Forward-Looking Statements

This presentation may contain forward-looking statements regarding future events, plans or the expected financial performance of our company, including our expectations regarding our products, technology, strategy, customers, markets, acquisitions and investments. These statements reflect management's current expectations, estimates and assumptions based on the information currently available to us. These forward-looking statements are not guarantees of future performance and involve significant risks, uncertainties and other factors that may cause our actual results, performance or achievements to be materially different from results, performance or achievements expressed or implied by the forward-looking statements contained in this presentation.

For additional information about factors that could cause actual results to differ materially from those described in the forward-looking statements made in this presentation, please refer to our periodic reports and other filings with the SEC, including the risk factors identified in our most recent quarterly reports on Form 10-Q and annual reports on Form 10-K, copies of which may be obtained by visiting the Splunk Investor Relations website at www.investors.splunk.com or the SEC's website at www.sec.gov. The forward-looking statements made in this presentation are made as of the time and date of this presentation. If reviewed after the initial presentation, even if made available by us, on our website or otherwise, it may not contain current or accurate information. We disclaim any obligation to update or revise any forward-looking statement based on new information, future events or otherwise, except as required by applicable law.

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. We undertake no obligation either to develop the features or functionalities described or to include any such feature or functionality in a future release.

splunk> .conf21

Getting To Know Your Data

An SPL Prerequisite
TRU1192B

Mary Cordova

Senior Manager |
Data Engineering & Analytics
Information Security

splunk> .conf21

Mary Cordova

Senior Manager | Data Engineering & Analytics Information Security

- Splunk Trust member & Splunk Certified Architect
- Nearly a decade in SIEM, SOAR, IR, & Data Analytics
- B.S. Information Systems, CCNA, 6xSANS, SSCP, ISC^2 exam developer
- Splunk talks for Shellcon, WSC, SD DFIR

splunk> .conf21

Agenda

1) Why do you need to know your data before you start your SPL?

2) Process & Method

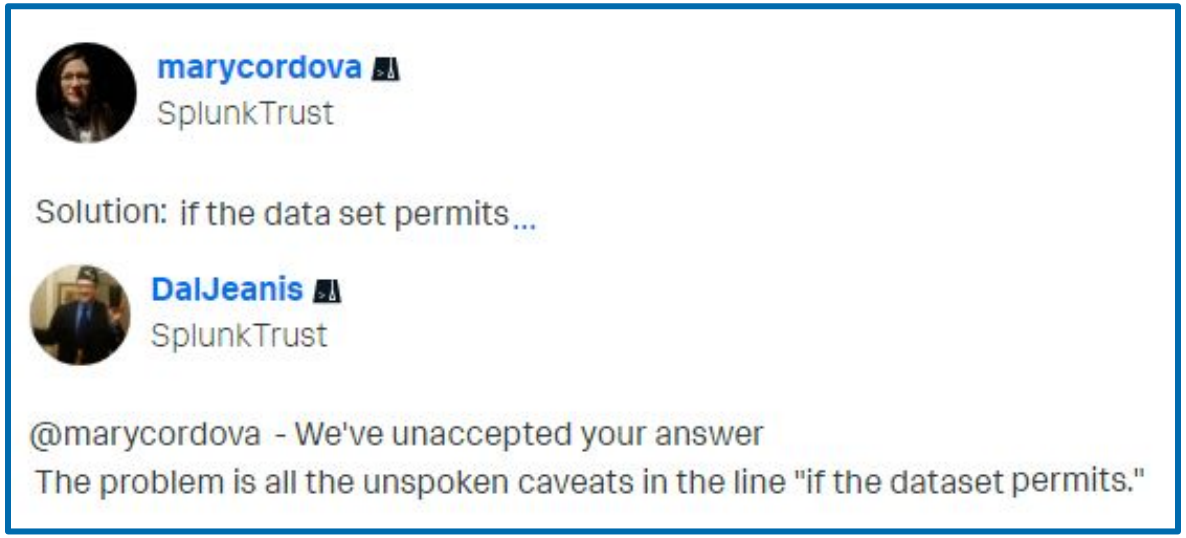
3) Demo

4) Wrap Up & Resources

splunk> .conf21

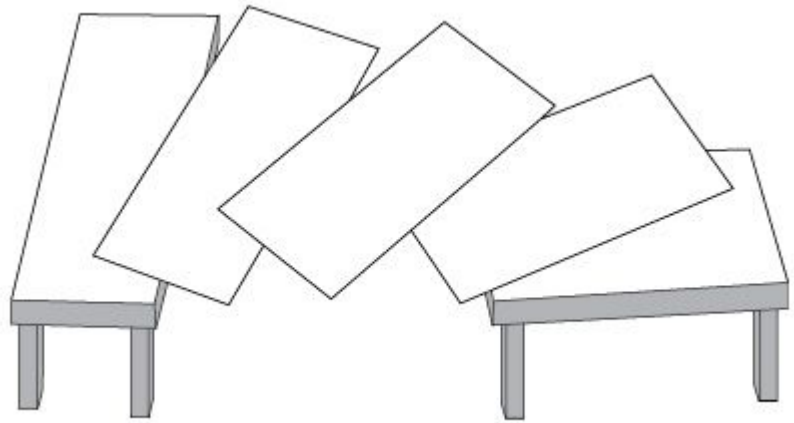
If the Dataset Permits…

- ~5k “Search” questions on Splunk Answers
- pick any question at random…
  - the answer is probably some version of “if the dataset permits”



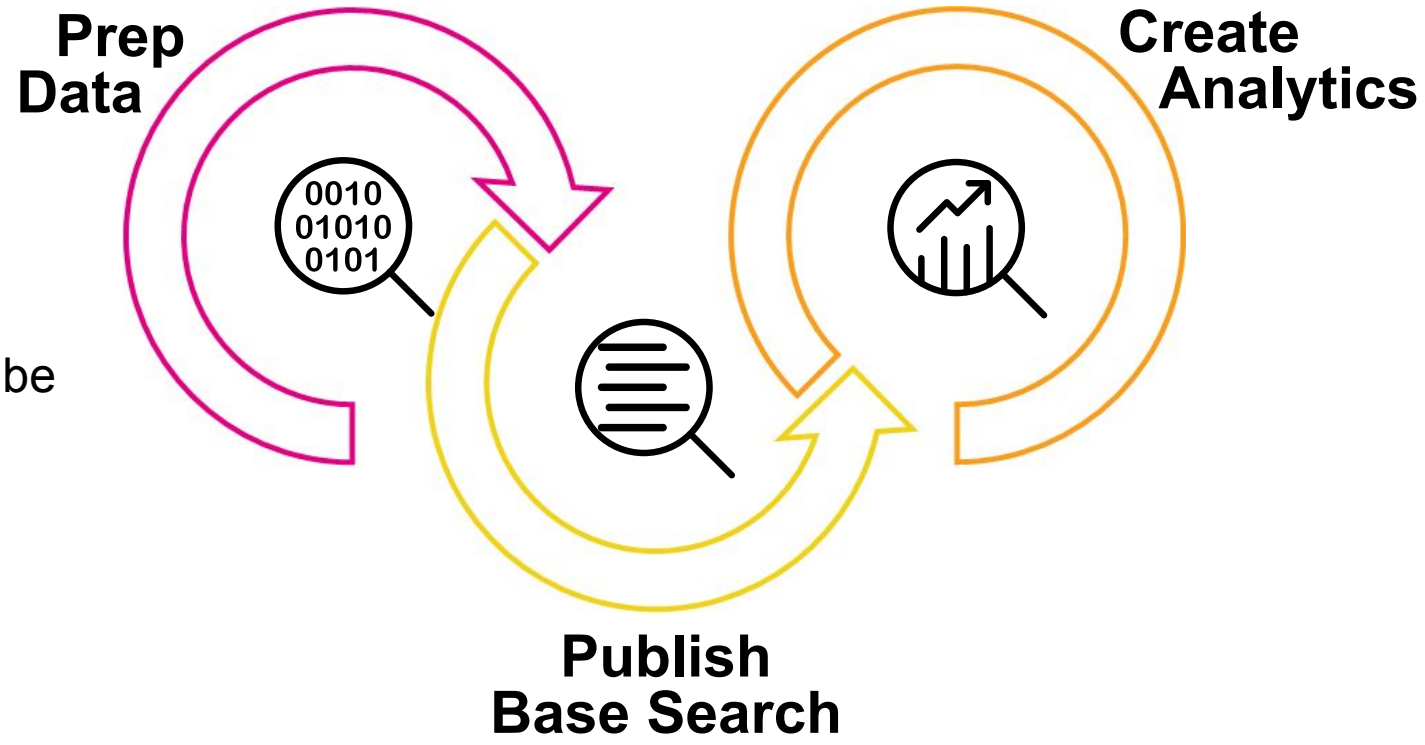
But There's an App (TA) for That...

- But what if there wasn't an App or TA?
- Or the log format changed and doesn't match the TA anymore?
- Or what if the fields aren't CIM compliant?
- What if you're developing a custom use case?
- What if it's a custom log source?
- What if you just don't want to table out the same fields over and over every time you work with a dataset?

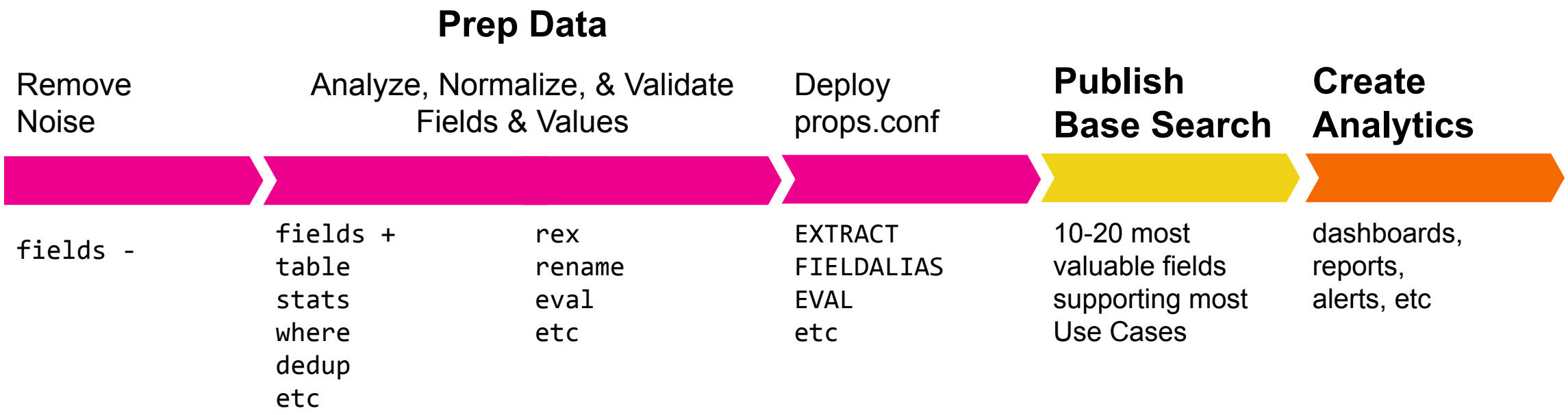


Process

- Prepping data takes time
  - but it will save time in the long run
- Base searches enable your users to be self-sufficient



Process > Method



Method > Setup

- verbose mode
- **static** time range
- representative data sample
- data source **Admin Guide**
- CIM data model(s) reference

Sysmon v13.22

06/22/2021 • 14 minutes to read • +2

By Mark Russinovich and Thomas Garnier

```
index=botsv1 sourcetype=XmlWinEventLog:Microsoft-Windows-Sysmon/Operational
```

during Wed, Aug 24, ...

✓ 167,894 events (8/24/16 12:00:00.000 AM to 8/25/16 12:00:00.000 AM) No Event Sampling Job Verbose Mode

Data models

- How to use the CIM data model reference tables
- CIM fields per associated data model
- Alerts
- Application State (deprecated)
- Authentication
- Certificates
- Change
- Change Analysis (deprecated)
- Data Access
- Databases
- Data Loss Prevention
- Email
- **Endpoint**
- Event Signatures
- Interprocess Messaging
- Intrusion Detection

Method > Remove Splunk Noise

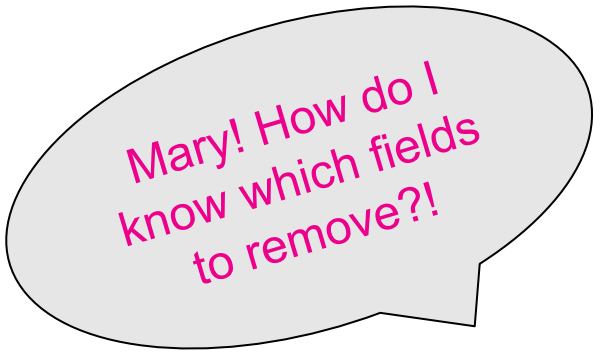
```
index=botsv1 sourcetype=XmlWinEventLog:Microsoft-Windows-Sysmon/Operational
| fields - date* time*pos index source sourcetype splunk_server linecount
punct tag tag::eventtype eventtype vendor_product
```

Method > Isolate Fields to Work With

- focus on fields with 100% coverage to start


```
index=botsv1 sourcetype=XmlWinEventLog:Microsoft-Windows-Sysmon/Operational

| fields - date* time*pos index source sourcetype splunk_server linecount
punct tag tag::eventtype eventtype vendor_product

| table action Computer direction dvc dvc_nt_host EventChannel EventCode
EventDescription EventID host Keywords Level Opcode process RecordID
SecurityID signature signature_id Task TimeCreated UtcTime Version
```

Method > Remove Event Noise




```
index=botsv1 sourcetype=XmlWinEventLog:Microsoft-Windows-Sysmon/Operational

| fields - date* time*pos index source sourcetype splunk_server linecount
punct tag tag::eventtype eventtype vendor_product

| fields - EventChannel EventID Keywords Level Opcode Task Version

| table action Computer direction dvc dvc_nt_host EventCode
EventDescription host process RecordID SecurityID signature signature_id
TimeCreated UtcTime
```

Method > Are You Sure?

- spot check with `stats values(*)`
- high cardinality fields can crash your browser :)

```
index=botsv1 sourcetype=XmlWinEventLog:Microsoft-Windows-Sysmon/Operational
| fields - date* time*pos index source sourcetype splunk_server linecount
punct tag tag::eventtype eventtype vendor_product
| table + EventChannel EventID Keywords Level Opcode Task Version
| stats values(*) as *
```

EventChannel	EventID	Keywords	Level	Opcode	Task	Version
Microsoft-Windows-Sysmon/Operational	1 2 3 5 6 7	0x8000000000000000	4	0	1 2 3 5 6 7	3 4 5

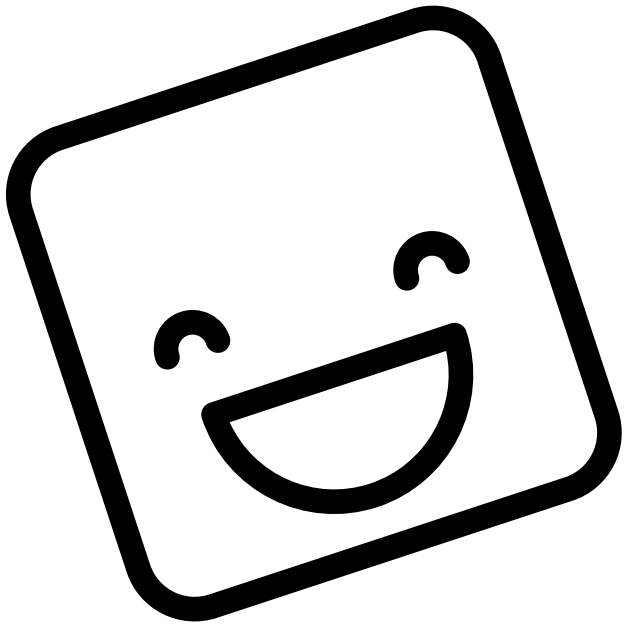
Method > Identify Important Fields

- analyze the remaining fields with `fields`, `table`, `stats`, & the `UI`
- find & group fields that have duplicate, similar, or related values

```
index=botsv1 sourcetype=XmlWinEventLog:Microsoft-Windows-Sysmon/Operational
| fields - date* time*pos index source sourcetype splunk_server linecount
punct tag tag::eventtype eventtype vendor_product
| fields - EventChannel EventID Keywords Level Opcode Task Version
| fields - TimeCreated UtcTime
| table action Computer direction dvc dvc_nt_host EventCode
EventDescription host process RecordID SecurityID signature signature_id
| table Computer dvc dvc_nt_host host
```

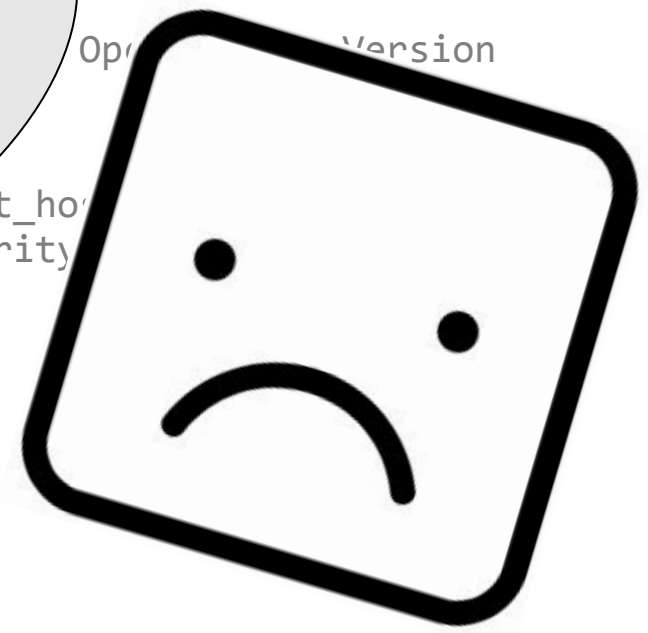
Method > We are Learning, not SPLing

- **don't worry about your search or SPL right now**



```
index=botsv1 sourcetype=XmlWin                    rosoft-Windows-Sysmon/Operational
| fields - date* time*                                    ype splunk_server linecount
punct tag tag::eve                                      ct
| fields - Ever                                          Op        Version
| fields - Ti
| table acti                                            _nt_ho
EventDescrip                                            curity
| table Computer                                  st host
```

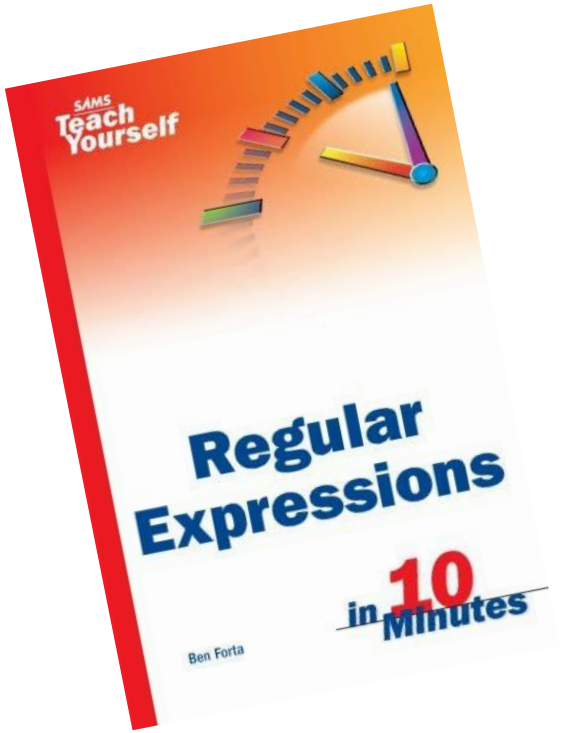
EWWWWW! Look at that search!



splunk> .conf21

Method > Normalize (with CIM)

- begin to craft normalized fields & values




```
index=botsv1 sourcetype=XmlWinEventLog:Microsoft-Windows-Sysmon/Operational

| fields - date* time*pos index source sourcetype splunk_server linecount
punct tag tag::eventtype eventtype vendor_product

| fields - EventChannel EventID Keywords Level Opcode Task Version

| fields - TimeCreated UtcTime

| table action Computer direction dvc dvc_nt_host EventCode
EventDescription host process RecordID SecurityID signature signature_id

| table Computer dvc dvc_nt_host host

| rename dvc_nt_host as dest_nt_host

| rex field=dvc "[\w\d]\.(?<dest_nt_domain>.*)$"

| fields - Computer dvc host
```

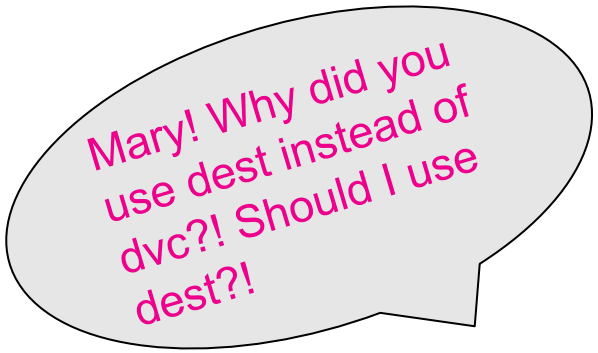
Method > Are You Sure You're Sure?

- spot check with

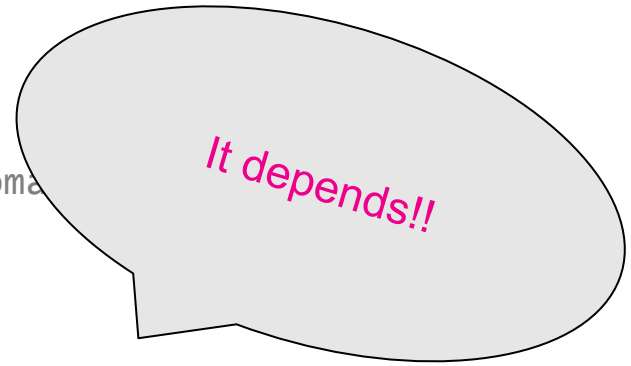
```
stats values()
where isnull()
where isnotnull()
dedup
```

```
index=botsv1 sourcetype=XmlWinEventLog:Microsoft-Windows-Sysmon/Operational
| <snip>
| rex field=dvc "[\w\d]\.(?<dest_nt_domain>.*)$"
| fields - Computer dvc host
| where isnull('dest_nt_host') OR 'dest_nt_host'==""
----------------------------------------
| where isnotnull('dest_nt_host') OR NOT 'dest_nt_host'==""
```

Method > A Little Art & a Little Science




```
index=botsv1 sourcetype=XmlWinEventLog:Microsoft-Windows-Sysmon/Operational
| fields - date* time*pos index source sourcetype splunk_server linecount
punct tag tag::eventtype eventtype vendor_product
| fields - EventChannel EventID Keywords Level Opcode Task Version
| fields - TimeCreated UtcTime
| table action Computer direction dvc dvc_nt_host EventCode
EventDescription host process RecordID SecurityID signature signature_id
| table Computer dvc dvc_nt_host host
| rename dvc_nt_host as dest_nt_host
| rex field=dvc "[\w\d]\.(?<dest_nt_doma
| fields - Computer dvc host
```



splunk> .conf21

Method > Iterate

- iterate, iterate iterate
- find, combine, extract, & normalize fields & values
- keep getting rid of fields that aren't really helpful
- start to look at fields with 90% coverage, then 50%

```
index=botsv1 sourcetype=XmlWinEventLog:Microsoft-Windows-Sysmon/Operational
| fields - date* time*pos index source sourcetype splunk_server linecount
punct tag tag::eventtype eventtype vendor_product
| fields - EventChannel EventID Keywords Level Opcode Task Version
| fields - TimeCreated UtcTime
| rename dvc_nt_host as dest_nt_host
| rex field=dvc "[\w\d]\.(?<dest_nt_domain>.*)$"
| fields - Computer dvc dvc_nt_host host
| fields - dest_nt_domain dest_nt_host
| table action direction EventCode EventDescription process RecordID
SecurityID signature signature_id
| table EventCode EventDescription signature signature_id
```

Method > Clean Up

- combine commands
- sort commands
- remove commands
- table remaining fields with the most important data to the left
  - timestamp
  - system generating event
  - type of event
  - outcome of event
  - then remaining contextual information

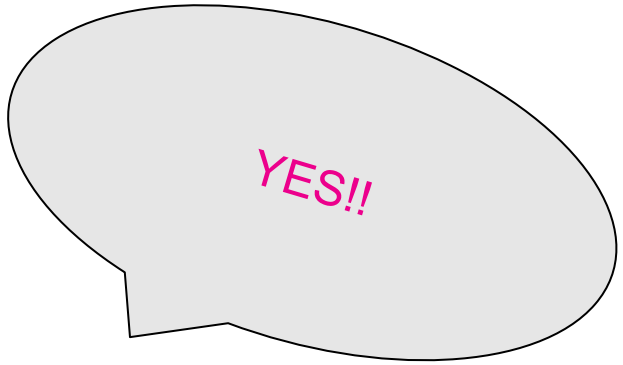
```
index=botsv1 sourcetype=XmlWinEventLog:Microsoft-Windows-Sysmon/Operational

| rename Initiated as initiated ParentCommandLine as parent_cmdline
CurrentDirectory as current_dir SHA256 as sha256 Signature as
image_signature LogonId as logon_id ImageLoaded as image_loaded

| eval src=mvdedup(mvsort(lower(mvappend('src','src_host','src_ip'))))

| eval dest=mvdedup(mvsort(lower(mvappend('dest','dest_host','dest_ip'))))

| eval process=if(isnull('app'),'process','app')

| eval
protocol=mvdedup(mvsort(lower(mvappend('protocol','SourcePortName','DestPor
tName'))))

| table _time dvc signature_id signature process_id process direction user
logon_id initiated src src_port protocol transport dest dest_port
image_loaded image_signature current_dir parent_process_id parent_process
parent_cmdline cmdline file_path sha256
```

Method > Validate

- spot check your base search
- expand your time range
- expand your data sample
- do you have null or empty values?
- do any of your fields have unexpected values?
- is your regex still working for every event?

```
index=botsv1 sourcetype=XmlWinEventLog:Microsoft-Windows-Sysmon/Operational

| <snip>

| table _time dvc signature_id signature process_id process direction user
logon_id initiated src src_port protocol transport dest dest_port
image_loaded image_signature current_dir parent_process_id parent_process
parent_cmdline cmdline file_path sha256

| where isnull('field') OR 'field'==""

----------------------------------------

| where isnotnull('field') OR NOT 'field'==""

----------------------------------------

| stats values(*) as *
```

Method > Deploy & Publish



- deploy your normalization to props.conf
- publish a base search for your users
  - try using macros for your base search catalog

```
|`sysmonbasesearch`

----------------------------------------

index=botsv1 sourcetype=XmlWinEventLog:Microsoft-Windows-Sysmon/Operational

| table _time dvc signature_id signature process_id process direction user
logon_id initiated src src_port protocol transport dest dest_port
image_loaded image_signature current_dir parent_process_id parent_process
parent_cmdline cmdline file_path sha256
```

	signature_id	signature	process_id	process	direction	user
pinc.local	7	Image Load	724	C:\Windows\System32\svchost.exe	inbound	
rpinc.local	5	Process Terminate	4012	C:\Windows\System32\wbem\WmiPrvSE.exe	inbound	
we1149srv.waynecorpinc.local	1	Process Create	2688	C:\Windows\System32\wbem\WmiPrvSE.exe	inbound	NT AUTHORITY\SYSTEM
we1149srv.waynecorpinc.local	3	Network Connect	4	System	outbound	NT AUTHORITY\SYSTEM

Wrap Up

- **Prep Data**
  - **Setup**
    - static time range
    - representative data sample
    - verbose mode
  - **Remove Noise**
    - remove splunk noise
    - remove event noise
  - **Analyze/Normalize Data**
    - identify important fields
    - combine & CIM normalize fields/values
    - reduce to 10-20 most useful fields for any use case
  - **Validate Work**
    - expanded time range and/or data sample
  - **Deploy props.conf**
- **Publish Base Search**
  - tabular format 10-20 most useful fields
- **Create Analytics**

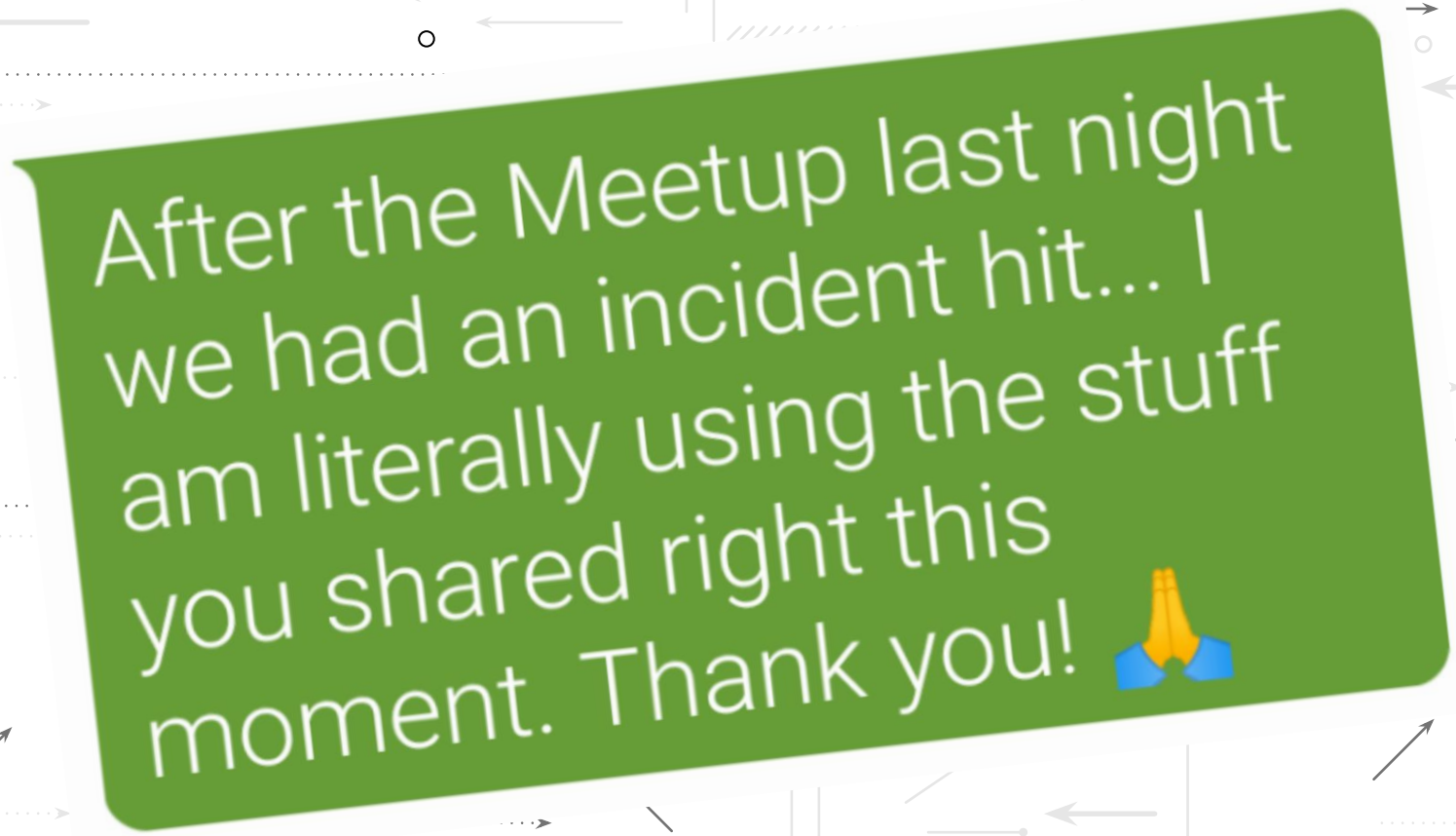
Useful Resources

- Boss of the SOC (BOTS) Workshop
  - https://bots.splunk.com/workshop/3JjlyhUc2P7hYfhkBW4OE3
- Security Dataset Project
  - Introducing the Security Dataset Project
  - Splunk Security Datasets Project
- BOTS Datasets
  - Boss of the SOC v3 Dataset Released!
  - splunk/securitydatasets: Home for Splunk security datasets
- BOTS VM
  - https://cyberdefenders.org/labs/
- Misc
  - Sams Teach Yourself Regular Expressions in 10 Minutes
- Blogs
  - Boss of the SOC V at .conf20
  - Hunting with Splunk: The Basics
  - This is NOT the Data You Are Looking For (OR is it) | Splunk

Education

As with previous years, we know that it can be scary to see new datasets that you've never been exposed to. With that in mind, we will be starting to release blogs, webinars, videos, and more to help you level-up to meet these new challenges. Follow @splunk on Twitter, and subscribe to Splunk Blogs for updates and webinar announcements. For extra points, follow @meansec, @daveherrald, @james_brodsky, and @stonerpsu on Twitter for "special" announcements. To be clear, these blogs will be VERY relevant to BOTS 5.0 at .conf20, so we highly recommend reading them. And of course, don't forget our handy dandy blog series, "Hunting with Splunk: The Basics," which was inspired by the questions customers have asked at BOTS events all over the world!

Finally, you can try out or practice these new techniques using our cloud-hosted "Security Datasets Project" that has the BOTSv1 dataset and more. If you'd rather set up a home lab and really dig into BOTS data, try out our BOTSv1, BOTSv2, and BOTSv3 open-source dataset and CTF scoring server app. If those seem scary, check out the work done by CyberDefenders. They have no affiliation with us, but they seem to have stood up an awesome instance of BOTS data for everyone in the world to play and learn from!



splunk> .conf21

Thank You

Please provide feedback via the

SESSION SURVEY

splunk> .conf21