Data On-Boarding

Andrew Duca

Principal Services Consultant

Splunk

splunk>

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 forward-looking 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.

About Me

- Principal Professional Services Consultant based in Boston, MA
- 15+ years of world wide Professional Services Consulting; the last three years at Splunk
- Deployed 20+ engagements ranging from 1GB to 5TB for Splunk Professional Services engagements

Agenda

- Data
- Splunk Components
- Index Data
- Proper Parsing
- Challenging Data
- Advanced Inputs
- Questions

You Are in The Right Room If…

- You have used Splunk once, or at least read about it
- You are interested in Splunk's best practices
- You like to use Splunk's default parsing rules
- You just took over a Splunk deployment and you're not sure what to do
- This is not an education class; it's best practice

Data

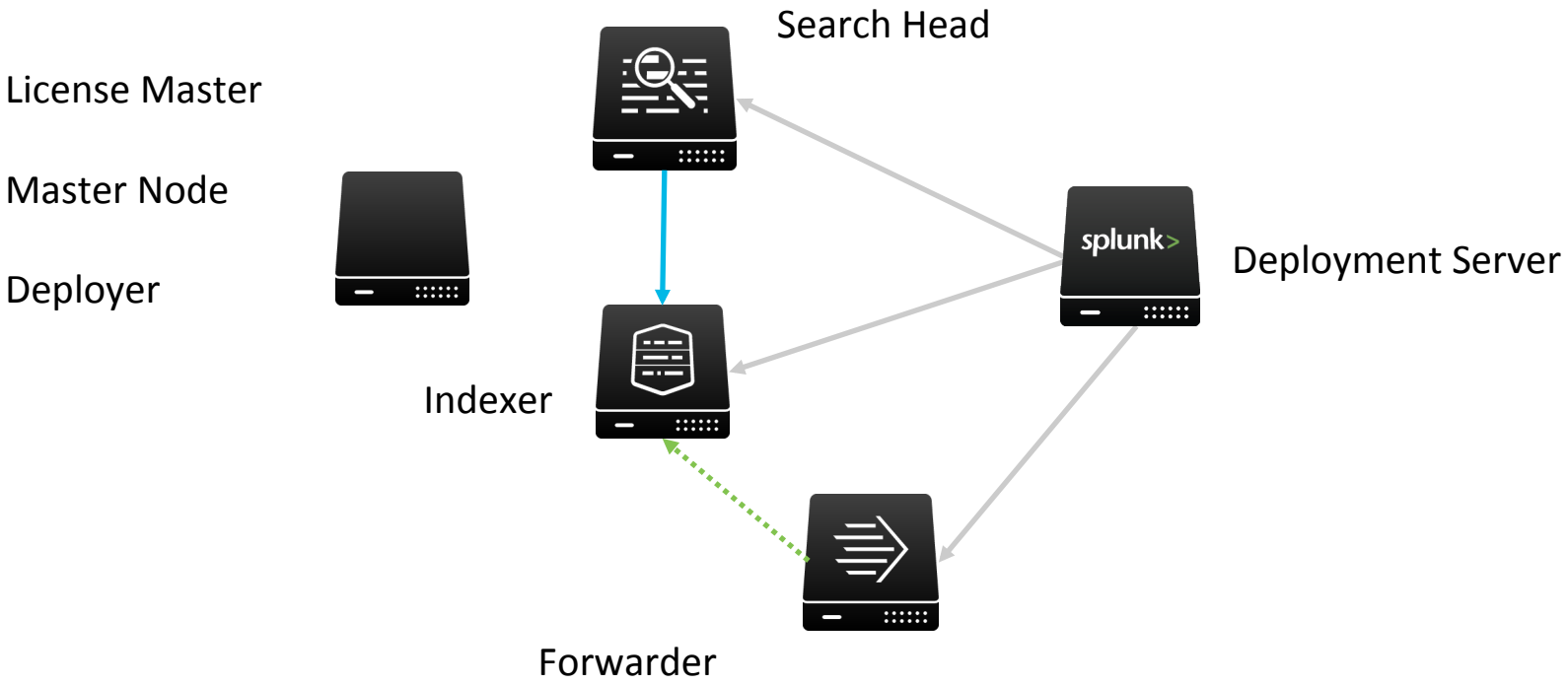
Splunk is the engine for machine data

- **Machine data is more than just logs** - it's configuration data, data from APIs and message queues, change events, the output of diagnostic commands and more
- **Log types:** Application, Web Access and Proxy, Call Detail Records (CDR), Clickstream, Message Queues, Packet, Database audit and tables, File audit, Syslog, WMI, PerfMon
- **Manual:** Getting Data In http://docs.splunk.com/Documentation/Splunk/latest/Data/WhatSplunkcanmonitor

Splunkbase Apps

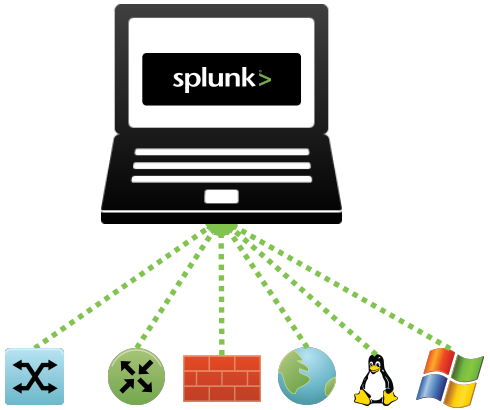
- Look to SplunkBase first and utilize Add-Ons: http://splunkbase.splunk.com/
- Applies the Common Information Model (CIM)
- CIM details the standard fields, event type tags, and host tags that Splunk uses when it processes most IT data
- Example Splunk Supported Apps and Add-ons:
  - Splunk App for Stream
  - Splunk Add-on for Microsoft Windows
- Example Premium Apps
  - Splunk App for Enterprise Security
  - Splunk App for Microsoft Exchange
  - Splunk app for VMware

Splunk Distributed Components



Test Environment

- Every Splunk deployment should have a test environment
- It can be a laptop, virtual machine or spare server
- Should have the same version of Splunk running in production
- Accessible to other Splunk developers and administrators



One Shot

- Easiest way to get data into your test environment
- Components of the oneshot:

  ./splunk add oneshot user_conf.txt –index indexname –sourcetype sourcetype name

- Where to find more information:
  http://docs.splunk.com/Documentation/Splunk/latest/Data/MonitorfilesanddirectoriesusingtheCLI

Data - Broken

i	Time	Event
>	7/30/15 9:00:01.000 AM	2015-09-15 09:00:01 msg="Data From Splunk .CONF 2015" title="Getting Data In, Correctly" author="Duca" city="Las Vegas" state="NV" 2015-09-15 09:02:01 msg="More Data" title="Getting Data In, Correctly" author="Duca" city="Las Vegas" state="NV" 2015-09-15 09:03:01 msg="The first step in using Splunk Enterprise is to feed it data. Once Splunk Enterprise gets some data, it immediately indexes it, so that it's available for searching." author="Duca" city="Las Vegas" state="NV" DEBUG INFO "extra data here" 2015-09-15 09:03:51 msg="Basically, you point Splunk Enterprise at data and in moments, you can start searching the data, or use it to create charts, reports, alerts, and other interesting outputs." author="Duca" city="Las Vegas" state="NV" 2015-09-15 09:04:09 msg="The data can be on the same machine as the Splunk Enterprise indexer (local data), or it can be on another machine altogether (remote data)." 2015-09-15 09:04:09 msg="Splunk has a large and growing variety of apps and add-ons that offer preconfigured inputs for various types of data sources. Take advantage of Splunk apps and free yourself from having to configure the inputs yourself." 2015-09-15 09:04:24 msg="As described earlier, Splunk provides tools to configure all sorts of data inputs, including many that are specific to particular application needs. Splunk also provides the tools to configure any arbitrary data input types." 2015-09-15 09:05:02 msg="You can index and search Windows data on a non-Windows instance of Splunk Enterprise, but you must first use a Windows instance to gather the data. You can do this with a Splunk forwarder running on Windows." Collapse host = aduca-mbp15-2.local ¦ source = /Users/aduca/Updated_Dropbox/Dropbox/Splunk/CONF/sample.txt ¦ sourcetype = user_conf_2015

Props

- Always set these six parameters

```
# USER CONFERENCE
[user_conf_2015]
TIME_PREFIX = ^
TIME_FORMAT = %Y-%m-%d %H:%M:%S
MAX_TIMESTAMP_LOOKAHEAD = 19
SHOULD_LINEMERGE = False
LINE_BREAKER = ([\n\r]+)\d{4}-\d{2}-\d{2}\s\d{2}:\d{2}:\d{2}
TRUNCATE = 10000
```

Props

- Defaults to empty

```
# USER CONFERENCE
[user_conf_2015]
TIME_PREFIX = ^
TIME_FORMAT = %Y-%m-%d %H:%M:%S
MAX_TIMESTAMP_LOOKAHEAD = 19
SHOULD_LINEMERGE = False
LINE_BREAKER = ([\n\r]+)\d{4}-\d{2}-\d{2}\s\d{2}:\d{2}:\d{2}
TRUNCATE = 10000
```

Props

- strptime style format

```
# USER CONFERENCE
[user_conf_2015]
TIME_PREFIX = ^
TIME_FORMAT = %Y-%m-%d %H:%M:%S
MAX_TIMESTAMP_LOOKAHEAD = 19
SHOULD_LINEMERGE = False
LINE_BREAKER = ([\n\r]+)\d{4}-\d{2}-\d{2}\s\d{2}:\d{2}:\d{2}
TRUNCATE = 10000
```

Props

- By default MAX_TIMESTAMP_LOOKAHEAD = 150 characters

```
# USER CONFERENCE
[user_conf_2015]
TIME_PREFIX = ^
TIME_FORMAT = %Y-%m-%d %H:%M:%S
MAX_TIMESTAMP_LOOKAHEAD = 19
SHOULD_LINEMERGE = False
LINE_BREAKER = ([\n\r]+)\d{4}-\d{2}-\d{2}\s\d{2}:\d{2}:\d{2}
TRUNCATE = 10000
```

Props

- By default set to True

```
# USER CONFERENCE
[user_conf_2015]
TIME_PREFIX = ^
TIME_FORMAT = %Y-%m-%d %H:%M:%S
MAX_TIMESTAMP_LOOKAHEAD = 19
SHOULD_LINEMERGE = False
LINE_BREAKER = ([\n\r]+)\d{4}-\d{2}-\d{2}\s\d{2}:\d{2}:\d{2}
TRUNCATE = 10000
```

Props

- By default set to ([\r\n]+); change to positive lookahead

```
# USER CONFERENCE
[user_conf_2015]
TIME_PREFIX = ^
TIME_FORMAT = %Y-%m-%d %H:%M:%S
MAX_TIMESTAMP_LOOKAHEAD = 19
SHOULD_LINEMERGE = False
LINE_BREAKER = ([\r\n]+)\d{4}-\d{2}-\d{2}\s\d{2}:\d{2}:\d{2}
TRUNCATE = 10000
```

Props

- By default set to 10000 bytes; set to 0 to never truncate

```
# USER CONFERENCE
[user_conf_2015]
TIME_PREFIX = ^
TIME_FORMAT = %Y-%m-%d %H:%M:%S
MAX_TIMESTAMP_LOOKAHEAD = 19
SHOULD_LINEMERGE = False
LINE_BREAKER = ([\n\r]+)\d{4}-\d{2}-\d{2}\s\d{2}:\d{2}:\d{2}
TRUNCATE = 10000
```

Data - Fixed

i	Time	Event
>	7/30/15 2:50:29.000 PM	2015-09-15 09:05:02 msg="You can index and search Windows data on a non-Windows instance of Splunk Enterprise, but you must first use a Windows instance to gather the data. You can do this with a Splunk forwarder running on Windows." host = aduca-mbp15-2.local ¦ source = /Users/aduca/Updated_Dropbox/Dropbox/Splunk/CONF/sample.txt ¦ sourcetype = user_conf_2015
>	7/30/15 2:50:29.000 PM	2015-09-15 09:04:24 msg="As described earlier, Splunk provides tools to configure all sorts of data inputs, including many that are specific to particular application needs. Splunk also provides the tools to configure any arbitrary data input types." host = aduca-mbp15-2.local ¦ source = /Users/aduca/Updated_Dropbox/Dropbox/Splunk/CONF/sample.txt ¦ sourcetype = user_conf_2015
>	7/30/15 2:50:29.000 PM	2015-09-15 09:04:09 msg="Splunk has a large and growing variety of apps and add-ons that offer preconfigured inputs for various types of data s ources. Take advantage of Splunk apps and free yourself from having to configure the inputs yourself." host = aduca-mbp15-2.local ¦ source = /Users/aduca/Updated_Dropbox/Dropbox/Splunk/CONF/sample.txt ¦ sourcetype = user_conf_2015
>	7/30/15 2:50:29.000 PM	2015-09-15 09:04:09 msg="The data can be on the same machine as the Splunk Enterprise indexer (local data), or it can be on another machine alt ogether (remote data)." host = aduca-mbp15-2.local ¦ source = /Users/aduca/Updated_Dropbox/Dropbox/Splunk/CONF/sample.txt ¦ sourcetype = user_conf_2015
>	7/30/15 2:50:29.000 PM	2015-09-15 09:03:51 msg="Basically, you point Splunk Enterprise at data and in moments, you can start searching the data, or use it to create c harts, reports, alerts, and other interesting outputs." author="Duca" city="Las Vegas" state="NV" host = aduca-mbp15-2.local ¦ source = /Users/aduca/Updated_Dropbox/Dropbox/Splunk/CONF/sample.txt ¦ sourcetype = user_conf_2015
>	7/30/15 2:50:29.000 PM	2015-09-15 09:03:01 msg="The first step in using Splunk Enterprise is to feed it data. Once Splunk Enterprise gets some data, it immediately indexes it, so that it's available for searching." author="Duca" city="Las Vegas" state="NV" DEBUG INFO "extra data here" host = aduca-mbp15-2.local ¦ source = /Users/aduca/Updated_Dropbox/Dropbox/Splunk/CONF/sample.txt ¦ sourcetype = user_conf_2015
>	7/30/15 2:50:29.000 PM	2015-09-15 09:02:01 msg="More Data" title="Getting Data In, Correctly" author="Duca" city="Las Vegas" state="NV" host = aduca-mbp15-2.local ¦ source = /Users/aduca/Updated_Dropbox/Dropbox/Splunk/CONF/sample.txt ¦ sourcetype = user_conf_2015
>	7/30/15 2:50:29.000 PM	2015-09-15 09:00:01 msg="Data From Splunk .CONF 2015" title="Getting Data In, Correctly" author="Duca" city="Las Vegas" state="NV" host = aduca-mbp15-2.local ¦ source = /Users/aduca/Updated_Dropbox/Dropbox/Splunk/CONF/sample.txt ¦ sourcetype = user_conf_2015

6.3 Splunk Web Data On-Boarding

splunk>

Why Use Splunk Web to Onboard?

Quick and easy way to…

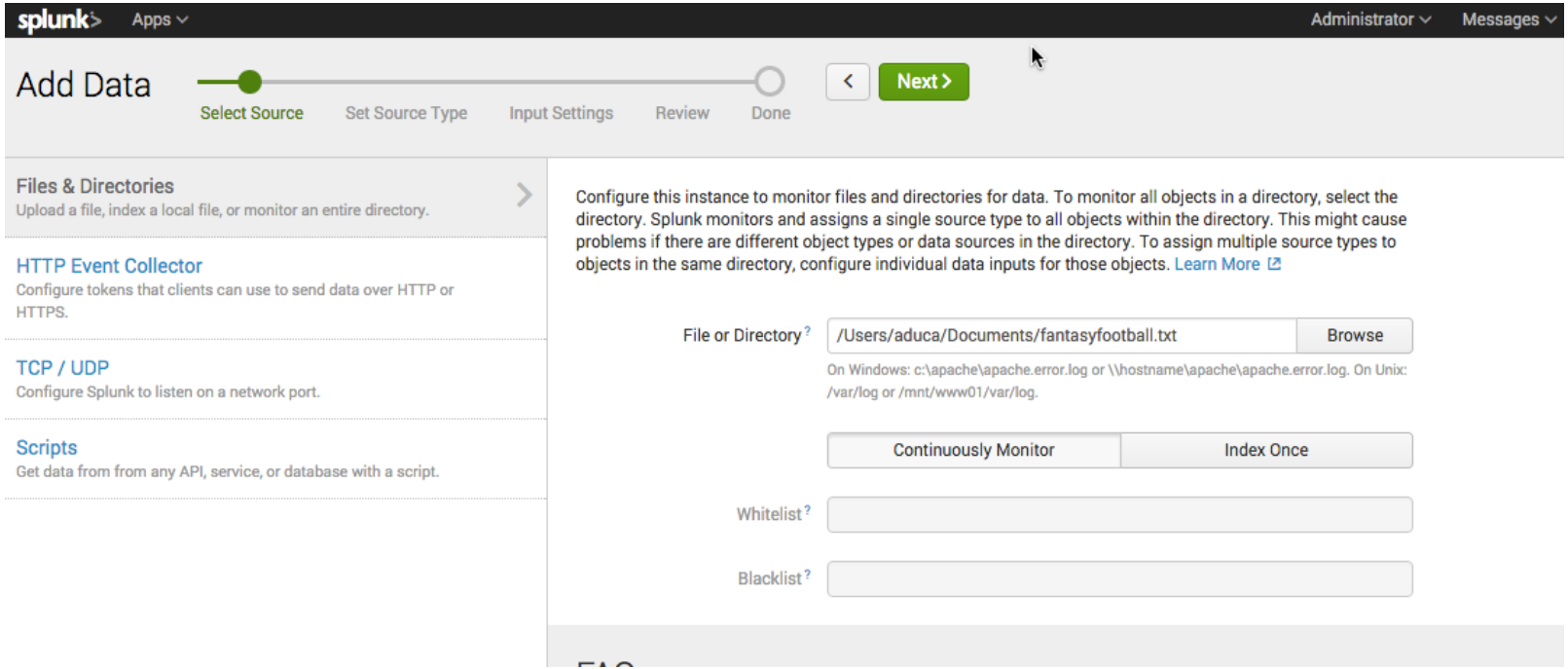
- Easily visualize the data into events rather then lines of text
- Quickly get the data properly broken into events
- Accurately get the timestamp extracted

All in a wicked cool GUI…

Once everything is good you take your PROPS settings and deploy

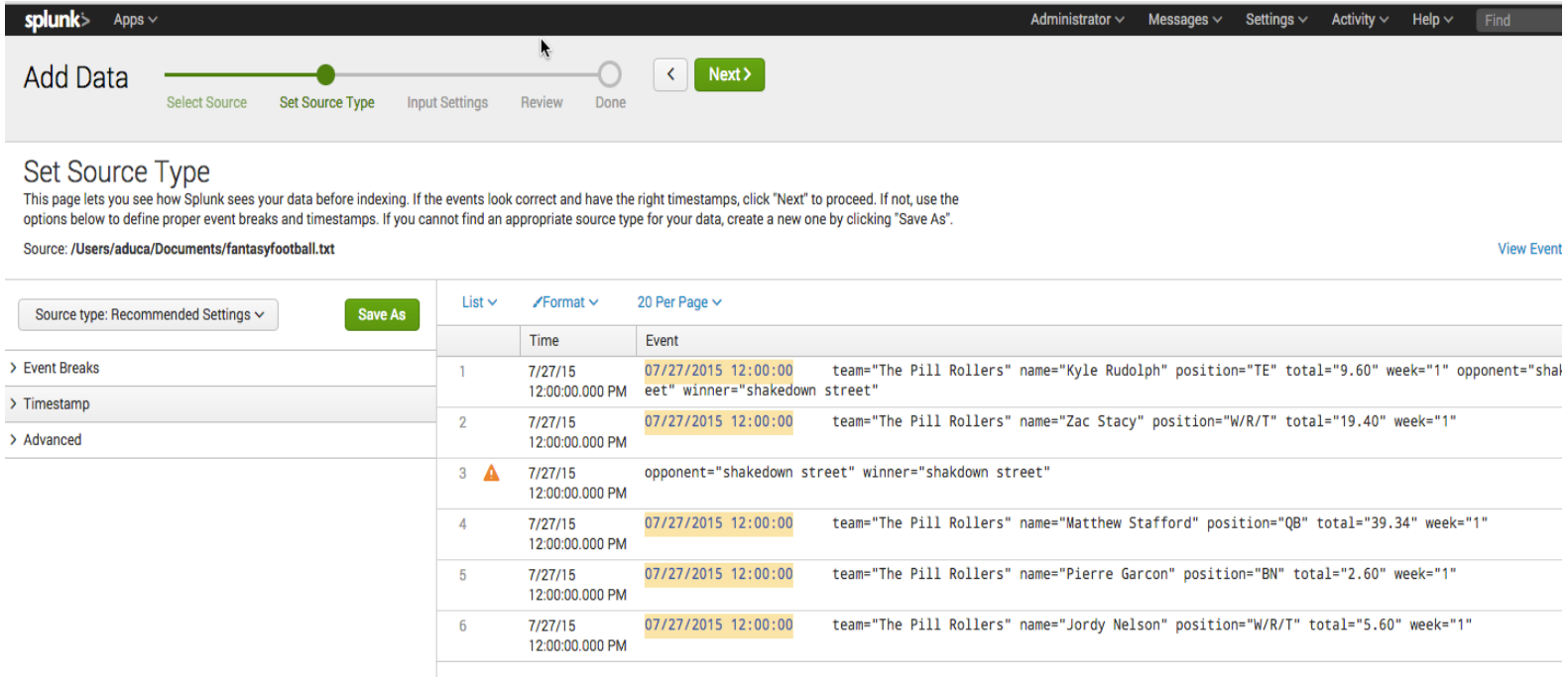
Splunk Web Data On-Boarding

- Locate the Source file on the Splunk Server's file system



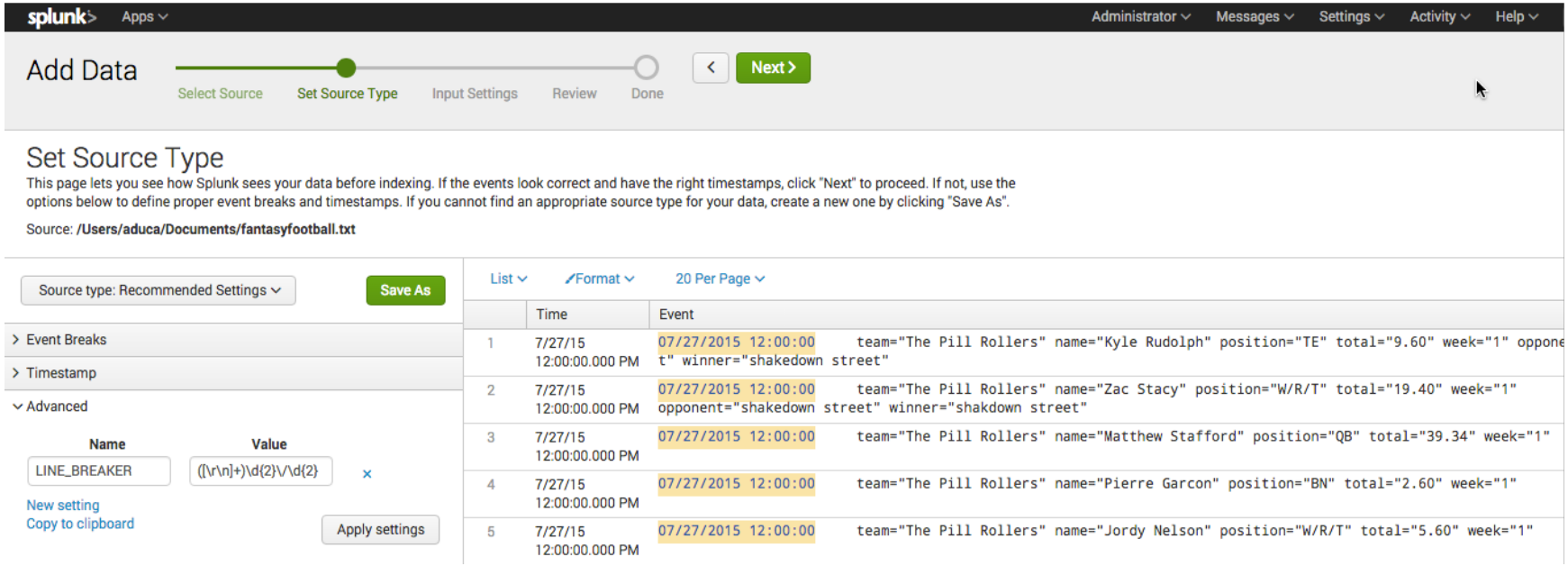
Splunk Web Data On-Boarding

- Validate Event breaking and Timestamp recognition



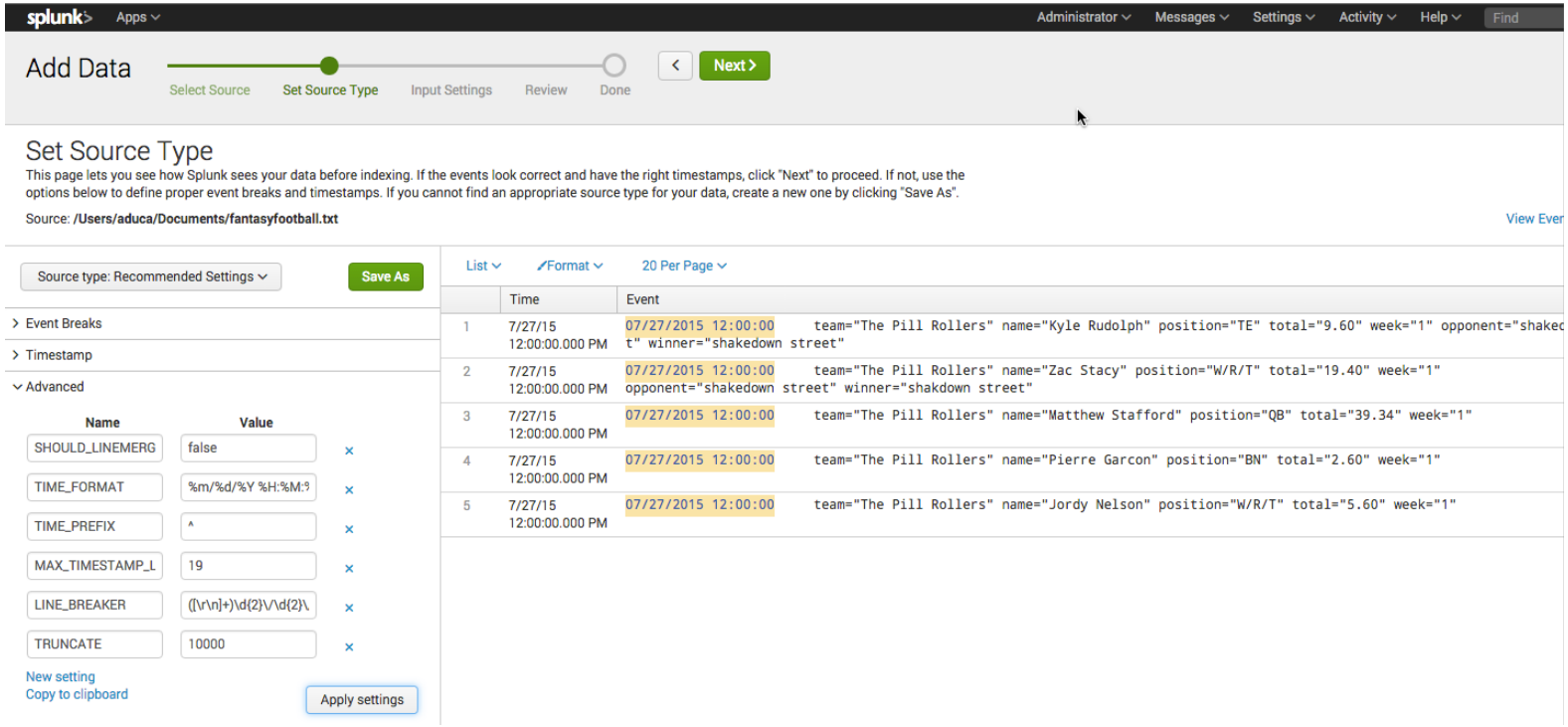
Splunk Web Data On-Boarding

- Resolve Event Breaking



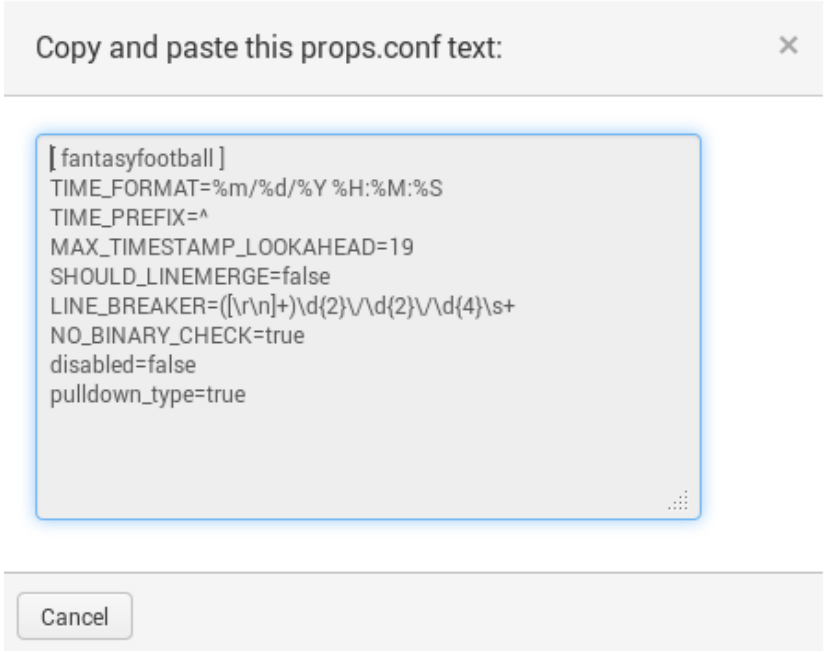
Splunk Web Data On-Boarding

- Set timestamp format even if Splunk figures it out automatically



Splunk Web Data On-Boarding

- Copy the props.conf settings and deploy in a custom app



Challenging Data

splunk>

Limit Indexed Data

- Anonymize data:

```
[source::.../accounts.log]
SEDCMD-accounts = s/ssn=\d{5}(\d{4})/ssn=xxxxx\1/g s/cc=(\d{4}-){3}(\d{4})/cc=xxxx-xxxx-xxxx-\2/g
```

- Rewrite raw data:

```
[source::.../sql.log]
SEDCMD-sqllog = s/(.*?)Command:EXECUTE[.\d\D\w\W]*/\1/g
```

- Discard events:

```
props
[source::/var/log/user_conf.txt]
TRANSFORMS-null= setnull
```

```
transforms
[setnull]
REGEX    = (?i)DEBUG
DEST_KEY = queue
FORMAT   = nullQueue
```

Limit Indexed Data

- Anonymize data:

```
[source::.../accounts.log]
SEDCMD-accounts = s/ssn=\d{5}(\d{4})/ssn=xxxxx\1/g s/cc=(\d{4}-){3}(\d{4})/cc=xxxx-xxxx-xxxx-\2/g
```

- Rewrite raw data:

```
[source::.../sql.log]
SEDCMD-sqllog = s/(.*?)Command:EXECUTE[.\d\D\w\W]*/\1/g
```

- Discard events:

```
props
[source::/var/log/user_conf.txt]
TRANSFORMS-null= setnull
```

```
transforms
[setnull]
REGEX    = (?i)DEBUG
DEST_KEY = queue
FORMAT   = nullQueue
```

Limit Indexed Data

- Anonymize data:

```
[source::.../accounts.log]
SEDCMD-accounts = s/ssn=\d{5}(\d{4})/ssn=xxxxx\1/g s/cc=(\d{4}-){3}(\d{4})/cc=xxxx-xxxx-xxxx-\2/g
```

- Rewrite raw data:

```
[source::.../sql.log]
SEDCMD-sqllog = s/(.*?)Command:EXECUTE[.\d\D\w\W]*/\1/g
```

- Discard events:

props

```
[source::/var/log/user_conf.txt]
TRANSFORMS-null= setnull
```

transforms

```
[setnull]
REGEX    =  (?i)DEBUG
DEST_KEY =  queue
FORMAT   =  nullQueue
```

Whitelist or Blacklist Windows Events

- This will selectively include or exclude events from collection on a Windows forwarder
- Available feature on 6.x or greater Windows forwarders
- All controlled through inputs.conf on the Windows forwarders
- Examples:

```
[WinEventLog://Security]
whitelist = 4,5,7,100-200
…
```

```
[WinEventLog://Security]
blacklist = EventCode=%^200$% User=%duca%
…
```

For more Info: http://docs.splunk.com/Documentation/Splunk/latest/Data/MonitorWindowsdata#Create_advanced_filters_with_.27whitelist.27_and_.27blacklist.27

Index Extractions

- Provides reliable and consistent indexing of data with headers
- Address issue on forwarder:

```
INDEX_EXTRACTIONS = {CSV | W3C | TSV | PSV | JSON}
```

- Supports custom header parsing and easy mode for common formats
- Extract IIS fields using Props.conf on Windows forwarder:

```
[iis]
INDEX_EXTRACTIONS = w3c
```

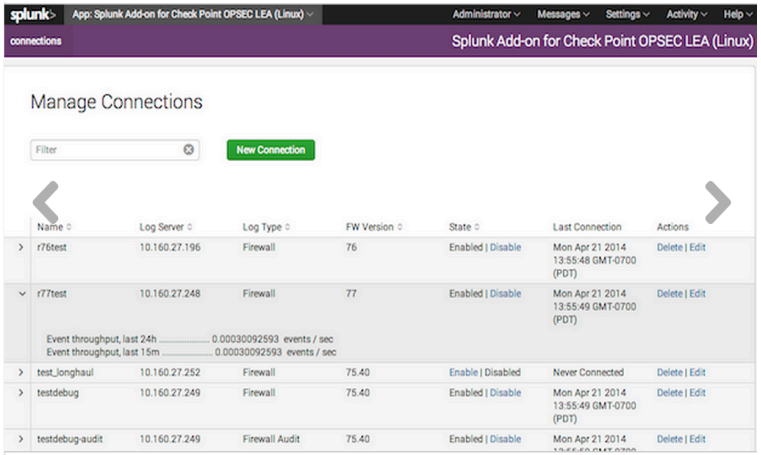
Modular and Scripted Inputs

splunk>

Modular Inputs

Splunk Enterprise app or add-on that extends the Splunk Enterprise framework to define a custom input capability

- Returned data must be properly structured to index into Splunk
- Features:
  - Configuration files and configuration settings inside Splunk and Splunk Web interfaces
  - Can also be configured via deployed .conf files and accessed via REST API
- Examples:
  - Checkpoint OPSEC, Twitter, Stream, Amazon S3 Online storage
- For more info:
  - http://docs.splunk.com/Documentation/Splunk/latest/AdvancedDev/ModInputsBasicExample



Scripted Inputs

- A scripted input is used to get data from application program interfaces (APIs) and other remote data interfaces and message queues.

- Features
  - Scripted Inputs are deployed specific to the OS deployed on where Modular Inputs can support multiple.
  - Almost any program that can output text can be used to index data.

- Examples
  - VMStat, Top, iostat

- For more info
  - http://docs.splunk.com/Documentation/Splunk/6.2.4/Data/Setupcustominputs

Scripted Inputs Example

- Shell script saved in /opt/splunk/bin/scripts/ OR in a specific App
- Allows you to execute any program on Splunk Forwarder and index STDOUT data.
- Utilizing Key Value Pairs makes for easier searching.

```
sh-3.2# sh /Applications/Splunk/bin/scripts/FantastyFootball.sh
09/07/2015 12:00:00 team="Little Lebowski" name="Andrew Luck" position="QB" total="38.70" week="1"
09/07/2015 12:00:00 team="Little Lebowski" name="Antonio Brown" position="WR"
09/07/2015 12:00:00 team="Little Lebowski" name="Alshon Jeffery" position="WR" week="1"
09/07/2015 12:00:00 team="Little Lebowski" name="Matt Forte" position="RB" week="1"
09/07/2015 12:00:00 team="Little Lebowski" name="Montee Ball" position="RB" week="1"
09/07/2015 12:00:00 team="Little Lebowski" name="Jordan Reed" position="TE" week="1"
09/07/2015 12:00:00 team="Little Lebowski" name="Dennis Pitta" position="TE" week="1"
09/07/2015 12:00:00 team="Little Lebowski" name="Arian Foster" position="W/R/T" week="1"
09/07/2015 12:00:00 team="Little Lebowski" name="Cordarrell Patterson" position="BN" week="1"
09/07/2015 12:00:00 team="Little Lebowski" name="Ryan Mathews" position="BN" week="1"
09/07/2015 12:00:00 team="Little Lebowski" name="Tavon Austin" position="BN" week="1"
09/07/2015 12:00:00 team="Little Lebowski" name="Reggie Bush" position="BN" week="1"
09/07/2015 12:00:00 team="Little Lebowski" name="Steven Hauschka" position="K" week="1"
09/07/2015 12:00:00 team="Little Lebowski" name="Cincinnati" position="DEF" week="1"
```

Sample output from custom script /Applications/Splunk/bin/scripts/FantasyFootball.sh

Scripted Inputs Example

Shell script calls local system binary programs and can provide configuration options.

```
aduca-mbp15:Yahoo aduca$ more /Applications/Splunk/bin/scripts/FantasyFootball.sh
#!/bin/sh

/usr/bin/php /Users/aduca/Desktop/Yahoo/Score_Easy.php
```

Use Inputs.conf to define INDEX, SOURCETYPE, and INTERVAL for the scripted input

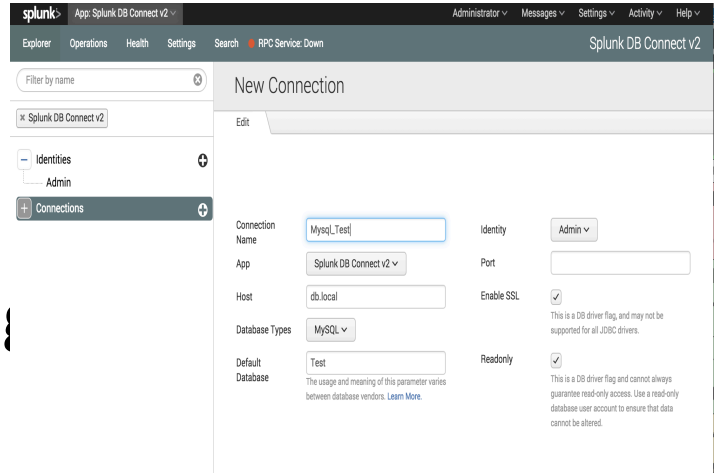
```
[script:///Applications/Splunk/bin/scripts/FantasyFootball.sh]
disabled = 0
index = FantasyFootball
interval = 30
sourcetype = FantasyFootball
```

Database Data

splunk>

Splunk DB Connect

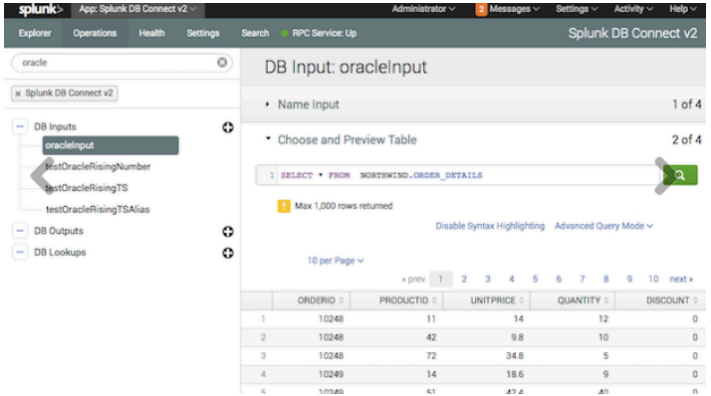
- Allows for indexing data directly from database queries.
- Allows for adding meta data to events from Database sources using lookups.
- Example use cases:
  - Symantec Endpoint Protection data
  - Custom CMDB Databases



Supported DBs

DB2, Informix, MemSQL, Microsoft SQL Server MySQL, Oracle, PostgreSQL, SAP SQL Anywhere, Sybase ASE, Sybase IQ, Teradata

DB Connect Best Practices



- Normalize timestamps natively inside the SQL Query
- Filter results down in SQL Query to reduce garbage in Splunk Index
- Repeated DBLookups should be converted to static lookup
- Search Head Pooling requires encrypted password replication
- Search Head Clustering Supported

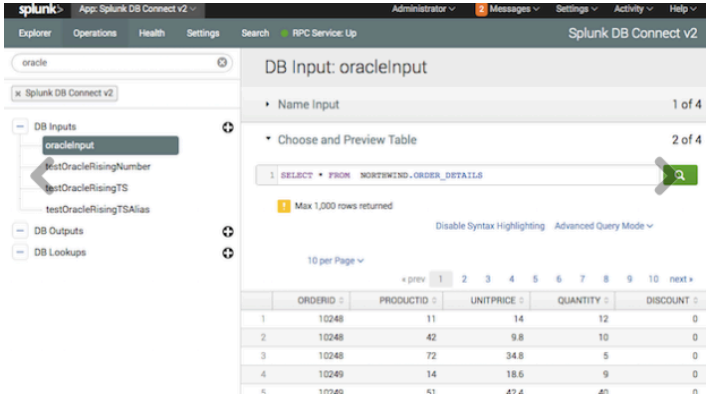
Network Data

splunk>

Splunk App For Stream

- Provides the ability to capture real-time streaming wire data from anywhere in your datacenter or from any public Cloud infrastructure.
- Supports Local collection and via SPAN/TAP
- Stream Forwarder runs on the following OS:
  - Linux
  - Windows
  - Mac OS

DB Connect Best Practices



- Normalize timestamps natively inside the SQL Query
- Filter results down in SQL Query to reduce garbage in Splunk Index
- Repeated DBLookups should be converted to static lookup
- Search Head Pooling requires encrypted password replication
- Search Head Clustering supported

Splunk Stream DNS Capture

- Full DNS Queries without logging enabled

```
C:\Users\Administrator>nslookup
Default Server:  UnKnown
Address:  172.27.1.111

> www.splunk.com
Server:  [172.27.1.111]
Address:  172.27.1.111

Non-authoritative answer:
Name:    d2n3e195vjrw9z.cloudfront.net
Addresses:  204.246.169.197
          54.230.36.216
          54.192.36.139
          54.230.37.123
          54.230.36.23
          54.230.37.60
          54.192.36.115
          54.192.38.162
Aliases:  www.splunk.com
```

Resources

- Get educated: http://www.splunk.com/view/education/SP-CAAAAH9
- Download Splunk applications: http://apps.splunk.com/
- Hire Splunk Professional Services: http://www.splunk.com/view/professional-services/SP-CAAABH9
- Watch some videos: http://www.splunk.com/videos

THANK YOU

splunk>