Security Ninjutsu Part Six

Campfire Stories of Demons and Bad People

David Veuve

Principal Security Strategist | Splunk

@davidveuve

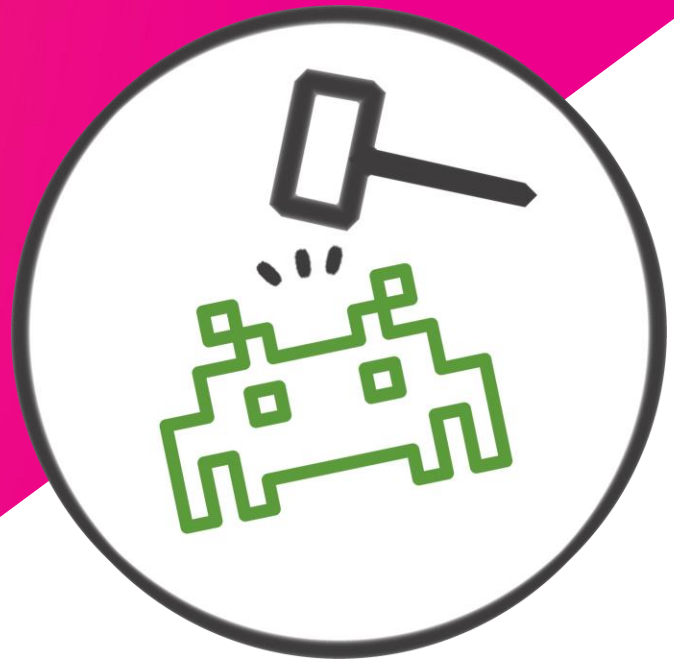
splunk> .conf19

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or plans of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results may differ materially. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, it may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements made herein.

In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only, and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionalities described or to include any such feature or functionality in a future release.

Past Ninjutsus



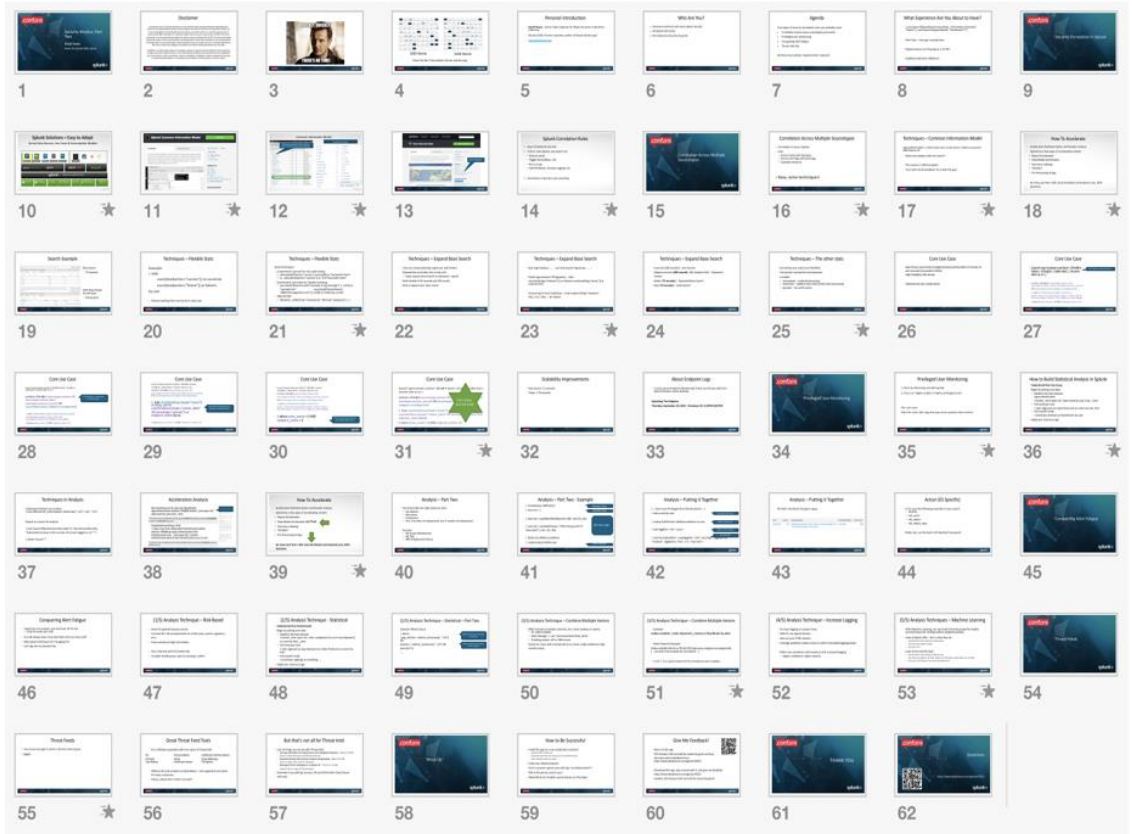
.conf19

splunk>

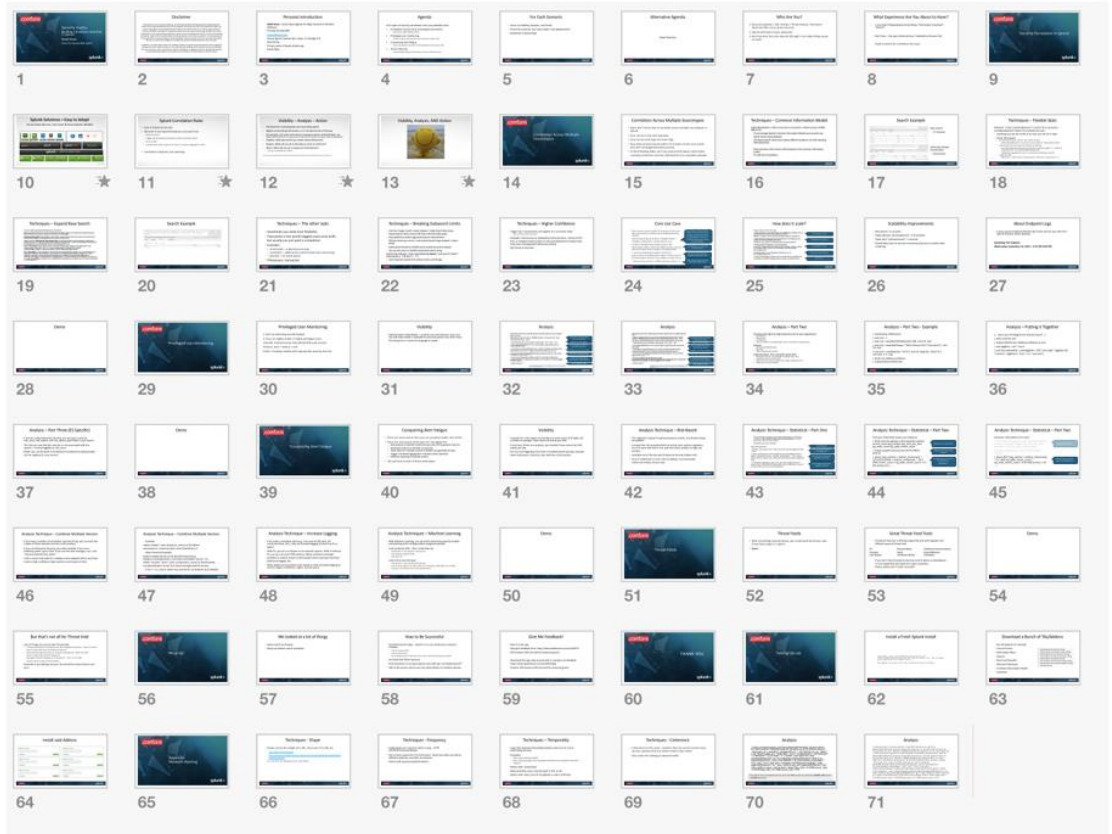
Lots of Ninjutsus

Security Ninjutsu Series

- .conf18 -- Security Ninjutsu Part Five: Our SPL Goes to 12.. 11 Isn't Enough
- .conf2017 -- Security Ninjutsu Part Four: Attackers Be Gone in 45 Minutes of Epic SPL
- .conf2016 -- Security Ninjutsu Part Three: Real-World Correlation Searches
- .conf2015 -- Security Ninjutsu Part Two: More Security Analytics, Correlation and Action!
- .conf2014 -- Security Ninjutsu: Using Splunk for Advanced Correlation, Anomaly Detection and Response Automation



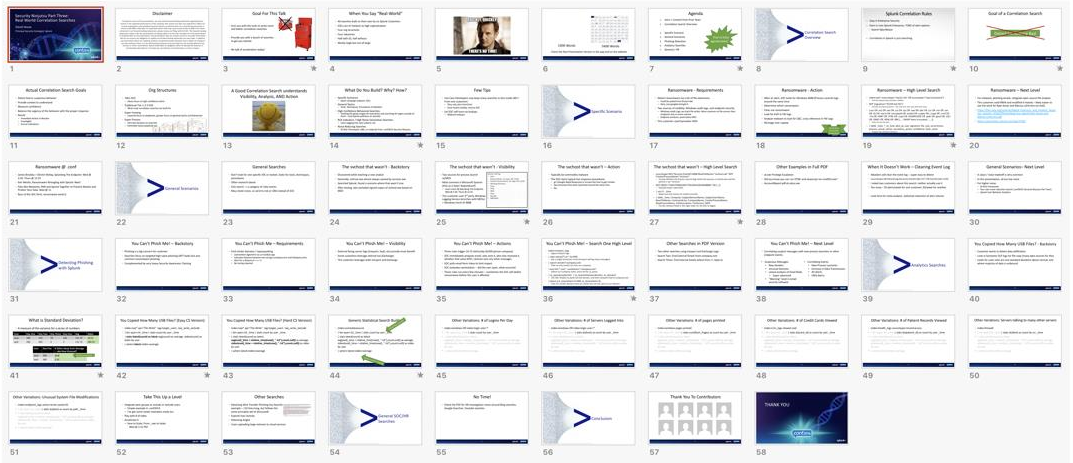
1800 Words



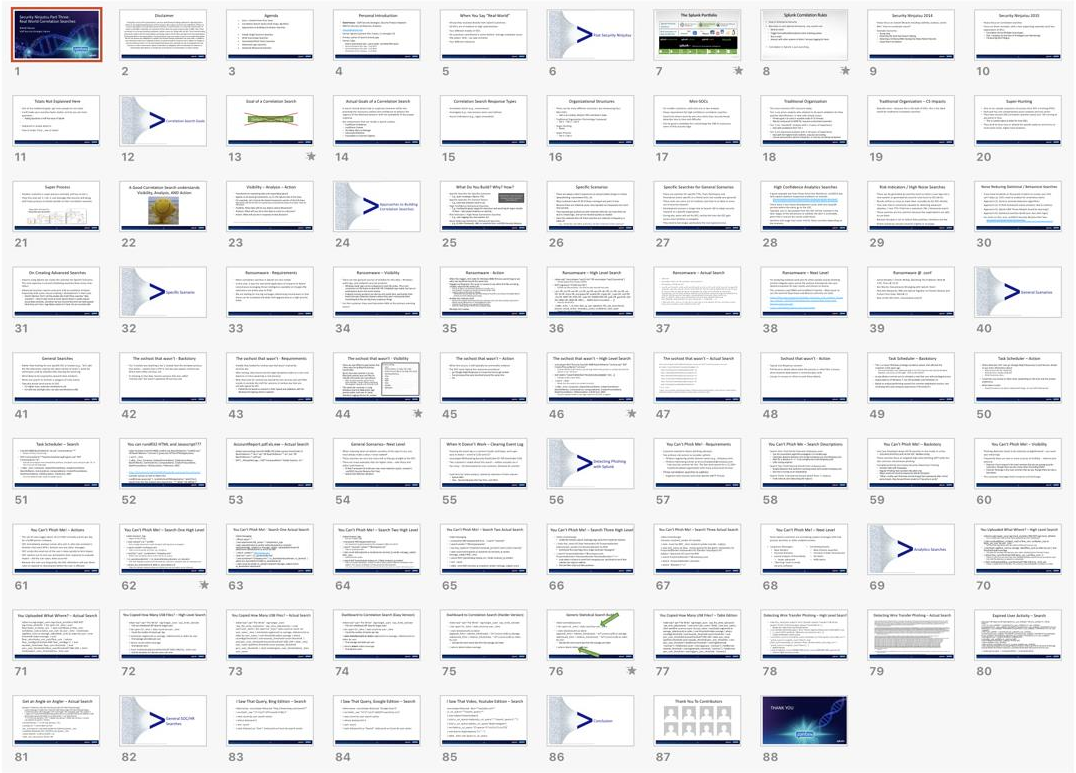
3200 Words

Check the Non-Presentation Version and the App

#TBT 2015



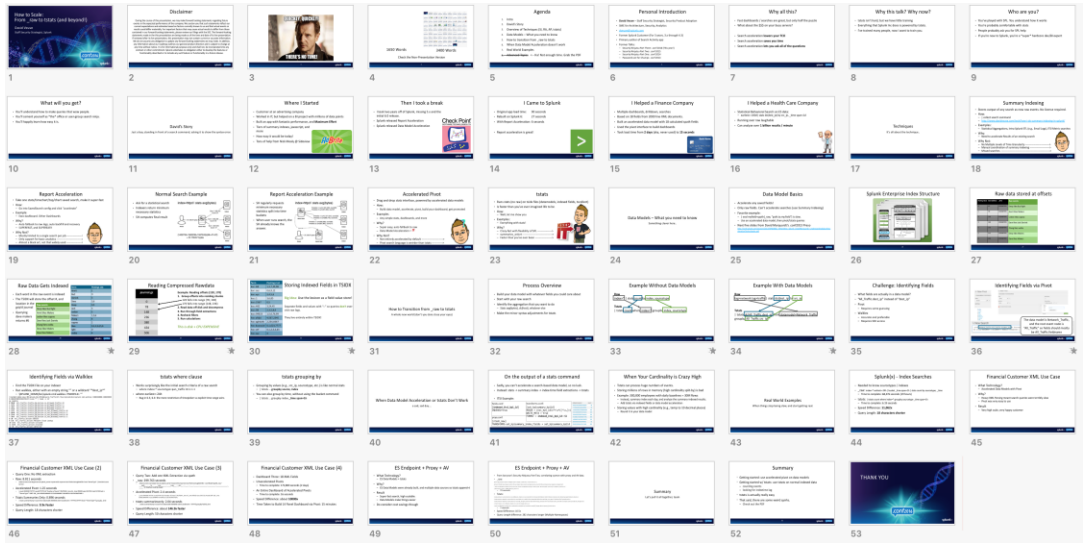
1900 Words



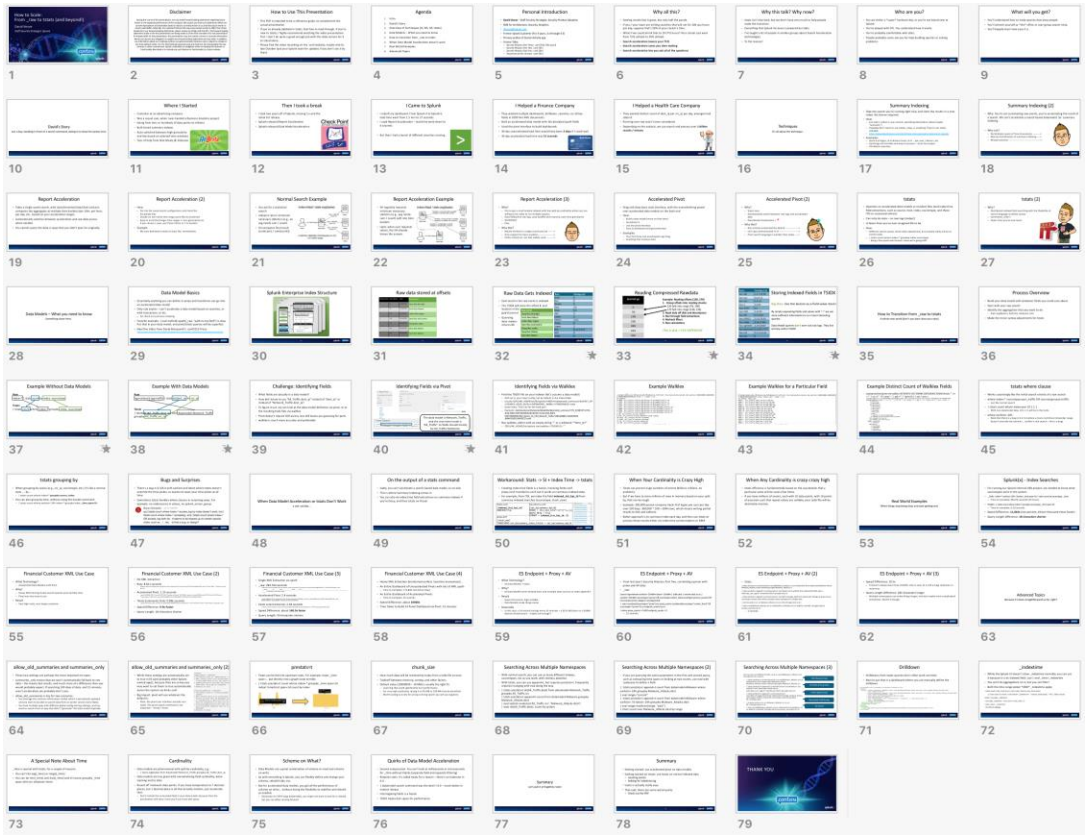
5600 Words

Check the Non-Presentation Version in the app and on the website

#TBT 2016



2400 Words

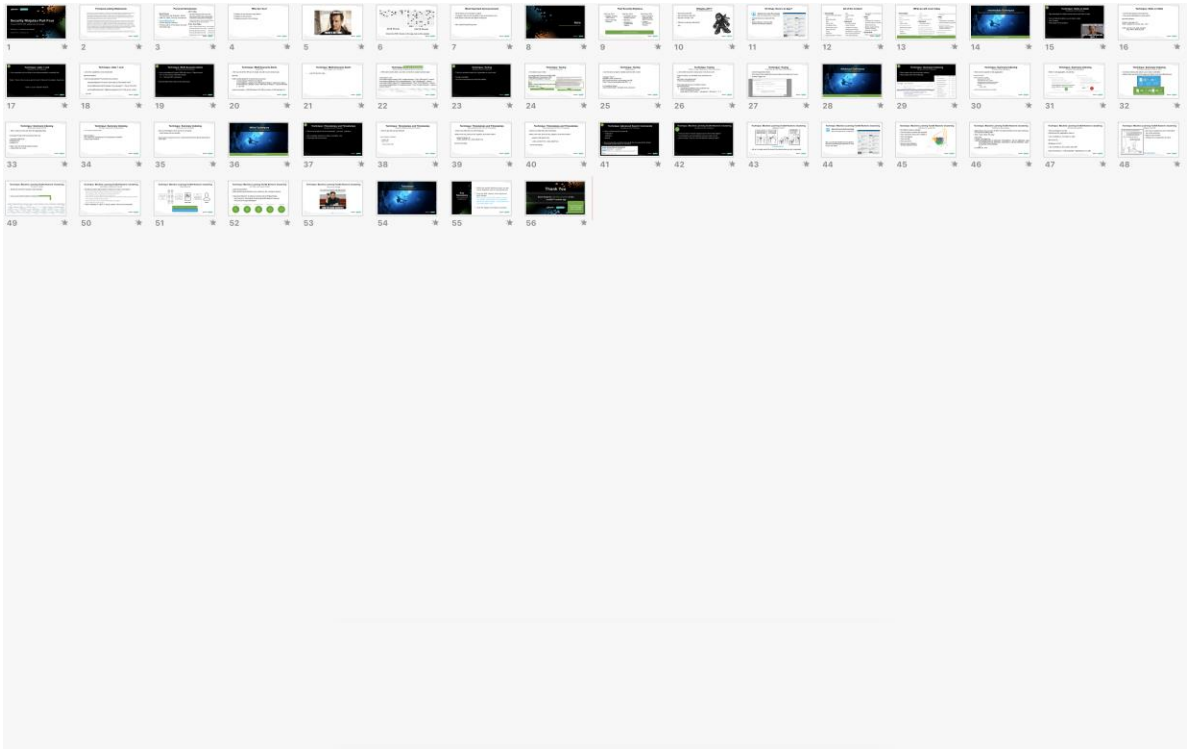


4500 Words

Check the Non-Presentation (Whitepaper) Version in the app and on the website

#TBT 2017

splunk> .conf2017

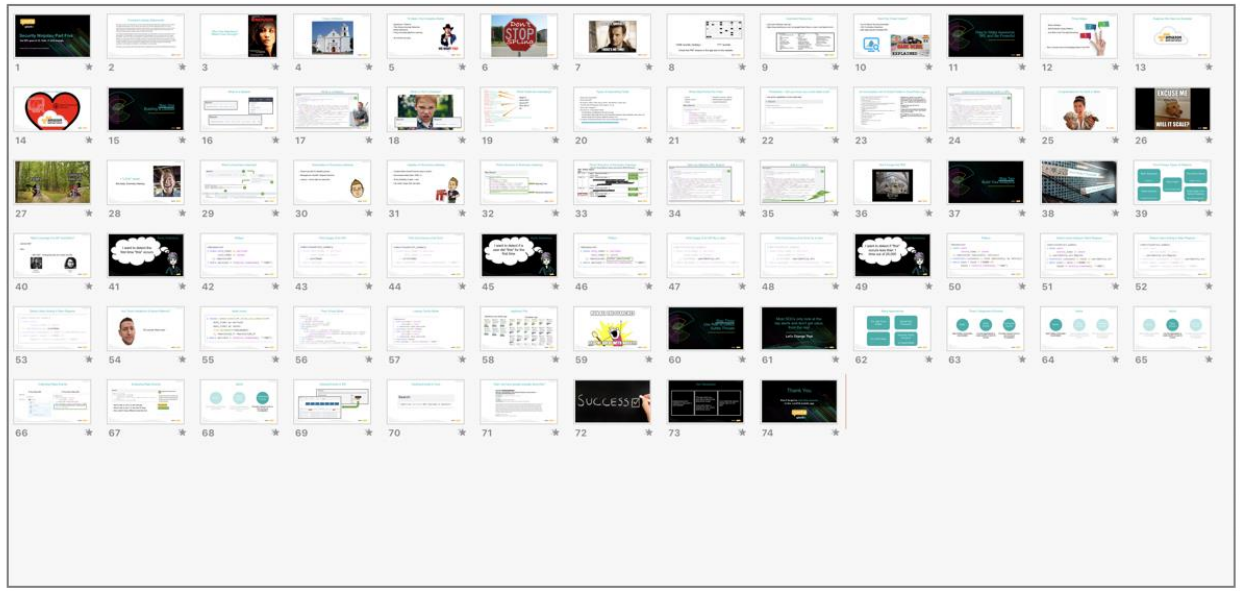


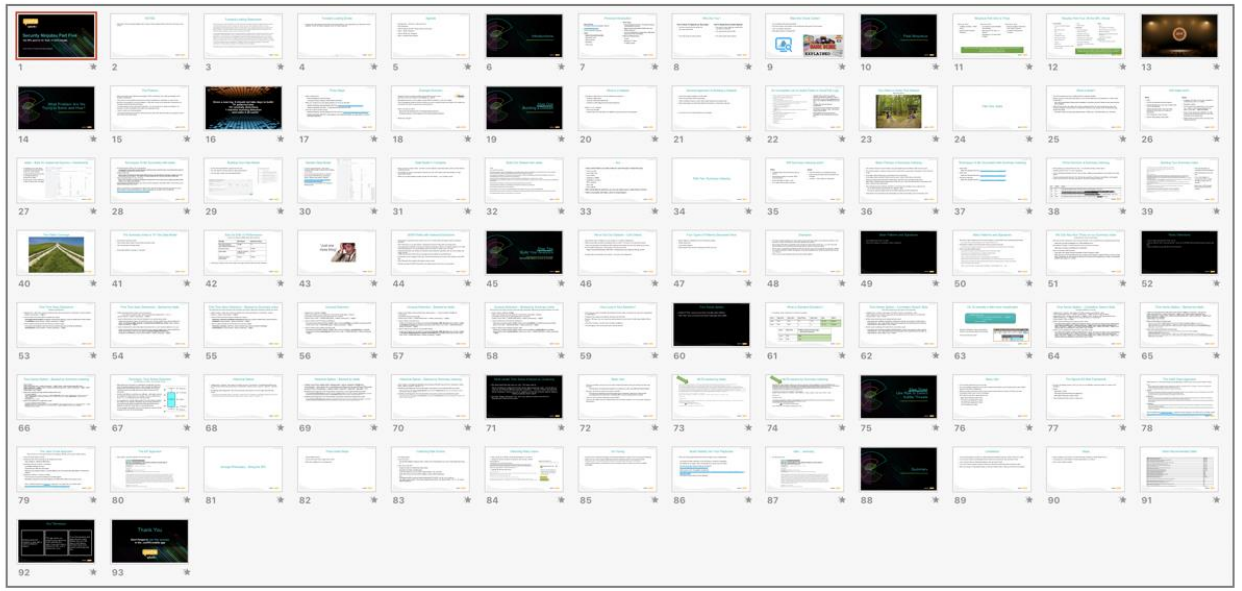
2296 Words

24418 Words

Check the PDF Version in the app and on the website

#TBT 2017





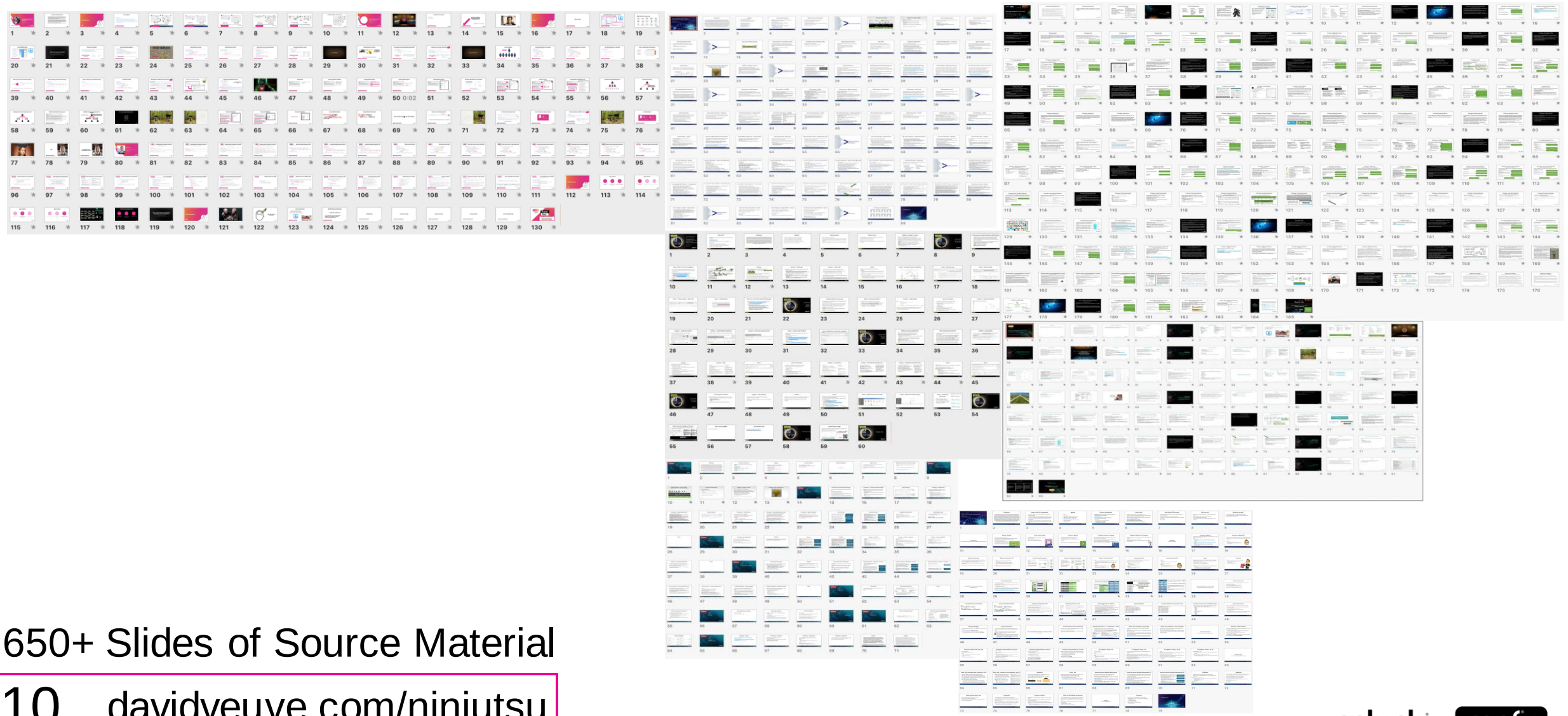
??? words (today)

??? words

Check the PDF Version in the app and on www.davidveuve.com

#TBT 2018

2019 Version (plus a few new ones)



650+ Slides of Source Material

10 davidveuve.com/ninjutsu

Security Ninjutsu Part Six

ALL OF THE SPL

David Veuve
Principal Security Strategist | Splunk

@davidveuve

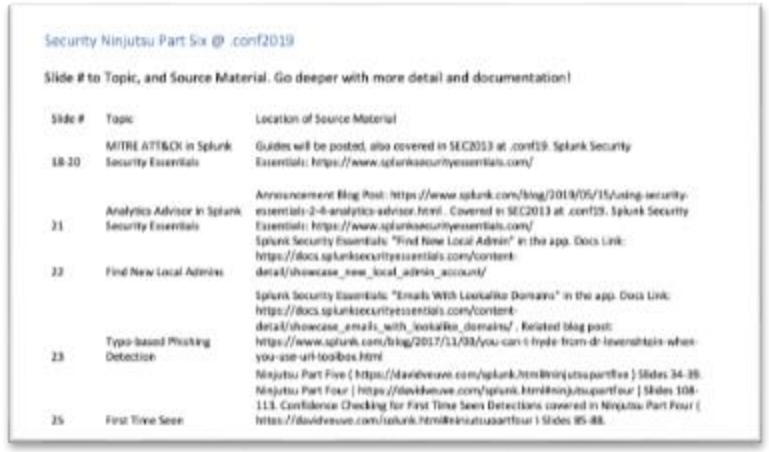
splunk> .conf19

splunk> BOSS OF THE SOC

2019

What Should You Expect?

- Lots of SPL
- Ideas > Syntax
- Full docs available



Security Ninjutsu Part Six @ .conf2019

Slide # to Topic, and Source Material. Go deeper with more detail and documentation!

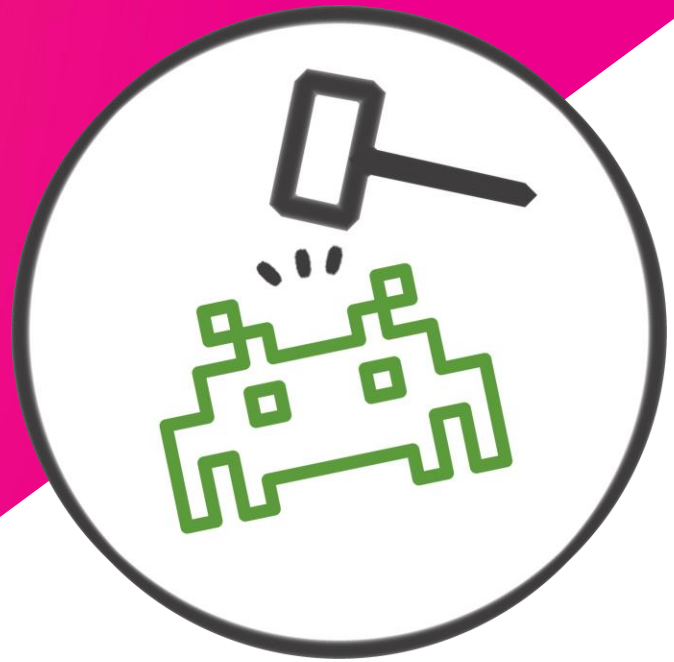
Slide #	Topic	Location of Source Material
18-20	MITRE ATT&CK in Splunk Security Essentials	Guides will be posted, also covered in SEC2013 at .conf19. Splunk Security Essentials: https://www.splunksecurityessentials.com/
21	Analytics Advisor in Splunk Security Essentials	Announcement Blog Post: https://www.splunk.com/blog/2019/05/15/using-security-essentials-2-0-analytics-advisor.html . Covered in SEC2013 at .conf19. Splunk Security Essentials: https://www.splunksecurityessentials.com/
22	Find New Local Admins	Splunk Security Essentials: "Find New Local Admin" in the app. Docs Link: https://docs.splunksecurityessentials.com/content-detail/showcase_new_local_admin_account/
23	Typo-based Phishing Detection	Splunk Security Essentials: "Emails With Lookalike Domains" in the app. Docs Link: https://docs.splunksecurityessentials.com/content-detail/showcase_emails_with_lookalike_domains/ . Related blog post: https://www.splunk.com/blog/2017/11/03/you-can-t-hyde-from-dr-levenshtein-when-you-use-url-toolbox.html
25	First Time Seen	Ninjutsu Part Five (https://davidveuve.com/splunk.html#ninjutsupartfive) Slides 34-39. Ninjutsu Part Four (https://davidveuve.com/splunk.html#ninjutsupartfour) Slides 108-113. Confidence Checking for First Time Seen Detections covered in Ninjutsu Part Four (https://davidveuve.com/splunk.html#ninjutsupartfour) Slides 85-88.

Want the full details? Take the Slide Number to davidveuve.com

Standard Slide Rate: 120 seconds / slide
This Presentation: 20 seconds / slide

14 davidveuve.com/ninjutsu

Begin!



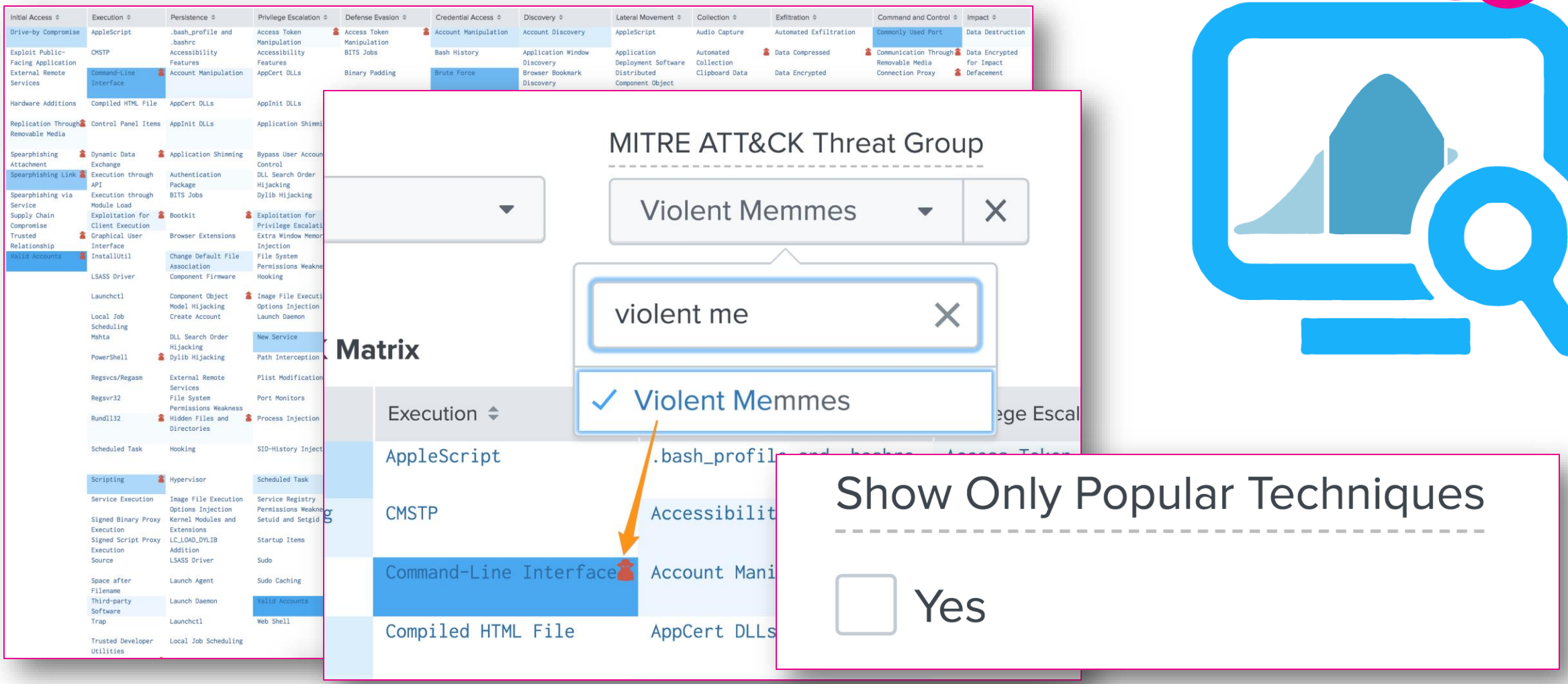
.conf19

splunk>

Where to Start?

Use MITRE ATT&CK To Choose Detections

New!



davidveuve.com/ninjutsu

MITRE ATT&CK Throughout App

Enrich Enterprise Security

ATT&CK Descriptions in Incident Review and risk framework

MITRE Threat Groups

View which detections handle techniques used by which Threat Groups, w/ MITRE's evidence

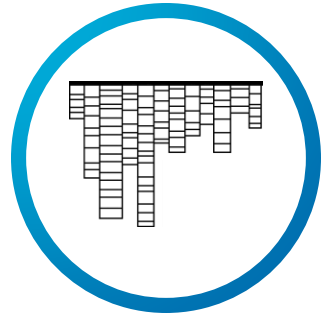
MITRE-based Content Advice

Content Recommendations tied to techniques popular amongst many threat groups

Analyze ES Risk w/ ATT&CK

Drilldown to a customized ATT&CK Matrix, correlate risky events across Tactics, Techniques

View Your ATT&CK Coverage



ATT&CK Matrix highlighting gaps and showing content you can enable for free with existing data

splunk> .conf19

Not Just MITRE Though

	Data Source Category	Total	Active	Available
1	Process Launch	49	5	44
2	AWS Cloudtrail	32	4	28
3	Windows Security Logs	30	0	0
4	Basic Traffic Logs	27	1	26
5	Object Change	23	0	23
6	Malware Detected	18	0	0
7	Process Launch with CLI	17	4	13
8	Web server access logs	13	0	13
9	Proxy Requests	12	0	0
10	Access logs	8	0	0

Splunk Security Essentials

The free app that helps makes security easier.

splunk> .conf19

Find New Local Admins

OOTB with SSE

```
index=* source="*WinEventLog:Security" EventCode=4720 OR (EventCode=4732 Administrators)
| transaction Security_ID maxspan=180m connected=false
| search EventCode=4720 (EventCode=4732 Administrators)
| table _time EventCode Account_Name Target_Account_Name Message
```

Detect a user created (4720) and added to local group (4732)

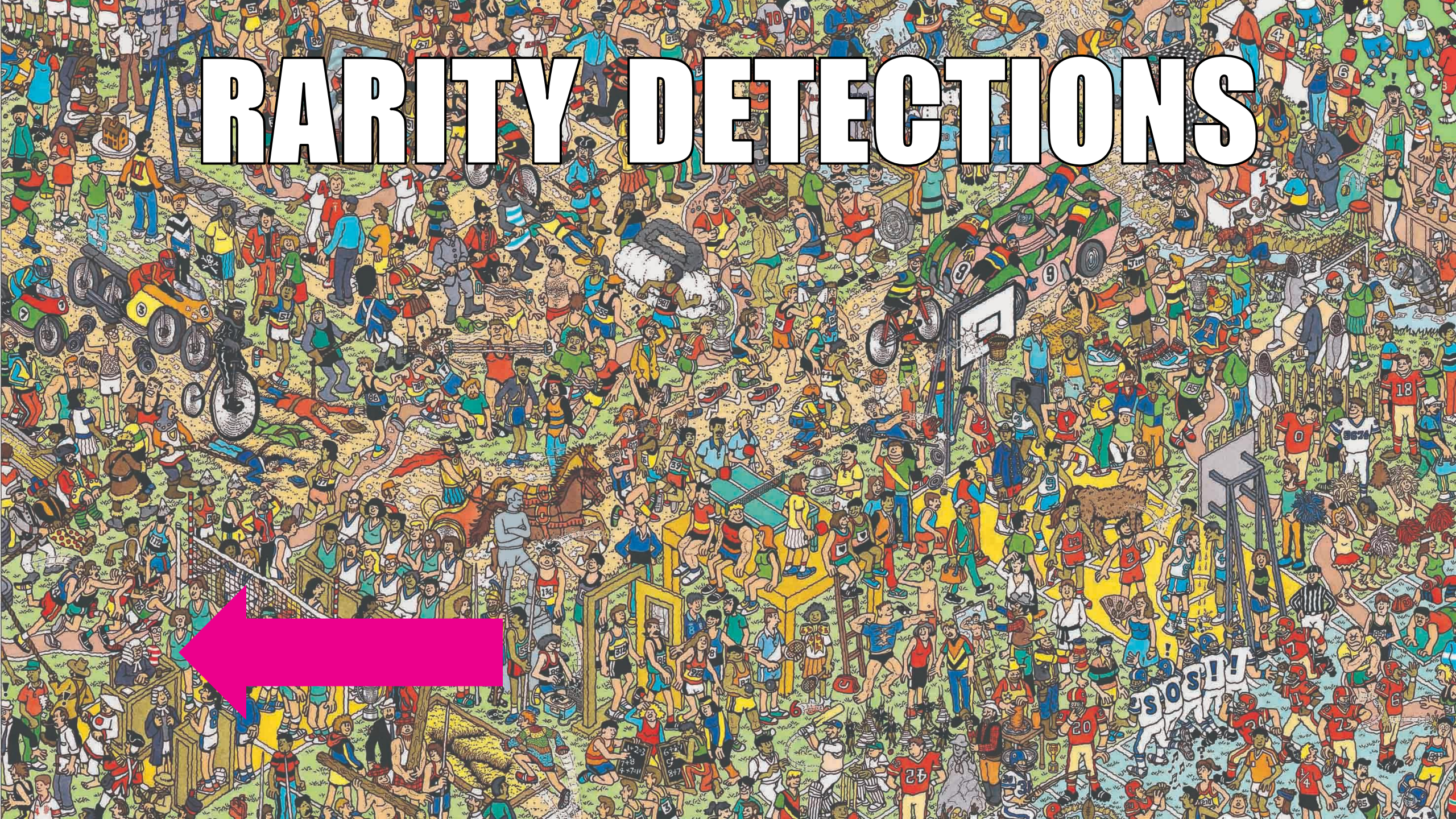
```
New Account:
        Security ID:            AGRADY-L\svc_print
        Account Name:           svc_print
        Account Domain:         AGRADY-L
```

Typo-based Phishing Detection


```
index=ironport_logs
| stats values(*) as * by MID
| search sender!=company.com
| eval list=".com", ourdomain="company.com"
|`ut_parse(sender,list)` |`ut_levenshtein(ourdomain, ut_domain)`
| where (ut_levenshtein>0 AND ut_levenshtein<3)
```

URL Toolbox (splunkbase) easily finds this

No results found. Try expanding the time range.


RARITY DETECTIONS

Rarity Detections are Easy

```
index=clouddata
| stats min(_time) as earliest
        max(_time) as latest
    by errorCodes, user
| where earliest > relative_time(now(), "-1d@d")
```

Rarity Detections are Easy

```
index=cloudtrail
| stats min(_time) as earliest
        max(_time) as latest
    by errorCodes, user
| where earliest > relative_time(now(), "-1d@d")
```

Did you know we can cache that lookup?


```
tag=authentication
| stats earliest(_time) as earliest
    latest(_time) as latest
    by user, dest
| inputlookup append=t login_tracker.csv
| stats min(earliest) as earliest
    max(latest) as latest
    by user, dest
| where latest > relative_time(now(), "-90d")
| outputlookup sample_cache_group.csv
| where earliest >= relative_time(now(), "-1d@d")
```

Find New Processes for a Host

OOTB with SSE

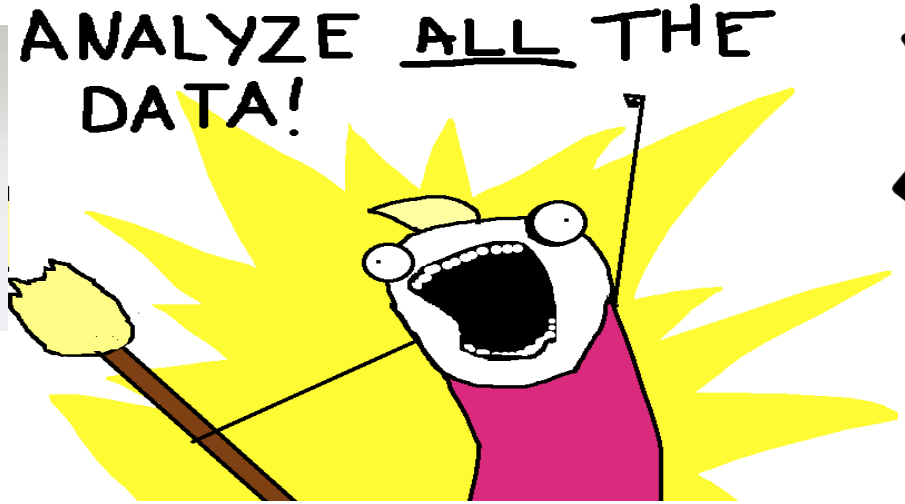
```
sourcetype=*sysmon* process_name!=""
| stats
    min(_time) as earliest
    max(_time) as latest
   by process_name host
| where earliest>relative_time(now(), "-1d@d")
```

Detect processes launched for the first time in the past day

process_name	host
PickerHost.exe	FMALTEKESKO-L
SenseIR.exe	FMALTEKESKO-L
SystemSettings.exe	FMALTEKESKO-L
more.com	FMALTEKESKO-L
notepad++.exe	FMALTEKESKO-L
whoami.exe	FMALTEKESKO-L

Interlude

Visibility Analysis and Action



Make sure you fully think through the problem

Give Analysts Descriptive Fields


```
...
| eval risk_score_reason = case(
    one_day_risk>threshold_1day,
                "One Day Risk > " . threshold_1day,
    thirty_day_risk>threshold_30day,
                strftime(now(), "%m-%d-%Y") . " 30 Day Risk > " . threshold_30day,
    1=1,
                "Key Words Flagged")
```

Tell Analysts Why They Should Care

risk_score	risk_score_reason
300	One Day Risk > 500

Combine Lots of Context in One Field

New!

Use Aaron Kohler's Field Stuffing Technique

```
...
| lookup dest_information dest OUTPUT dest_information_*
| eval info = null
| foreach dest_information_*
   [eval info = if(isnotnull('<<FIELD>>'), mvappend(info, '<<FIELD>>'), info) ]
| fields - dest_information_*
```

Provide Information Density in Standardized Fields

Courtesy of Aaron Kohler

dest	info
10.1.1.1	Recorded Future Threat Intel Hit Past Notables! Current Open Notables!

</ Interlude >

Time Series Spikes

Time Series Detections are Also Easy

```
index=cloudtrail
| bucket _time span=1d
| stats count by _time user
| stats latest(count) as latest
        avg(count)    as avg
        stdev(count)  as stdev
    by user
| where latest > avg + 4 * stdev
```

Time Series Detections are Also Easy

```
index=cloudtrail
| bucket _time span=1d
| stats count by _time user
| stats latest(count) as latest
        avg(count)    as avg
        stdev(count)  as stdev
    by user
| where latest > avg + 4 * stdev
```



Time Series Detections are Also Easy

```
index=cloudtrail
| bucket _time span=1d
| stats count by _time user
| stats latest(count) as latest
        avg(count)    as avg
        stdev(count)  as stdev
    by user
| where latest > avg + 4 * stdev
```

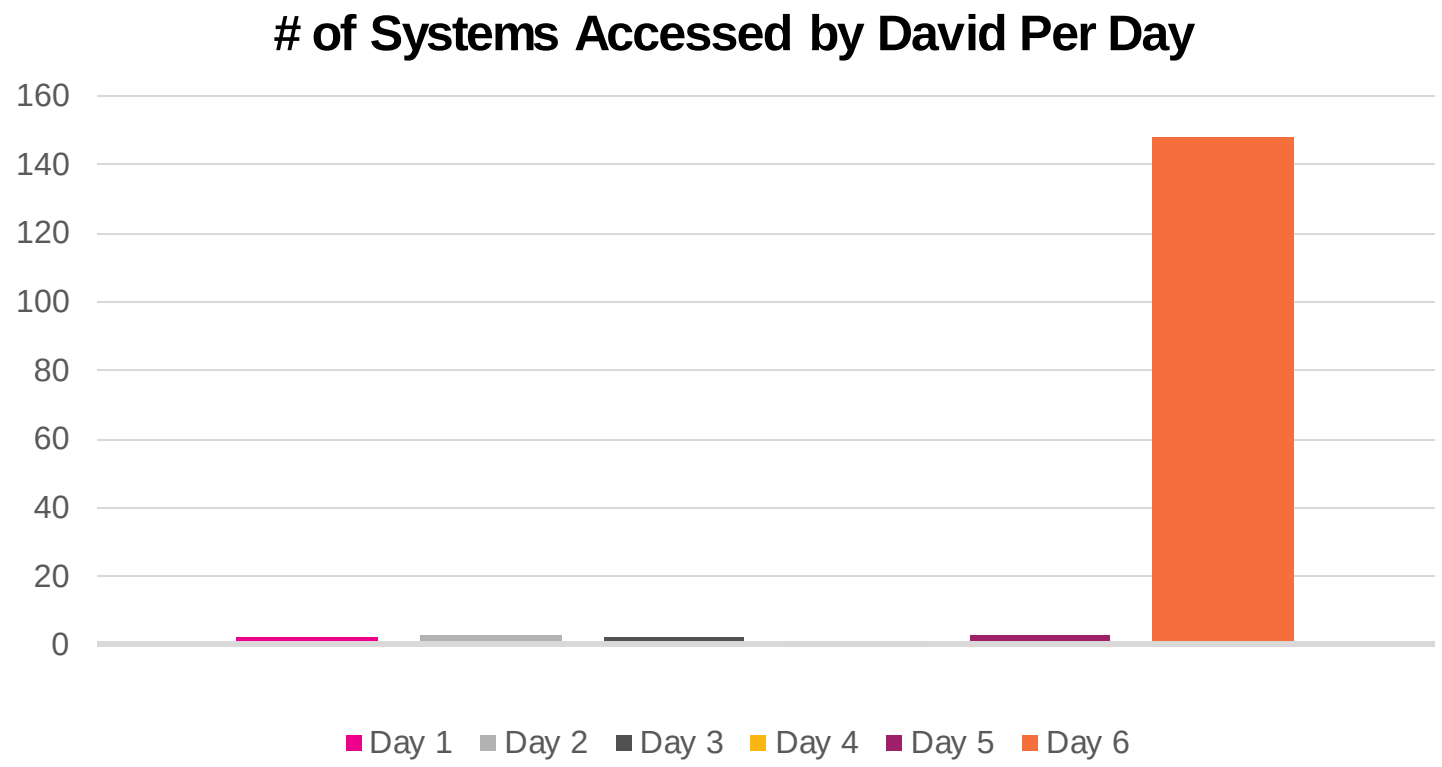
There's a Problem with Calculating Avg

First Five Days

- Mean: 2.2
- Median: 2

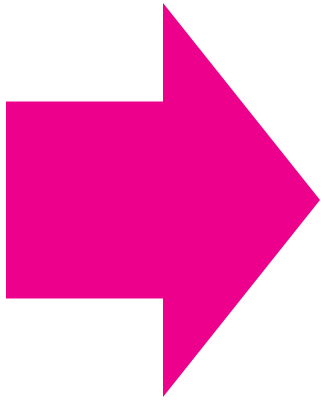
Entire Data Set

- Mean: **26.5**
- Median: 2.5



Stats + Eval is key for Time Series Accuracy

```
avg(count)
```



```
avg(eval(
    if(_time<relative_time(now(), "-1d@d"),
        count,
        null)
)) as avg
```

Stats + Eval allows you to embed any eval into a stats command. (No tstats ☹)

Stats + Eval is key for Time Series Accuracy


```
index=cloudtrail
| bucket _time span=1d
| stats count by _time user
| stats max(eval(if(_time>=relative_time(now(), "-1d@d"), count, null)))  as latest
        avg(eval(if(_time<relative_time(now(), "-1d@d"), count, null)))   as avg
        stdev(eval(if(_time<relative_time(now(), "-1d@d"), count, null))) as stdev
    by user
| where latest > avg + 4 * stdev
```

Stats + Eval is key for Time Series Accuracy

```
index=cloudtrail
| bucket _time span=1d
| stats count by _time user
| stats max(eval(if(_time>=relative_time(now(), "-1d@d"), count, null)))  as latest
        avg(eval(if(_time<relative_time(now(), "-1d@d"), count, null)))   as avg
        stdev(eval(if(_time<relative_time(now(), "-1d@d"), count, null))) as stdev
    by user
| where latest > avg + 4 * stdev
```

Use IQR If You Want

New!

```
<datasource>
| bucket _time span=1d
| stats count by <monitored>
| eventstats perc25(count) as perc25
             perc75(count) as perc75
        by <monitored>
| where count > perc75 + (perc75 - perc25) * 1.5
```

MLTK Magic w/ Probability Density Function

New!

New Search

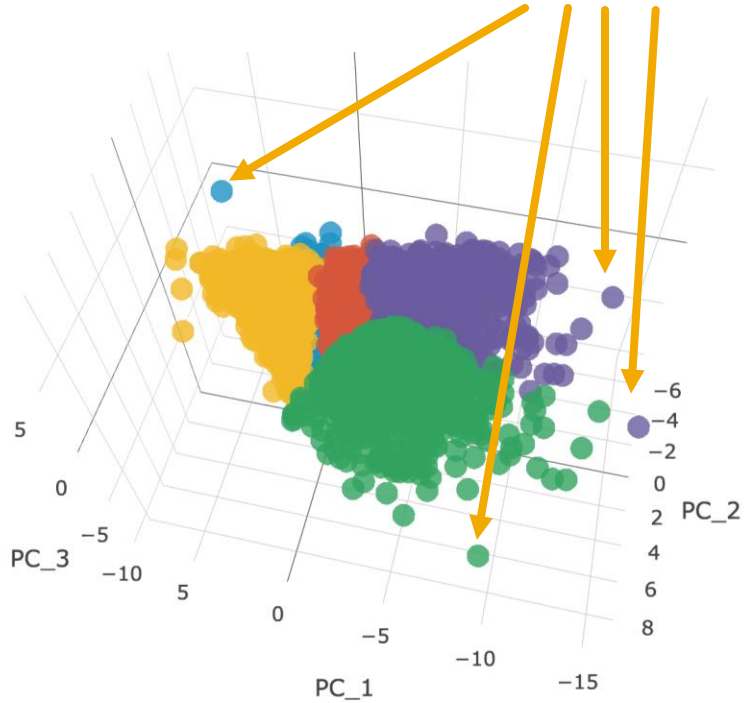
```
index=pan_logs sourcetype=pan:traffic dest_port=443 user!=unknown bytes_out>10000
| eval time=strftime(_time, "%A") . " - " . strftime(_time, "%H") . ":00 Hour"
| stats sum(bytes_out) by app, user, time
| fit DensityFunction sum_bytes_out by time show_density=true
```

Add One Line

Low Cardinality
Easy ML Baselines

ProbabilityDensity(sum_bytes_out)	app	sum_bytes_out
0.0	dropbox-base	4185958717
0.0	dropbox-base	22658418879

MLTK with All the Bells & Whistles



New Search

```
index=sfdc
| bucket _time span=1d
| stats dc(eval(if(like(URI_ID_DERIVED, "00140000%"), URI_ID_DERIVED, null)))
        dc(eval(if(like(URI_ID_DERIVED, "0063300%"), URI_ID_DERIVED, null)))
        sum(ROWS_PROCESSED)
        […]
by USER_ID _time
```

New Search

```
1  | inputlookup sfdc_aggregated_data.csv
2  | `use_eventstats_to_normalize_to_stdev`
3  | fit PCA k=5 Z_*
4  | fit KMeans k=5 PC_*
5  | `use_eventstats_to_find_ouliers_with_IQR`
6  | where NumEntries < 5 OR clusterDist > MaxDistance_For_IQR
```

Apply Similar Techniques w/o Time


```
index=* sourcetype="*sysmon" EventCode=1
| eval cmdlen=len(CommandLine)
| eventstats stdev(cmdlen) as stdev
             avg(cmdlen)   as avg
    by host
| stats values(cmdlen) as maxlen,
        values(stdev)  as stdevperhost,
        values(avg)    as avgperhost
    by host, CommandLine
| where maxlen > ( 10 * stdevperhost + avgperhost )
```

Find Processes Much Longer than Normal for that Host

maxlen	CommandLine
7478	"C:\Windows\System32\WindowsPowerShell\v1.0\powershell.exe" -nop -w hidden -e aQBmACgAWwBJAG4AdABQAHQAcgBdADoAOgBTAGkAegBlACAALQBlAHEAIAA0ACkAewAkAGIAPQAnAHAAbwB3AGUAcgBzAGgAZQBsAGwALgBlAHgAZQAnAH0AZQBsAHMAZQ
2334	"C:\WINDOWS\System32\WindowsPowerShell\v1.0\powershell.exe" -nop -w hidden -noni -c "&([scriptblock]::create((New-Object Syste


Damaged
Why
so
SLOW?

Example Data

Sample Events

4.14.104.185 Doug completes a triathlon

198.179.87.243 David completes some coding

Let's compare two technologies:

1. Databases
2. Splunk

Ingesting Data Into a Database

Sample Events

4.14.104.185 Doug completes a triathlon

198.179.87.243 David completes some coding

Database Storage

Person	Activity
Doug	triathlon
David	coding

Ingesting Data Into Splunk

Sample Events

4.14.104.185 Doug completes a triathlon

198.179.87.243 David completes some coding

> Splunk Storage

Posting 1:
4.14.104.185, a, completes, doug, triathlon

Posting 2:
198.179.87.243, coding, completes, david, some

Ingesting Data Into Splunk

Sample Events

4.14.104.185 Doug completes a triathlon

198.179.87.243 David completes some coding

4.14.104.185 a completes doug triathlon

198.179.87.243 coding david some

Ingesting Data Into Splunk

Postings

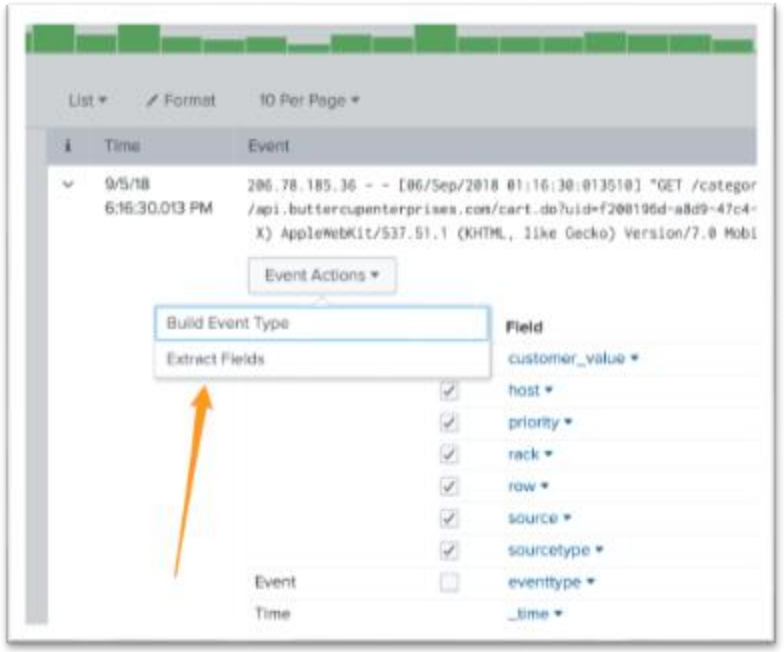
#1: 4.14.104.185 Doug completes a triathlon

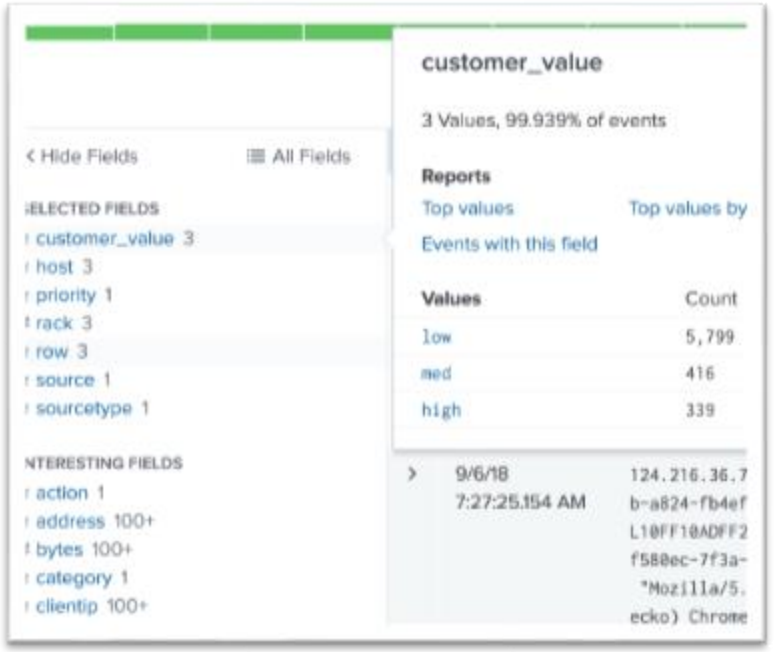
#2: 198.179.87.243 David completes some coding

Time-Series Index

Term	Posting #
4.14.104.185	1
198.179.87.243	2
a	2
coding	2
completes	1, 2
david	2
doug	1
some	2
triathlon	1

We Extract Fields When You Click Search





Field Extractions = Schema
Splunk does schema-on-read
aka schema-on-the-fly, aka late binding schema

How Search Works

Postings

#2

#1: 4.14.104.185 Doug completes a triathlon

#2: 198.179.87.243 David completes some coding

Potential Field Extractions

(via Schema-On-Read)

#3

User	Status	Activity	IP
Doug	Success	triathlon	4.14.104.185
David	Success	coding	198.179.87.243

Time-Series Index

#1

Term	Posting #
4.14.104.185	1
198.179.87.243	2
a	2
coding	2
completes	1, 2
david	2
doug	1
some	2
triathlon	1

Search For "doug"

Postings

#2

#1: 4.14.104.185 Doug completes a triathlon

#2: 198.179.87.243 David completes some coding

Potential Field Extractions

(via Schema-On-Read)

#3

User	Status	Activity	IP
Doug	Success	triathlon	4.14.104.185
David	Success	coding	198.179.87.243

Time-Series Index

#1

Term	Posting #
4.14.104.185	1
198.179.87.243	2
a	2
coding	2
completes	1, 2
david	2
doug	1
some	2
triathlon	1

Knowing What You Will Want to Ask

Four Easy Fields:

4.14.104.185 Doug completes a triathlon

198.179.87.243 David completes some coding

Every field extracted costs money, but could be crucial for an incident.

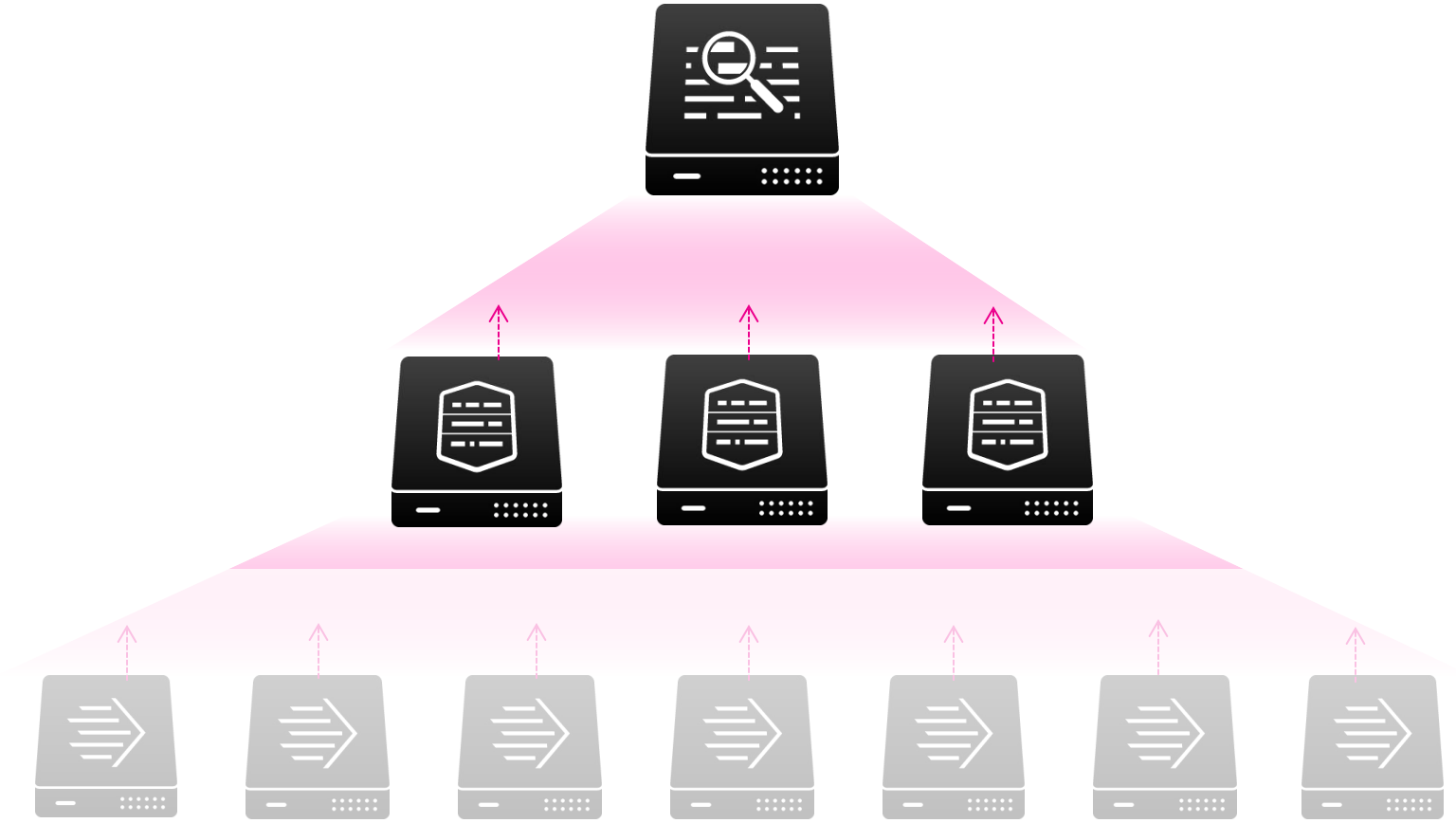
Splunk lets you extract fewer upfront, and pay less money.

55 davidveuve.com/ninjutsu

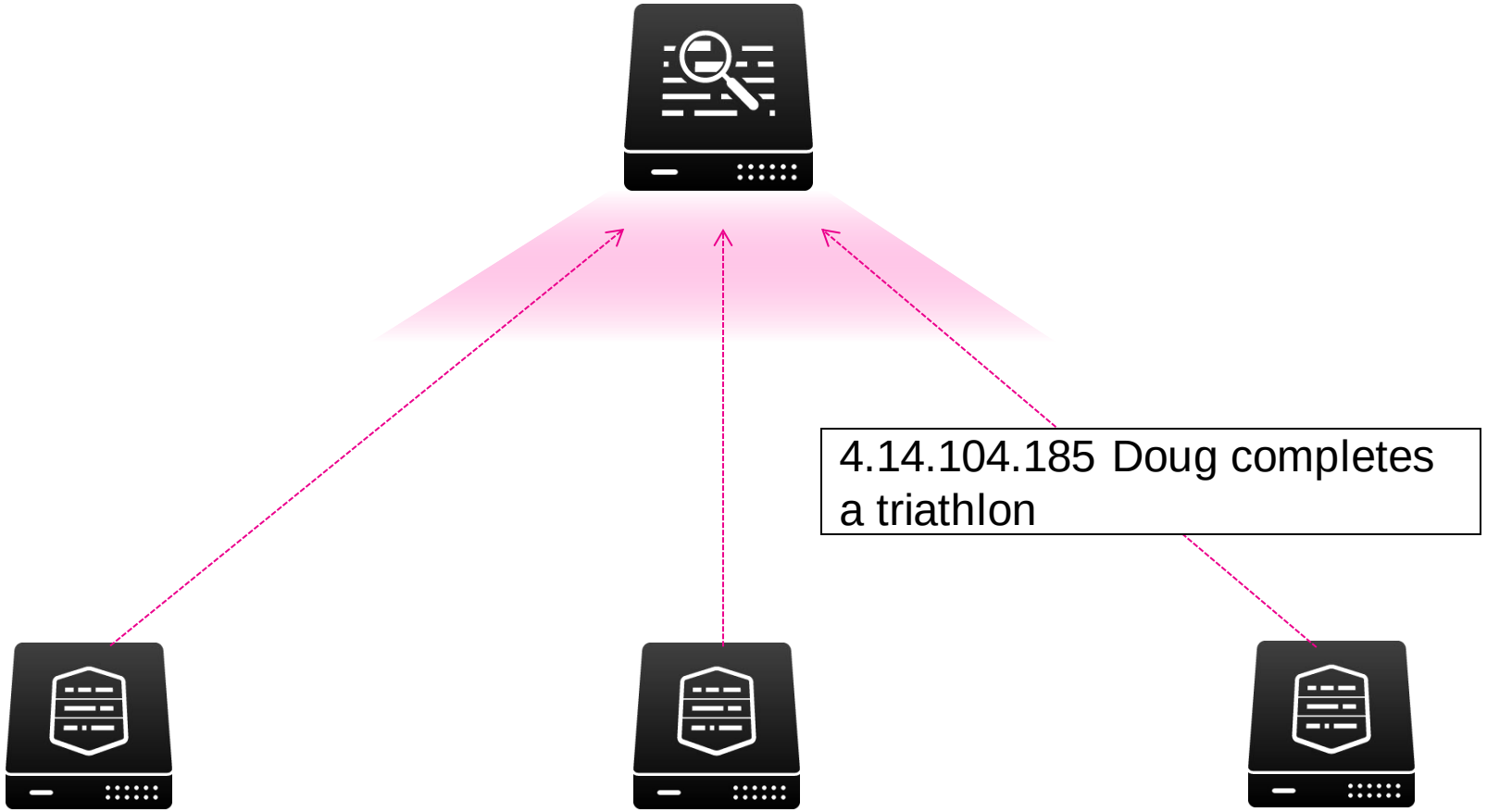
Many potential fields:

```
06-Sep-2018 23:30:29:194786 000000c6 DataStoreCont E com.ibm.wps.datastore.impl.DataStoreContext handleException EJPDB0001E: Error occurred during
 database access. Last SQL statement is [SELECT OID, CREATED, MODIFIED, PORT_DESC_OID, PORT_DESC_SL, SCOPE_OID, WSC_INST_HANDLE, WSP_IS_PROVIDED,
WSP_PROD_OFF_INST, TYPE, PARENT_OID, PARENT_SL, PREF_SCOPE_TYPE, PREF_SCOPE_UID, VALIDATION_BASE, VALIDATION_STATE, JSR_DATA FROM customiz.PORT_IN
ST WHERE (OID = ?)].
        com.ibm.wps.datastore.domains.DomainUnavailableException: EJPDB0101E: Database domain [Domain: cust] is currently unavailable.
        com.ibm.wps.datastore.domains.DomainUnavailableException: EJPDB0101E: Database domain [Domain: cust] is currently unavailable.
        at com.ibm.wps.datastore.impl.DataStoreContext.handleException(DataStoreContext.java:315)
        at com.ibm.wps.datastore.impl.DataStoreContext.handleException(DataStoreContext.java:315)
        at com.ibm.wps.datastore.impl.ResourcePersister.findInternal2(ResourcePersister.java:981)
        at com.ibm.wps.datastore.impl.ResourcePersister.findInternal2(ResourcePersister.java:981)
        at com.ibm.wps.datastore.impl.PortletInstancePersister.findInternal2(PortletInstancePersister.java:373)
        at com.ibm.wps.datastore.impl.PortletInstancePersister.findInternal2(PortletInstancePersister.java:373)
        at com.ibm.wps.datastore.impl.ResourcePersister.findInternal(ResourcePersister.java:880)
        at com.ibm.wps.datastore.impl.ResourcePersister.findInternal(ResourcePersister.java:880)
        at com.ibm.wps.datastore.impl.ResourcePersister.findSingleObject(ResourcePersister.java:1283)
        at com.ibm.wps.datastore.impl.ResourcePersister.findSingleObject(ResourcePersister.java:1283)
        at com.ibm.wps.datastore.impl.ResourcePersister.findInternal(ResourcePersister.java:1456)
        at com.ibm.wps.datastore.impl.ResourcePersister.findInternal(ResourcePersister.java:1456)
        at com.ibm.wps.datastore.impl.ResourcePersister.find(ResourcePersister.java:1312)
        at com.ibm.wps.datastore.impl.ResourcePersister.find(ResourcePersister.java:1312)
        at com.ibm.wps.datastore.impl.ResourceHomeImpl.find(ResourceHomeImpl.java:55)
        at com.ibm.wps.datastore.impl.ResourceHomeImpl.find(ResourceHomeImpl.java:55)
        at com.ibm.wps.datastore.impl.federation.ResourceHomeImpl.find(ResourceHomeImpl.java:446)
        at com.ibm.wps.datastore.impl.federation.ResourceHomeImpl.find(ResourceHomeImpl.java:446)
        ... 12 more
Caused by: com.ibm.websphere.ce.cm.StaleConnectionException: [IBM][CLI Driver] SQL1224N  The database manager is not able to accept new requests,
has terminated all requests in progress, or has terminated your particular request due to a problem with your request.  SQLSTATE=55032
Caused by: com.ibm.websphere.ce.cm.StaleConnectionException: [IBM][CLI Driver] SQL1224N  The database manager is not able to accept new requests,
has terminated all requests in progress, or has terminated your particular request due to a problem with your request.  SQLSTATE=55032
        at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native Method)
        at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native Method)
        at sun.reflect.NativeConstructorAccessorImpl.newInstance(NativeConstructorAccessorImpl.java:67)
        at sun.reflect.NativeConstructorAccessorImpl.newInstance(NativeConstructorAccessorImpl.java:67)
        at sun.reflect.DelegatingConstructorAccessorImpl.newInstance(DelegatingConstructorAccessorImpl.java:45)
        at sun.reflect.DelegatingConstructorAccessorImpl.newInstance(DelegatingConstructorAccessorImpl.java:45)
        at java.lang.reflect.Constructor.newInstance(Constructor.java:522)
        at java.lang.reflect.Constructor.newInstance(Constructor.java:522)
        at com.ibm.websphere.rsadapter.GenericDataStoreHelper.mapExceptionHelper(GenericDataStoreHelper.java:525)
        at com.ibm.websphere.rsadapter.GenericDataStoreHelper.mapExceptionHelper(GenericDataStoreHelper.java:525)
        at com.ibm.websphere.rsadapter.GenericDataStoreHelper.mapException(GenericDataStoreHelper.java:580)
        at com.ibm.websphere.rsadapter.GenericDataStoreHelper.mapException(GenericDataStoreHelper.java:580)
        at com.ibm.ws.rsadapter.jdbc.WSJdbcUtil.mapException(WSJdbcUtil.java:909)
        at com.ibm.ws.rsadapter.jdbc.WSJdbcUtil.mapException(WSJdbcUtil.java:909)
        at com.ibm.ws.rsadapter.jdbc.WSJdbcPreparedStatement.executeQuery(WSJdbcPreparedStatement.java:689)
        at com.ibm.ws.rsadapter.jdbc.WSJdbcPreparedStatement.executeQuery(WSJdbcPreparedStatement.java:689)
        at com.ibm.wps.datastore.impl.DataStoreContext.executeQuery(DataStoreContext.java:538)
        at com.ibm.wps.datastore.impl.DataStoreContext.executeQuery(DataStoreContext.java:538)
        at com.ibm.wps.datastore.impl.ResourcePersister.findInternal2(ResourcePersister.java:914)
```

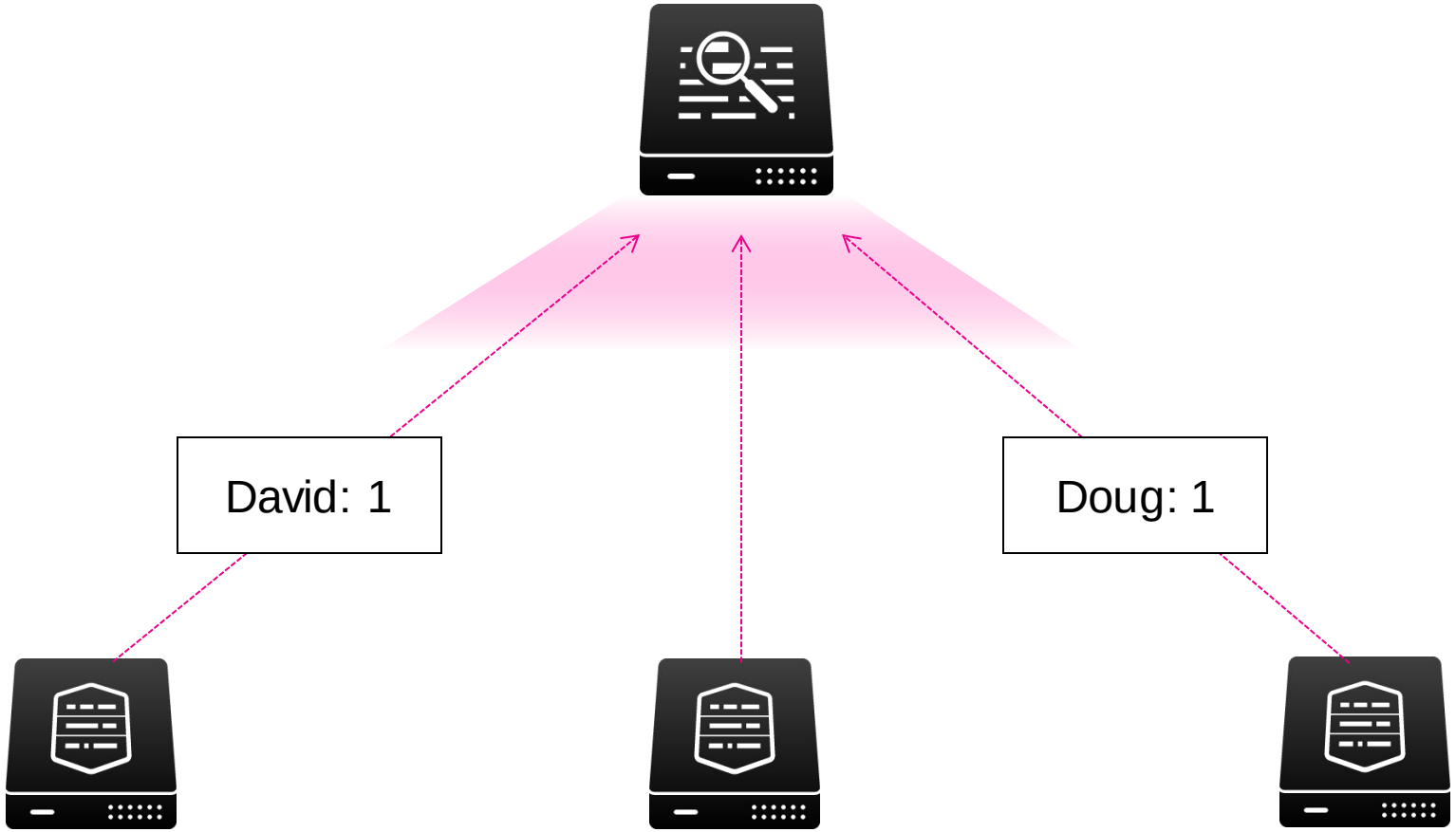
How Search Works With Many Servers



Events with user=Doug



How Many Successful Completions by User?



What if We Had More Interesting Logs?

Sample Events

09/05/2018 -- 4.14.104.185 Doug completes a triathlon in 1.25 hours

09/03/2018 -- 198.179.87.243 David completes some coding in 5 hours

09/02/2018 -- 198.133.81.178 David completes some coding in 7 hours

08/25/2018 -- 39.195.31.8 David completes some coding in 9 hours

08/09/2018 -- 4.14.104.185 Doug completes a triathlon in 1.15 hours

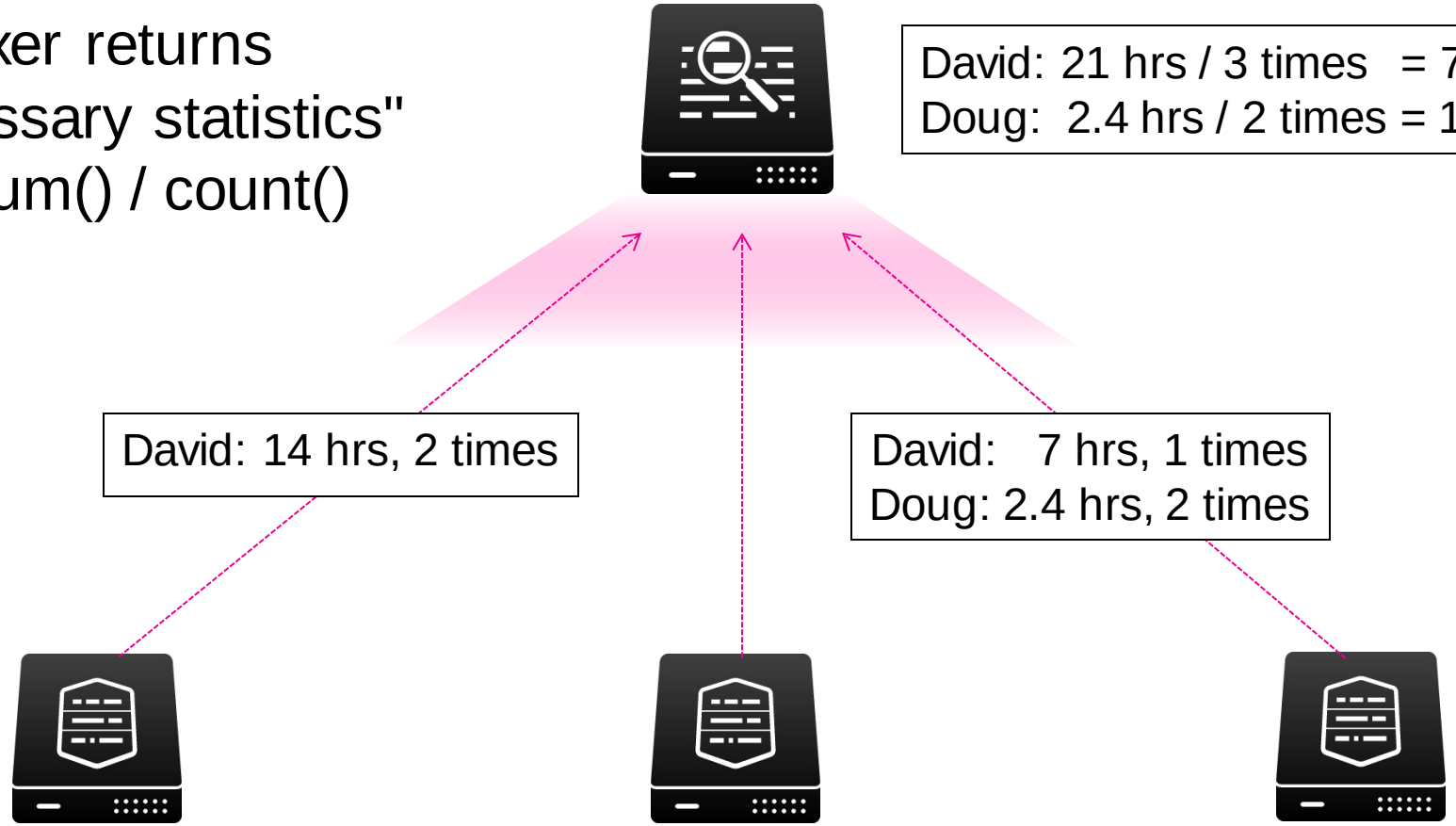
Multiple Events for each person

Duration Fields

Timestamps

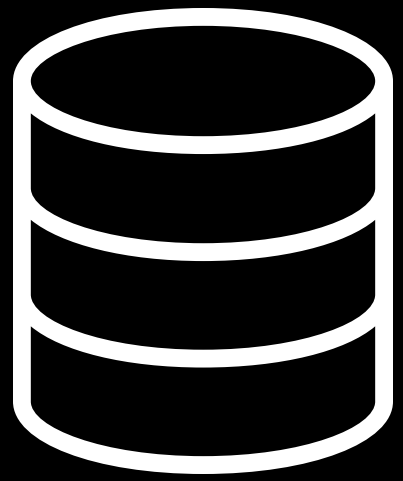
What Was The Average Completion Time per User?

Each Indexer returns
"minimum necessary statistics"
average() = sum() / count()

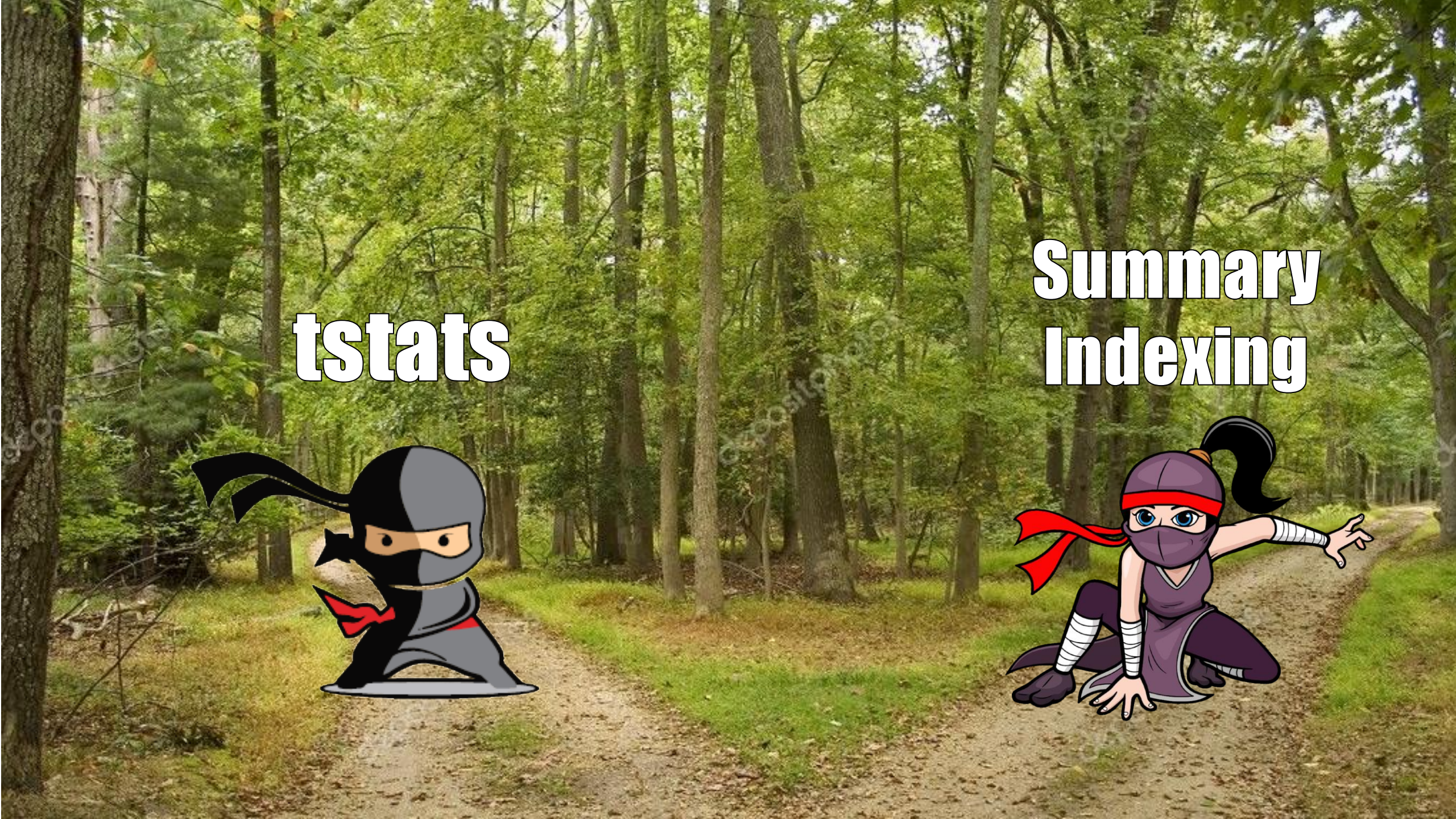


Here's a Secret About Basic Splunk…

For well understood and modeled problems…



Databases Are Faster


tstats
Summary
Indexing

Summary Indexing

Structured Data Models

Postings

#1: 4.14.104.185 Doug completes a triathlon

#2: 198.179.87.243 David completes some coding

Potential Field Extractions

(via Schema-On-Read)

User	Status	Activity	IP
Doug	Success	triathlon	4.14.104.185
David	Success	coding	198.179.87.243

Time-Series Index

Term	Posting #
4.14.104.185	1
198.179.87.243	2
a	2
coding	2
completes	1, 2
david	2
doug	1
some	2
triathlon	1

Structured Data Models

Postings

#1: 4.14.104.185 Doug completes a triathlon

#2: 198.179.87.243 David completes some coding

Potential Field Extractions

(via Schema-On-Read)

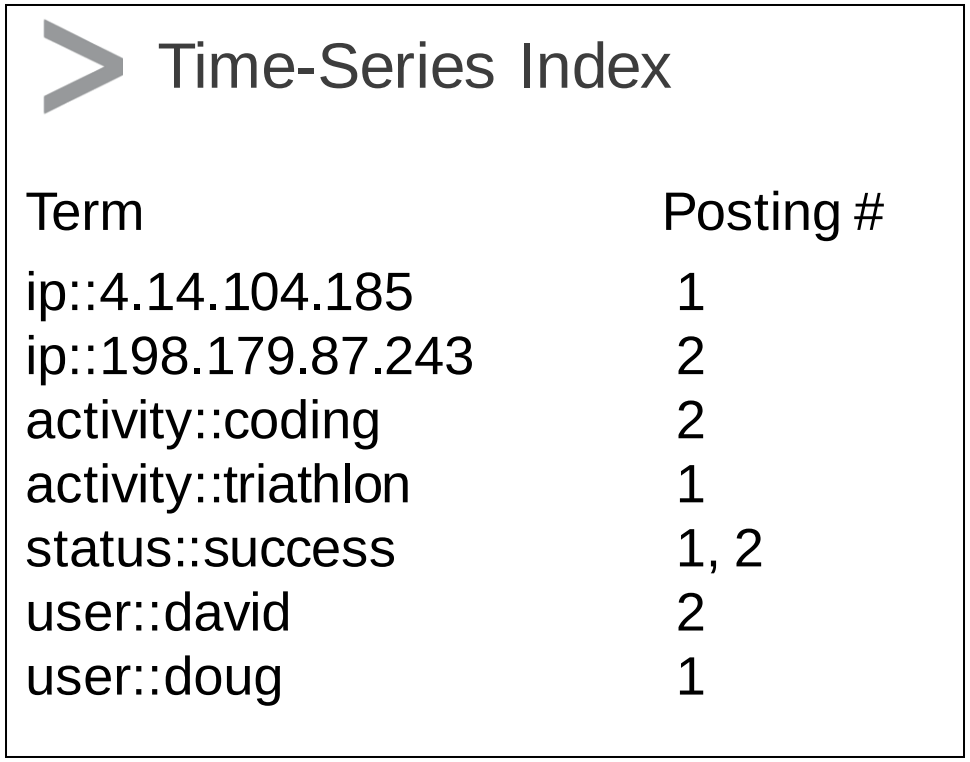
User	Status	Activity	IP
Doug	Success	triathlon	4.14.104.185
David	Success	coding	198.179.87.243

Time-Series Index

Term	Posting #
ip::4.14.104.185	1
ip::198.179.87.243	2
activity::coding	2
activity::triathlon	1
status::success	1, 2
user::david	2
user::doug	1

Structured Data Models

10-100-1000x Faster



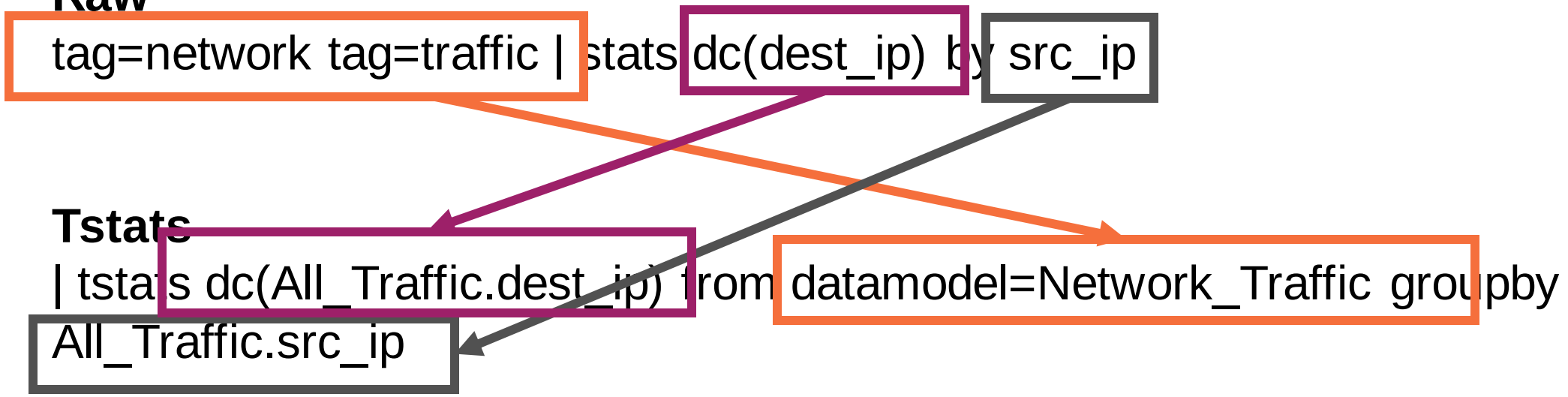
Example With Data Models

Raw

```
tag=network tag=traffic | stats dc(dest_ip) by src_ip
```

Tstats

```
| tstats dc(All_Traffic.dest_ip) from datamodel=Network_Traffic groupby
All_Traffic.src_ip
```



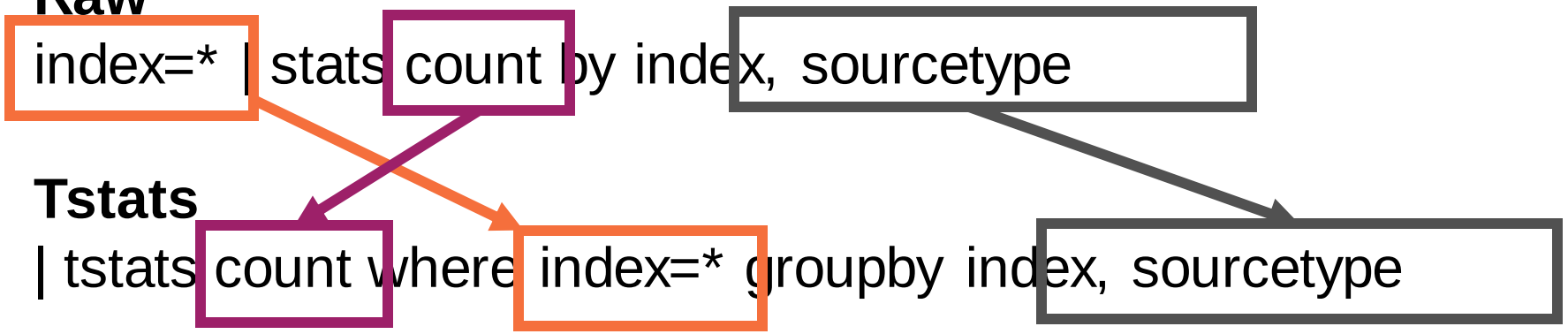
Example Without Data Models

Raw

```
index=* | stats count by index, sourcetype
```

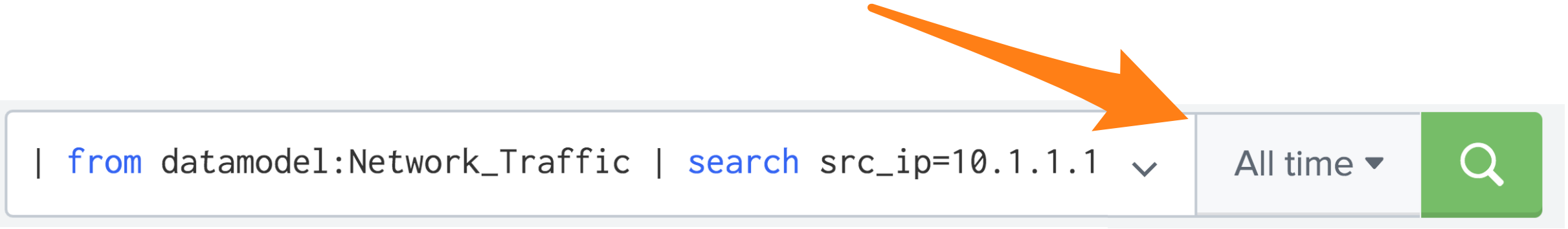
Tstats

```
| tstats count where index=* groupby index, sourcetype
```



Aww, but tstats is hard and only returns statistics ☹

Schema-Accelerated Event Search



Pull Raw Events at Data Model Speeds using | from or | datamodel!

tstats

Summary Indexing

Use Summary Indexing for Speed

```
| makeresults | eval status="Hello World" | collect index=summary
```

Generate New Events!

Power of Summary Indexing for Speed

Billions of Raw Logs

Thousands of Stats Rows

Summary Indexing in JSON? You Fancy.

New!

```
| makeresults
| eval _raw="{\"status\": \"Hello World\"}"
| collect index=summary
```

i	Time	Event
>	10/17/19 8:58:25.000 PM	{ [-] status: Hello World } Show as raw text

Combine with JSON Index-time Extractions!

tstats + Summary Indexing

For Ultimate Speed

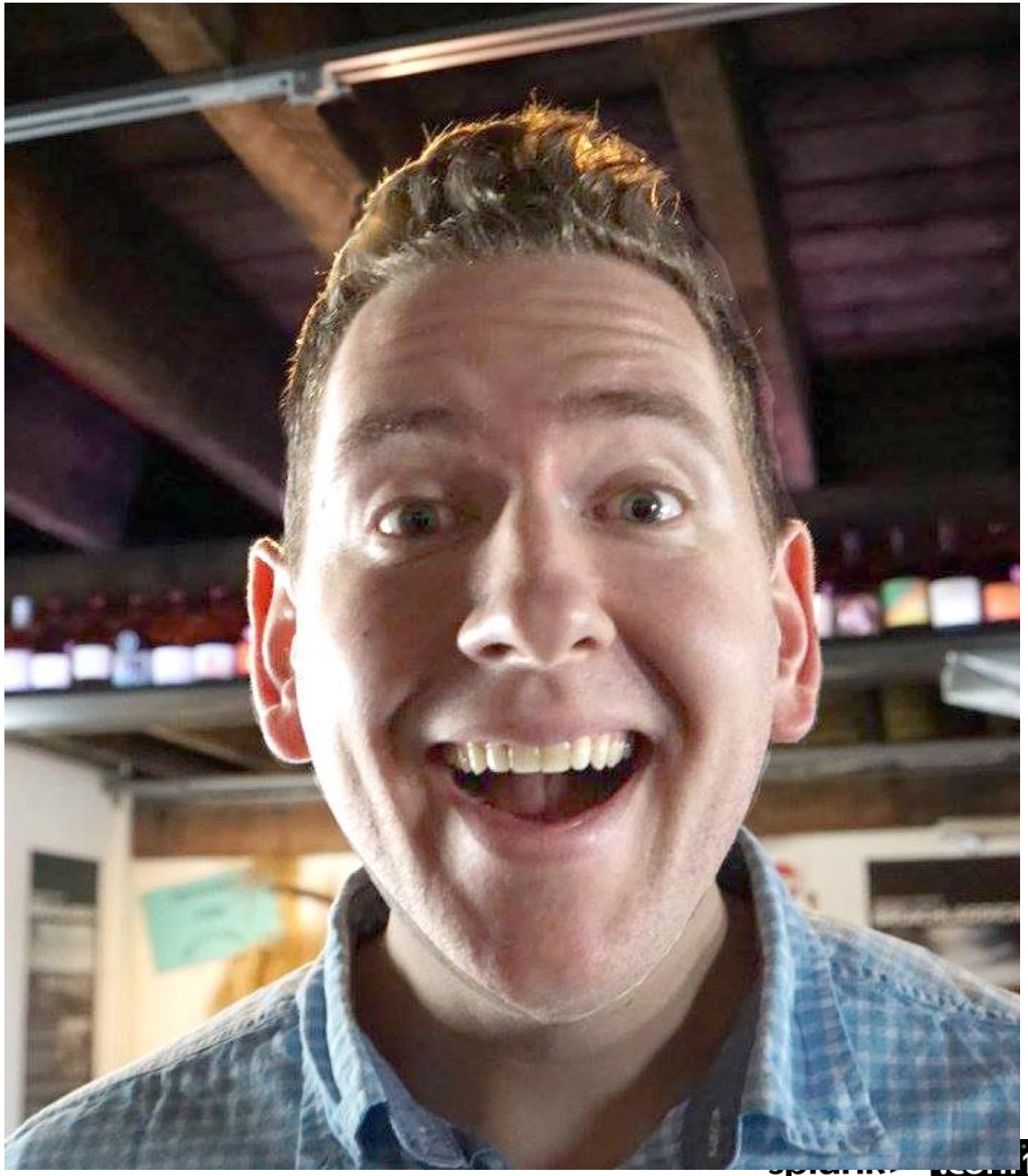
Billions of Raw Logs

Thousands of Stats Rows

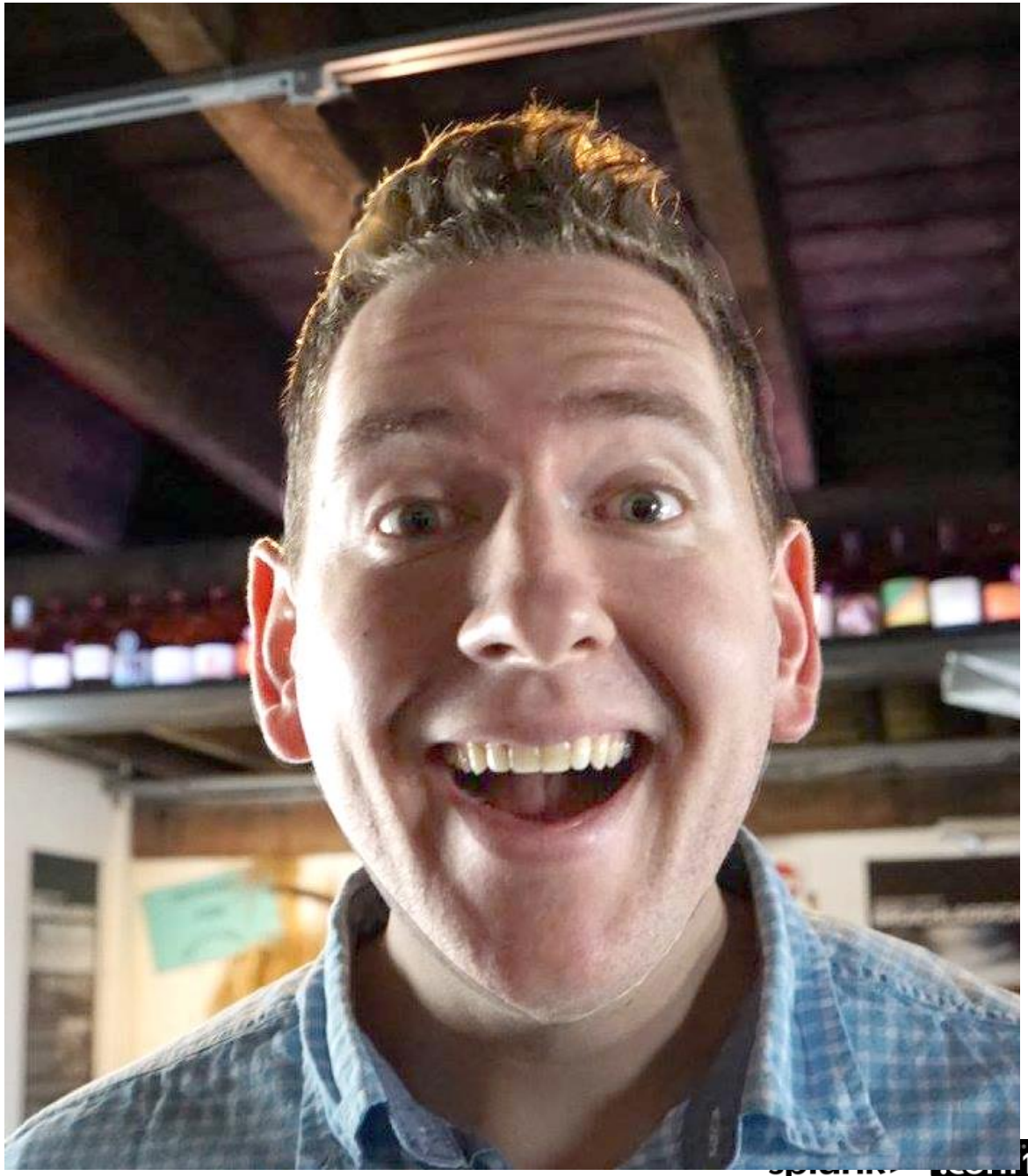
Data Model Acceleration


I'M TIRED OF THEORY!
I WANT SPL!

YES!



Consolidation of all these queries done by



Shout Out!

You're about to see a lot of SPL on the next slides. This was put together by

David Wells

Sr Manager | PWC Canada

splunk> .conf19

Risk

Baselining with Confidence Checks (I)

This verifies that the user is 3x their standard deviation AND there are at least 7 previous days worth of risk scores

```
<Pull risk scores>
| bucket _time span=1d
| stats sum(risk_score) as risk_score by user, _time
| stats count as num_data_samples max(eval(if(_time >= relative_time(now(), "-1d@d"), risk_score,null))) as
    latest
  avg(eval(if(_time<relative_time(now(), "-1d@d"), risk_score,null))) as avg stdev(eval(if(_time<relative_time
      (now(), "-1d@d"), risk_score,null))) as
  stdev by user
| where latest > avg + stdev * 3 AND num_data_samples > 7 AND latest > avg * 2
```


Risk

Baselining with Confidence Checks (II)

Identify When A User's # of Risk Kill Chain (or category) is Above 2 and the Number of Unique Risk Signatures is Above 1

```
<Pull risk scores>
| stats sum(risk_score) as risk_score_aggregate values(risk_id) as risk_id values(risk_description) as
    risk_description values(risk_stage) as
  risk_stage by risk_object
| where mvcount(risk_stage)>2 AND mvcount(risk_id)>1
```


OOTB with SSE

Risk

Baselining with Confidence Checks (III)

Calculate a User's 30 Day Risk Score As a Baseline and Identify When Today's is 3x Higher Than the Average

```
index=risk earliest=-30d
| stats values(source) as search_names sum(risk_score) as thirty_day_risk sum(eval(if(_time > relative_time(now
    (), "-1d"),risk_score,0))) as
  one_day_risk by risk_object
    | eval threshold_1day = 500, threshold_30day = 1200
    | eventstats avg(thirty_day_risk) as avg_thirty_day_risk stdev(thirty_day_risk) as stdev_thirty_day_risk
| where one_day_risk>threshold_1day OR thirty_day_risk>threshold_30day OR thirty_day_risk> (avg_thirty_day_risk
    + 3 * stdev_thirty_day_risk)
```


Risk

Baselining with Confidence Checks (IV)

Calculate if a User is Above the One Day Risk Threshold, the 30 Day Risk Threshold or More Than 3x Its Own Standard Deviation

```
index=risk earliest=-30d
| stats values(source) as search_names     sum(risk_score) as thirty_day_risk sum(eval(if(_time >
    relative_time(now(), "-1d"),risk_score,0))) as one_day_risk by     risk_object
| eval threshold_1day = 500, threshold_30day = 1200
| eventstats avg(thirty_day_risk) as avg_thirty_day_risk     stdev(thirty_day_risk) as stdev_thirty_day_risk
| where one_day_risk>threshold_1day OR     thirty_day_risk>threshold_30day OR     thirty_day_risk
    >(avg_thirty_day_risk + 3 * stdev_thirty_day_risk)
| eval risk_score_reason = case(one_day_risk>threshold_1day, "One Day Risk Score above " . threshold_1day,
    thirty_day_risk>threshold_30day   . " on " . strftime(now(), "%m-%d-%Y"), "Thirty Day Risk Score above " .
    threshold_30day, 1=1, "Thirty Day Risk Score more than three standard deviations above normal (>" . round
    ((avg_thirty_day_risk + 3 * stdev_thirty_day_risk),2) . ")")
| fields - avg* stdev*
```


OOTB with SSE

Spikes

Standard Deviation Anomaly (Good)

Identify When Something Is X Times Past Their Standard Deviation

```
<datasource> | bucket _time span=1d | stats count by <monitored> _time
| stats max(eval(if(_time >= relative_time(now(), "-1d@d"),count, null))) as latest
        avg(eval(if(_time < relative_time(now(), "-1d@d"),count, null))) as avg
        stdev(eval(if(_time < relative_time(now(), "-1d@d"),count, null))) as stdev
    by <monitored>
| where latest > avg + 6*stdev
```

Spikes

Standard Deviation Anomaly (Better)

Adding Relative Filters to Statistical Assessments

```
tag=authentication
| bucket _time span=1d
| stats dc(dest) as count by user, _time
| stats count as num_data_samples
        max(eval(if(_time >= relative_time(now(), "-1d@d"), count,null))) as latest
        avg(eval(if(_time<relative_time(now(), "-1d@d"), count,null))) as avg
        stdev(eval(if(_time<relative_time(now(), "-1d@d"), count,null))) as stdev
    by user
| where latest > avg + stdev * 3 AND num_data_samples > 7 AND latest > avg * 2
```


Spikes

Standard Deviation Anomaly (CC #s)

Look for users who view more credit cards than they typically do

```
index=crm_logs viewed card
| bin span=1d _time
| stats dc(card_id) as count by user _time
| stats count as num_data_samples
        max(eval(if(_time >= relative_time(now(), "1d"), count, null))) as latest
        avg(eval(if(_time < relative_time(now(), "-1d"),count,null))) as average
        stdev(eval(if(_time < relative_time(now(), "-1d"),count,null))) as stdev
    by user
| where latest> 2*stdev+average AND num_data_samples>7 AND latest > avg * 2
```


Spikes

IQR Anomaly (Raw Data)

```
<datasource>
| bucket _time span=1d
| stats count by <monitored>
| eventstats perc25(count) as perc25
             perc75(count) as perc75
        by <monitored>
| where count > perc75 + (perc75 - perc25) * 1.5
```

Spikes

IQR Anomaly (tstats Data)

```
| tstats count from datamodel=<datamodel>
                where earliest=-30d@d
        by <monitored> _time span=1d
| eventstats perc25(count) as perc25
             perc75(count) as perc75
        by <monitored>
| where count > perc75 + (perc75 - perc25) * 1.5
```

Ratios

Detect Rare Events by Ratio (Raw Data)

```
<datasource> earliest=-30d@d
| stats count latest(_time) as latest
    by <monitored> [ optionally: <entity>]
| eventstats sum(count) as total
    [ optionally: by <entity>]
| where count / total < 1/20000 AND
        latest > relative_time(now(), "-1d@d")
```

Ratios

Detect Rare Events by Ratio (tstats Data)

```
| tstats count latest(_time) as latest
      from datamodel=<…>
      where earliest=-30d@d
      by <monitored>
| eventstats sum(count) as total
| where count / total < 1/20000 AND
        latest > relative_time(now(), "-1d@d")
```

Ratios

Detect Rare Events by Ratio (Example I)

Looking for unusual errors in AWS Logs

```
| tstats count latest(_time) as latest
        from datamodel=Example_AWS_Security
        where earliest=-30d@d
        by cloudtrail.errorCode
| eventstats sum(count) as total
| where count / total < 1/20000 AND
        latest > relative_time(now(), "-1d@d")
```

Ratios

Detect Rare Events by Ratio (Example II)

Looking for unusual MFA anomalies across the env in CloudTrail logs

```
| tstats count latest(_time) as latest
      from datamodel=Example_AWS_Security
      where earliest=-30d@d
      by cloudtrail.mfaAuthenticated, cloudtrail.userIdentity.arn
| eventstats sum(count) as total
| where count / total < 1/20000 AND
        latest > relative_time(now(), "-1d@d")
```

Ratios

Detect Rare Events by Ratio (Example III)

Looking for unusual APIs per user in CloudTrail logs

```
sourcetype=aws:cloudtrail earliest=-30d@d
| stats count
    by eventName, userIdentity.arn
| eventstats sum(count) as total
    by userIdentity.arn
| where count / total < 1/20000 AND
        latest > relative_time(now(), "-1d@d")
```

First Time

Detect New (Rare) Events (Raw Data)

```
<datasource>
| stats earliest(_time) as earliest
        latest(_time) as latest
    by <field(s)>
| where _time > relative_time(now(), "-1d@d")
```

First Time

Detect New (Rare) Events (tstats Data)

```
| tstats summariesonly=t allow_old_summaries=t
        min(_time) as earliest
        max(_time) as latest
    from datamodel=<..>
    by <..>
| where earliest > relative_time(now(), "-1d@d")
```

First Time

Detect New (Rare) Events (Example)

Detect When Users Take High Risk Actions From A New Country

```
| tstats summariesonly=t allow_old_summaries=t
      min(_time) as earliest max(_time) as latest
      from datamodel=Example_AWS_Security
      where cloudtrail.HighRiskAPICalls>0
      by cloudtrail.sourceIPAddress_Country
| where earliest > relative_time(now(), "-1d@d")
```

First Time

Detect New (Rare) Events (Examples++)

First Logon to New Server
```
sourcetype=win*security
| stats earliest(_time) as earliest latest(_time) as latest by user, dest
| eval isOutlier=if(earliest >= relative_time(now(), "-1d@d"), 1, 0)
```

Authentication against a New Domain Controller
```
sourcetype=win*security
| stats earliest(_time) as earliest latest(_time) as latest by user, dc
| eval isOutlier=if(earliest >= relative_time(now(), "-1d@d"), 1, 0)
```

First Access to a New Source Code Repository
```
sourcetype=source_code_access
| stats earliest(_time) as earliest latest(_time) as latest by user, repo
| eval isOutlier=if(earliest >= relative_time(now(), "-1d@d"), 1, 0)
```

First External Email Claiming to be Internal from Server
```
sourcetype=cisco:esa src_user=*@mycompany.com src!=10.0.0.0/8
| stats earliest(_time) as earliest latest(_time) as latest by user, src
| eval isOutlier=if(earliest >= relative_time(now(), "-1d@d"), 1, 0)
```

Familiar Filename on a New Path
```
Sourcetype=win*security EventCode=4688 `IncludeMicrosoftFiles`
| stats earliest(_time) as earliest latest(_time) as latest by filename, path
| eval isOutlier=if(earliest >= relative_time(now(), "-1d@d"), 1, 0)
```

New Database Table Accessed
```
sourcetype=database
| stats earliest(_time) as earliest latest(_time) as latest by user, table
| eval isOutlier=if(earliest >= relative_time(now(), "-1d@d"), 1, 0)
```

New Interactive Logon by Service Account
```
sourcetype=win*security user=srv_* Logon_Type=2 OR .. 11 .. 12
| stats earliest(_time) as earliest latest(_time) as latest by user, dest
| eval isOutlier=if(earliest >= relative_time(now(), "-1d@d"), 1, 0)
```

New Parent Process for cmd.exe
```
sourcetype=win*security EventCode=4688 filename=4688
| stats earliest(_time) as earliest latest(_time) as latest by parent_process
| eval isOutlier=if(earliest >= relative_time(now(), "-1d@d"), 1, 0)
```

First Time

Peer Group Analysis

Detect actions that are new not just for a user, but for an entire peer group

```
<datasource>
    | stats earliest(_time) as earliest latest(_time) as latest by user, dest
    | inputlookup append=t sample_cache_group.csv
    | stats min(earliest) as earliest max(latest) as latest by user, dest
    | outputlookup sample_cache_group.csv
    | lookup peer_group.csv user OUTPUT peergroup
    | makemv peergroup delim=","
    | multireport
    [| stats values(*) as * by user dest ]
    [| stats values(eval(if(earliest>=relative_time(now(),"-1d@d"),dest ,null))) as peertoday
      values(eval(if(earliest<relative_time(now(),"-1d@d"),dest ,null))) as peerpast by peergroup dest ]
    | eval user=coalsce(user, peergroup)
    | fields - peergroup
    | stats values(*) as * by user dest
    | where isnotnull(earliest) AND earliest>=relative_time(now(),"-1d@d") AND isnull(peerpast),1,0)
```


Rarity

Combined Anomaly Detection Methods I

Identify Higher-risk IP Addresses Based On The Uniqueness of the IPS signature

```
tag=ids tag=attack
| bucket _time span=1d
| stats count by severity signature dest _time
| stats sum(count) as count
        avg(count) as avg
        stdev(count) as stdev
        sum(eval(if(_time > relative_time(now(), "-1d"), count, 0))) as recent_count
        min(_time) as earliest
    by severity signature dest
| eventstats avg(avg) as avg_num_per_dest
             avg(earliest) as avg_earliest
             sum(count) as sig_wide_count
             sum(recent_count) as sig_wide_recent_count
        by signature
| where NOT (avg_earliest < relative_time(now(), "-1y") AND
             sig_wide_recent_count / sig_wide_recent_count < 0.05 AND
             priority <=3)
```

Rarity

Combined Anomaly Detection Methods II

Alert When Users Who Usually Log Into Few Systems Suddenly Log Into Many

```
| tstats summariesonly=true count from datamodel=Authentication where earliest=-30d@d groupby
    Authentication.dest Authentication.user _time span=1d
| rename Authentication.dest as dest Authentication.user as user
| eval isRecent=if(_time>relative_time(now(),"-1d"), "yes", "no")
| stats avg(eval(if(isRecent="no",count,null))) as avg
        first(count) as recent
    by user, dest
| eventstats count(eval(if(avg>0,"yes",null))) as NumServersHistorically
            count(eval(if(recent>0,"yes",null))) as NumServersRecently
    by user
| eval Cause=if(isnull(avg) AND NumServersHistorically!=0, "This is the first logon to this server", "")
| eval Cause=if(NumServersRecently>3 AND NumServersHistorically * 3 < NumServersRecently,
    mvappend(Cause,"Substantial increase in the number of servers logged on to"), Cause)
| where Cause!=""
```

Combined

Multiple Data Sources (CIM)

```
tag=authentication
| chart count over src by action
| where success>0 AND failure>10
```

Combined

Multiple Data Sources (Multiple CIM)

```
(tag=malware action=allowed) OR (tag=proxy category=uncategorized)
| eval mydest=if(tag="malware", dest, src)
| stats count(eval(tag="malware")) as malware
        count(eval(tag="proxy")) as proxy
    by mydest
| where malware>0 AND proxy>0
```

Combined

No Transaction, Yes Eventstats + Stats

Ironport logs are a canonical transaction example, but searches are faster without

```
sourcetype=ironport OR sourcetype=cisco:esa
| eventstats values(TLS) as TLS
             values(src_ip) as src_ip
             values(…) as …
    by ICID
| stats values(icid) AS icid
        values(src*) AS src*
    by mid
| eval recipient_count=mvcount(recipient)
```

Combined

Persist into Summary Index!

Provide a single consolidated record with all relevant values

```
sourcetype=cisco:esa* earliest=-20m
| eventstats values(sending_server) as sending_server values(sending_server_dns_status) as
    sending_server_dns_status values(sending_server_dkim) as sending_server_dkim
    values(sending_server_tls_status) as sending_server_tls_status values(sending_server_tls_cipher) as
        sending_server_tls_cipher values(sending_server_whitelist) as sending_server_whitelist by
    icid
| stats min(_time) as _time max(_time) as email_processing_complete_time count(eval(searchmatch("Message
    Finished MID"))) as complete_count count(eval(searchmatch("Start MID"))) as
    start_count values(d) as d values(message_id) as message_id values(message_subject) as message_subject
        values(mid) as mid values(recipient) as recipient values(sender) as sender
    values(spam_status) as spam_status values(encoding) as encoding values(subject) as subject values
        (attachment) as attachment values(queue) as queue values(message_scan_error) as
    message_scan_error values(message_size) as message_size values(sending_server) as sending_server values
        (sending_server_dns_status) as sending_server_dns_status
    values(sending_server_dkim) as sending_server_dkim values(sending_server_tls_status) as
        sending_server_tls_status values(sending_server_tls_cipher) as sending_server_tls_cipher
    values(sending_server_whitelist) as sending_server_whitelist values(icid) as icid values(dcid) as dcid by
        mid
| where complete_count > 1 AND start_count > 1 AND email_processing_complete_time >= relative_time(now(), "
    -7m@m") AND email_processing_complete_time < relative_time(now(),
    "-2m@m")
| collect index=parsed_emails
```

Combined

Multiple Data Sources (tstats)

```
| tstats prestats=t summariesonly=t count(Malware_Attacks.src)
        from datamodel=Malware
        where Malware_Attacks.action=allowed
        by Malware_Attacks.src
| tstats prestats=t append=t summariesonly=t count(web.src)
        from datamodel=Web
        where web.http_user_agent="shockwave flash"
        groupby web.src
| rename web.src as src Malware_Attacks.src as src
| stats count(Malware_Attacks.src) as malwarehits
        count(web.src) as webhits by src
| where malwarehits > 0 AND
        webhits > 0
```

Misc

Superman Analysis

Find users moving faster than a jet engine flies

```
index=* sourcetype=aws:cloudtrail user=*
| sort 0 user, _time | streamstats window=1 current=f values(_time) as last_time values(src) as last_src by
    user
| where last_src != src AND _time - last_time < 8*60*60
| iplocation last_src | rename lat as last_lat lon as last_lon | eval location = City . "|" . Country . "|" .
    Region
| iplocation src
| eval rlat1 = pi()*last_lat/180, rlat2=pi()*lat/180, rlat = pi()*(lat-last_lat)/180, rlon= pi()*(lon-last_lon
    )/180 | eval a = sin(rlat/2) * sin(rlat/2) + cos(rlat1) * cos(rlat2) * sin(rlon/2) * sin(rlon/2) | eval c =
    2 * atan2(sqrt(a), sqrt(1-a)) | eval distance = 6371 * c, time_difference_hours = round((_time - last_time)
    / 3600,2), speed=round(distance/ ( time_difference_hours),2) | fields - rlat* a c
| eval day=strftime(_time, "%m/%d/%Y")
| stats values(accountId) values(awsRegion) values(eventName) values(distance) values(eval(mvappend
    (last_Country, Country))) as Country values(eval(mvappend(last_City, City))) as City values(eval(mvappend
    (last_Region, Region))) as Region  values(lat) values(lon)  values(userAgent) max(speed) as max_speed_kph
    min(time_difference_hours) as min_time_difference_hours by day user distance
```


Misc

Randomness Detection w/ URL Toolbox

```
index=* sourcetype=aws:cloudtrail user=*
| sort 0 user, _time | streamstats window=1 current=f values(_time) as last_time values(src) as last_src by
    user
| where last_src != src AND _time - last_time < 8*60*60
| iplocation last_src | rename lat as last_lat lon as last_lon | eval location = City . "|" . Country . "|" .
    Region
| iplocation src
| eval rlat1 = pi()*last_lat/180, rlat2=pi()*lat/180, rlat = pi()*(lat-last_lat)/180, rlon= pi()*(lon-last_lon
    )/180 | eval a = sin(rlat/2) * sin(rlat/2) + cos(rlat1) * cos(rlat2) * sin(rlon/2) * sin(rlon/2) | eval c =
    2 * atan2(sqrt(a), sqrt(1-a)) | eval distance = 6371 * c, time_difference_hours = round((_time - last_time)
    / 3600,2), speed=round(distance/ ( time_difference_hours),2) | fields - rlat* a c
| eval day=strftime(_time, "%m/%d/%Y")
| stats values(accountId) values(awsRegion) values(eventName) values(distance) values(eval(mvappend
    (last_Country, Country))) as Country values(eval(mvappend(last_City, City))) as City values(eval(mvappend
    (last_Region, Region))) as Region  values(lat) values(lon)  values(userAgent) max(speed) as max_speed_kph
    min(time_difference_hours) as min_time_difference_hours by day user distance
```

Context

Add in Organizational Information

```
...
| table user *
| inputlookup LDAPSearch user OUTPUT title department num_reports
```

user	action	department	num_reports	title
dveuve	Non-work Web Browsing	Global Security Strategist Team	0	Principal Security Strategist

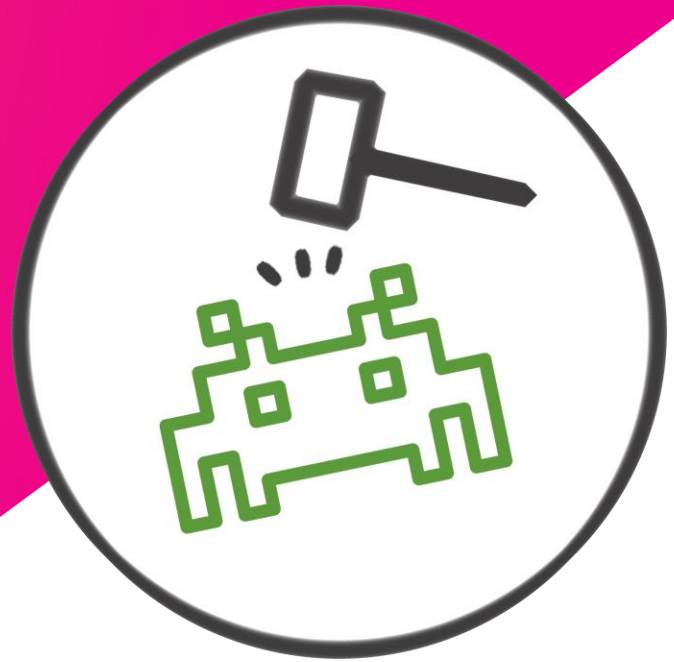
Context

Calculate Risk based on HR Data

Estimate level of exposure numerically

```
...
| inputlookup LDAPSearch user
| eval risk = 1
| eval risk = if(NumWhoReportIn>100, risk+10, risk)
| eval risk = if(like(Groups, "%,OU=IT Security,%"), risk + 10, risk)
| eval risk = case(like(title, "VP %"), risk+10,
                   like(title, "Chief %"), risk+100,
                   1=1, risk)
| fields risk sAMAccountName
| outputlookup RiskPerUser
```

Alert Overload!



.conf19
splunk>

Types of Events

High Fidelity, Actionable, Automatable

Can Be Aggregated to create alerts or used for context

Provides relevant hints to use during an investigation

Types of Events

High Fidelity, Actionable, Automatable

Can Be Aggregated to create alerts or used for context

Provides relevant hints to use during an investigation

Types of Events

High Fidelity, Actionable, Automatable

Can Be Aggregated to create alerts or used for context

Provides relevant hints to use during an investigation

Types of Events

High Fidelity, Actionable, Automatable

Can Be Aggregated to create alerts or used for context

Provides relevant hints to use during an investigation

We Call it Risk-Based Alerting

SEC 1556 – Building Behavioral Detections: Cross-Correlating Suspicious Activity with the MITRE ATT&CK Framework

Tuesday, October 22, 01:45 PM - 02:30 PM

SEC 1803 – Modernize and Mature Your SOC with Risk-Based Alerting

Tuesday, October 22, 03:00 PM - 03:45 PM

SEC 1538 - Getting started with Risk-Based Alerting and MITRE

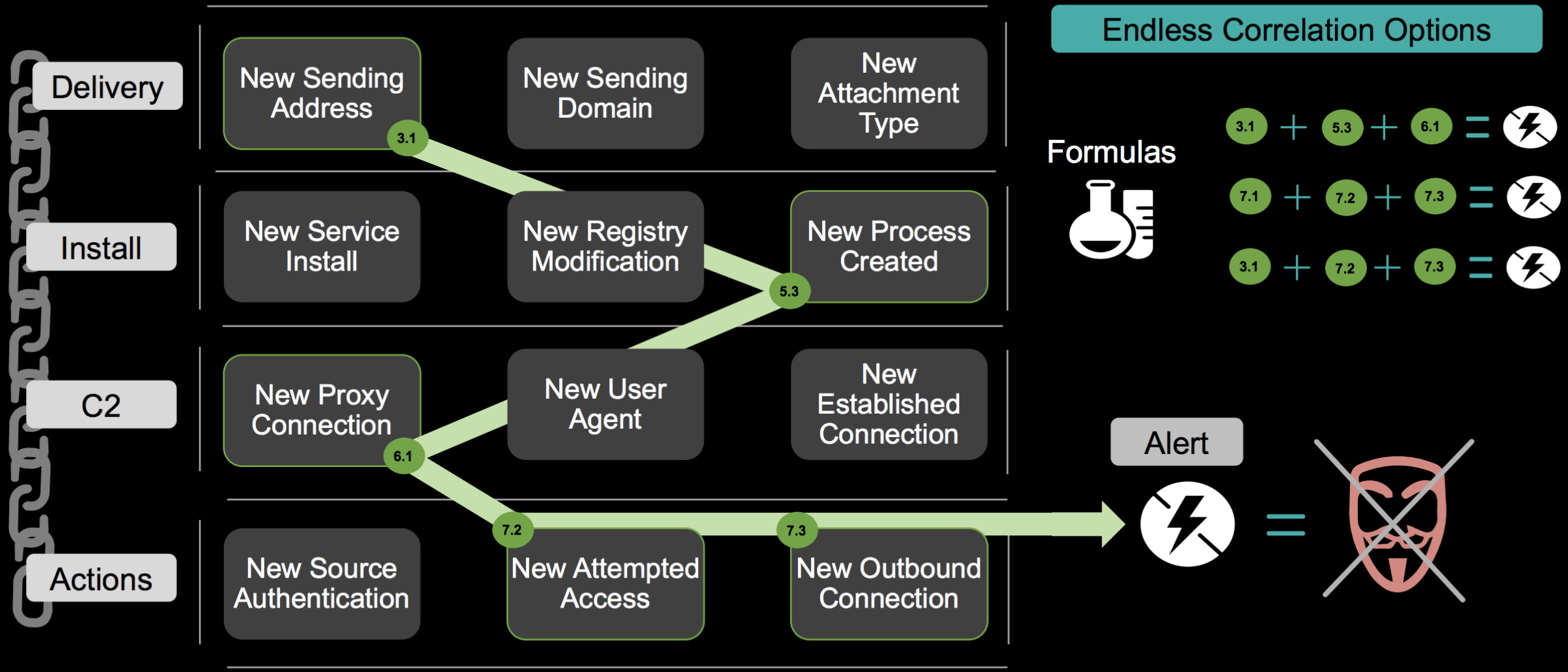
Wednesday, October 23, 12:30 PM - 01:15 PM

SEC 1908 – Tales from a Threat Team: Lessons and Strategies for Succeeding with a Risk-Based Approach

Wednesday, October 23, 03:00 PM - 03:45 PM

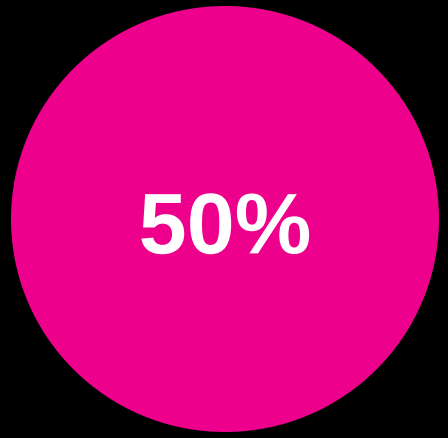
Birds of a Feather – Meet the RBA Community

SUGARCANE Raw Bar Grill – Tuesday 6:30-8:30

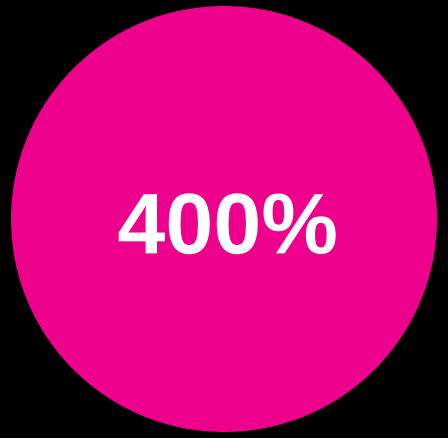


AmFam Success Metrics Seen

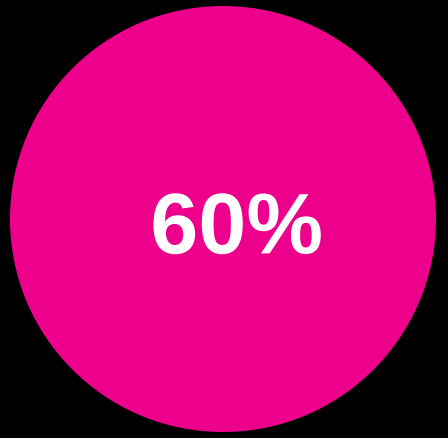
American Family Insurance – Comparing Alert Driven to Risk Driven



Decrease in Notable Events Generated



Increase in # of Security Detections

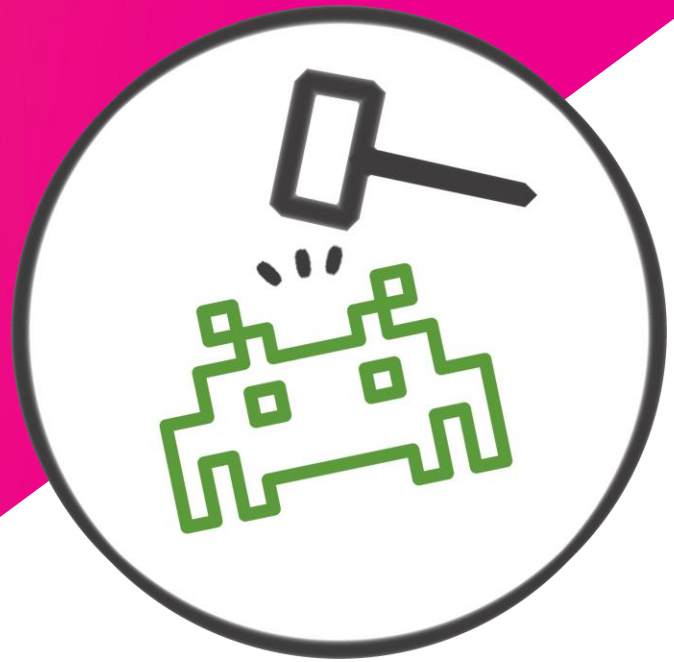


True Positives, up from 10%!

“60% of the Risk-based Alerting incidents are true positives.

Security staff at company that rolled out RBA 5 months ago

Let's Summarize



.conf19
splunk>


TOO MUCH SPL!

How will you ever remember?

Use The Cheat Codes

Splunk Security Essentials

www.splunksecurityessentials.com

The free app to make security easier

123 davidveuve.com/ninjutsu

All SSE Slide Decks All Available!

davidveuve.com/splunk.html

- Copy-paste SPL from the Presentations
- Slide # to Source Material mapping
- David Wells' Shortcut Preso

And With That!

davidveuve.com/splunk.html
splunksecurityessentials.com

You Have The Power

davidveuve.com/splunk.html
splunksecurityessentials.com

You Have The Tools

davidveuve.com/splunk.html
splunksecurityessentials.com

You Have The SPL

davidveuve.com/splunk.html
splunksecurityessentials.com

Go Find The Bad People

davidveuve.com/splunk.html
splunksecurityessentials.com

Thank You!

davidveuve.com/splunk.html

splunksecurityessentials.com

Please Rate This Session!