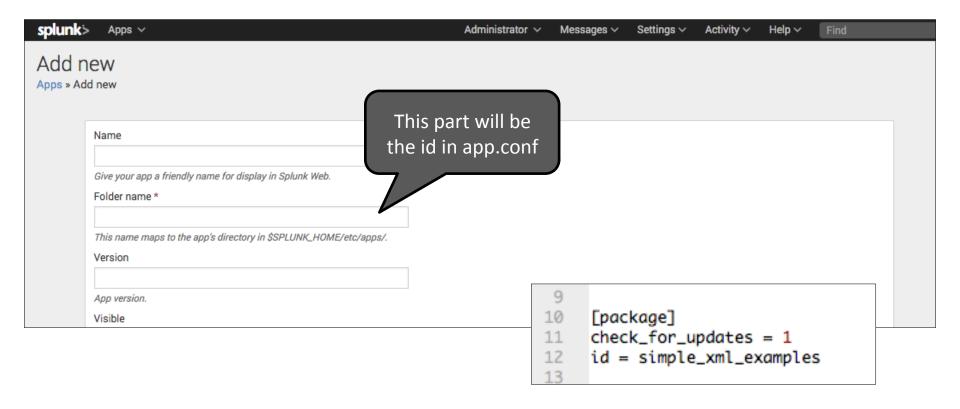
To the size you of the edge, which this has been a second as a second and a second a second and a second a second and a second a second and a second and a second a second and a second a second a second and a second a second and a second a second a second and a second a second and a second a sec







Naming the <u>Directory</u> for Your App or Add-on



Naming the <u>Directory</u> for Your App or Add-on

- ☐ For applications (dashboards, forms, alerts, etc.):
 - Vendor-app-product (example = acme-app-widget)
- ☐ For add-ons (data collection with no dashboards):
 - TA_vendor-product (example: TA_acme-widget)
- ☐ For Enterprise Security add-ons:
 - TA-<datasource> (example: TA-snort)

Note: you may see some other naming standards such as SA or DA out there.

Naming Your App or Add-on



Note: after uploading an application to Splunkbase, the directory name and the "id" parameter in app.conf cannot be changed.

The actual name of the application displayed on the Splunk start screen and on Splunkbase is controlled by a file named app.conf and is independent of the directory name mentioned previously.

App naming guidelines -> http://docs.splunk.com/Documentation/Splunkbase/latest/Splunkbase/Namingguidelines

Should You Break Up Your App?



Consolidated App



Distributed App

Should You Break Up Your App?

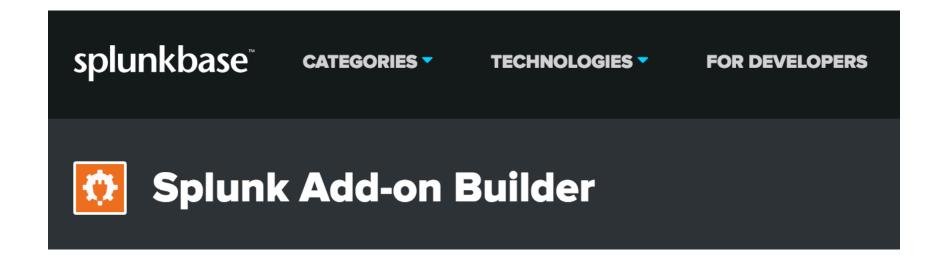
- Do you need to collect data from forwarders?
- Need to share knowledge objects with multiple apps?
- Distributed Environment?
- The Splunk App for AWS is a good example



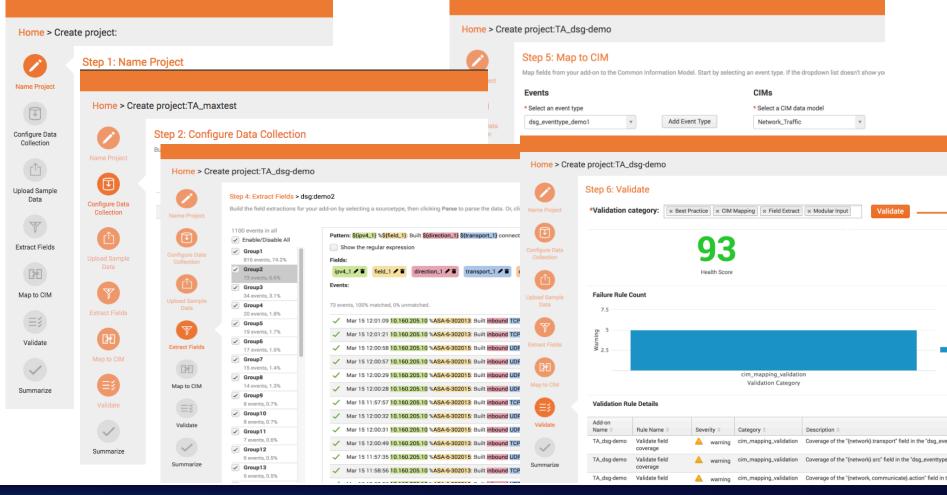
depends on



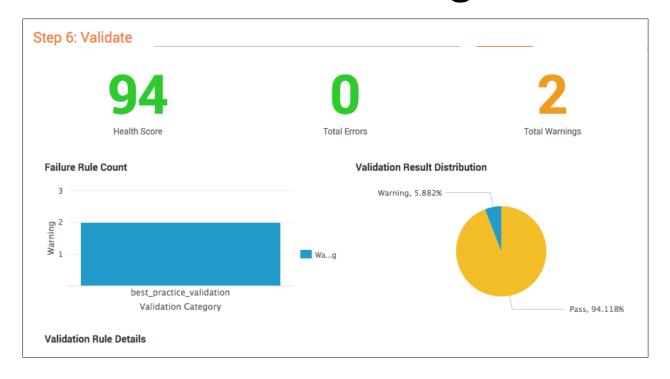
Quick Start = Splunk Add-on Builder



https://splunkbase.splunk.com/app/2962/



Use the Builder on Existing Content Too



Note: you may get some inapplicable warnings for apps since this version is mainly about add-ons.

Getting Data In .conf2016 splunk 3

Getting Data In

Reaching out to get data	Listening for data
 Reading files on a disk Windows Inputs Perfmon Event Logs Registry WMI Scripts* Modular inputs* 	 TCP UDP HTTP Stream Scripts* Modular inputs*

^{*} Scripts and modular inputs can really do either depending on what you code

Best Practices for Logging Data to be Consumed by Splunk

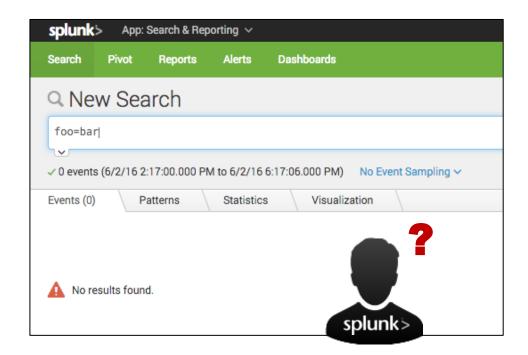
- Log in text format
- Start the log line event with a time stamp
- Use clear key-value pairs
- Create events that humans can read
- Use unique identifiers
- Keep multi-line events to a minimum
- Use JSON (JavaScript Object Notation) format

http://dev.splunk.com/view/logging-best-practices/SP-CAAADP6

Best Practices for Writing Data to an Index

Write to the default "main" index

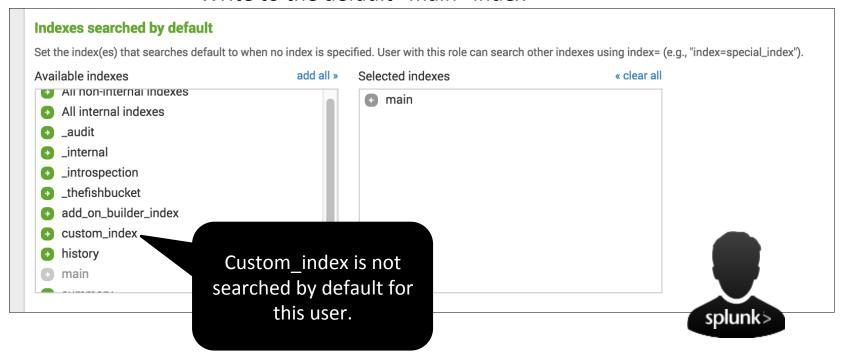


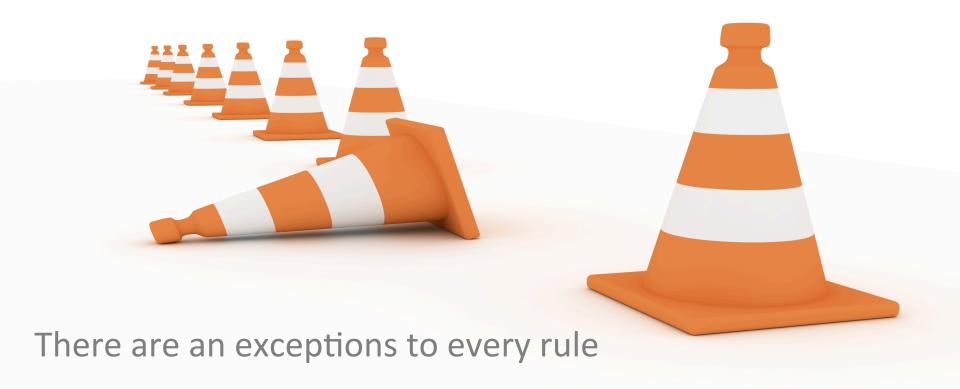




Best Practices for Writing Data to an Index

Write to the default "main" index



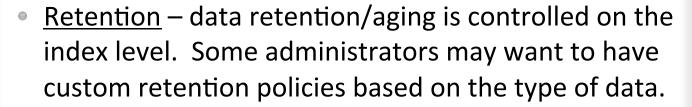


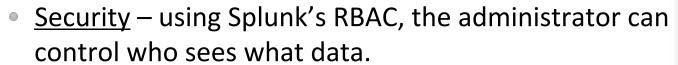
Exceptions to using the "main" Index

- Testing writing data to a test index during development allows the developer to quickly and easily clear out all events in the index without impacting other events elsewhere.
 \$SPLUNK HOME/bin/splunk clean eventdata custom index
- <u>Retention</u> data retention/aging is controlled on the index level.
 Some administrators may want to have custom retention policies based on the type of data.
- <u>Security</u> using Splunk's RBAC, the administrator can control who sees what data.

Exceptions to using the "main" Index

 Testing – writing data to a test index during development allows the developer to quickly and easily clear out all events in the index without impacting other events elsewhere

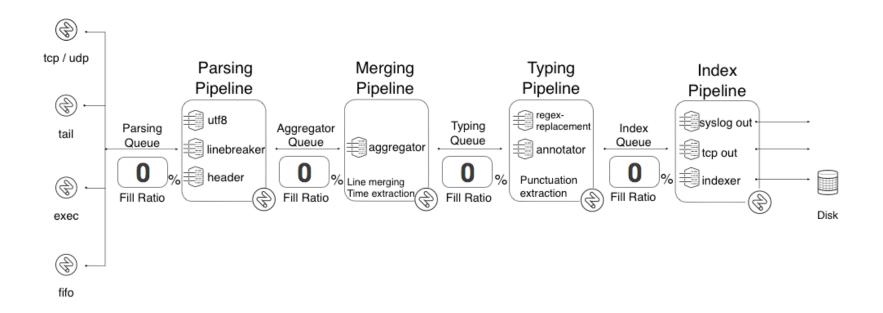






The last 2
exception
decisions
should be made
by the Splunk
admin – not the
developer.

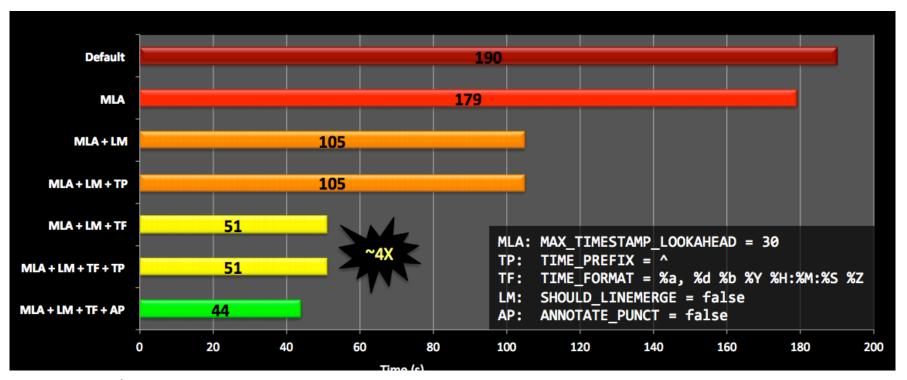
Get to Know Your Pipelines



Useful Index Time Processing Attributes

```
LINE BREAKER
Event
                            <where to break the stream>
Breaking
          SHOULD LINEMERGE <enable/disable merging>
          MAX TIMESTAMP_LOOKAHEAD <# chars in to look for ts>
Timestamp
          TIME PREFIX
                          <pattern before ts>
Extraction
          TIME FORMAT
                          <strptime format string to extract ts>
          ANNOTATE PUNCT <enable/disable punct:: extraction>
Typing
```

Useful Index Time Processing Attributes



HT: Dritan Bitincka

Adding Inputs



Scripted versus Modular Inputs

Feature	Scripted Inputs	Modular Inputs
End user configuration	Often requires editing text configuration files	User interface provided in the Splunk Web interface. This makes the input "look and feel" as if it were a native Splunk feature.
Multi-platform support	No	Yes You can package your script to include versions for separate platforms.
Custom REST endpoints	No	Yes Modular inputs can be access and manipulated using Splunk REST endpoints.
Endpoint permissions	N/A	Access implemented using Splunk Enterprise capabilities.

More complete information can be found on the Splunk documentation page.

Scripted versus Modular Inputs

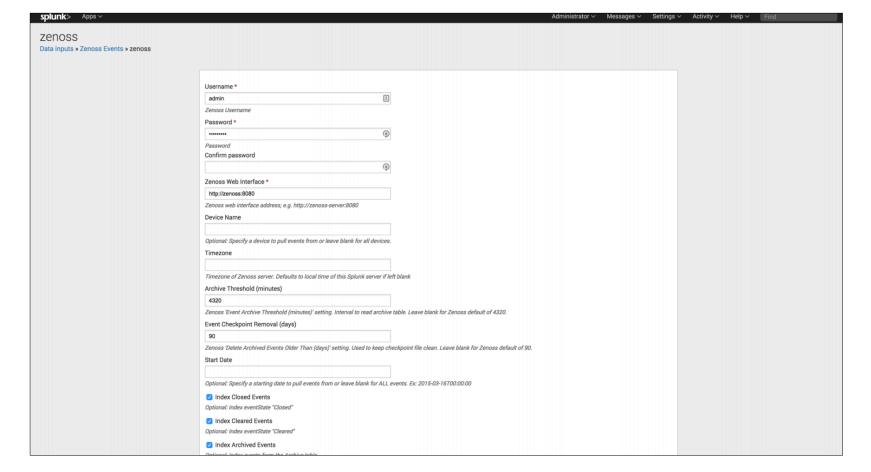
<u>Scripted inputs</u> are more suited for trivial tasks such as running an OS command (like **top** for *nix or **Get-Process** from Windows PowerShell) and sending the output to Splunk.

Modular inputs are more suited for tasks that require end user setup or more advanced event processing. Calling a REST API with parameters is a good example of when to use a modular input.

This or...

```
script://./bin/zenoss_wrapper.sh -u admin -p password -a h8p://
zenoss:8080 -z America/Los_Angeles -t 4320 -r 90 -s
2015-03-16T00:00:00 -index-closed-events 1 -index-cleared-events 1 -
index-archived-events 1 -index-suppressed-events 1 -index-
repeatevents 1]
```

```
sourcetype = zenoss-events
interval = 60
index = zenoss
```



HT: Scott Haskell

Scripted and Modular Input Best Practices

Do not hard code paths

Example (Python):

os.path.join(os.environ["SPLUNK_HOME"],'etc','apps', APP_NAME)

Example (PowerShell):

Join-Path -path (get-item env:\SPLUNK_HOME).value "Splunk\etc \apps"

Scripted and Modular Input Best Practices

<u>Use Error Trapping</u> (so that you can search them in the _internal index)

```
import logging
try:
     Some code that may fail like opening a
file
except IOError, err:
     logging.error('%s - ERROR - File may not
exist %s\n' % (time.strftime("%Y-%m-%d %H:%M:
%S"), str(err)))
     pass
```

Scripted and Modular Input Best Practices

Error Trapping (you can use stderr too)

```
try:
     Some code that may fail like opening a
file
except IOError, err:
     sys.stderr.write('%s - ERROR - File may
not exist %s\n' % (time.strftime("%Y-%m-%d %H:
%M:%S"), str(err)))
     pass
```

Scripted and Modular Input Best Practices

Use Splunk methods to read cascaded settings

```
Example (Python):
```

```
import splunk.clilib.cli_common

def __init__(self,obj):
    self.object = obj
    self.settings =
splunk.clilib.cli_common.getConfStanza("acme","default")
```

- Give more explanation on previous slide
- Mention someone trying to read from default and write to local
- Maybe mention btool too

Scripted and Modular Input Best Practices

Disable any inputs by default

inputs.conf:

```
[my_stanza]
disabled = 1
```

Scripted Inputs Best Practices

Test Scripts using Splunk CMD

Mac:

```
/Applications/Splunk/bin/splunk cmd python /Applications/Splunk/etc/apps/<your app>/bin/<your script>
```

Windows:

```
C:\Program Files\Splunk\bin\splunk.exe cmd C:\Program Files
\Splunk\etc\apps\<your app>\bin\<your script>
```

<u>Use Splunk SDKs</u> (these abstract a lot of code for you)

Python http://dev.splunk.com/view/python-sdk/SP-CAAAER3

C# http://dev.splunk.com/view/csharp-sdk/SP-CAAAEQH

Java http://dev.splunk.com/view/java-sdk/SP-CAAAER2

Modular Input SDKs

Before = 453 lines

```
438
           return val_data
439
       if __name__ == '__main__':
440
441
442
           if len(sys.argv) > 1:
               if sys.argv[1] == "--scheme":
443
444
                   do_scheme()
               elif sys.argv[1] == "--validate-ar
445
446
                   do_validate()
447
               else:
448
                   usage()
449
           else:
450
               do_run()
451
452
           sys.exit(0)
```

After = 92 lines

```
except Exception as e:
raise e

if __name__ == "__main__":
exitcode = MyScript().run(sys.argv)
sys.exit(exitcode)
```

Modular Input SDK Logging

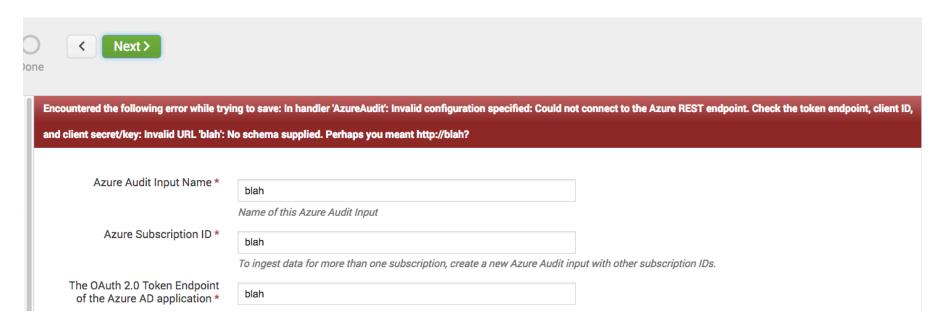
By default, only INFO and higher events are logged to _internal.



Validate User Input

```
# Try to connect to the Azure API to validate the given arguments
work
try:
    access token = get token from client credentials (
        endpoint = val data["token endpoint"],
        client id = val data["client id"],
        client secret = val data["client secret"])
except Exception, e:
    raise Exception, "Could not connect to the Azure REST endpoint.
Check the token endpoint, client ID, and client secret/key: %s" %
str(e)
```

Validate User Input



Test Inputs using Splunk CMD

Example (real):

```
/Applications/Splunk/bin/splunk cmd splunkd print-modinput-config AzureDiagnostics AzureDiagnostics://gsa1892 | / Applications/Splunk/bin/splunk cmd python /Applications/Splunk/etc/apps/TA Azure/bin/AzureDiagnostics.py
```

Test Inputs using Splunk CMD

Name of the input

Instance of the input

Example (real):

```
/Applications/Sprenk/bin/splunk cmd lunkd print-modinput-config AzureDiagnostics AzureDiagnostics://gsa1892 | / Applications/Splunk/bin/splunk cmd python /Applications/Splunk/etc/apps/TA Azure/bin/AzureDiagnostics.py
```

Input code

Use the checkpoint parameter to persist data



1st run: position = 0

2nd run: position = 5

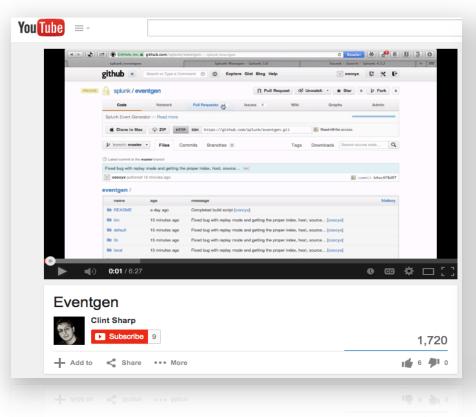
3rd run: position = 10

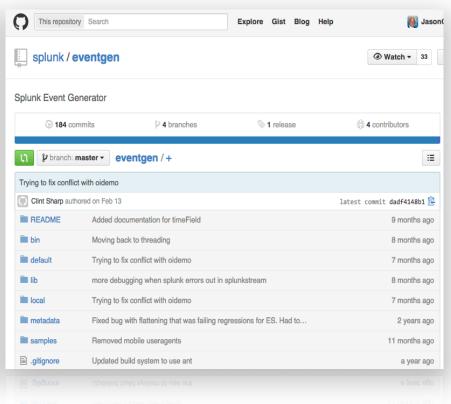
 N^{th} run: position = x

1st run: returned 0 – 4 2nd run: returned 5 – 9 3rd run: returned 10 - x

http://blogs.splunk.com/2016/05/11/splunking-continuous-rest-data/

Eventgen





Anonymize Eventgen Samples

Regex Find and Replace Tools are Your Friend!

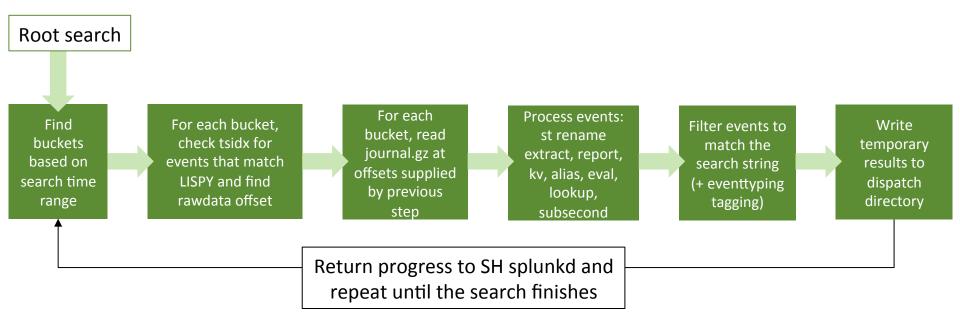




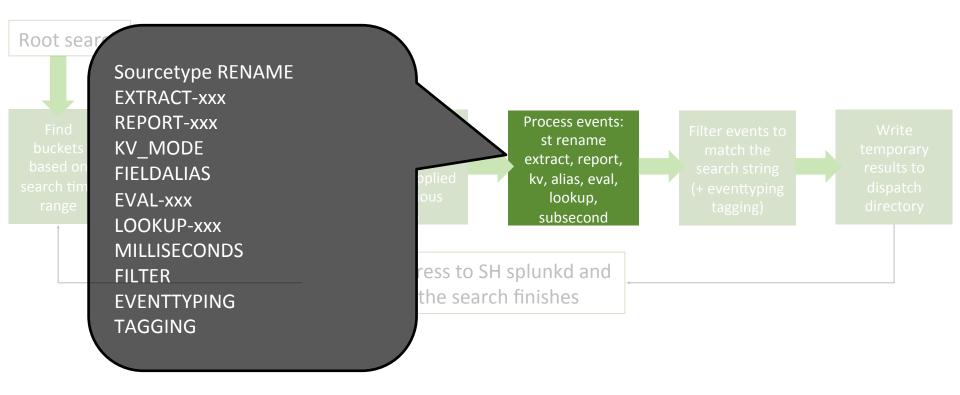


Asking Questions of your Data .conf2016 splunk⁵₅₁

Get to Know the Search Pipeline



Get to Know the Order of Operations



Parameterize Root Searches

marcros.conf example:

```
[acme_index]
definition = index=acme
```

Example search using macro:

```
`acme_index` sourcetype=widiget |
stats count
```

Remember that main index thing earlier?



Get to Know Distributed Search

macros.conf

[my_index]
definition = index=main

eventtypes.conf

[my_eventtype]
search = `my_index` sourcetype="foo"

Example search: eventtype=my_eventtype | stats count

This will not work in a distributed environment by default.

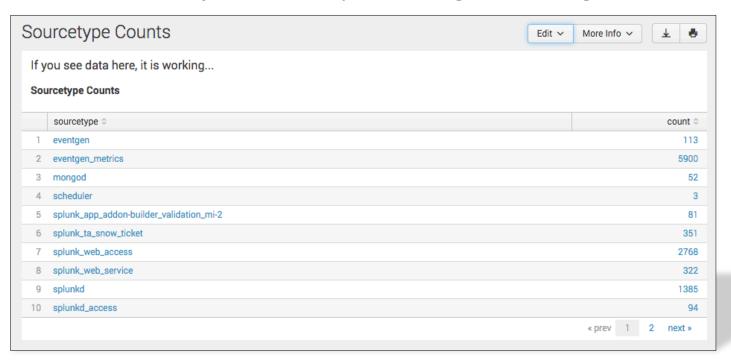
Get to Know the Big Book of Search

www.bbosearch.com

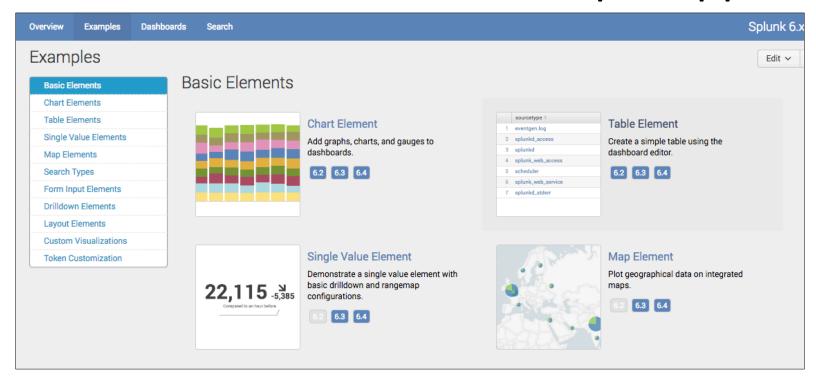
```
Create List Pretty SPL About
                                                                                                                          jconger@splunk.com (logout)
Speed / Distance Login Anomaly
   iplocation clientip
    sort _time
    strcat lat "," lon latlon
    streamstats current=f global=f window=1 last(latlon) as last_latlon
    eval last_latlon=if(isnull(last_latlon), latlon, last_latlon)
    streamstats current=f global=f window=1 last( time) as last ts
    eval time_since_last = _time - last_ts
    eval time_since_last=if(isnull(time_since_last), 0, time_since_last)
    haversine originField=last_latlon outputField=distance units=mi latlon
    eval speed=if(time_since_last=0, 0, (distance/(time_since_last/60/60)))
    where speed > 500
    strcat speed " MPH" speed
    table user, distance, _time, time_since_last, speed, _raw
    toggle comments copy w/out comments edit
Find those tuples of events where the speed needed to cover distance in time between events is greater than 500MPH
requirements:
haversine app clientip
comments:
apburdell
table (docs) where (docs) sort (docs) eval (docs) iplocation (docs) streat (docs) haversine (docs) streamstats (docs)
```

Include Prebuilt Panels

Even if it just to verify the thing is working

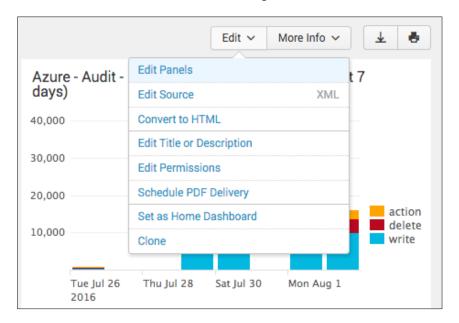


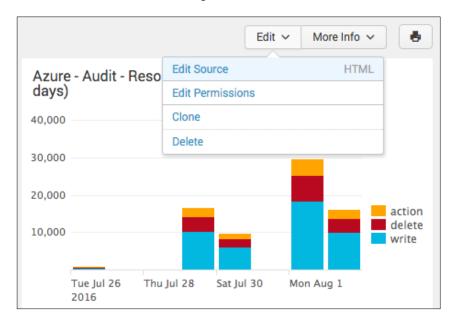
Use the Dashboards Example App



https://splunkbase.splunk.com/app/1603/

Use Simple XML as much as possible



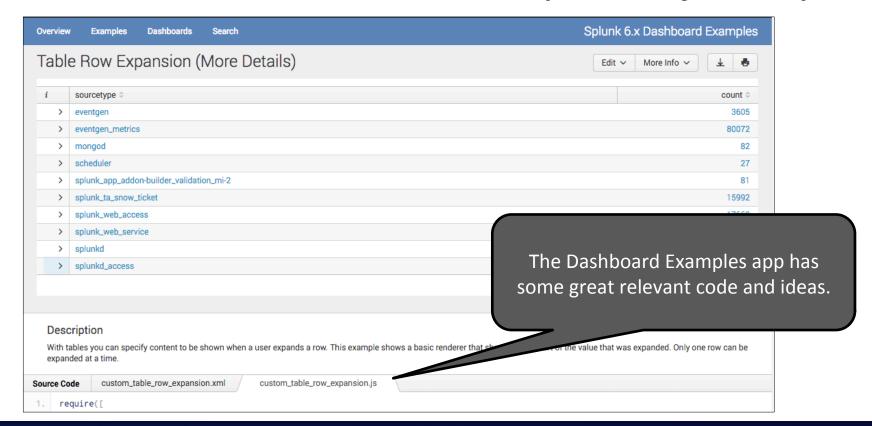


Simple XML HTML

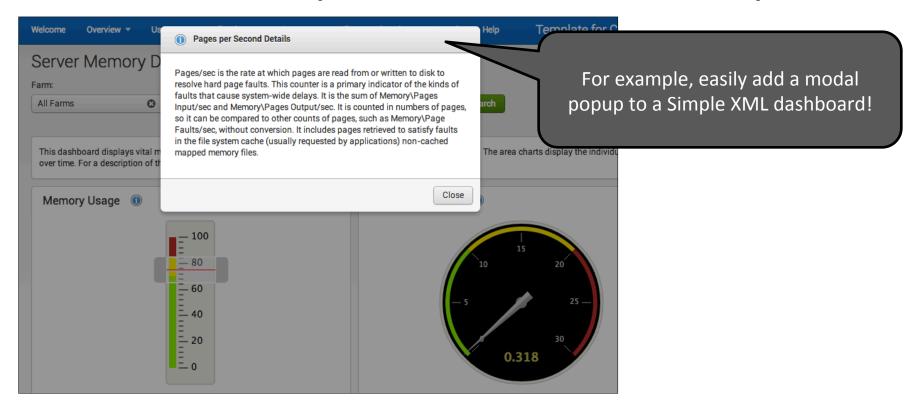
Use Simple XML as much as possible



Get to know JavaScript and jQuery



Bootstrap can add functionality

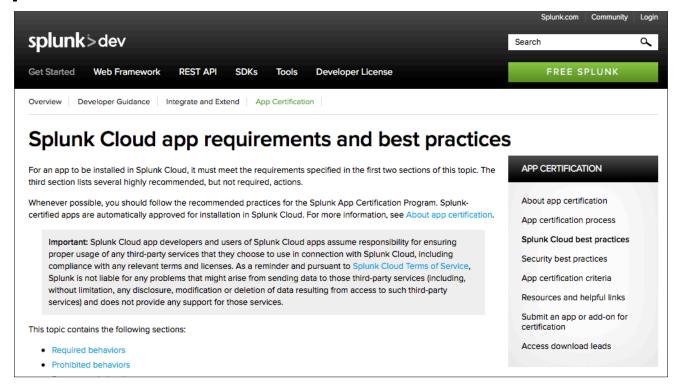


Get to know CSS

- All Splunk elements have an id now
- Check out Firefox's 3D view for layering



Splunk Cloud has Best Practices too



http://dev.splunk.com/view/app-cert/SP-CAAAE85

Do's and Don'ts - Packaging Applications

Do	Don't
Follow the guidelines found at	Leave any hidden files in the app such as Mac's files.
Include a screen shot of your application in the correct location.	
Let the user choose which inputs are enabled for their environment.	Enable all inputs by default if not necessary.
Use a build automation tool such as Apache Ant if necessary to ensure a clean build/package.	Leave anything in: \$SPLUNK_HOME/etc/apps/ <app>/local directory \$SPLUNK_HOME/etc/apps/<app>/metadata/local.meta</app></app>
Ensure the appropriate settings are set in app.conf	
Document your app with a README.txt file	
Test your application on a clean system	

Do's and Don'ts - Data Collection

Do	Don't
Support multiple platforms.	Code for a single OS.
Use scripting language utilities such as os.path.join() and the special environment variable \$SPLUNK_HOME to construct paths in scripts.	Hard code script paths.
Write data to the "main" index. This ensures that your data is searchable by default.	Hard code index names in searches if you must use a custom index.
Use key=value pairs in writing to log files (if you have control of the logging output).	Use name abbreviations.
Throttle how much data is collected at one time from an API.	Overwhelm a system by pulling exorbitant amounts of data at one time from an API.
Use logging and error trapping in scripts and inputs.	

Do's and Don'ts - Applications

Do	Don't
Use setup.xml or a mod input configs to allow the end user to configure the app	Make users manually enter information such as API credentials into configuration files.
Encrypt user input passwords.	Store clear text passwords in .conf files.
Parameterize indexes so that they can be easily changed	Hard code indexes in your searches
Use the CIM add-on	
Place all .conf files in default	Leave any content in
\$SPLUNK_HOME/etc/apps/ <your_app>/default</your_app>	\$SPLUNK_HOME/etc/apps/ <your_app>/local</your_app>
Set default permissions in: \$SPLUNK_HOME/etc/apps/ <your_app>/metadata/default.meta</your_app>	Have a local.meta file located in: \$SPLUNK_HOME/etc/apps/ <your_app>/metadata</your_app>

THANK YOU

