Splunk® Data Onboarding Cheat Sheet

V2.1
https://www.aplura.com/ocs

6 props.conf Settings You Should Have

For greater efficiency at getting data into Splunk, use these six props.conf settings when you define a source type.

```
[mysourcetype]
TIME_PREFIX = regex of the text that leads up to the timestamp
MAX_TIMESTAMP_LOOKAHEAD = how many characters for the timestamp
TIME_FORMAT = strptime format of the timestamp
SHOULD_LINEMERGE = false (always false)
LINE_BREAKER = regular expression for event breaks
TRUNCATE = 999999 (always a high number)
```

Useful strptime Directives

Year (four digit/two digit)	%Y/%y
Month (number/name/abbr)	%m/%B/%b
Day of month (leading zero/no zero)	%d/%e
Hour (24 hour/12 hour)	%H/%I
Minute	%M
Second/Millisecond	%S/%3N
Epoch time	%s
Time zone (UTC offset/name)	%z/%Z
AM/PM	%p

Useful Regular Expressions

IP Address	`\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}`
Syslog-ng header	`[\r\n]+|^\w{3}\s+\d+\s+[\d:]{8}\s+\S+\s+`
Match to the first pipe (negated character class)	`[^|]+`

Metadata Rewrites (to use, add TRANSFORMS-<classname> to a sourcetype stanza in props.conf, then add rewrite to transforms.conf)

Host

```
[rewrite_host]
REGEX = ^Message\s+from\s+(\S+)
DEST_KEY = MetaData:Host
FORMAT = host::$1
```

Sourcetype

```
[rewrite_sourcetype]
REGEX = this\s+is\s+another\s+sourcetype
DEST_KEY = MetaData:Sourcetype
FORMAT = sourcetype::other_sourcetype
```

Index

```
[rewrite_index]
REGEX = this\s+should\s+go\s+elsewhere
DEST_KEY = _MetaData:Index
FORMAT = other_index
```

Field Extractions

Using EXTRACT

In props.conf:

```
[mysourcetype]
EXTRACT-user_src = \s(?<user>\S+)\s+logged\s+in IN source_field
```

Using REPORT

In props.conf:

```
[mysourcetype]
REPORT-user_src = mysourcetype_user_source
```

In transforms.conf:

```
[mysourcetype_user_source]
REGEX = \s(\S+)\s+logged\s+in\s+from\s+(\S+)
FORMAT = src::$1 user::$2
```

Lookups

props.conf

```
[mysourcetype]
LOOKUP-mysourcetype-actions = my_lookup event_field OUTPUT lookup_field
```

transforms.conf

```
[my_lookup]
filename = mysourcetype_actions.csv
case_sensitive_match = false
max_matches = 1
```

Field Aliases, SED Commands, Calculated Fields (add to sourcetype stanzas in props.conf)

Field alias

```
FIELDALIAS-myalias = my_field AS new_field my_field AS new_field2
```

SED command

```
SEDCMD-abc_to_xyz = s/abc/xyz/g
```

Calculated field

```
EVAL-total_bytes = bytes_in + bytes_out
```

Search-Time Operation Order

EXTRACT → REPORT → KV_MODE → FIELDALIAS → EVAL → LOOKUP

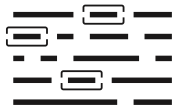
= search-time
= index-time
gray items are optional

Getting Data Into Accelerated Data Models

Review The Data

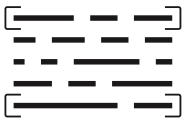
After you have correctly onboarded your data (correct meta data, line breaking, and time stamping), review the events to determine which data models the events match. A single sourcetype can contain events that are appropriate for different data models. For example, a proxy feed can have authentication events for users logging in, web proxy events showing traffic, and configuration changes as administrators adjust settings.

Extract Fields



Configure field extractions to populate as many of the data model objects (fields) as you can. See the Splunk Common Information Model Add-on Manual to learn what the field contents and names should be.

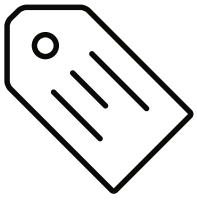
Configure Event Types



Configure event types for the data. Event types should use searches that capture all of the events you expect to fill in a particular data model. For example, to capture all login events (both successes and failures), you might use a search like:

```
sourcetype=my_sourcetype “Login for user” (“failed” OR “succeeded”)
```

Tag The Event Types

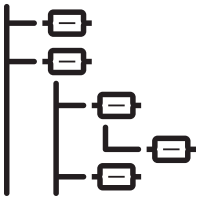


Tag the event types you just created. The CIM Add-on Manual tells you the tags which should be used for the data model you are aiming for. While tagging can be done in other ways, the current best practice is to attach the tags to event types.

Review Index Constraints

Newer versions of the CIM Add-on use index constraints to improve performance and let you control what data to accelerate. Use the CIM Add-on Setup page to confirm that the constraints include the indexes that contain the data you are working with.

Preview The Data Model



While the data model acceleration might take a while to process, you can preview the data with the datamodel command. A template for this search looks like:

```
| datamodel <data model name> <data model child object> search | search sourcetype=<new sourcetype> | table <data model name>.*
```


v2.1.5