

The Power of Data Normalization

A look at the Common Information Model

Mark Bonsack, CISSP

| Staff Sales Engineer, Splunk

Vladimir Skoryk, CISSP, CCFE, CHFI, CISA, CISM, RGTT

| PS Supreme Architect, Splunk

September 27, 2017

| Washington, DC

Introduction

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.

Splunk, Splunk>, Listen to Your Data, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2017 Splunk Inc. All rights reserved.

We will discuss:

This is where the subtitle goes

1. Why data normalization is important
2. The Splunk Common Information Model
3. Tools and Resources on the road to CIM compliance

Vulnerability Center

Edit ↓ ↓ ↓

Severity: All Business Unit: Category: All Last 90 days **Submit**

[Edit](#)

VULNS PER SYSTEM
Average Count

6.7 ↑ +6.7

VULNERABLE SYSTEMS
Percent Vulnerable

100 %

VULNERABLE SYSTEMS
System Count

106 ↑ +106

TOTAL VULNS
Count

708 ↑ +708

VULNERABILITY AGE
Average Days

15.2

AUTH. APPS
Distinct Count

19 0

AUTH. DEST'S
Distinct Count

9k ↓ -82

AUTH. SOURCES
Distinct Count

9k ↑ +22

AUTH. USERS
Distinct Count

4k ↓ -47

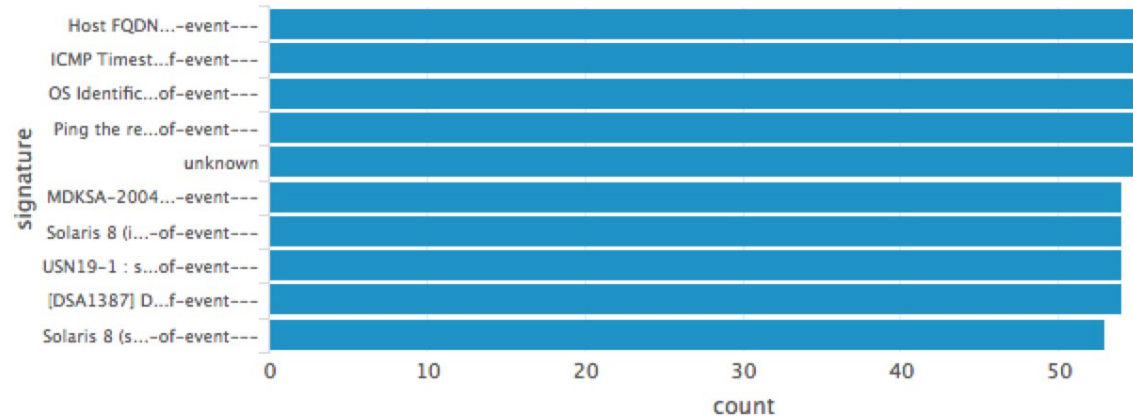
DEFAULT ACCOUNTS
Distinct Accounts

12 ↑ +1

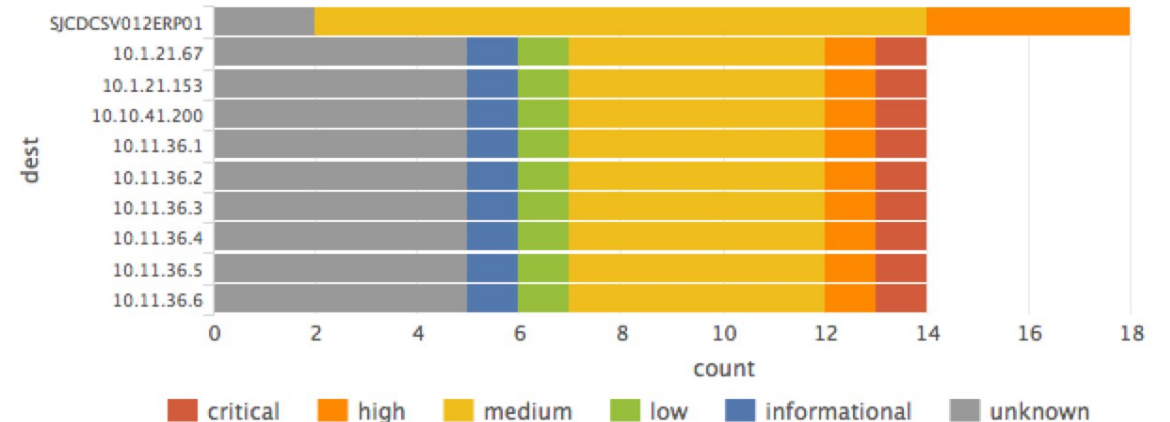
AUTH. ATTEMPTS
Total Count

medium ↓ decreasing minimally
Currently is: 4M

Top Vulnerabilities



Most Vulnerable Hosts



Data Normalization

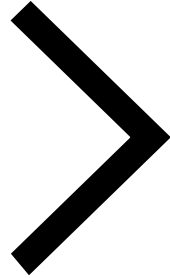
Making the Most of Schema on the Fly

The Value of Normalization

Makes things easier for a search user
Simple apps can play nicely together
Complex apps become far more useful

Normalization: Not just a dirty word

(tag=malware tag=attack
action=blocked)



(sourcetype=SYMC “Delete failed”) OR
(product=“VirusScan Enterprise” action=would*) OR
(SourceName=“Trend Micro OfficeScan Server” “Action: * cannot *”)

- ▶ Normalizing at index time is pretty lame
- ▶ Normalizing the data before it's stored is VERY lame
- ▶ Normalizing with tags and fields at search time is very AWESOME

The Splunk Common Information Model

Unlocking the Power of Data Normalization

[Edit](#)**HIGH SEV. ATTACKS**

Count

16k ↗ +187**ATTACK CATEGORIES**

Unique Count

73 0**ATTACK SIGNATURES**

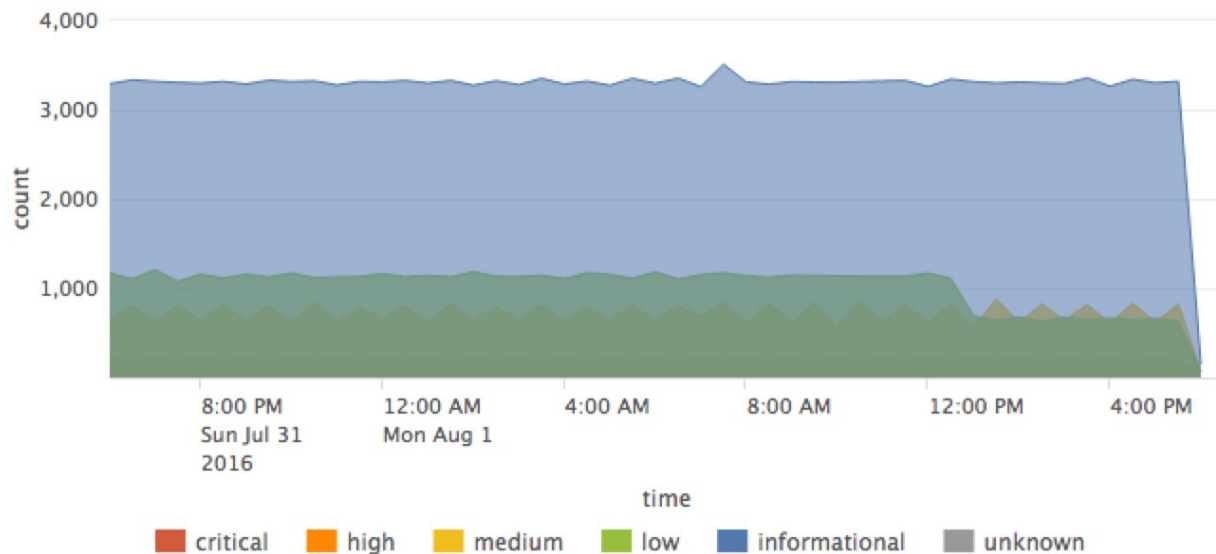
Unique Count

337 0**ATTACK SOURCES**

Unique Count

110k ↘ -374**ATTACK DESTINATIONS**

Unique Count

91k ↗ +4k**Attacks Over Time By Severity****Top Attacks**

signature ↕	src_count ↕	dest_count ↕	count ↕
URL Filtering log(9999)	61845	61977	135931
File type detection(52020)	13605	13553	29256
Common Standard Protection:Prevent termination of McAfee processes	18355	15	18355
Data filtering detection(60000)	46	15	12830
WEB-CGI calendar access	693	1	8361
WEB-MISC SSLv3 invalid data version attempt	693	1	6079
DoS: Oracle.9i.TNS.OneByte.DoS	66	66	5763
File type detection(52060)	5186	5186	5186
ATTACK-RESPONSES 403 Forbidden	8	66	2930
FTP:AUDIT:REP-INVALID-REPLY	60	54	2656

Different products, same view!

Scanning Activity (Many Attacks)**New Attacks - Last 30 Days**

firstTime ↕	ids_type ↕	signature ↕	vendor_product ↕
04/14/2016 22:32:39	host	Web Attack: Mass Iframe Injection Website 17	Symantec Endpoint Protection
04/14/2016 21:29:01	host	Web Attack : Malvertisement Website Redirect	Symantec Endpoint Protection
04/14/2016 21:25:35	host	Web Attack: Facebook Manual Share 44	Symantec Endpoint Protection
04/14/2016 21:16:35	host	Web Attack: Angler Exploit Kit Website 6	Symantec Endpoint Protection
04/14/2016 20:51:53	unknown	Spyware phone home detection(12620)	Palo Alto Networks Firewall
04/14/2016 20:49:41	unknown	Spyware phone home detection(13024)	Palo Alto Networks Firewall

Splunk did *not* invent the CIM...

The screenshot shows the top portion of the DMTF website. At the top right, there are buttons for 'Workspace' and 'Members Area'. Below these are buttons for 'DMTF中国' and 'DMTF 日本'. The main navigation bar includes a home icon and links for 'About DMTF', 'Standards & Technology', 'News & Events', 'Learning Center', 'Conformance', and 'Join Us'. Below the navigation bar, a breadcrumb trail reads 'Home > DMTF Releases Version Three of Common Information Model (CIM) Standard'.

DMTF Releases Version Three of Common Information Model (CIM) Standard

DMTF recently released Version 3 of the [Common Information Model \(CIM\)](#) standard and the Managed Object Format (MOF) schema description language.

CIM was initially developed in 1997 as a modeling language and as a schema that describes a set of conceptual models to define the components of managed computing and networking environments. The CIM schema has since expanded to include models for new markets (including cloud infrastructure management, virtualization management, peripherals, network components and applications) and collectively has evolved to become one of the most widely implemented system and network management information models to-date.

The CIM standard enables a common definition of information for any management domain, including systems, networks, applications and services. It also allows for vendor extensions. CIM's common definitions enable vendors to exchange semantically rich management information between systems throughout the network.

As part of the CIM release, a number of enhancements and additions are introduced through new versions of the Schema including ongoing improvements to support products and alliance partners, and to support new DMTF Profiles and Management Initiatives. The new CIM Version 3.0 standard provides the following schema description enhancements:

- Enumerations (both global and local)
- Structures (both global and local)
- Improved support for the specification of Methods
 - Addition of parameters
 - Default value of parameters
 - Method Return Values can be arrays or void
- Support for the use of complex types, including by reference and by value

To download the latest version of CIM or to learn more, visit <http://dmtf.org/standards/cim>.

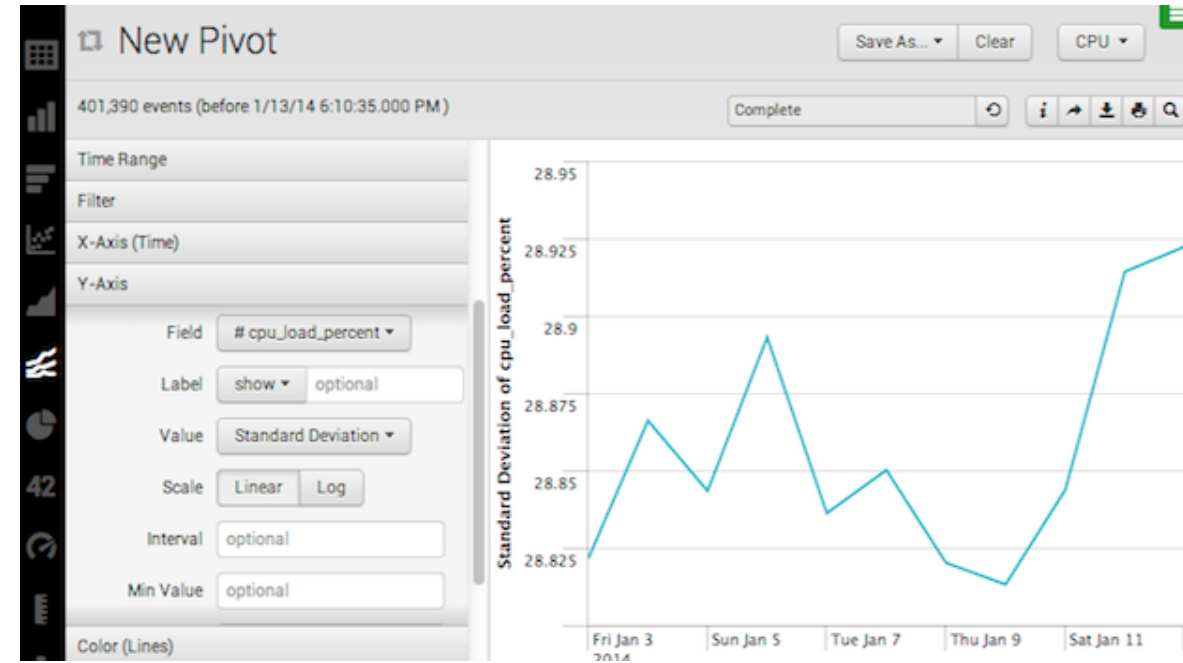
How to Get Started

The path to CIM compliance

The CIM Data Models

► Splunk_SA_CIM

- Packaged in premium apps
- Domain specific datamodel definitions
 - Search>Pivot
 - Dataset structure
 - Acceleration
- Provides CIM “dictionary”



► <https://splunkbase.splunk.com/app/1621/> (Splunk App)

► <http://docs.splunk.com/Documentation/CIM/latest/User/Overview> (Documentation)

CIM Data Model Details

- ▶ Alerts
- ▶ Application State
- ▶ Authentication
- ▶ Certificates
- ▶ Change Analysis
- ▶ Databases
- ▶ Data Loss Prevention New!
- ▶ Email
- ▶ Interprocess Messaging
- ▶ Intrusion Detection
- ▶ Inventory
- ▶ Java Virtual Machines
- ▶ Malware
- ▶ Network Resolution
- ▶ Network Sessions
- ▶ Network Traffic
- ▶ Performance
- ▶ Splunk Audit Logs
- ▶ Ticket Management
- ▶ Updates
- ▶ Vulnerabilities
- ▶ Web

ITSI Data models (Trial for CIM)

- ▶ Operating System
- ▶ Database
- ▶ Virtualization New!
- ▶ Load Balancer
- ▶ Application Server
- ▶ Web Server New!

Why CIM?

- ▶ CIM != Data Models
- ▶ Universal way to refer to an object
- ▶ Consider
 - destination_ip
 - d-ip
 - dstip
 - dest_ip
 - dst_ip
 - bob
- ▶ CIM solves this, **dest_ip**



Splunk_SA_CIM

Check your compliance to CIM Data Models

- ▶ **CIM Validation (S.o.S) datamodel**
 - Basic tools to spot untagged or partially parsed data

The screenshot displays the Splunk SA-CIM interface. At the top, there is a search bar with the query `datamodel Splunk_CIM_Validation Missing_Extractions_Network_Traffic search`. Below the search bar, the page title is "CIM Validation (S.o.S.)" with the sub-header "Splunk_CIM_Validation". A navigation bar includes buttons for "Edit", "Download", "Pivot", and "Documentation". On the left, a sidebar shows "Objects" with an "Add Object" button and a list of items: "SEARCHES", "Alerts", and "Application State". The main content area shows the details for the "Missing Extractions - Network Traffic" datamodel, including its name "Missing_Extractions_Network_Traffic" and a table of constraints.

New Search Save As Close

| datamodel Splunk_CIM_Validation Missing_Extractions_Network_Traffic search Last 24 hours Q

CIM Validation (S.o.S.)

Splunk_CIM_Validation Edit Download Pivot Documentation

[Back to Data Models](#)

Objects

Add Object

SEARCHES

Alerts

Application State

Missing Extractions - Network Traffic

Missing_Extractions_Network_Traffic Rename Delete

CONSTRAINTS

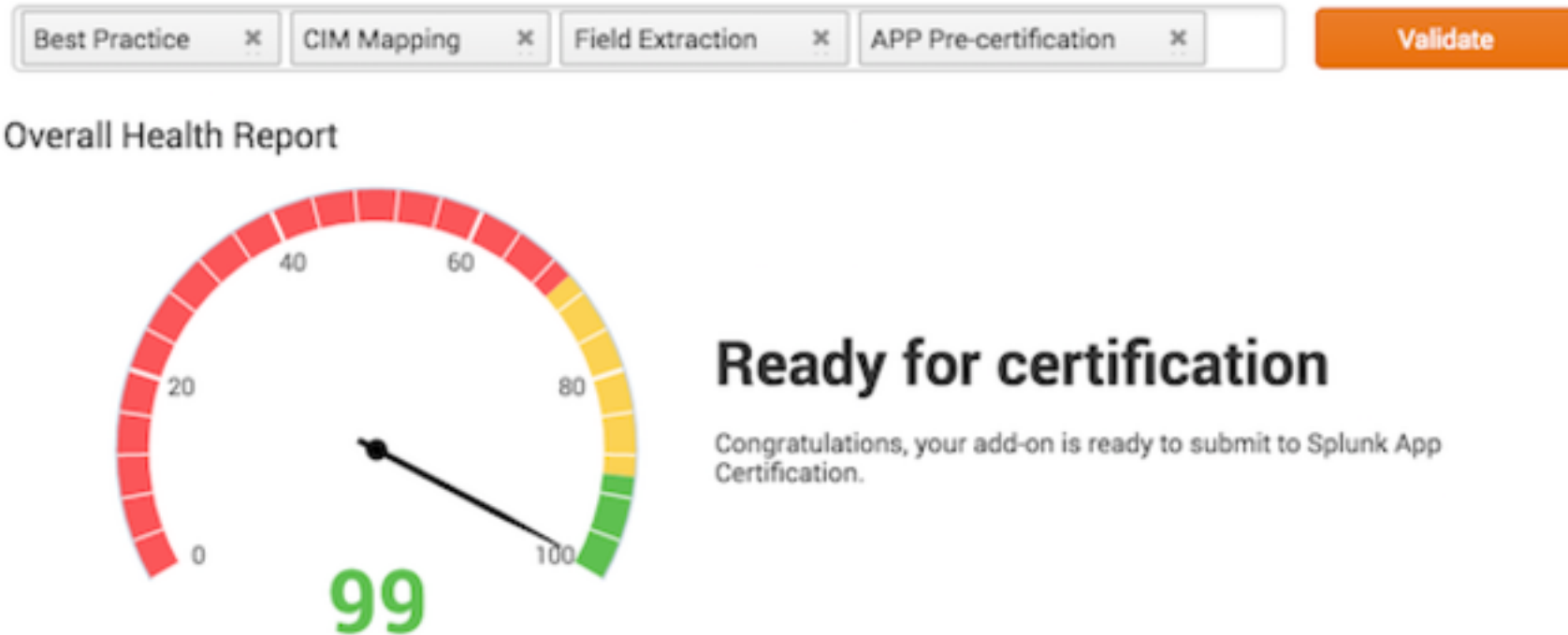
(action="unknown" OR dvc="unknown" OR rule="unknown" OR transport="unknown" OR src="unknown" OR src_port=0 OR dest="unknown")	Constraint	Edit
---	------------	------

Add-on Builder

Build CIM-compliant Add-ons and publish to Splunkbase!

► Helpful when developing new content

- Point and click
 - TA creation
 - Field extractor
 - CIM mapper
 - Branding
- Best practice validator



SA-cim_vladiator

Data Model Network_Traffic (and sub models) uses these fields:

Check for unexpected values

	field	total_events	distinct_value_count	percent_coverage	field_values	is_cim_valid
1	action	2268	5	100.00	48.15% allowed 33.33% NONE 11.11% DROP 3.7% IGNORE 3.7% TRAFFIC_IPACTION_NOTIFY	⚠ found 4 unexpected values (NONE, DROP, IGNORE, TRAFFIC_IPACTION_NOTIFY)
2	app	2268	5	70.37	29.63% NONE 18.52% NULL 11.11% SSL 7.41% HTTP 3.7% DNS	⚠ event coverage less then 90%
3	bytes	0				❌ no extracted values found
4	bytes_in	2268				⚠ event coverage less then 90%
5	bytes_out	2268	1	18.52	0.18% 0	⚠ event coverage less then 90%
6	channel	0	0	0		❌ no extracted values found
7	dest	2268	75	100.00	0.79% 10.11.36.43 0.57% 10.11.36.49 0.57% 10.11.36.11 0.57% 10.11.36.9 0.53% 10.11.36.10 0.53% 10.11.36.27 0.53% 10.11.36.40 0.49% 10.11.36.24 0.49% 10.11.36.32 0.49% 10.11.36.13 0.49% 10.11.36.47 0.44% 10.11.36.15 0.44% 10.11.36.12 0.44% 10.11.36.30 0.44% 10.11.36.42	✅ looking good!
8	dest_interface	0	0	0		❌ no extracted values found
9	dest_ip	2268	75	100.00	0.79% 10.11.36.43 0.57% 10.11.36.49 0.57% 10.11.36.11	✅ looking good!

Check for missing extractions

Check for extraction coverage

Eat a cookie!

SA-cim_vladiator: What else?

- ▶ Ability to assign score to dataset
- ▶ Ability to monitor score over time
- ▶ Ability to detect data format changes
 - Oh, remember that code upgrade last month? Log format changed...
 - Alert me!

Total fields	Issue fields	% CIM Compliance
48	40	17%

```

[07/Jan 18:10:57:153] "GET /category.screen?category_id=GIFTS&JSESSIONID=5D15L9FF10ADFF10 HTTP 1.1" 404 720 "http://buttercup-shopping.com/cart.do?action=view&itemId=EST-6&product_id=FI-SW-01"
[07/Jan 18:10:57:123] "GET /product.screen?product_id=FL-DSH-01&JSESSIONID=5D35L7FF6ADFF0 HTTP 1.1" 404 3322 "http://buttercup-shopping.com/cart.do?action=purchase&itemId=EST-26&product_id=KQ-CW-01"
[07/Jan 18:10:56:156] "GET /oldlink?item_id=EST-26&JSESSIONID=5D55L9FF1ADFF3 HTTP 1.1" 200 1318 "http://buttercup-shopping.com/cart.do?action=changequantity&itemId=EST-18&product_id=AV-CB-01&JSESSIONID=5D55L9FF1ADFF3"
[07/Jan 18:10:55:187] "GET /category.screen?category_id=FLOWERS&JSESSIONID=5D55L9FF1ADFF3 HTTP 1.1" 200 3865 "http://buttercup-shopping.com/cart.do?action=remove&itemId=EST-1"
[07/Jan 18:10:55:189] "GET /category.screen?category_id=FLOWERS&JSESSIONID=5D55L9FF1ADFF3 HTTP 1.1" 200 3865 "http://buttercup-shopping.com/cart.do?action=remove&itemId=EST-1"

```

Key Takeaways

1. CIM sets you up for success as your Splunk environment grows in size and sophistication
2. CIM != Data Models
3. Use Add-ons and the available tools to make your life easier

UX “Getting Data In” Survey

<http://tinyurl.com/GDI-conf2017>

Don't forget to **rate this session** in the
.conf2017 mobile app

splunk> **.conf2017**