

This guide describes key concepts and features, as well as commonly used commands and functions for Splunk Cloud and Splunk Enterprise.

## **Concepts**

#### **Events**

An event is a set of values associated with a timestamp. It is a single entry of data and can have one or multiple lines. An event can be a text document, a configuration file, an entire stack trace, and so on. This is an example of an event in a web activity log:

173.26.34.223 - - [01/Mar/2021:12:05:27 -0700]
"GET /trade/app?action=logout HTTP/1.1" 200 2953

You can also define transactions to search for and group together events that are conceptually related but span a duration of time. Transactions can represent a multistep business-related activity, such as all events related to a single customer session on a retail website.

#### **Metrics**

A metric data point consists of a timestamp and one or more measurements. It can also contain dimensions. A measurement is a metric name and corresponding numeric value. Dimensions provide additional information about the measurements. Sample metric data point:

Timestamp: 08-05-2020 16:26:42.025 -0700 Measurement: metric\_name:os.cpu.user=42.12,

metric\_name:max.size.kb=345

Dimensions: hq=us-west-1, group=queue, name=azd

Metric data points and events can be searched and correlated together, but are stored in separate types of indexes.

#### Host, Source, and Source Type

A host is the name of the physical or virtual device where an event originates. It can be used to find all data originating from a specific device. A source is the name of the file, directory, data stream, or other input from which a particular event originates. Sources are classified into source types, which can be either well known formats or formats defined by the user. Some

common source types are HTTP web server logs and Windows event logs.

Events with the same source types can come from different sources. For example, events from the file source=/var/log/messages and from a syslog input port source=UDP:514 often share the source type, sourcetype=linux\_syslog.

#### **Fields**

Fields are searchable name and value pairings that distinguish one event from another. Not all events have the same fields and field values. Using fields, you can write tailored searches to retrieve the specific events that you want. When Splunk software processes events at index-time and search-time, the software extracts fields based on configuration file definitions and user-defined patterns.

Use the Field Extractor tool to automatically generate and validate field extractions at search-time using regular expressions or delimiters such as spaces, commas, or other characters.

#### **Tags**

A tag is a knowledge object that enables you to search for events that contain particular field values. You can assign one or more tags to any field/value combination, including event types, hosts, sources, and source types. Use tags to group related field values together, or to track abstract field values such as IP addresses or ID numbers by giving them more descriptive names.

#### **Index-Time and Search-Time**

During index-time processing, data is read from a source on a host and is classified into a source type. Timestamps are extracted, and the data is parsed into individual events. Line-breaking rules are applied to segment the events to display in the search results. Each event is written to an index on disk, where the event is later retrieved with a search request.

When a search starts, referred to as search-time, indexed events are retrieved from disk. Fields are extracted from the raw text for the event.

#### **Indexes**

When data is added, Splunk software parses the data into individual events, extracts the timestamp, applies line-breaking rules, and stores the events in an index. You can create new indexes for different inputs. By default, data is stored in the "main" index. Events are retrieved from one or more indexes during a search.

#### **Core Features**

#### Search

Search is the primary way users navigate data in Splunk software. You can write a search to retrieve events from an index, use statistical commands to calculate metrics and generate reports, search for specific conditions within a rolling time window, identify patterns in your data, predict future trends, and so on. You transform the events using the Splunk Search Process Language (SPLTM). Searches can be saved as reports and used to power dashboards.

#### Reports

Reports are saved searches. You can run reports on an ad hoc basis, schedule reports to run on a regular interval, or set a scheduled report to generate alerts when the results meet particular conditions. Reports can be added to dashboards as dashboard panels.

#### **Dashboards**

Dashboards are made up of panels that contain modules such as search boxes, fields, and data visualizations.

Dashboard panels are usually connected to saved searches. They can display the results of completed searches, as well as data from real-time searches.

#### **Alerts**

Alerts are triggered when search results meet specific conditions. You can use alerts on historical and real-time searches. Alerts can be configured to trigger actions such as sending alert information to designated email addresses or posting alert information to a web resource.

## **Additional Features**

#### **Datasets**

Splunk allows you to create and manage different kinds of datasets, including lookups, data models, and table datasets. Table datasets are focused, curated collections of event data that you design for a specific business purpose. You can define and maintain powerful table datasets with Table Views, a tool that translates sophisticated search commands into simple UI editor interactions. It's easy to use, even if you have minimal knowledge of Splunk SPL.

#### **Data Model**

A data model is a hierarchically-organized collection of datasets. You can reference entire data models or specific datasets within data models in searches. In addition, you can apply data model acceleration to data models. Accelerated data models offer dramatic gains in search performance, which is why they are often used to power dashboard panels and essential on-demand reports.

#### **Apps**

Apps are a collection of configurations, knowledge objects, and customer designed views and dashboards. Apps extend the Splunk environment to fit the specific needs of organizational teams such as Unix or Windows system administrators, network security specialists, website managers, business analysts, and so on. A single Splunk Enterprise or Splunk Cloud installation can run multiple apps simultaneously.

#### **Distributed Search**

A distributed search provides a way to scale your deployment by separating the search management and presentation layer from the indexing and search retrieval layer. You use distribute search to facilitate horizontal scaling for enhanced performance, to control access to indexed data, and to manage geographically dispersed data.

## **System Components**

#### **Forwarders**

A Splunk instance that forwards data to another Splunk instance is referred to as a forwarder.

#### Indexer

An indexer is the Splunk instance that indexes data. The indexer transforms the raw data into events and stores the events into an index. The indexer also searches the indexed data in response to search requests. The search peers are indexers that fulfill search requests from the search head.

#### **Search Head**

In a distributed search environment, the search head is the Splunk instance that directs search requests to a set of search peers and merges the results back to the user. If the instance does only search and not indexing, it is usually referred to as a dedicated search head.

## **Search Processing Language (SPL)**

A Splunk search is a series of commands and arguments. Commands are chained together with a pipe "|" character to indicate that the output of one command feeds into the next command on the right.

```
search | command1 arguments1 |
command2 arguments2 | ...
```

At the start of the search pipeline, is an implied search command to retrieve events from the index. Search requests are written with keywords, quoted phrases, Boolean expressions, wildcards, field name/value pairs, and comparison expressions. The AND operator is implied between search terms. For example:

```
sourcetype=access_combined error | top 5 uri
```

This search retrieves indexed web activity events that contain the term "error." For those events, it returns the top 5 most common URI values.

Search commands are used to filter unwanted events, extract more information, calculate values, transform, and statistically analyze the indexed data. Think of the search results retrieved from the index as a dynamically created table. Each indexed event is a row. The field

values are columns. Each search command redefines the shape of that table. For example, search commands that filter events will remove rows, search commands that extract fields will add columns.

#### **Time Modifiers**

You can specify a time range to retrieve events inline with your search by using the latest and earliest search modifiers. The relative times are specified with a string of characters to indicate the amount of time (integer and unit) and an optional "snap to" time unit. The syntax is:

[+|-]<integer><unit>@<snap\_time\_unit>

The search "error earliest=-1d@d latest=-h@h" retrieves events containing "error" that occurred yesterday snapping to the beginning of the day (00:00:00) and through to the most recent hour of today, snapping on the hour.

The snap to time unit rounds the time down. For example, if it is 11:59:00 and you snap to hours (@h), the time used is 11:00:00 not 12:00:00. You can also snap to specific days of the week using @w0 for Sunday, @w1 for Monday, and so on.

#### **Subsearches**

A subsearch runs its own search and returns the results to the parent command as the argument value. The subsearch is run first and is contained in square brackets. For example, the following search uses a subsearch to find all syslog events from the user that had the last login error:

sourcetype=syslog [ search login error | return 1
user ]

#### **Optimizing Searches**

The key to fast searching is to limit the data that needs to be pulled off disk to an absolute minimum. Then filter that data as early as possible in the search so that processing is done on the minimum data necessary.

Partition data into separate indexes, if you will rarely perform searches across multiple types of data. For example, put web data in one index, and firewall data in another.

Limit the time range to only what is needed. For example -1h not -1w, or earliest=-1d.

Search as specifically as you can. For example, fatal\_error not \*error\*

Use post-processing searches in dashboards.

Use summary indexing, and report and data model acceleration features.

#### **Machine Learning Capabilities**

Splunk's Machine Learning capabilities are integrated across our portfolio and embedded in our solutions through the Splunk Machine Learning Toolkit.

#### SPL2

Several Splunk products use a new version of SPL, called SPL2, which makes the search language easier to use, removes infrequently used commands, and improves the consistency of the command syntax. See the SPL2 Search Reference.

Common Search Comm	nands
Command	Description
chart/timechart	Returns results in a tabular output for (time-series) charting.
dedup	Removes subsequent results that match a specified criterion.
eval	Calculates an expression. See COMMON EVAL FUNCTIONS.
fields	Removes fields from search results.
head/tail	Returns the first/last N results.
lookup	Adds field values from an external source.
rename	Renames a field. Use wildcards to specify multiple fields.
rex	Specifies regular expression named groups to extract fields.
search	Filters results to those that match the search expression.
sort	Sorts the search results by the specified fields.
stats	Provides statistics, grouped optionally by fields. See COMMON STATS FUNCTIONS.
mstats	Similar to stats but used on metrics instead of events.
table	Specifies fields to keep in the result set. Retains data in tabular format.
top/rare	Displays the most/least common values of a field.
transaction	Groups search results into transactions.
where	Filters search results using eval expressions. Used to compare two different fields.

Explore our full suite of products, or investigate the table below to find the specific starting point for your journey.

Or dive right in: Download the free trial and see for yourself what the Splunk platform can do for your data strategy.

## **Common Eval Functions**

The eval command calculates an expression and puts the resulting value into a field (e.g. "...| eval force = mass \* acceleration"). The following table lists some of the functions used with the eval command. You can also use basic arithmetic operators (+ - \* / %), string concatenation (e.g., "...| eval name = last . "," . first"), and Boolean operations (AND OR NOT XOR < > <= >= ! = = = LIKE).

<pre>abs(number)  case(error == 404, "Not found", error == 500,"Internal Server Error", error == 200, "OK")  ceil(1.9)  cidrmatch("123.132.32.0/25",ip)  coalesce(null(), "Returned val", null()) n=cos(0)</pre>
<pre>500,"Internal Server Error", error == 200 "0K")  ceil(1.9)  cidrmatch("123.132.32.0/25",ip)  coalesce(null(), "Returned val", null())</pre>
<pre>cidrmatch("123.132.32.0/25",ip) coalesce(null(), "Returned val", null())</pre>
<pre>coalesce(null(), "Returned val", null())</pre>
n=cos(0)
exact(3.14*num)
exp(3)
if(error==200, "OK", "Error")
if(in(status, "404","500","503"),"true", "false")
isbool(field)
isint(field)
isnull(field)
isstr(field)
len(field)
like(field, "addr%")
log(number,2)
lower(username)

Function	Description	Examples
ltrim(X,Y)	Returns X with the characters in Y trimmed from the left side. Y defaults to spaces and tabs.	ltrim(" ZZZabcZZ ", " Z")
match(X,Y)	Returns if X matches the regex pattern Y.	match(field, "^\d{1,3}\.\d\$")
max(X,)	Returns the maximum.	max(delay, mydelay)
md5(X)	Returns the MD5 hash of a string value X.	md5(field)
min(X,)	Returns the minimum.	min(delay, mydelay)
mvcount(X)	Returns the number of values of X.	mvcount(multifield)
mvfilter(X)	Filters a multi-valued field based on the Boolean expression X.	mvfilter(match(email, "net\$"))
mvindex(X,Y,Z)	Returns a subset of the multivalued field X from start position (zero-based) Y to Z (optional).	mvindex(multifield, 2)
mvjoin(X,Y)	Given a multi-valued field X and string delimiter Y, and joins the individual values of X using Y.	<pre>mvjoin(address, ";")</pre>
now()	Returns the current time, represented in Unix time.	now()
null()	This function takes no arguments and returns NULL.	null()
nullif(X,Y)	Given two arguments, fields X and Y, and returns the X if the arguments are different. Otherwise returns NULL.	nullif(fieldA, fieldB)
random()	Returns a pseudo-random number ranging from 0 to 2147483647.	random()
relative_time(X,Y)	Given epochtime time X and relative time specifier Y, returns the epochtime value of Y applied to X.	relative_time(now(),"-1d@d")
replace(X,Y,Z)	Returns a string formed by substituting string Z for every occurrence of regex string Y in string X.	Returns date with the month and day numbers switched, so if the input was $4/30/2022$ the return value would be $30/4/2022$ : replace(date, "^(\d{1,2})/(\d{1,2})/", "\2/\1/")
round(X,Y)	Returns X rounded to the amount of decimal places specified by Y. The default is to round to an integer.	round(3.5)
rtrim(X,Y)	Returns X with the characters in Y trimmed from the right side. If Y is not specified, spaces and tabs are trimmed.	rtrim(" ZZZZabcZZ ", " Z")

Function	Description	Examples
split(X,"Y")	Returns X as a multi-valued field, split by delimiter Y.	split(address, ";")
sqrt(X)	Returns the square root of X.	sqrt(9)
strftime(X,Y)	Returns epochtime value X rendered using the format specified by Y.	strftime(_time, "%H:%M")
strptime(X,Y)	Given a time represented by a string X, returns value parsed from format Y.	strptime(timeStr, "%H:%M")
substr(X,Y,Z)	Returns a substring field X from start position (1-based) Y for Z (optional) characters.	substr("string", 1, 3)
time()	Returns the wall-clock time with microsecond resolution.	time()
tonumber(X,Y)	Converts input string X to a number, where Y (optional, defaults to 10) defines the base of the number to convert to.	tonumber("0A4",16)
tostring(X,Y)	Returns a field value of X as a string. If the value of X is a number, it reformats it as a string. If X is a Boolean value, reformats to "True" or "False." If X is a number, the second argument Y is optional and can either be "hex" (convert X to hexadecimal), "commas" (formats X with commas and 2 decimal places), or "duration" (converts seconds X to readable time format HH:MM:SS).	This example returns: foo=615 and foo2=00:10:15:   eval foo=615   eval foo2 = tostring(foo "duration")
typeof(X)	Returns a string representation of the field type.	<pre>   eval n=typeof(12) + typeof("string") + typeof(1==2) + typeof(badfield)</pre>
urldecode(X)	Returns the URL X decoded.	urldecode("http%3A%2F%2Fwww.splunk. com%2Fdownload%3Fr%3Dheader")
validate (X,Y,)	Given pairs of arguments, Boolean expressions X and strings Y, returns the string Y corresponding to the first expression X that evaluates to False and defaults to NULL if all are True.	<pre>validate(isint(port), "ERROR: Port is not an integer", port &gt;= 1 AND port &lt;= 65535, "ERROF Port is out of range")</pre>

## **Common Stats Functions**

Common statistical functions used with the chart, stats, and timechart commands. Field names can be wildcarded, so avg(\*delay) might calculate the average of the delay and xdelay fields.

avg(X)	Returns the average of the values of field X.
count(X)	Returns the number of occurrences of the field X. To indicate a specific field value to match, format X as eval(field="value").
dc(X)	Returns the count of distinct values of the field X.
earliest(X)	Returns the chronologically earliest seen value of X.
latest(X)	Returns the chronologically latest seen value of X.
max(X)	Returns the maximum value of the field X. If the values of X are non-numeric, the max is found from alphabetical ordering.
median(X)	Returns the middle-most value of the field X.
min(X)	Returns the minimum value of the field X. If the values of X are non-numeric, the min is found from alphabetical ordering.
mode(X)	Returns the most frequent value of the field X.
perc <x>(Y)</x>	Returns the X-th percentile value of the field Y. For example, perc5(total) returns the 5th percentile value of a field "total".
range(X)	Returns the difference between the max and min values of the field X.
stdev(X)	Returns the sample standard deviation of the field X.
stdevp(X)	Returns the population standard deviation of the field X.
sum(X)	Returns the sum of the values of the field X.
sumsq(X)	Returns the sum of the squares of the values of the field X.
values(X)	Returns the list of all distinct values of the field X as a multi-value entry. The order of the values is alphabetical.
var(X)	Returns the sample variance of the field X.

# **Search Examples**

Filter Results	
Returns X rounded to the amount of decimal places specified by Y. The default is to round to an integer.	round(3.5)
Returns X with the characters in Y trimmed from the right side. If Y is not specified, spaces and tabs are trimmed.	rtrim(" ZZZZabcZZ ", " Z")
Returns X as a multi-valued field, split by delimiter Y.	split(address, ";")
Given pairs of arguments, Boolean expressions X and strings Y, returns the string Y corresponding to the first expression X that evaluates to False and defaults to NULL if all are True.	<pre>validate(isint(port), "ERROR: Port is not an integer", port &gt;= 1 AND port &lt;= 65535, "ERROR: Port is out of range")</pre>

Group Results	
Cluster results together, sort by their "cluster_count" values, and then return the 20 largest clusters (in data size).	cluster t=0.9 showcount=true   sort limit=20
Group results that have the same "host" and "cookie," occur within 30 seconds of each other, and do not have a pause greater than 5 seconds between each event into a transaction.	transaction host cookie maxspan=30s maxpause=5s
Group results with the same IP address (clientip) and where the first result contains "signon," and the last result contains "purchase."	<pre>   transaction clientip startswith="signon" endswith="purchase"</pre>

Order Results	
Return the first 20 results.	head 20
Reverse the order of a result set.	reverse
Sort results by "ip" value (in ascending order) and then by "url" value (in descending order).	sort ip, -url
Return the last 20 results in reverse order.	tail 20

Reporting	
Return the average and count using a 30 second span of all metrics ending in cpu.percent split by each metric name.	mstats avg(_value), count(_value) WHERE metric_ name="*.cpu.percent" by metric_name span=30s
Return max(delay) for each value of foo split by the value of bar.	chart max(delay) over foo by bar
Return max(delay) for each value of foo.	chart max(delay) over foo
Count the events by "host."	stats count by host
Create a table showing the count of events and a small line chart.	stats sparkline count by host
Create a timechart of the count of from "web" sources by "host."	timechart count by host
Calculate the average value of "CPU" each minute for each "host."	timechart span=1m avg(CPU) by host
Return the average for each hour, of any unique field that ends with the string "lay" (e.g., delay, xdelay, relay, etc).	stats avg(*lay) by date_hour
Return the 20 most common values of the "url" field.	top limit=20 url
Return the least common values of the "url" field.	rare url

Advanced Reporting	
Compute the overall average duration and add 'avgdur' as a new field to each event where the 'duration' field exists.	eventstats avg(duration) as avgdur
Find the cumulative sum of bytes.	<pre>   streamstats sum(bytes) as bytes_total   timechart max(bytes_total)</pre>
Find anomalies in the field 'Close_Price' during the last 10 years.	sourcetype=nasdaq earliest=-10y   anomalydetection Close_Price
Create a chart showing the count of events with a predicted value and range added to each event in the time-series.	timechart count   predict count
Computes a five event simple moving average for field 'count' and write to new field 'smoothed_count.'	"   timechart count   trendline sma5(count) as smoothed_count"

Metrics	
List all of the metric names in the "_metrics" metric index.	mcatalog values(metric_name) WHERE index=_metrics
See examples of the metric data points stored in the "_metrics" metric index.	mpreview index=_metrics target_per_timeseries=5
Return the average value of a metric in the "_metrics" metric index. Bucket the results into 30 second time spans.	mstats avg(aws.ec2.CPUUtilization) WHERE index=_ metrics span=30s

Add Fields	
Set velocity to distance / time.	eval velocity=distance/time
Extract "from" and "to" fields using regular expressions.  If a raw event contains "From: Susan To: David," then from=Susan and to=David.	rex field=_raw "From: (? <from>.*) To: (?<to>.*)"</to></from>
Save the running total of "count" in a field called "total_count."	accum count as total_count
For each event where 'count' exists, compute the difference between count and its previous value and store the result in 'countdiff.'	delta count as countdiff

Filter Fields	
Keep only the "host" and "ip" fields, and display them in that order.	fields + host, ip
Remove the "host" and "ip" fields from the results.	fields - host, ip

## **Lookup Tables (Splunk Enterprise only)**

For each event, use the lookup table usertogroup to locate ... | lookup usertogroup user output group the matching "user" value from the event. Output the group field value to the event.

Read in the usertogroup lookup table that is defined in the ... | inputlookup usertogroup transforms.conf file.

Write the search results to the lookup file "users.csv." ... | outputlookup users.csv

## **Modify Fields**

Rename the "\_ip" field as "IPAddress." ... | rename \_ip as IPAddress

## Regular Expressions (Regexes)

Regular Expressions are useful in multiple areas: search commands regex and rex; eval functions match() and replace(); and in field extraction.

Regex	Note	Example	Explanation
\s	white space	\d\s\d	digit space digit
\s	not white space	\d\S\d	digit non-whitespace digit
\d	digit	\d\d-\d\d-\d\d\d\d	SSN
\D	not digit	\D\D\D	three non-digits
\w	word character (letter, number, or _)	\w\w\w	three word chars
\W	not a word character	\w\w\w	three non-word chars
[]	any included character	[a-z0-9#]	any char that is a thru z, 0 thru 9, or #
[^]	no included character	[^xyz]	any char but x, y, or z
*	zero or more	\w*	zero or more words chars
+	one or more	\d+	integer
?	zero or one	\d\d\d-?\d\d-?\d\d\d\d	SSN with dashes being optional
I	or	\w \d	word or digit character
(?P <var>)</var>	named extraction	(?P <ssn>\d\d\d-\d\d- \d\d\d\d)</ssn>	pull out a SSN and assign to 'ssn' field
(?: )	logical or atomic grouping	(?:[a-zA-Z] \d)	alphabetic character OR a digit
^	start of line	^\d+	line begins with at least one digit
\$	end of line	\d+\$	line ends with at least one digit
{}	number of repetitions	\d{3,5}	between 3-5 digits
\	escape	]/	escape the [ character

B 4	4: 1		ued	
MILL	110	/ai		

Combine the multiple values of the recipients field into a single value.	nomv recipients
Separate the values of the "recipients" field into multiple field values, displaying the top recipients.	makemv delim="," recipients   top recipients
Create new results for each value of the multivalue field "recipients."	mvexpand recipients
Find the number of recipient values.	eval to_count = mvcount(recipients)
Find the first email address in the recipient field.	eval recipient_first = mvindex(recipient,0)
Find all recipient values that end in .net or .org.	<pre>   eval netorg_recipients = mvfilter match(recipient,"\.net\$") OR match(recipient,"\.org\$"))</pre>
Find the index of the first recipient value match "\.org\$"	eval orgindex = mvfind(recipient, "\.org\$")

#### **Common Date and Time Formatting**

Use these values for eval functions strftime() and strptime(), and for timestamping event data.

	%H	24 hour (leading zeros) (00 to 23)		
Time	%I	12 hour (leading zeros) (01 to 12)		
		Minute (00 to 59)		
		Second (00 to 61)		
	%N	subseconds with width (%3N = millisecs, %6N = microsecs, %9N =		
		nanosecs)		
	%p	AM or PM		
	%Z	Time zone (EST)		
	%z	Time zone offset from UTC, in hour and minute: +hhmm or -hhmm. (-0500 for EST)		
	%s	Seconds since 1/1/1970 (1308677092)		
	%d	Day of month (leading zeros) (01 to 31)		
	%j	Day of year (001 to 366)		
Days	%w	Weekday (0 to 6)		
	%a	Abbreviated weekday (Sun)		
	%A	Weekday (Sunday)		
	%b	Abbreviated month name (Jan)		
Months	%B	Month name (January)		
	%m	Month number (01 to 12)		
V	%y	Year without century (00 to 99)		
Years	%Y	Year (2022)		
Examples	%Y-%m-%d	2022-12-31		
	%y-%m-%d	22-12-31		
	%b %d, %Y	Jan 24, 2022		
	%B %d, %Y	January 24, 2022		
	q %d %b '%y = %Y-%m-%d	g 25 Feb '22 = 2022-02-25		

