


Get More Value From Splunk With the Common Information Model

A how to guide to normalize data for SOC, security and data analysts



splunk>



IT and security analysts need to find incidents and cyberthreats easily and quickly. However, inconsistencies in data from different vendors make it difficult. Field and event names don't match. Data and source types aren't all the same. Getting the right insights from each one means writing queries that are sprawling, tedious, and prone to simple errors, increasing the likelihood of chasing threats in the wrong places because of inaccurate search results.

Data isn't useful unless you can understand it. For example, you cannot detect anomalies or security threats (with high efficacy and accuracy) based on data and events without understanding each and every event.

You cannot address and respond to security threats without understanding the events that go with them. The ability to

understand comes from data normalization. So if the data is normalized it's easy to understand, and its meaning is shared with the Splunk user — whether they are a customer, content writer, SOC analyst, threat researcher, forensic investigator or something else.

Splunk delivers a **Common Information Model** (CIM) which is a semantic model focused on extracting values from data. CIM is a common lexicon that operates from an easily downloadable add-on from Splunk for a simpler, faster, more complete query experience.

In this book, you will learn:

- **What** is the Splunk Common Information Model?
- **Who** should use the Splunk Common Information Model?
- **Why** should you use the Splunk Common Information Model?
- **How** to get started with the Splunk Common Information Model.

What is Splunk CIM?

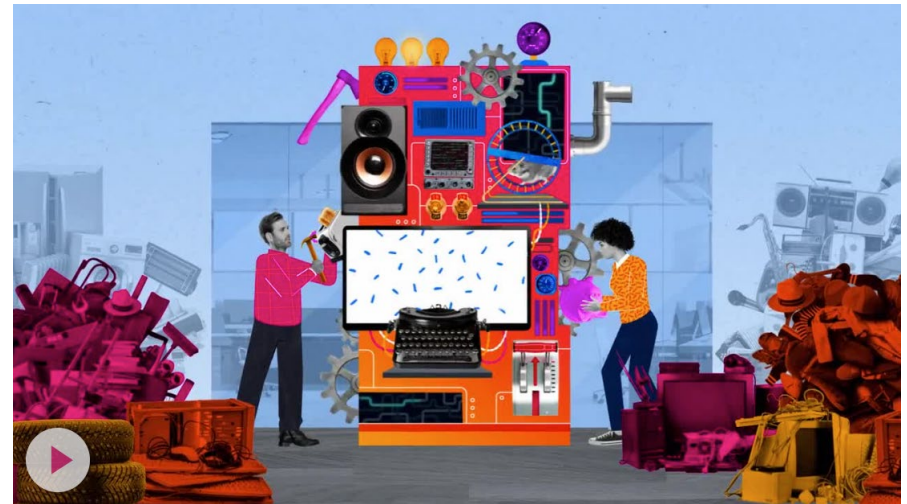
The Splunk Common Information Model is a taxonomy schema that maps from vendor fields to CIM fields, and those CIM fields and their meanings are the same for each data source and each vendor product. CIM acts as a translator from the vendor languages — their vendor-specific field names and data types — to the Splunk language and that is why CIM is also referred to as a schema.

The CIM schema has a collection of predefined data models which are applied or mapped to your data. Each CIM data model has a set of field names and tags that correspond to a domain of interest. For example, authentication, or endpoint, or domain name system (DNS). The data models further consist of datasets to define a more specific type of data to make your searches faster. An example of a data set is a privilege escalation authentication, which is a subset of all authentication events.

CIM can be used as a synonym for data normalization when events are tagged with CIM tags and fields are extracted, and mapped to CIM fields.

Thanks to data model acceleration (DMA), searches are faster with CIM than without. DMA is a method to speed up data searches by creating a summary index for that data. CIM-based searches and security detections are significantly faster than non-CIM searches.

CIM controls what data models and fields and data sets actually exist and how they are used in Splunk.



Security and IT analysts need to be able to find threats and issues without having to write complex search queries. The Splunk Common Information Model (CIM) delivers a common lexicon of field names and event types across different vendor data sources making them consistent so that analysts can write clearer queries and get better results with more true positives and fewer false positives. [Watch this video to learn more.](#)

Who should use CIM?

Splunk administrators are the primary target users for Splunk CIM, since CIM is delivered as an add-on which needs to be installed and maintained.

Developers of Splunk add-ons and apps who write their own product-specific add-ons will also use CIM so that the data they deliver is compliant with the CIM model and thus consistent with all other data using the same model.

Security content writers, along with threat researchers, use CIM to write CIM-based detection content. And finally the actual investigators — the SOC analysts who are analyzing the results of the CIM-based detections.



Why should you use CIM?

There are several advantages to using CIM.

First, searches with **CIM data is faster** in several ways than searches without. Data model acceleration is a complex topic but at its simplest, it's a way of creating a summary index of normalized data that can be queried very quickly compared to searching across raw unstructured data.

Because CIM uses standardized field names, **you can query across many source types** for a specific value without needing to know what you're going to need to search across. You don't need to know that your environment has three different firewall vendors, or who they are, or that your environment has two different EDR products — one of which you may not even know exists in your environment, or that your organization has three different varieties of web servers. All that you need to know is that you're looking for a SourceIP of n.n.n.n.

Using CIM is easy. Not only is it faster to use a CIM field but it's also easier. You don't need to memorize what every vendor calls a given IP field – is that client IP, actor IP, source IP, the first IP listed in the event, or something else entirely. You just need to know that the CIM field name is SourceIP.

CIM is more accurate because you don't need to know what 24 different vendors call any given field. You can write more accurate searches and have confidence that you're still catching what you want to catch — that is to say true positives, and not catching what you don't want to catch, those being false positives that waste your analyst time.

Even more important than all of those reasons is that **CIM makes less work for you** when you use Splunk. Because you only need to build searches, dashboards, detections, a single time rather than continuously, revising them as new feeds are onboarded.

Blocked Malware

Search without CIM

```
- (sourcetype=symantec:ep:* "Virus found" AND  
"Actual action: Cleaned by deletion")  
(sourcetype="mcafee:epo:syslog" AND  
"<ThreatActionTaken>blocked")  
(sourcetype="crowdstrike:events:sensor"  
"event_simpleName":"ProcessBlocked")
```

Search with CIM

```
- (tag=malware tag=attack action=blocked)
```

Windows Process Started

Search without CIM

```
- (sourcetype=crowdstrike:events:sensor AND  
"event_simpleName":"ProcessRollup2")  
(sourcetype=Perfmon:Process OR  
sourcetype=WMI:LocalProcesses AND  
"A new process has been created.")  
(sourcetype=XmlWinEventLog  
AND "<EventID>1")
```

Search with CIM

```
- (tag=processes tag=report action=allowed)
```

Take a look at the examples of two famous searches placed to the left. The top search in each example uses non-CIM vendor specific fields. The bottom searches are using the CIM fields to achieve the same results. You can see these CIM-based searches are simpler and you can see the big difference in length for the searches as well.

The searches and detections are also easier because you only need to type one short line instead of multiple complex statements with several hundred characters.

The CIM-based searches are also less error prone. It's easier to make a typo when the search is so long. And these CIM searches are faster since they are DMA accelerated. On top of that, these searches and security detections are more complete since they guarantee to cover all data sources collected by Splunk owned add-ons.

How to get started with CIM

First and foremost, start by using technical add-ons (TAs) that already support CIM mapping. If you're using Splunk supported TAs, the work has already been done for you and your events should automatically get mapping applied to them.

Check [the documentation](#) and be sure to install the [CIM add-on from Splunkbase](#). Pay special attention to the [data model section](#) and study the field names. Knowing these field names will let you build your searches without having to understand the underlying logs.

You can search for a field name even if you don't remember what data model it belongs to. If you're using [Splunk Enterprise Security](#), much of the work has already been done for you. If you're using custom content or you're building your own searches, you need to consider refactoring your searches and other content to use CIM fields instead of a raw event or a source type search and you'll be able to get those all the advantages we've been describing.

JVM

JVM Edit Download Pivot Documentation

< All Data Models

Datasets Add Dataset

EVENTS

JVM

Threading Dataset hierarchy

Runtime

OS

Compilation

Classloading

Memory Inherited fields

Extracted fields

Calculated fields

Runtime

Runtime Tags and other constraints Rename Delete

CONSTRAINTS

(cim_JVM_indexes) tag=jvm Inherited

tag=runtime Constraint Edit

Bulk Edit Data types Add Field

INHERITED

<input type="checkbox"/>	_time	Time	
<input type="checkbox"/>	host	String	Override
<input type="checkbox"/>	jvm_description	String	Override
<input type="checkbox"/>	source	String	Override
<input type="checkbox"/>	sourcetype	String	Override
<input type="checkbox"/>	tag	String	Override

EXTRACTED

<input type="checkbox"/>	process_name	String	Edit
<input type="checkbox"/>	start_time	Time	
<input type="checkbox"/>	uptime	Number	Edit
<input type="checkbox"/>	version	String	Edit

CALCULATED

<input type="checkbox"/>	vendor_product	String	Eval Expression Edit
--------------------------	----------------	--------	-----------------------------------

Calculated fields are processed in the order above, so ensure any dependent fields are defined first. Drag to rearrange.

A screenshot from the CIM app for the JVM data model showing available field names and data types.

Get Started with CIM today and go faster!

[Download](#) Splunk Common Information Model from Splunkbase. Or [read about the Common Information Model](#) in Splunk docs. And [don't forget to watch](#) the .conf22 session DEV1603 "Finding Threats Better With Splunk® Common Information Model (CIM) in Your Searches and Custom Add-ons."

