Critical Syslog Tricks, Part II

(That No One Seems to Know About)

Jonathan Margulies | Citadel LLC, Former Splunk Professional Services Consultant

George Barrett, PhD | Splunk Professional Services Consultant at Rational Cyber

October 2018

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Citadel Disclaimer

- This presentation reflects the analysis and views of the authors. No recipient should interpret this presentation to represent the general views of Citadel or its personnel. Facts, analysis, and views presented in this presentation have not been reviewed by, and may not reflect information known to, other Citadel professionals.

- Assumptions, opinions, views, and estimates constitute the authors’ judgment as of the date given and are subject to change without notice and without any duty to update. Citadel is not responsible for any errors or omissions contained in this presentation and accepts no liability whatsoever for any direct or consequential loss arising from your use of this presentation or its contents.

A Quick Note About These Slides

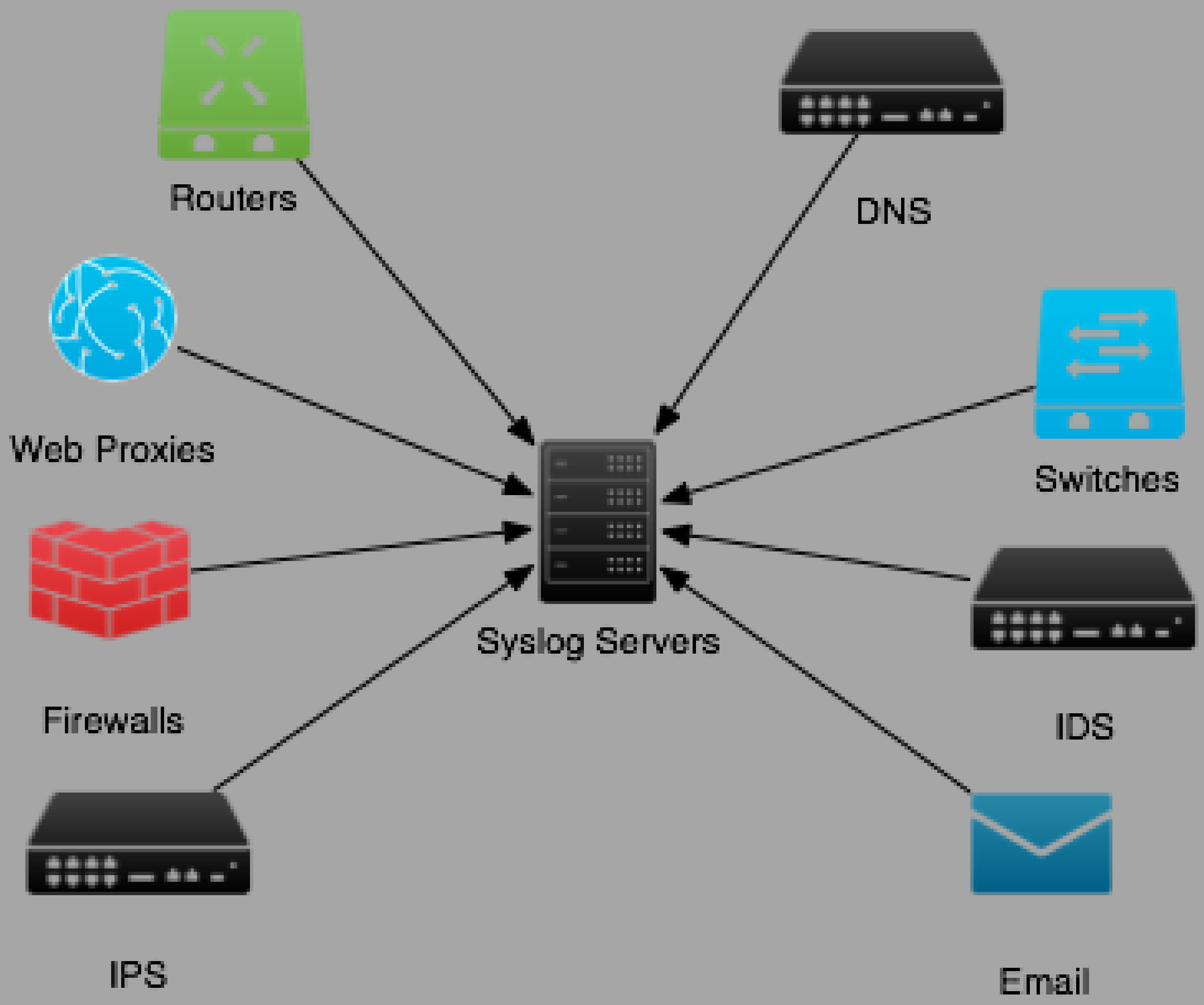
- Last year, we presented a talk on our bulletproof method for collecting syslog data into Splunk using syslog-ng
- This year, we have so much new stuff that we have no time to go over the old stuff
- But our new stuff builds so heavily on last year's presentation that, especially for people viewing these slides after the conference, we wanted to put it all together as one cohesive guide to collecting syslog data in the enterprise
- With that in mind, any slides that don't contain new material will have a gray background, and slides that do contain new material will have a white background

Do You Have a Syslog Collection Problem?

You Might Have a Syslog Collection Problem If…

- Your syslog data arrives in Splunk more than a few seconds after the event time
- Syslog data that comes in while Splunk is restarting gets dropped
- You notice gaps or missing events in your syslog data feeds
- You need a new listening port every time you get a new syslog data source
- Your indexers or heavy forwarders have to look in raw events to figure out what index, sourcetype, or host to assign to those events
- Multiple hosts' syslog data are being aggregated under the same host because they came through the same syslog server
- Your IT people use grep instead of Splunk to troubleshoot live issues

Syslog Brings in Your Most Important Logs



splunk> .conf18

What You'll Learn From This Presentation

- Last year:
  - How to configure syslog-ng to collect all your syslog data for Splunk
  - How to architect your syslog collection infrastructure
  - How to configure Splunk to collect all the data from syslog-ng and index it in about 3 seconds
  - How to find and troubleshoot syslog collection problems quickly
- This year:
  - Updates and fixes!
  - Our new rsyslog.conf for syslog-ng haters
  - Automatic sourcetyping and timezoning
  - Our new monitoring app
  - How to use everything we've built
  - HEC and Kafka

syslog-ng and rsyslog

A Few Things to Note About syslog-ng

- It’s free. There’s a paid version, but this presentation assumes you didn’t buy it.
- We recommend version 3.5 or higher, as that supports multithreading and some other useful features
- You can do everything we’re recommending using rsyslog instead, but we don’t recommend it
  - syslog-ng handles poorly formatted syslog events more gracefully
  - **That said, we’ll show you how to do it anyway**

~~https://syslog-ng.org/~~

~~https://www.balabit.com/documents/syslog-ng-ose-latest-guides/en/syslog-ng-ose-guide-admin/html/index.html~~

This link doesn’t work anymore because someone not named Splunk bought Balabit. New link: https://www.syslog-ng.com/technical-documents/doc/syslog-ng-open-source-edition/3.16/administration-guide

Configuring syslog-ng (options)

```
options {
    flush_lines (100);
    time_reopen (10);
    log_fifo_size (1000);
    chain_hostnames (off);
    use_dns (no);
    use_fqdn (no);
    create_dirs (yes);
    keep_hostname (yes);
    threaded (yes);
};
```

https://gitlab.com/rationalcyber/syslog-ng-configuration/blob/master/syslog-ng.conf

Configuring syslog-ng (Listening and Writing)

```
source s_aggregation {
    network(ip(0.0.0.0) transport(tcp) port(514));
    network(ip(0.0.0.0) transport(udp) port(514));
};

destination d_splunkf {
     file(“/mnt/$LOGHOST/log/$R_YEAR-$R_MONTH-$R_DAY/$HOST_FROM/$HOST/$FACILITY.log" dir-owner("splunk") dir-group("splunk") owner("splunk") group("splunk"));
};
```

This Is The Most Important Line!

```
file(“/mnt/$LOGHOST/log/$R_YEAR-$R_MONTH-$R_DAY/$HOST_FROM/$HOST/$FACILITY.log” dir-owner("splunk") dir-group("splunk") owner("splunk") group(“splunk"));
```

- **/$LOGHOST**
  - Essentially, “the hostname of this syslog-ng server.” You’re going to be collecting syslog on more than one server, so this will help with troubleshooting.
- **/log/$R_YEAR-$R_MONTH-$R_DAY**
  - This is important for log rotation. We’ll explain that on its own slide.
- **/$HOST_FROM**
  - “The host I received this feed from.” It may be the same as the originating host, or it may be an intermediate syslog server. In the latter case, helps with troubleshooting.

The Rest of That Line

```
file(“/mnt/$LOGHOST/log/$R_YEAR-$R_MONTH-$R_DAY/$HOST_FROM/$HOST/$FACILITY.log” dir-owner("splunk") dir-group("splunk") owner("splunk") group(“splunk"));
```

- **/$HOST**
  - “The hostname from the syslog header.” This may be an actual hostname, FQDN, or IP address, but it’s always the most reliable source of the logs’ originating host.
- **/$FACILITY.log**
  - “The syslog facility setting.” This generally isn’t useful by itself, but it can almost always be used in combination with $HOST to separate different sourcetypes from the same host.
- **dir-owner("splunk") dir-group("splunk") owner("splunk") group(“splunk”)**
  - Splunk should never be running as root! Make sure the splunk user can read and rotate all the log files.

Rotating Logs

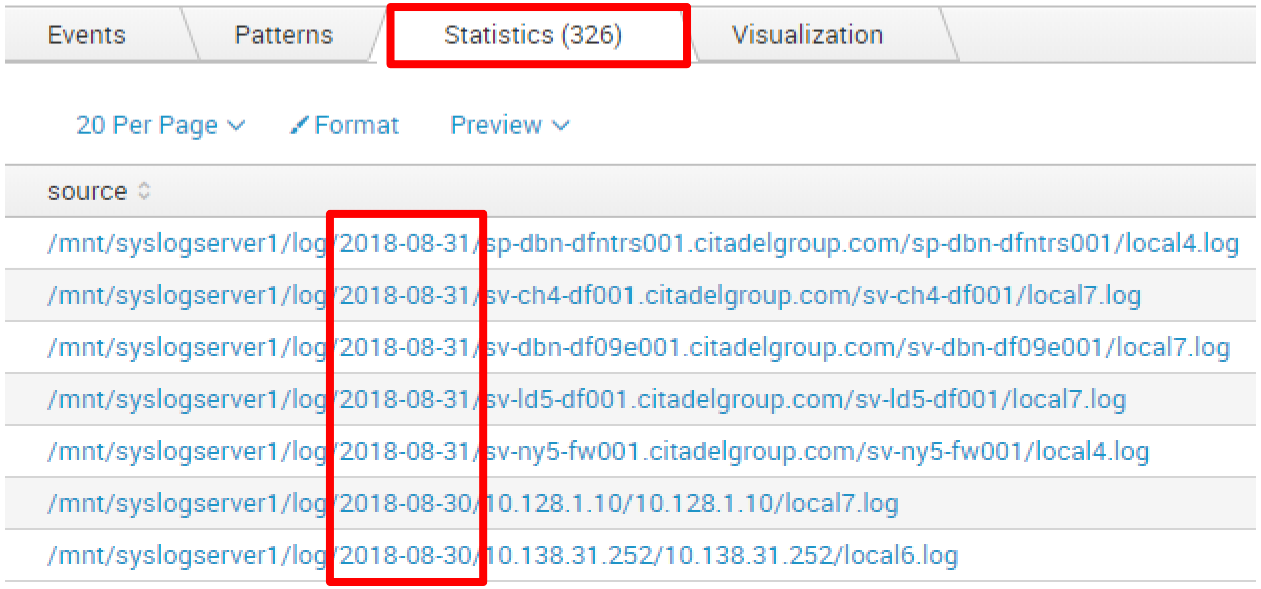
- Do not use logrotate on a syslog server
  - It will restart syslog-ng and you’ll lose a couple of seconds of logs
  - https://www.syslog-ng.com/technical-documents/doc/syslog-ng-open-source-edition/3.16/administration-guide/86#TOPIC-956714
  - ~~https://www.balabit.com/documents/syslog-ng-ose-latest-guides/en/syslog-ng-ose-guide-admin/html/example-logrotate.html~~
- Use these cron jobs instead (adjust the times as needed):

```
#cron job 1: at 5am, find yesterday's logs, and move them to old_logs
0 5 * * * /usr/bin/find /mnt/*/log/????-??-?? -maxdepth 0 -type d ! -mmin -300 -exec bash
-c 'dir={}; old=${dir/\/log\//\/old_logs\/}; mv ${dir} ${old}' \;

#cron job 2: find any files older than 5 days, 23 hours, and delete them
0 4 * * * /usr/bin/find /mnt/*/old_logs/????-??-?? -maxdepth 0 -type d ! -mmin -8580 -exec
rm -rf {} \;
```

A Small Improvement to Our Directory Structure

- The flaw in our old date-handling is that caused the source field in Splunk to proliferate with otherwise identical sources being separated by different dates

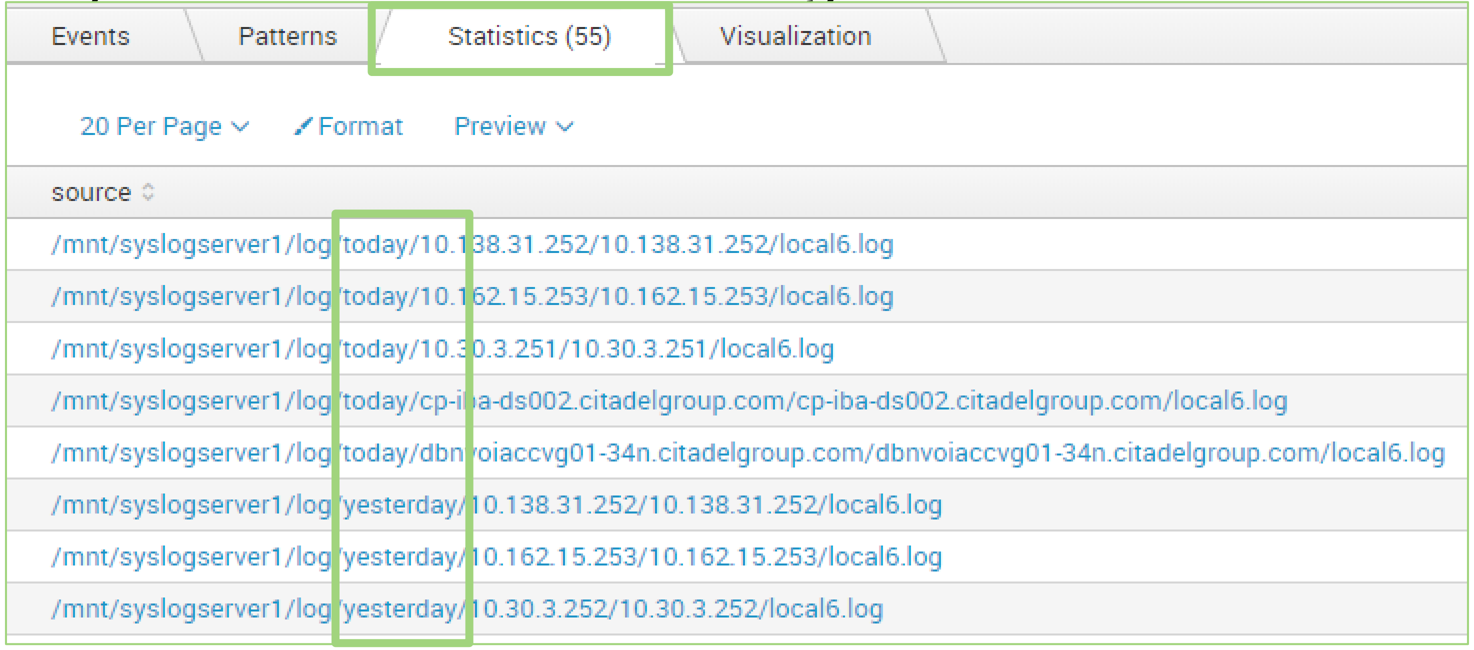


source
/mnt/syslogserver1/log/2018-08-31/sp-dbn-dfntrs001.citadelgroup.com/sp-dbn-dfntrs001/local4.log
/mnt/syslogserver1/log/2018-08-31/sv-ch4-df001.citadelgroup.com/sv-ch4-df001/local7.log
/mnt/syslogserver1/log/2018-08-31/sv-dbn-df09e001.citadelgroup.com/sv-dbn-df09e001/local7.log
/mnt/syslogserver1/log/2018-08-31/sv-ld5-df001.citadelgroup.com/sv-ld5-df001/local7.log
/mnt/syslogserver1/log/2018-08-31/sv-ny5-fw001.citadelgroup.com/sv-ny5-fw001/local4.log
/mnt/syslogserver1/log/2018-08-30/10.128.1.10/10.128.1.10/local7.log
/mnt/syslogserver1/log/2018-08-30/10.138.31.252/10.138.31.252/local6.log

- The solution: Use symlinks to alias the directory named for today's date to "today" (and the same for "yesterday") and have Splunk monitor the "today" and "yesterday" directories instead

New Cron Job

-
  ```
  0 0 * * * rm -f /mnt/*/log/today /mnt/*/log/yesterday;
      ln -fs `date +\%Y-\%m-\%d` /mnt/*/log/today; ln -fs
      `date -d yesterday +\%Y-\%m/-%d` /mnt/*/log/yesterday
  ```
- The result is way fewer and more meaningful sources:



source
/mnt/syslogserver1/log/today/10.138.31.252/10.138.31.252/local6.log
/mnt/syslogserver1/log/today/10.162.15.253/10.162.15.253/local6.log
/mnt/syslogserver1/log/today/10.30.3.251/10.30.3.251/local6.log
/mnt/syslogserver1/log/today/cp-iba-ds002.citadelgroup.com/cp-iba-ds002.citadelgroup.com/local6.log
/mnt/syslogserver1/log/today/dbnvoiaccvg01-34n.citadelgroup.com/dbnvoiaccvg01-34n.citadelgroup.com/local6.log
/mnt/syslogserver1/log/yesterday/10.138.31.252/10.138.31.252/local6.log
/mnt/syslogserver1/log/yesterday/10.162.15.253/10.162.15.253/local6.log
/mnt/syslogserver1/log/yesterday/10.30.3.252/10.30.3.252/local6.log

Our Rsyslog Configs

- Everything we’re doing in rsyslog is intended to achieve the same results as our syslog-ng configuration—just different syntax
- Warning: When it comes to automatic directory creation (which is critical to our method of Splunk syslog collection), rsyslog does not handle raw data (i.e., any log message that lacks a proper syslog header) as gracefully
  - We have not found a fix for this
  - If using rsyslog, test improperly formatted syslog data carefully
  - If rsyslog receives raw data using our configs, it may create directory names based on seemingly random portions of the message that it mistakes for hostnames

Rsyslog Config Highlights

- Configures the proper directory hierarchy, as discussed earlier for syslog-ng

```
template(name="DateHostFacility" type="string" string="/mnt/%$myhostname%/log/%$YEAR%-
%$MONTH%-%$DAY%/%FROMHOST%/%HOSTNAME%/%syslogfacility-text%.log")
```

- Apply the directory template from the bullet above, and set the proper file permissions

```
ruleset(name="splunk") {
    action(
            type="omfile"
            dirCreateMode="0755"
            fileGroup="splunk"
            dirGroup="splunk"
            fileOwner="splunk"
            dirOwner="splunk"
            DynaFile="DateHostFacility")
            template="FileFormat"
    }
```

Rsyslog Configuration Highlights

- Tell rsyslog to listen on TCP/UDP port 514 and apply the “splunk” ruleset defined on the previous slide

```
input(
        name="syslog_tcp"
        type="imtcp"
        port="514"
        ruleset="splunk"
)
input(
        name="syslog_udp"
        type="imudp"
        port="514"
        ruleset="splunk"
)
```

Rsyslog Configuration Highlights

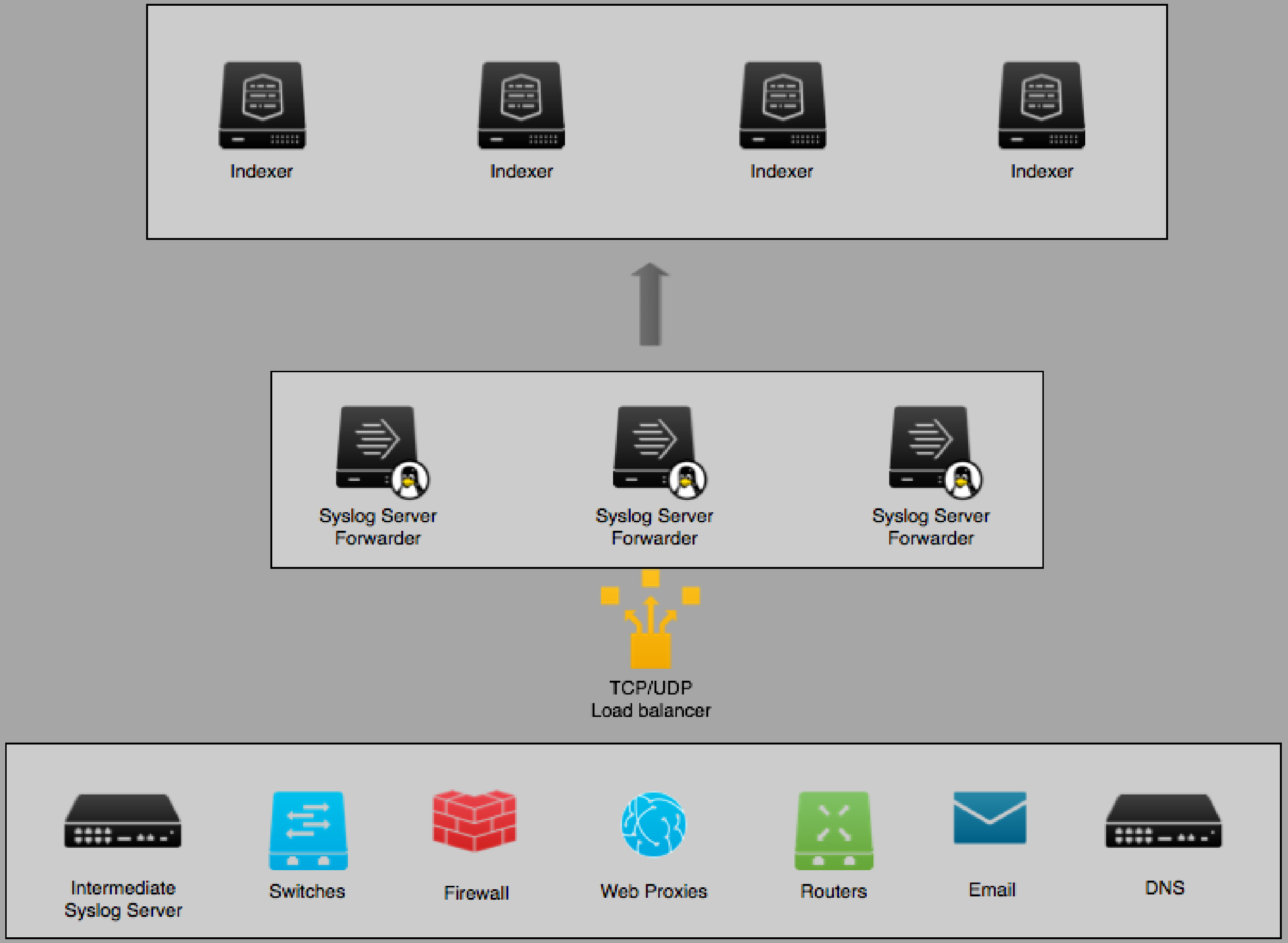
- Tell rsyslog not to append a syslog header to the front of events with invalid syslog headers
  - When rsyslog adds its own syslog header, it breaks timezones and adds worthless data to your license

```
template(name="FileFormat" type="list") {
  property(name="rawmsg-after-pri")
  constant(value="\n")
}
```

Architecting Syslog Infrastructure for Splunk

splunk> .conf18

Network Architecture



What Kind of Forwarder?

Heavy vs Universal

Heavy Forwarder Advantages

- Can handle timezone conversions
  - Keep your props and inputs together
- Takes load off your indexers
- PII masking
- Minimize indexer restarts on config changes

Universal Forwarder Advantages

- Need a lot less bandwidth to the indexing tier
  - Less metadata
- Less processor/memory load on the syslog servers

Configuring the Forwarder

inputs.conf

```
[monitor:///mnt/*/log/*day/*/fireeye*/local2.log]
host_segment = 6
index = idps
sourcetype = fe_cef_syslog

[monitor:///mnt/*/log/*day/*/mail*/*]
host_segment = 6
index = mail
sourcetype = sendmail_syslog
```

props.conf

```
[source::/mnt/*/log/*/*/fireeye*/local2.log]
SHOULD_LINEMERGE = false
TZ = UTC

[source::/mnt/*/log/*/*/mail*/*]
SHOULD_LINEMERGE = false
TZ = US/Eastern
```

outputs.conf

- Most of Splunk’s pipeline queues default to a maximum size of 512KB. That may be fine for a universal forwarder monitoring local OS logs, but not for a syslog server.
- Is your output queue too small?

```
index=_internal host=<syslog_server> source=*metrics.log
group=queue name=tcpout* | eval
output_queue_pct=current_size/max_size*100 | timechart
perc95(output_queue_pct) by host | eval Bad=80
```

- A 64MB output queue works well for many enterprise syslog servers, but you may need more (if your 64MB queue is filling up) or less (if your RAM is filling up)
- Outputs.conf contents:

```
[tcpout]

maxQueueSize = 64MB
```

server.conf

- Like the outputs queue, most Splunk queues default to a maximum size of 512KB, which is often insufficient for a syslog server.
- Contents of server.conf:

```
 [queue]

maxSize = 64MB
```

- Be sure to take into account the number of active pipelines!
  - We discuss parallel ingestion pipeline configuration in a later slide

limits.conf

- Don’t forget to configure your limits.conf!
- If you use a universal forwarder, the maximum speed (per pipeline) defaults to 256kbps
  - That value will likely be insufficient for your syslog monitoring, so remove restrictions:

```
[thruput]
maxKBps = 0
```

- The maximum number of file descriptors that an ingestion pipeline in Splunk will keep open (per pipeline) defaults to 100
  - If you’re listening to this talk, this won’t be enough; you’ll likely need thousands
  - Use `` `find . | wc -l` `` in the `/log` directory to help you determine what this should be, but a few thousand is often a safe bet

```
[inputproc]
max_fd = <integer>
```

Better Balance Across Indexers

- For better load balancing, have the forwarders change indexers often and mid-stream
- outputs.conf:

```
[tcpout]
autoLBFrequency = 5
autoLBVolume = 100000000
forceTimebasedAutoLB = true
```

- On UFs running 6.5+, do not use forceTimebasedAutoLB. Add this to props.conf for each data source (or in `[default]`) instead:

```
EVENT_BREAKER_ENABLE = true
```

https://www.splunk.com/blog/2014/03/18/time-based-load-balancing.html

Parallel Ingestion Pipelines

- Parallel ingestion pipelines allow Splunk to use more resources so it can ingest multiple streams of data at once
- Since these syslog servers are dedicated to Splunk data collection, they’re excellent candidates for this feature
- The number of pipelines you set will depend on your hardware capacity and data rates. See notes on side effects of this setting at https://docs.splunk.com/Documentation/Splunk/latest/Admin/Serverconf
- Enabling parallel ingestion pipelines in server.conf:

```
[general]
parallelIngestionPipelines = 2
```

Automation

Scaling your Sysloging Splunk Infrastructure as Code

Syslog at Scale

- In a large enterprise, do not build syslog inputs and props manually!
  - With thousands of syslog feeds, they become impossible to manage
  - Small typos can cause massive failures

- We manage all of our syslog inputs in a CSV file
  - Human and machine readable
  - Easier to sort, group, find entries, and identify errors

- Find our script to auto-generate inputs.conf and props.conf for syslog servers at:
  - https://gitlab.com/rationalcyber/

Using a Catchall Index

- Sometimes upstream syslog sources start sending data you weren’t expecting
- You want this data in Splunk, but you don’t know what index or sourcetype to give it
- inputs.conf:

```
 [monitor:///mnt/log/*/*day]
blacklist = (fireeye.*/local2\.log)|(mail.*/.*)
index = catchall
```

- This blacklist regex becomes unmanageable quickly; the script on the previous slide auto-generates it for you
- It is often OK to assign the same index as `lastChanceIndex` even though this is a different usecase

Monitoring and Alerting

- Problems with one of the Splunk syslog servers (run every few minutes):

```
| tstats count where source=/mnt/*/log/* by source | rex field=source
"/mnt/(?<splunk_syslog_server>[^/]+)/" | stats sum(count), count by
splunk_syslog_server
```

- Problems with an upstream syslog server (run every few minutes):

```
| tstats count where source=/mnt/*/log/* by source | rex field=source
"/mnt/[^/]+/log/[^/]+/(?<upstream_syslog_server>[^/]+)/" | stats
sum(count), count by upstream_syslog_server
```

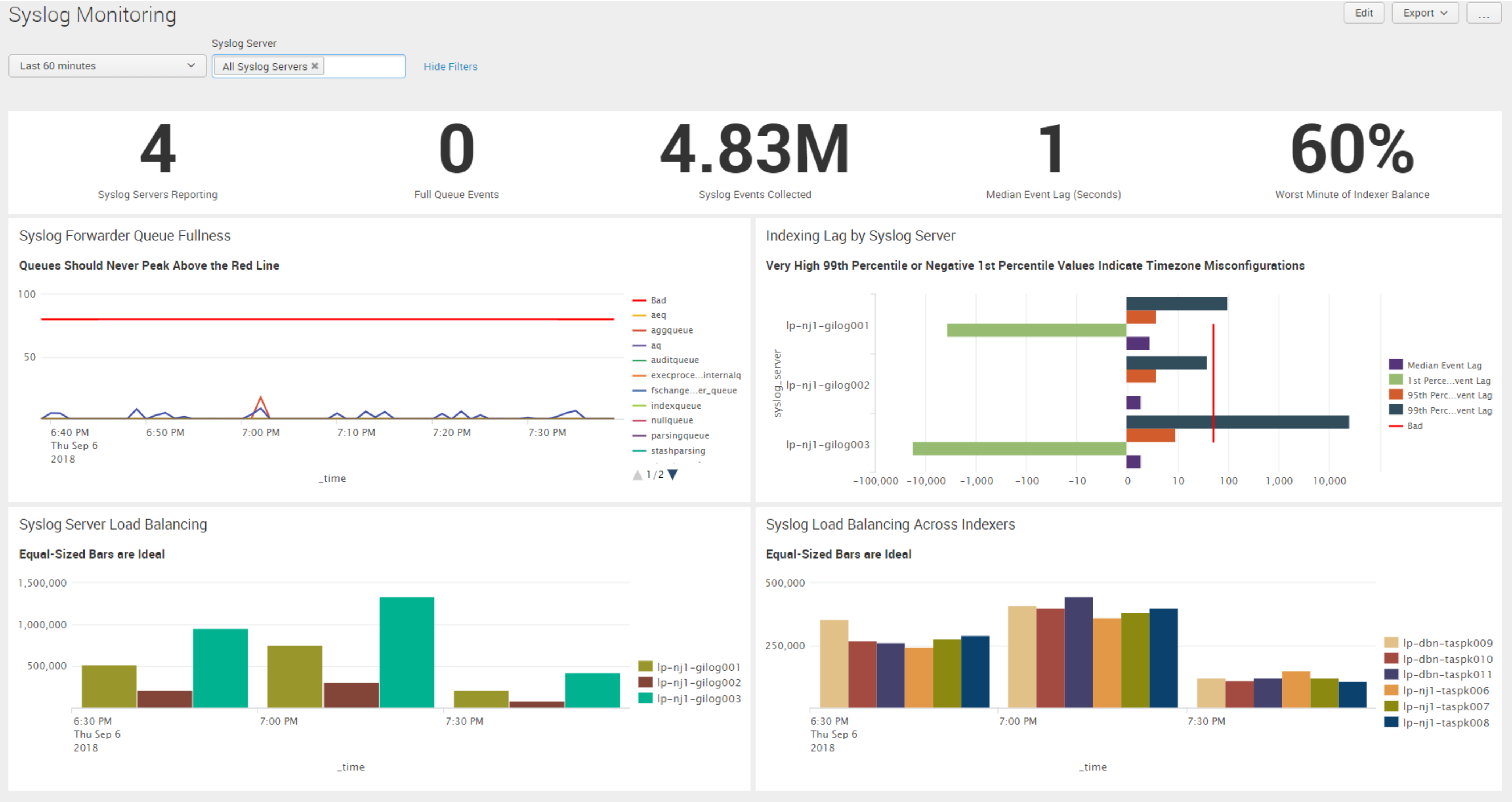
- Queues filling up and causing delays (observe daily—look for sustained issues):

```
index=_internal host=<syslog_server> source=*metrics.log group=queue
| eval queue_pct=if(isnull(current_size_kb), (current_size/max_size),
(current_size_kb/max_size_kb))*100 | timechart limit=50
perc99(queue_pct) by name | eval Bad=80
```

- Unknown syslog feeds (observe daily):

```
| tstats count where index=catchall by source
```

Our New “Syslog Tools for Splunk” App!



Syslog Monitoring

Syslog Server

Last 60 minutes

All Syslog Servers

4
Syslog Servers Reporting

0
Full Queue Events

4.83M
Syslog Events Collected

1
Median Event Lag (Seconds)

60%
Worst Minute of Indexer Balance

Syslog Forwarder Queue Fullness

Queues Should Never Peak Above the Red Line

Indexing Lag by Syslog Server

Very High 99th Percentile or Negative 1st Percentile Values Indicate Timezone Misconfigurations

Syslog Server Load Balancing

Equal-Sized Bars are Ideal

Syslog Load Balancing Across Indexers

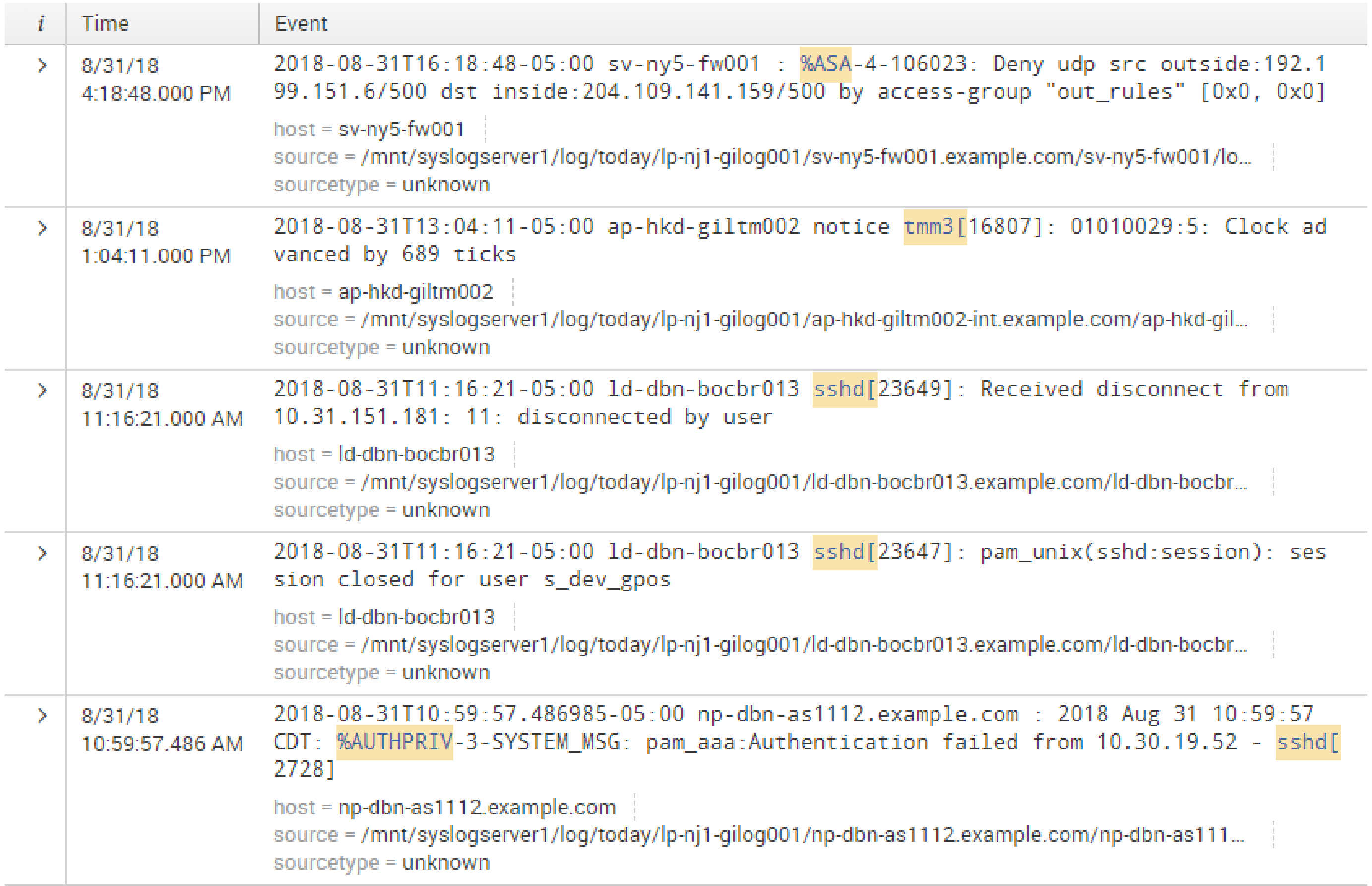
Equal-Sized Bars are Ideal

Automatic Sourcetyping

- Most of your syslog data sources are probably off-the-shelf products common to lots of Splunk environments
- A lot of the syslog data that ends up in your catchall index will probably be more of those same products—you just need to identify them
- For many of these products, Splunkbase already has add-ons you can drop in and get perfect parsing right away
- Can we take advantage of these facts?

Automatic Sourcetyping

- Trick #1: Keyword searches (note the yellow highlights)



i	Time	Event

>	8/31/18
	4:18:48.000 PM
		2018-08-31T16:18:48-05:00 sv-ny5-fw001 : %ASA-4-106023: Deny udp src outside:192.199.151.6/500 dst inside:204.109.141.159/500 by access-group "out_rules" [0x0, 0x0]
		host = sv-ny5-fw001
		source = /mnt/syslogserver1/log/today/lp-nj1-gilog001/sv-ny5-fw001.example.com/sv-ny5-fw001/lo...
		sourcetype = unknown

>	8/31/18
	1:04:11.000 PM
		2018-08-31T13:04:11-05:00 ap-hkd-giltm002 notice tmm3[16807]: 01010029:5: Clock advanced by 689 ticks
		host = ap-hkd-giltm002
		source = /mnt/syslogserver1/log/today/lp-nj1-gilog001/ap-hkd-giltm002-int.example.com/ap-hkd-gil...
		sourcetype = unknown

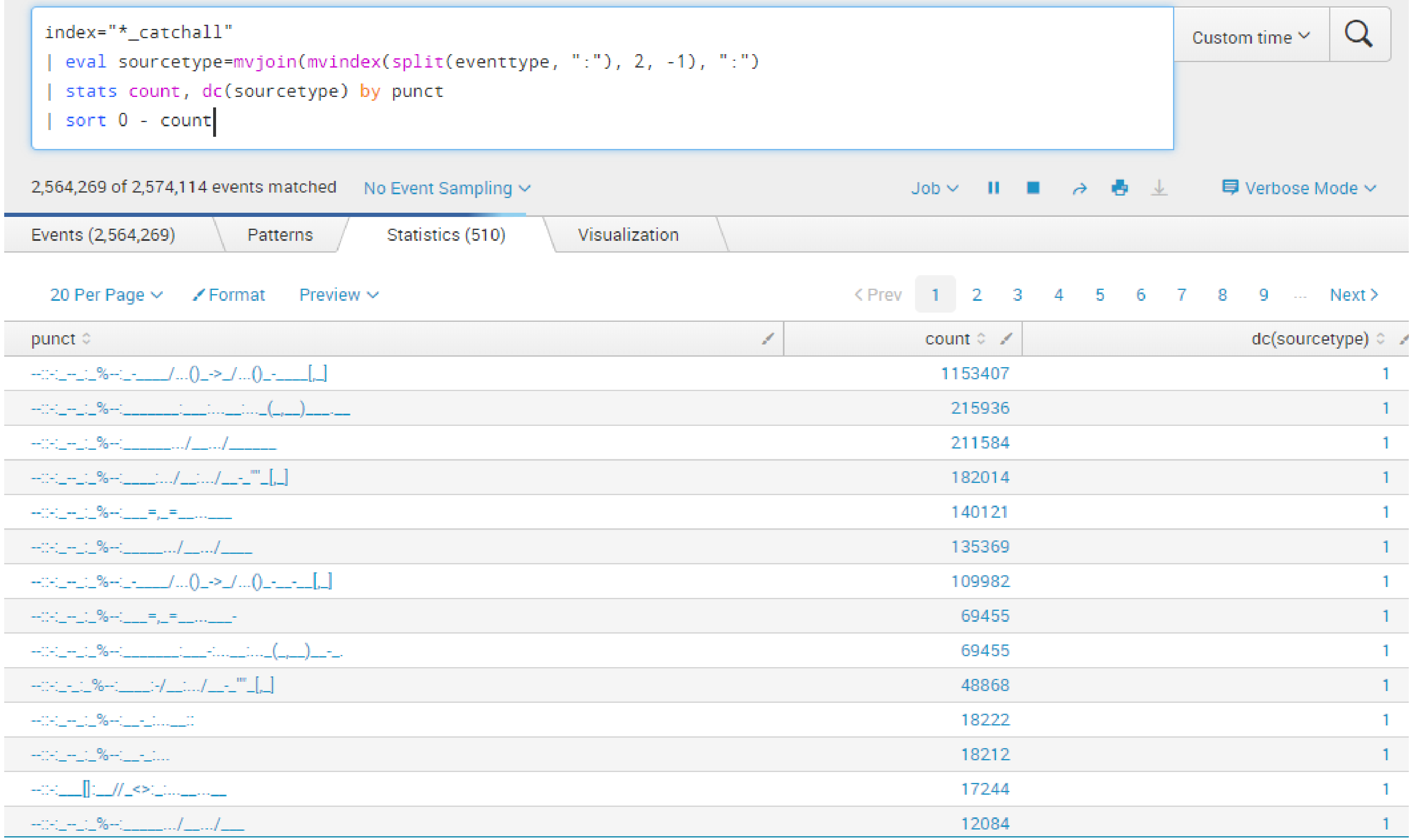
>	8/31/18
	11:16:21.000 AM
		2018-08-31T11:16:21-05:00 ld-dbn-bocbr013 sshd[23649]: Received disconnect from 10.31.151.181: 11: disconnected by user
		host = ld-dbn-bocbr013
		source = /mnt/syslogserver1/log/today/lp-nj1-gilog001/ld-dbn-bocbr013.example.com/ld-dbn-bocbr...
		sourcetype = unknown

>	8/31/18
	11:16:21.000 AM
		2018-08-31T11:16:21-05:00 ld-dbn-bocbr013 sshd[23647]: pam_unix(sshd:session): session closed for user s_dev_gpos
		host = ld-dbn-bocbr013
		source = /mnt/syslogserver1/log/today/lp-nj1-gilog001/ld-dbn-bocbr013.example.com/ld-dbn-bocbr...
		sourcetype = unknown

>	8/31/18
	10:59:57.486 AM
		2018-08-31T10:59:57.486985-05:00 np-dbn-as1112.example.com : 2018 Aug 31 10:59:57 CDT: %AUTHPRIV-3-SYSTEM_MSG: pam_aaa:Authentication failed from 10.30.19.52 - sshd[2728]
		host = np-dbn-as1112.example.com
		source = /mnt/syslogserver1/log/today/lp-nj1-gilog001/np-dbn-as1112.example.com/np-dbn-as111...
		sourcetype = unknown

Automatic Sourcetyping

- Trick #2: Punct

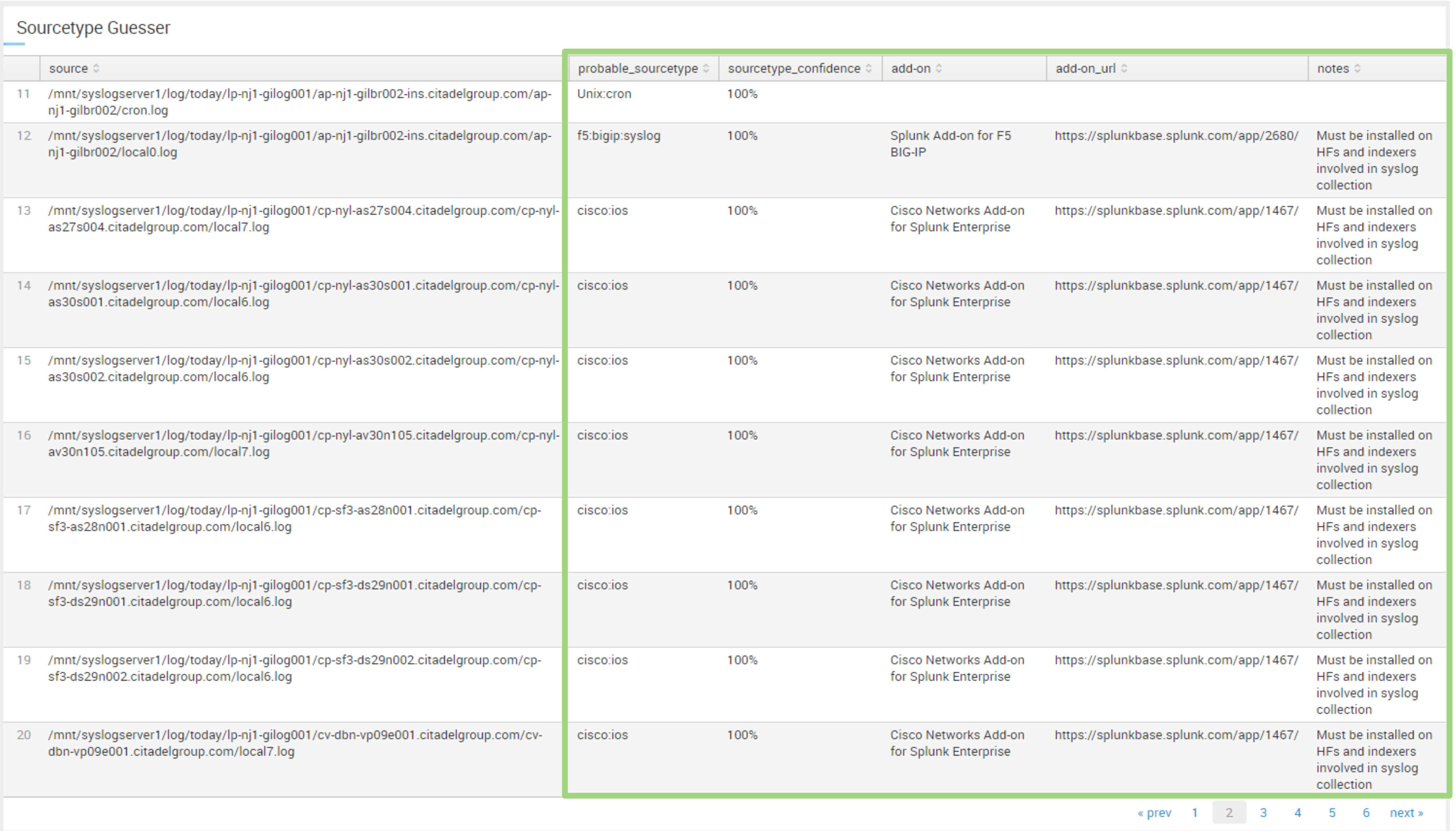


```
index="*_catchall"
| eval sourcetype=mvjoin(mvindex(split(eventtype, ":"), 2, -1), ":")
| stats count, dc(sourcetype) by punct
| sort 0 - count
```

punct	count	dc(sourcetype)
--::-:_--_:_%--:_-____/...()_->_/...()_-____[,_]	1153407	1
--::-:_--_:_%--:________:___:...__:..._(_,__)___-__	215936	1
--::-:_--_:_%--:_______.../__.../______	211584	1
--::-:_--_:_%--:_____:.../__:.../__-_""_[,_]	182014	1
--::-:_--_:_%--:____=,_=__...____	140121	1
--::-:_--_:_%--:______.../__.../____	135369	1
--::-:_--_:_%--:_-____/...()_->_/...()_-__-__[,_]	109982	1
--::-:_--_:_%--:____=,_=__...____-	69455	1
--::-:_--_:_%--:________:___-:...__:..._(_,__)__-_.	69455	1
--::-:_-_:_%--:_____:-/__:.../__-_""_[,_]	48868	1
--::-:_--_:_%--:__-_:...__::	18222	1
--::-:_--_:_%--:__-_:...	18212	1
--::-:___[]:__//_<>:_:...__...__	17244	1
--::-:_--_:_%--:______.../__.../___	12084	1

splunk> .conf18

Automatic Sourcetyping

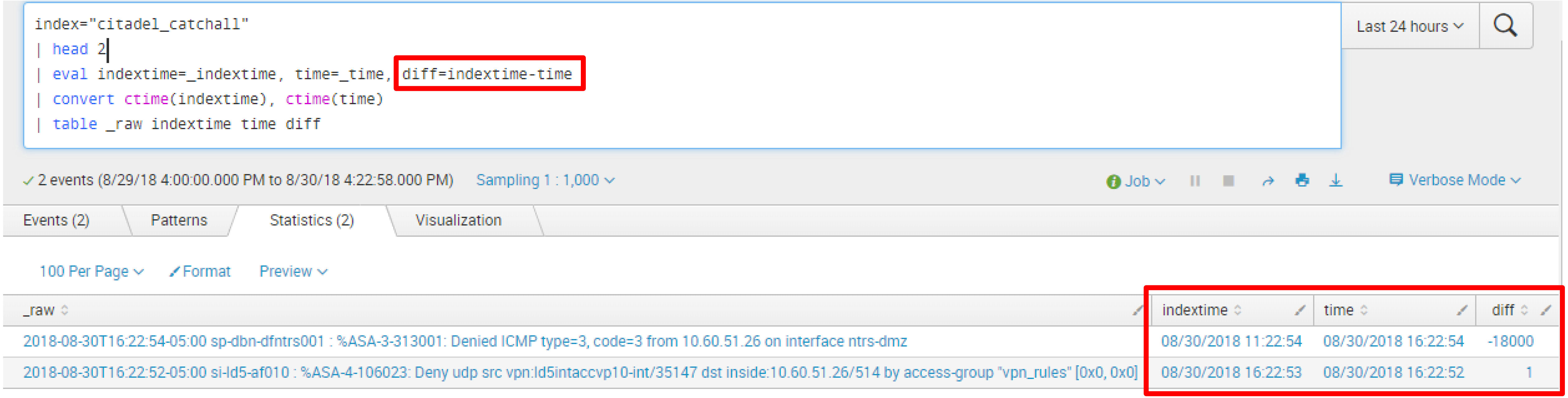


Sourcetype Guesser

	source	probable_sourcetype	sourcetype_confidence	add-on	add-on_url	notes
11	/mnt/syslogserver1/log/today/lp-nj1-gilog001/ap-nj1-gilbr002-ins.citadelgroup.com/ap-nj1-gilbr002/cron.log	Unix:cron	100%			
12	/mnt/syslogserver1/log/today/lp-nj1-gilog001/ap-nj1-gilbr002-ins.citadelgroup.com/ap-nj1-gilbr002/local0.log	f5:bigip:syslog	100%	Splunk Add-on for F5 BIG-IP	https://splunkbase.splunk.com/app/2680/	Must be installed on HFs and indexers involved in syslog collection
13	/mnt/syslogserver1/log/today/lp-nj1-gilog001/cp-nyl-as27s004.citadelgroup.com/cp-nyl-as27s004.citadelgroup.com/local7.log	cisco:ios	100%	Cisco Networks Add-on for Splunk Enterprise	https://splunkbase.splunk.com/app/1467/	Must be installed on HFs and indexers involved in syslog collection
14	/mnt/syslogserver1/log/today/lp-nj1-gilog001/cp-nyl-as30s001.citadelgroup.com/cp-nyl-as30s001.citadelgroup.com/local6.log	cisco:ios	100%	Cisco Networks Add-on for Splunk Enterprise	https://splunkbase.splunk.com/app/1467/	Must be installed on HFs and indexers involved in syslog collection
15	/mnt/syslogserver1/log/today/lp-nj1-gilog001/cp-nyl-as30s002.citadelgroup.com/cp-nyl-as30s002.citadelgroup.com/local6.log	cisco:ios	100%	Cisco Networks Add-on for Splunk Enterprise	https://splunkbase.splunk.com/app/1467/	Must be installed on HFs and indexers involved in syslog collection
16	/mnt/syslogserver1/log/today/lp-nj1-gilog001/cp-nyl-av30n105.citadelgroup.com/cp-nyl-av30n105.citadelgroup.com/local7.log	cisco:ios	100%	Cisco Networks Add-on for Splunk Enterprise	https://splunkbase.splunk.com/app/1467/	Must be installed on HFs and indexers involved in syslog collection
17	/mnt/syslogserver1/log/today/lp-nj1-gilog001/cp-sf3-as28n001.citadelgroup.com/cp-sf3-as28n001.citadelgroup.com/local6.log	cisco:ios	100%	Cisco Networks Add-on for Splunk Enterprise	https://splunkbase.splunk.com/app/1467/	Must be installed on HFs and indexers involved in syslog collection
18	/mnt/syslogserver1/log/today/lp-nj1-gilog001/cp-sf3-ds29n001.citadelgroup.com/cp-sf3-ds29n001.citadelgroup.com/local6.log	cisco:ios	100%	Cisco Networks Add-on for Splunk Enterprise	https://splunkbase.splunk.com/app/1467/	Must be installed on HFs and indexers involved in syslog collection
19	/mnt/syslogserver1/log/today/lp-nj1-gilog001/cp-sf3-ds29n002.citadelgroup.com/cp-sf3-ds29n002.citadelgroup.com/local6.log	cisco:ios	100%	Cisco Networks Add-on for Splunk Enterprise	https://splunkbase.splunk.com/app/1467/	Must be installed on HFs and indexers involved in syslog collection
20	/mnt/syslogserver1/log/today/lp-nj1-gilog001/cv-dbn-vp09e001.citadelgroup.com/cv-dbn-vp09e001.citadelgroup.com/local7.log	cisco:ios	100%	Cisco Networks Add-on for Splunk Enterprise	https://splunkbase.splunk.com/app/1467/	Must be installed on HFs and indexers involved in syslog collection

Automatic Timezoning

- What would need to figure out what timezone a host is logging in
  - What time does the host think it is right now?



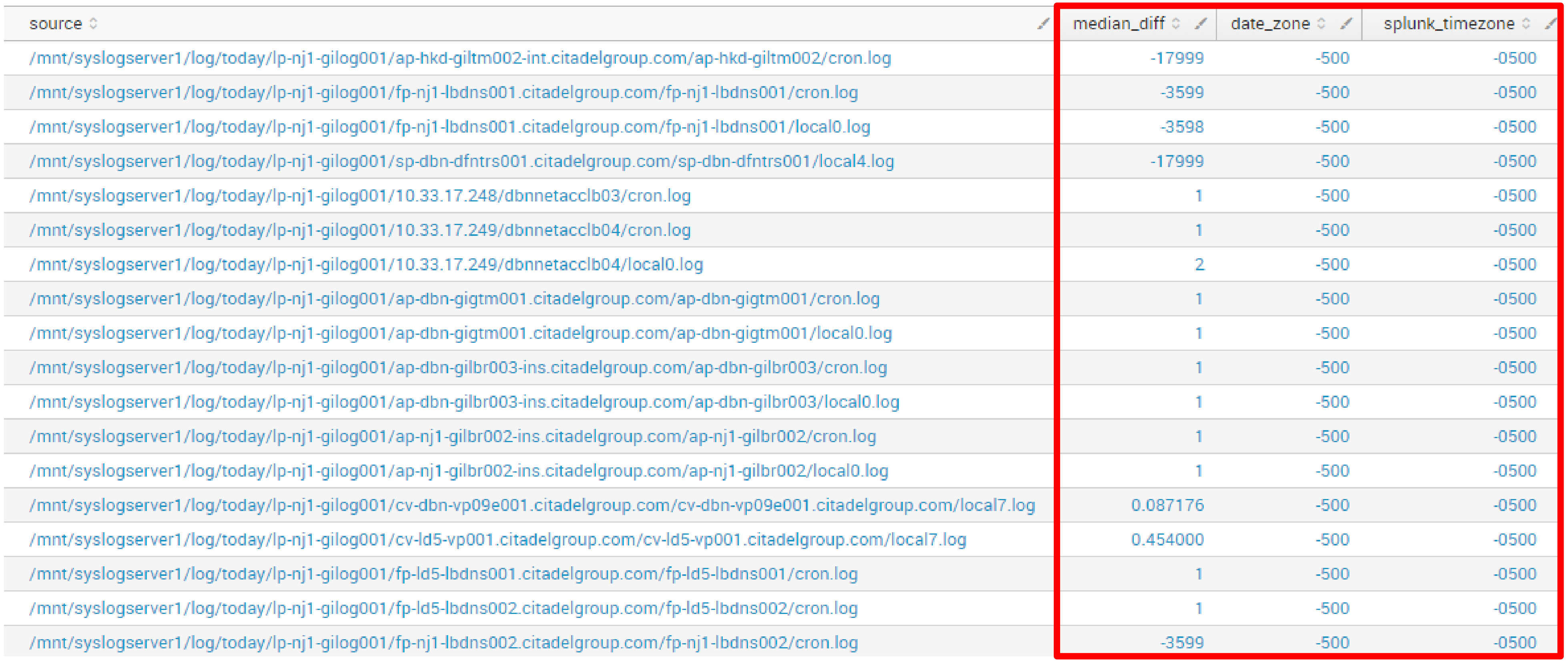
```
index="citadel_catchall"
| head 2
| eval indextime=_indextime, time=_time, diff=indextime-time
| convert ctime(indextime), ctime(time)
| table _raw indextime time diff
```

✓ 2 events (8/29/18 4:00:00.000 PM to 8/30/18 4:22:58.000 PM)

_raw	indextime	time	diff
2018-08-30T16:22:54-05:00 sp-dbn-dfntrs001 : %ASA-3-313001: Denied ICMP type=3, code=3 from 10.60.51.26 on interface ntrs-dmz	08/30/2018 11:22:54	08/30/2018 16:22:54	-18000
2018-08-30T16:22:52-05:00 si-ld5-af010 : %ASA-4-106023: Deny udp src vpn:ld5intaccvp10-int/35147 dst inside:10.60.51.26/514 by access-group "vpn_rules" [0x0, 0x0]	08/30/2018 16:22:53	08/30/2018 16:22:52	1

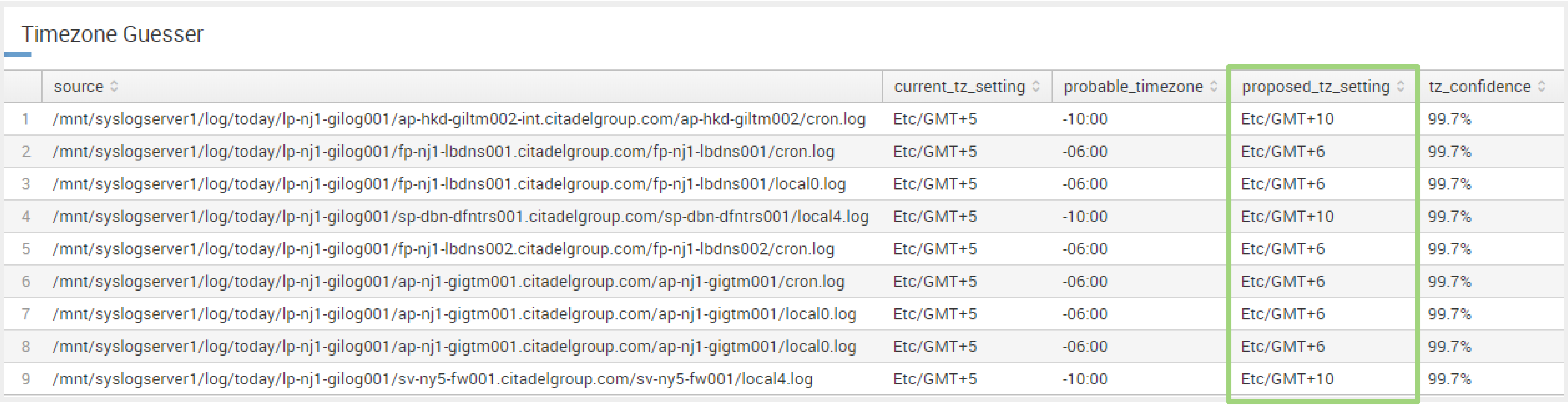
Automatic Timezoning

- What timezone does Splunk think the host is set to?
- What time does Splunk think it is right now?



source	median_diff	date_zone	splunk_timezone
/mnt/syslogserver1/log/today/lp-nj1-gilog001/ap-hkd-giltm002-int.citadelgroup.com/ap-hkd-giltm002/cron.log	-17999	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/fp-nj1-lbdns001.citadelgroup.com/fp-nj1-lbdns001/cron.log	-3599	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/fp-nj1-lbdns001.citadelgroup.com/fp-nj1-lbdns001/local0.log	-3598	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/sp-dbn-dfntrs001.citadelgroup.com/sp-dbn-dfntrs001/local4.log	-17999	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/10.33.17.248/dbnnetacclb03/cron.log	1	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/10.33.17.249/dbnnetacclb04/cron.log	1	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/10.33.17.249/dbnnetacclb04/local0.log	2	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/ap-dbn-gigtm001.citadelgroup.com/ap-dbn-gigtm001/cron.log	1	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/ap-dbn-gigtm001.citadelgroup.com/ap-dbn-gigtm001/local0.log	1	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/ap-dbn-gilbr003-ins.citadelgroup.com/ap-dbn-gilbr003/cron.log	1	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/ap-dbn-gilbr003-ins.citadelgroup.com/ap-dbn-gilbr003/local0.log	1	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/ap-nj1-gilbr002-ins.citadelgroup.com/ap-nj1-gilbr002/cron.log	1	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/ap-nj1-gilbr002-ins.citadelgroup.com/ap-nj1-gilbr002/local0.log	1	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/cv-dbn-vp09e001.citadelgroup.com/cv-dbn-vp09e001.citadelgroup.com/local7.log	0.087176	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/cv-ld5-vp001.citadelgroup.com/cv-ld5-vp001.citadelgroup.com/local7.log	0.454000	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/fp-ld5-lbdns001.citadelgroup.com/fp-ld5-lbdns001/cron.log	1	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/fp-ld5-lbdns002.citadelgroup.com/fp-ld5-lbdns002/cron.log	1	-500	-0500
/mnt/syslogserver1/log/today/lp-nj1-gilog001/fp-nj1-lbdns002.citadelgroup.com/fp-nj1-lbdns002/cron.log	-3599	-500	-0500

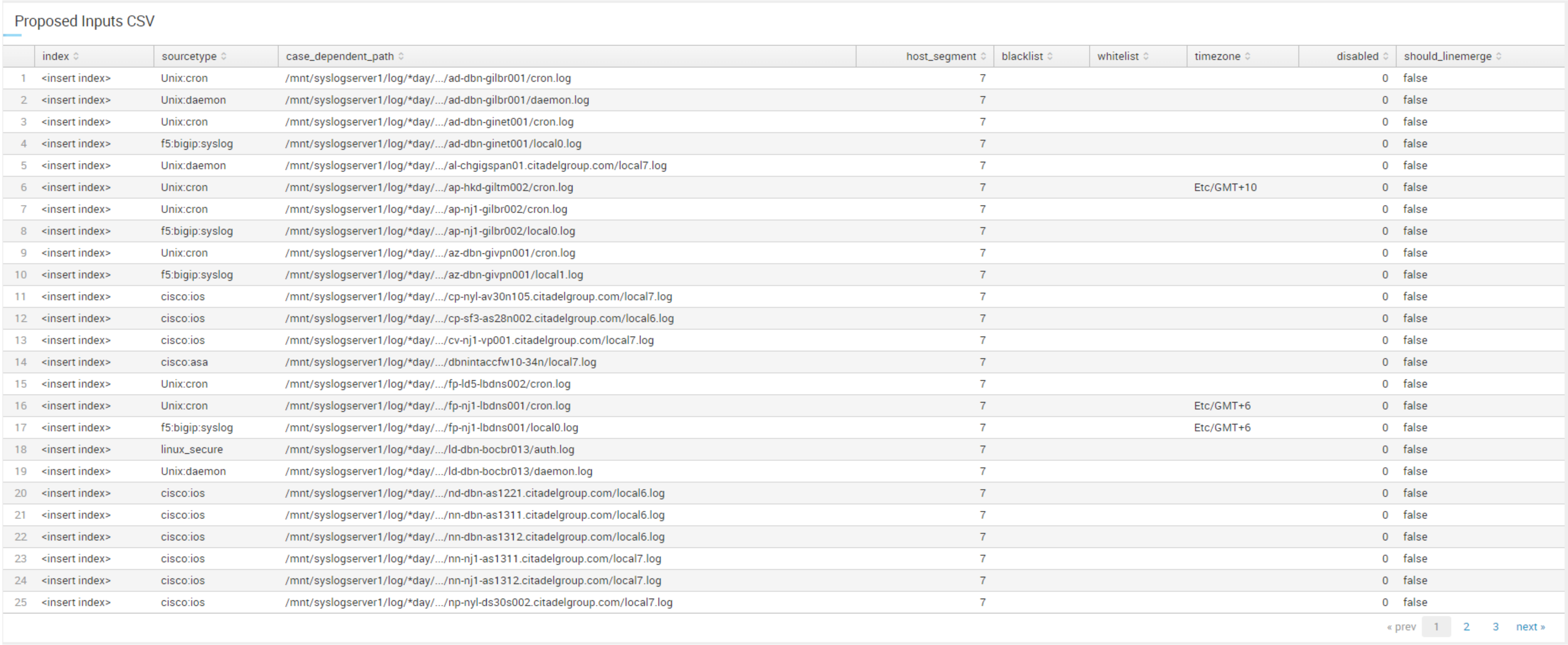
Automatic Timezoning



Timezone Guesser

	source	current_tz_setting	probable_timezone	proposed_tz_setting	tz_confidence
1	/mnt/syslogserver1/log/today/lp-nj1-gilog001/ap-hkd-giltm002-int.citadelgroup.com/ap-hkd-giltm002/cron.log	Etc/GMT+5	-10:00	Etc/GMT+10	99.7%
2	/mnt/syslogserver1/log/today/lp-nj1-gilog001/fp-nj1-lbdns001.citadelgroup.com/fp-nj1-lbdns001/cron.log	Etc/GMT+5	-06:00	Etc/GMT+6	99.7%
3	/mnt/syslogserver1/log/today/lp-nj1-gilog001/fp-nj1-lbdns001.citadelgroup.com/fp-nj1-lbdns001/local0.log	Etc/GMT+5	-06:00	Etc/GMT+6	99.7%
4	/mnt/syslogserver1/log/today/lp-nj1-gilog001/sp-dbn-dfntrs001.citadelgroup.com/sp-dbn-dfntrs001/local4.log	Etc/GMT+5	-10:00	Etc/GMT+10	99.7%
5	/mnt/syslogserver1/log/today/lp-nj1-gilog001/fp-nj1-lbdns002.citadelgroup.com/fp-nj1-lbdns002/cron.log	Etc/GMT+5	-06:00	Etc/GMT+6	99.7%
6	/mnt/syslogserver1/log/today/lp-nj1-gilog001/ap-nj1-gigtm001.citadelgroup.com/ap-nj1-gigtm001/cron.log	Etc/GMT+5	-06:00	Etc/GMT+6	99.7%
7	/mnt/syslogserver1/log/today/lp-nj1-gilog001/ap-nj1-gigtm001.citadelgroup.com/ap-nj1-gigtm001/local0.log	Etc/GMT+5	-06:00	Etc/GMT+6	99.7%
8	/mnt/syslogserver1/log/today/lp-nj1-gilog001/ap-nj1-gigtm001.citadelgroup.com/ap-nj1-gigtm001/local0.log	Etc/GMT+5	-06:00	Etc/GMT+6	99.7%
9	/mnt/syslogserver1/log/today/lp-nj1-gilog001/sv-ny5-fw001.citadelgroup.com/sv-ny5-fw001/local4.log	Etc/GMT+5	-10:00	Etc/GMT+10	99.7%

Automatically Generate Inputs CSVs



Proposed Inputs CSV

	index	sourcetype	case_dependent_path	host_segment	blacklist	whitelist	timezone	disabled	should_linemerge
1	<insert index>	Unix:cron	/mnt/syslogserver1/log/*day/.../ad-dbn-gilbr001/cron.log	7				0	false
2	<insert index>	Unix:daemon	/mnt/syslogserver1/log/*day/.../ad-dbn-gilbr001/daemon.log	7				0	false
3	<insert index>	Unix:cron	/mnt/syslogserver1/log/*day/.../ad-dbn-ginet001/cron.log	7				0	false
4	<insert index>	f5:bigip:syslog	/mnt/syslogserver1/log/*day/.../ad-dbn-ginet001/local0.log	7				0	false
5	<insert index>	Unix:daemon	/mnt/syslogserver1/log/*day/.../al-chgigspan01.citadelgroup.com/local7.log	7				0	false
6	<insert index>	Unix:cron	/mnt/syslogserver1/log/*day/.../ap-hkd-giltm002/cron.log	7			Etc/GMT+10	0	false
7	<insert index>	Unix:cron	/mnt/syslogserver1/log/*day/.../ap-nj1-gilbr002/cron.log	7				0	false
8	<insert index>	f5:bigip:syslog	/mnt/syslogserver1/log/*day/.../ap-nj1-gilbr002/local0.log	7				0	false
9	<insert index>	Unix:cron	/mnt/syslogserver1/log/*day/.../az-dbn-givpn001/cron.log	7				0	false
10	<insert index>	f5:bigip:syslog	/mnt/syslogserver1/log/*day/.../az-dbn-givpn001/local1.log	7				0	false
11	<insert index>	cisco:ios	/mnt/syslogserver1/log/*day/.../cp-nyl-av30n105.citadelgroup.com/local7.log	7				0	false
12	<insert index>	cisco:ios	/mnt/syslogserver1/log/*day/.../cp-sf3-as28n002.citadelgroup.com/local6.log	7				0	false
13	<insert index>	cisco:ios	/mnt/syslogserver1/log/*day/.../cv-nj1-vp001.citadelgroup.com/local7.log	7				0	false
14	<insert index>	cisco:asa	/mnt/syslogserver1/log/*day/.../dbnintaccfw10-34n/local7.log	7				0	false
15	<insert index>	Unix:cron	/mnt/syslogserver1/log/*day/.../fp-ld5-lbdns002/cron.log	7				0	false
16	<insert index>	Unix:cron	/mnt/syslogserver1/log/*day/.../fp-nj1-lbdns001/cron.log	7			Etc/GMT+6	0	false
17	<insert index>	f5:bigip:syslog	/mnt/syslogserver1/log/*day/.../fp-nj1-lbdns001/local0.log	7			Etc/GMT+6	0	false
18	<insert index>	linux_secure	/mnt/syslogserver1/log/*day/.../ld-dbn-bocbr013/auth.log	7				0	false
19	<insert index>	Unix:daemon	/mnt/syslogserver1/log/*day/.../ld-dbn-bocbr013/daemon.log	7				0	false
20	<insert index>	cisco:ios	/mnt/syslogserver1/log/*day/.../nd-dbn-as1221.citadelgroup.com/local6.log	7				0	false
21	<insert index>	cisco:ios	/mnt/syslogserver1/log/*day/.../nn-dbn-as1311.citadelgroup.com/local6.log	7				0	false
22	<insert index>	cisco:ios	/mnt/syslogserver1/log/*day/.../nn-dbn-as1312.citadelgroup.com/local6.log	7				0	false
23	<insert index>	cisco:ios	/mnt/syslogserver1/log/*day/.../nn-nj1-as1311.citadelgroup.com/local7.log	7				0	false
24	<insert index>	cisco:ios	/mnt/syslogserver1/log/*day/.../nn-nj1-as1312.citadelgroup.com/local7.log	7				0	false
25	<insert index>	cisco:ios	/mnt/syslogserver1/log/*day/.../np-nyl-ds30s002.citadelgroup.com/local7.log	7				0	false

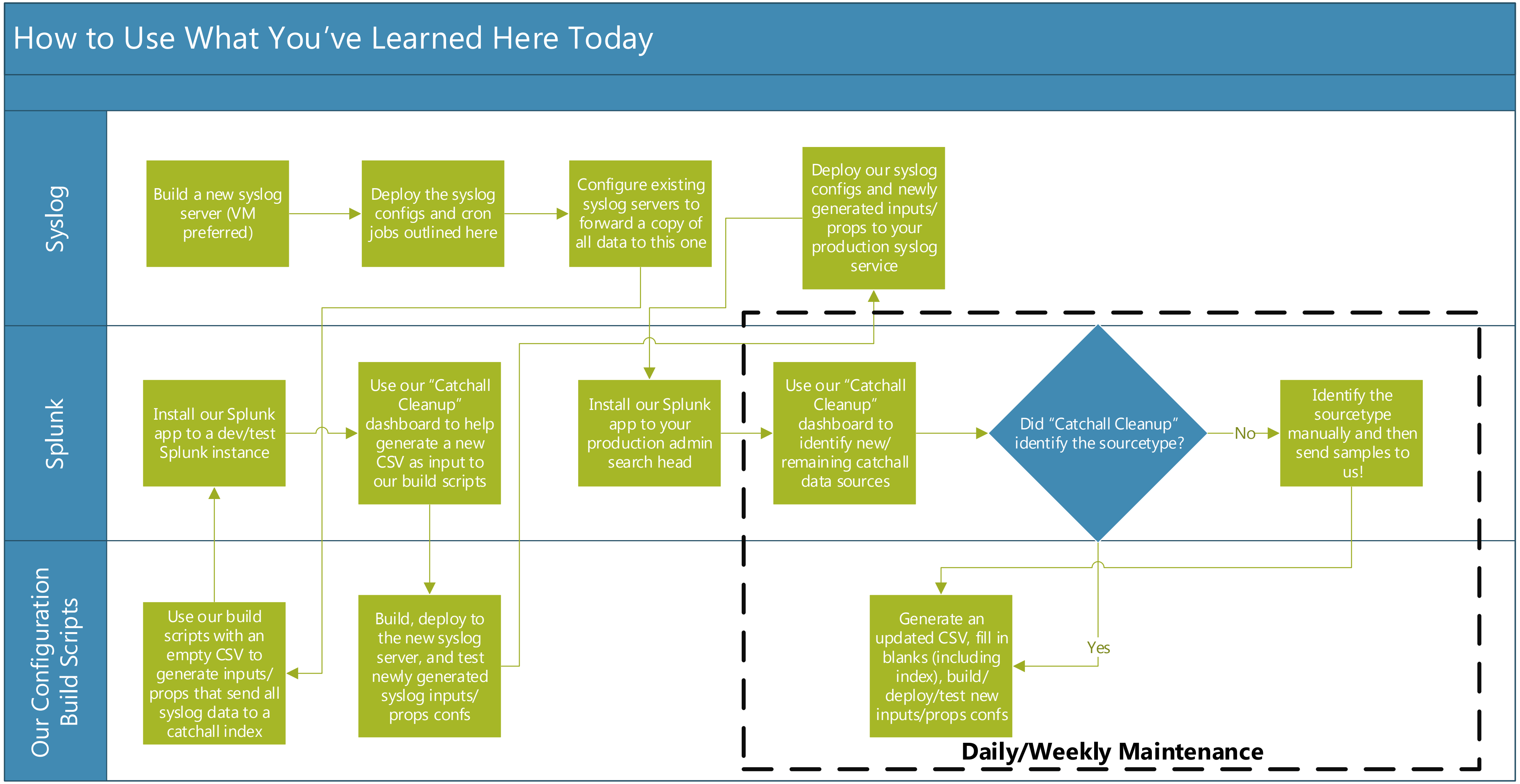
We Need Your Help!

- The source type automation relies on a library of keywords and punctuation
- That library is stored in a CSV file in our open source Syslog Tools for Splunk App

  https://gitlab.com/rationalcyber

- Please submit your samples!
  - If there's a unique keyword (or multiple unique keywords) in a given sourcetype's events, please include it
  - If there are no reliable keywords that distinguish a sourcetype, please submit common samples of punct
    - | tstats count where index=* sourcetype=example by punct | sort 10 - count
    - Please replace environment-specific portions of the punct, such as timestamp formatting driven by your syslog server, with asterisks

How Should I Plan to Deploy All of This?



How to Use What You've Learned Here Today

Syslog

Build a new syslog server (VM preferred)

Deploy the syslog configs and cron jobs outlined here

Configure existing syslog servers to forward a copy of all data to this one

Deploy our syslog configs and newly generated inputs/ props to your production syslog service

Splunk

Install our Splunk app to a dev/test Splunk instance

Use our "Catchall Cleanup" dashboard to help generate a new CSV as input to our build scripts

Install our Splunk app to your production admin search head

Use our "Catchall Cleanup" dashboard to identify new/ remaining catchall data sources

Did "Catchall Cleanup" identify the sourcetype?

No

Identify the sourcetype manually and then send samples to us!

Our Configuration Build Scripts

Use our build scripts with an empty CSV to generate inputs/ props that send all syslog data to a catchall index

Build, deploy to the new syslog server, and test newly generated syslog inputs/ props confs

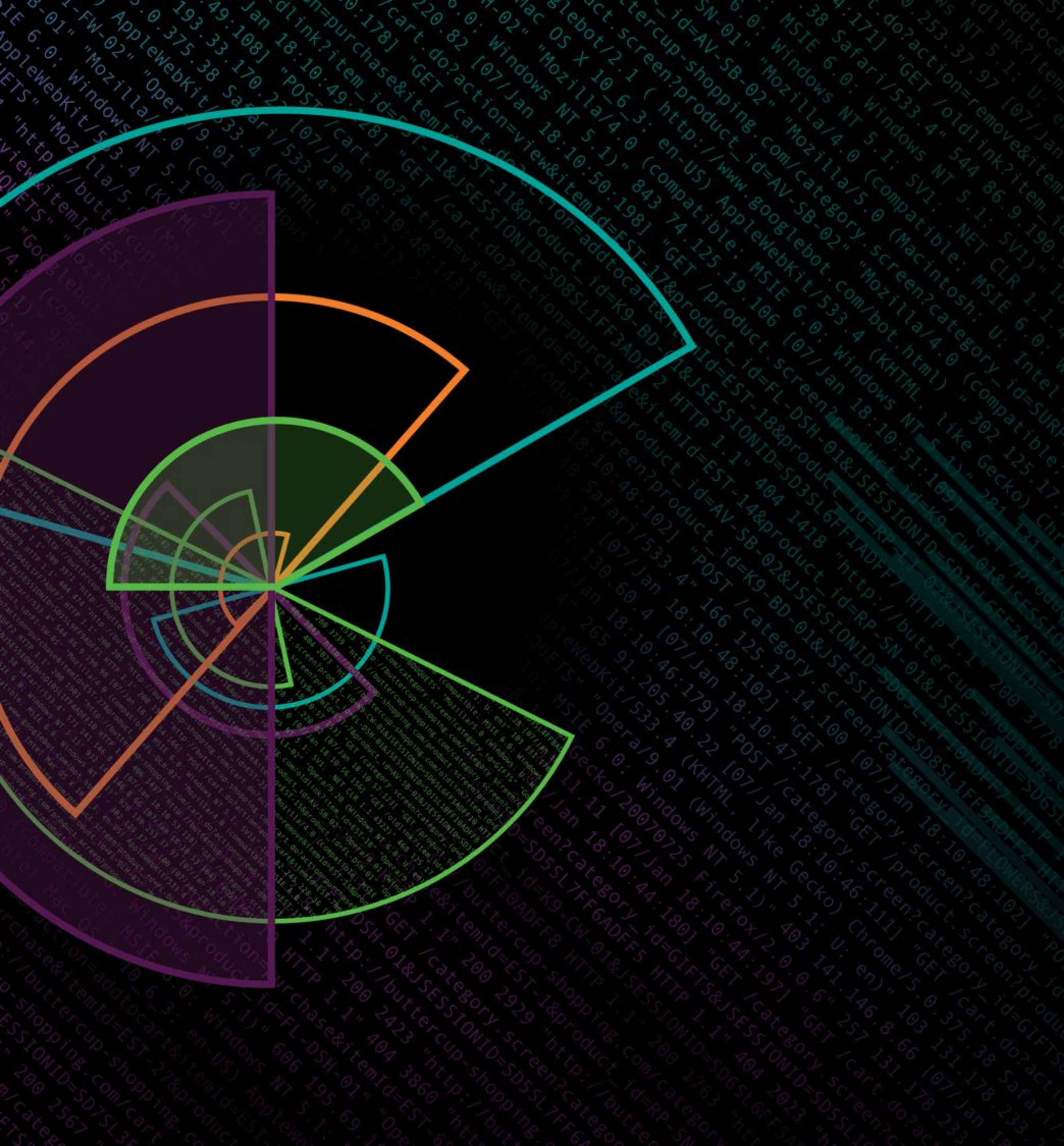
Yes

Generate an updated CSV, fill in blanks (including index), build/ deploy/test new inputs/props confs

Daily/Weekly Maintenance

Previous Slide Summarized

- Assume your syslog deployment has issues
- It’s easier to start over than to fix half a problem
- Stand up a new syslog server using our syslog configs and forward it all the logs from your existing syslog servers
- Send all the logs to a catchall index and use our app to autogenerate clean new configs with accurate sourcetypes and timezones
- Deploy and test the clean configs
- Rebuild/reconfigure all your Splunk syslog servers to act like the new one



HEC/Kafka

Scaling your syslog intermediaries

To HEC (and Kafka?) With Syslog!

- Last year, Mark Bonsack and Ryan Faircloth discussed coupling HEC to syslog for scalable aggregated data collection
  - “HEC” is **HTTP Event Collection** at the indexers
  - Primary use case dealt with improving the distribution of events across indexers
  - The standard script is at https://bitbucket.org/rfaircloth-splunk/rsyslog-omsplunk

- More recently, Splunk released **Splunk Connect for Kafka** on splunkbase
  - “Kafka” is the Apache project for building real-time data pipelines
  - Some folks are considering using Kafka stream processors to parse syslog and send it to Splunk and elsewhere
  - https://docs.splunk.com/Documentation/KafkaConnect/1.0.0/User/ConfigureSplunkKafkaConnect

Simplify Syslog Configs and Scale at the Same Time

- The main issue we address is scaling the connection of syslog with other intermediary technologies in large Splunk environments
  - I.e., many types of syslog data from many hosts in multiple time zones need to end up in the correct indexes with all the right metadata assigned
  - Spawning hundreds of HEC python scripts or hundreds of Kafka topics may not be reasonable (syslog-ng program() destination kicks off once)

- Solution:
  - Reuse the lessons we’ve discussed for syslog and forwarders!

Best of Both Worlds for Syslog and Intermediary Streamers

- Keep the complexity out of syslog-ng.conf
- Have syslog-ng pass the relevant metadata to the python glue script (json FTW!)
- Use the /event endpoint instead of /raw
- Make the automation we’ve discussed more dynamic
  - Python script maps event with its `$HOST` and `$FACILITY` metadata to sourcetype and index; fixes timezone issues with `$S_UNIXTIME` offset
  - Script behaves just like legacy omsplunkhec.py (“To HEC with syslog!”) if no lookup exists for mapping

Syslog-ng Config for HEC/Kafka with Python Script

```
destination d_program_json {
    program("/usr/local/bin/syslog2all.py <token> <server> \
          --index=catchall \
          --sourcetype=stash \
          --lookup=/usr/local/bin/splunk_map.csv"
          template("$(format-json event=$MESSAGE \
                host=$HOST \
                host_from=$HOST_FROM \
                facility=$FACILITY \
                time=$S_UNIXTIME)\n")
    );
};
```

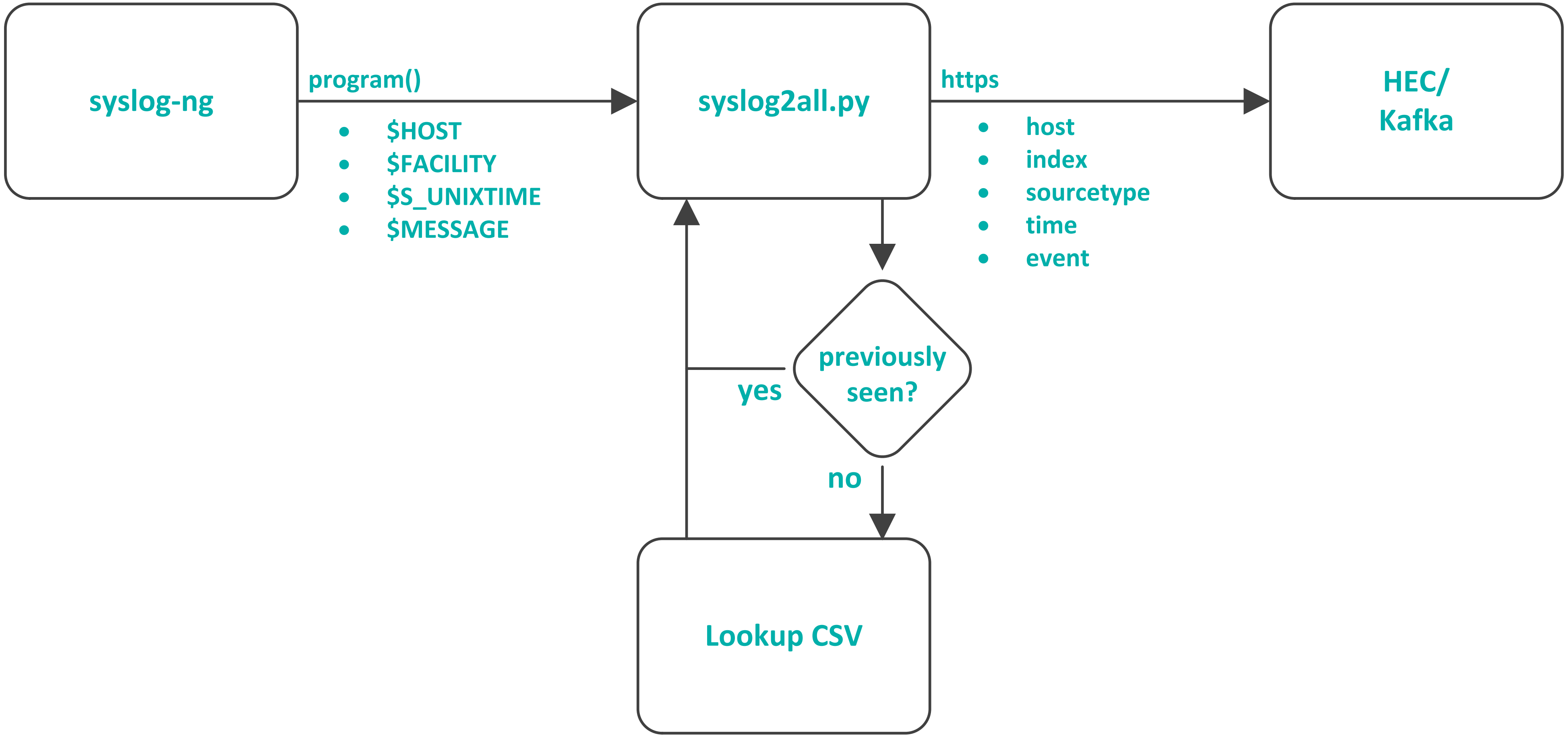
This is the most important line!

template("$(**format-json** event=**$MESSAGE** host=**$HOST**
host_from=**$HOST_FROM** facility=**$FACILITY** time=**$S_UNIXTIME**)\n")

- Does this metadata look familiar?
  - It’s the same info we used to build our standardized directory tree for writing syslog to files

- **format-json :** Used to safely package the data for the python script
- **$HOST_FROM :** "The host I received this feed from." It may be the same as the originating host, or it may be an intermediate syslog server. In the latter case, helps with troubleshooting.
- **$HOST and $FACILITY :** Used to lookup index and sourcetype
- **$S_UNIXTIME :** Used, along with an offset, to set /event endpoint time attribute (timestamps are not parsed from the message text when using the /event endpoint!)

syslog2all.py



How to Choose Your Collection Method

	I have tightly limited bandwidth between syslog servers and indexers	I'm only collecting data from a single time zone	I can't get my syslog load balanced across indexers even with autoLBFrequency=5	My company is already using Kafka to collect logs, or we'd benefit from the pub/sub model for other log consumers	I need to organize my syslog data into tens or hundreds of indexes to enforce least privilege	I have storage to safely keep hours or days of raw logs on my syslog servers	My syslog server is barely keeping up as it is
Yes	Universal Forwarder, HEC, or Kafka	Forwarder, HEC, or Kafka	HEC or Kafka to an HTTP load balancer with persistence turned off	Forwarder or Kafka	Forwarder, or HEC/Kafka with syslog4all.py	Forwarder, HEC, or Kafka	Forwarder (but really, enhance your server if you can)
No	Heavy Forwarder, HEC, or Kafka	Forwarder, or HEC/Kafka with syslog4all.py	Forwarder, HEC, or Kafka	Forwarder, HEC, or Kafka	Forwarder, HEC, or Kafka	HEC or Kafka	Forwarder, HEC, or Kafka

- For most use cases, we still recommend the forwarder, but we love the work Mark Bonsack, Scott Haskell, and Ryan Faircloth are doing on HEC/Kafka

Thank you!

- george@rationalcyber.com
  - Twitter: @RationalCyber
  - Slack usergroup: @chulobo
- Jonathan.Margulies@citadel.com
  - Twitter: @UnsaltedHash
- All of our open source projects, including all of our syslog resources:
  - https://gitlab.com/rationalcyber/

Want to learn more?

- HEC Yeah!! How Priceline Uses HEC to Ingest 4TB of Data Every Day
  - Wednesday, Oct 03, 4:30 p.m. - 5:15 p.m.
    - Mukund Murthy, Software Engineer, Priceline.com
    - Jagadeesh Motamarri, Senior Software Engineer, priceline.com LLC
- Old Meets New: Syslog and Splunk Connect for Kafka
  - Tuesday, Oct 02, 2:15 p.m. - 3:00 p.m.
    - Scott Haskell, Principal SE Architect, Splunk
    - Mark Bonsack, Staff Sales Engineer, Splunk
- Security Use Cases in Record Time With Splunk and Kafka
  - Wednesday, Oct 03, 3:15 p.m. - 4:00 p.m.
    - Nikolas Macroglou, Sales Engineer, Splunk
    - Lock Langdon, Global Director - Security Analytics at McKesson, McKesson

The Critical Tricks to Getting Syslog Into Splunk the Right Way

Don't forget to rate this session in the .conf18 mobile app