**Splunk Notes**

* Splunk – It is a software designed for analyzing, monitoring and analyzing the machine generated data from various devices in an IT organization.
* It ingests and index the data from a wide source include log, event, metrics and other data.
* We can filter and visualize data using spl query.

**Splunk installation**

* To forward the data from one machine to another we need a UF and Indexer. In addition to parsing or filter we need an Heavy forwarder.
* Splunk should be installed in separate user for access management.

Splunk installation commands:

NOTE: if you use instance change root password before installation

sudo su

adduser splunker

passwd splunker

wget -O splunk-9.1.2-b6b9c8185839-Linux-x86\_64.tgz "https://download.splunk.com/products/splunk/releases/9.1.2/linux/splunk-9.1.2-b6b9c8185839-Linux-x86\_64.tgz"

chown -R splunker:splunker splunk-9.1.2-b6b9c8185839-Linux-x86\_64.tgz

su splunker

tar xvf splunk-9.1.2-b6b9c8185839-Linux-x86\_64.tgz -C /opt/

chmod -R splunker:splunker /opt/splunk/

./splunk enable boot-start systemd-managed 1 --accept-license

* After installation enable receiving port

./splunk enable listen 9997

./splunk display listen

**Similarly, installation process for Universal Forwarder in instance:**

For UF we need configure the port and adding forward server (indexer IP:port)

And we need to mention log path.

./splunk add forward-server ip:port in UF

./splunk add monitor /var/log in UF

./splunk list forward-server in UF

**Installing splunk or UF in instance:**

While installing in instance we need to set the inbound and outbound rules for appropriate port such 9997 for data forwarding port , 8000 for web port.

**Basic SPL query:**

SPL command are used to get the desired results from the dataset .

Component of SPL :

* search terms ( keyword or phase looking for. EX: “splunk” looking for event containing word splunk)
* commands – head 5 (it will show the first five events )
* functions – stats avg(number) it will give the average of number.

Job Inspector:

It shows overall stats of search.

Used to troubleshoot problems and understanding the knowledge object on processing.

Components:

Header – Execution costs – search job properties.

Visualization:

Representing the data visually to understand more easily.

Some visualize type statistics table, chart: line column, pie,, single value, maps, etc.

Sequence of related data only represented as visualize format.

Command:

Multi-series table – chart or timechart

index="\_internal" | chart count over component

TOP – most repeated top values

RARE – shows the lowest repeated value

Stats count – viewing result as single value

Eval – eval must

Index =\* | fields -id == it exclude the id field from the event

Index =\* | rename host as “Hoster”, count as “counts”

| erex , |rex = we can extract unextracted field from raw for temporary time period with provided sample value ------ erex don’t require regex , rex require regex

A screenshot of a computer

Description automatically generated

Field Alias :

We can use this when multiple event having different field name incase 50 event have username as US and 40 event has users – you can assign us=username and users=username

Assign an alternative name for an field.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

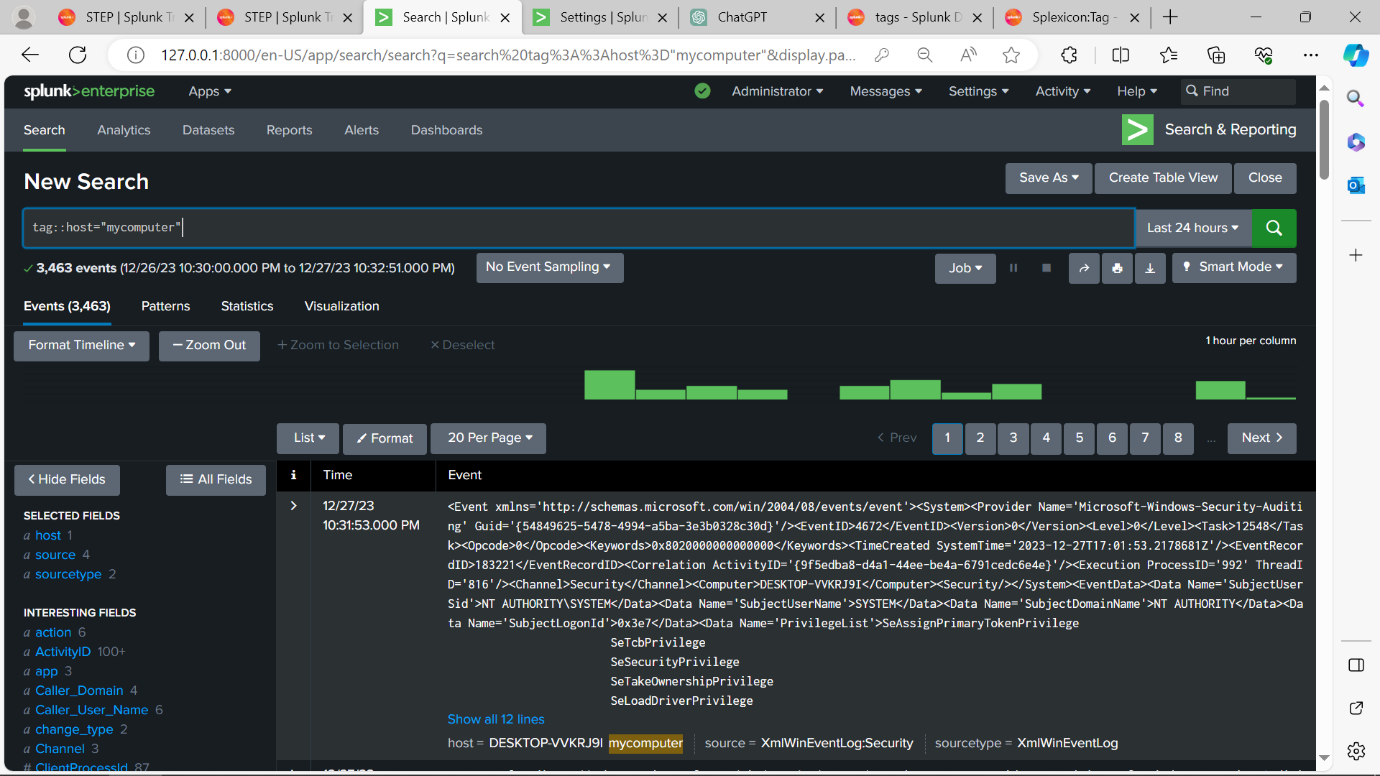
Description automatically generated

TAG:

It is an alternative name for host , source , source type, event type.

A screenshot of a computer

Description automatically generated



Event type:

Creating event types allows you to define a set of conditions that an event must meet to be included in that type.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

Only it can be used if query does not have pipeline

We can build an with event type builder.

**Macros:**

With macros we can write query and pass parameter in the query with this we can use this in many searches without writing the same query again.

**Creating workflow Actions:**

We can pass field values as parameters to websites by post or get method and it can be used for search too.

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

**Scheduling and Reporting:**

Report can be saved and run at anytime it will fetch fresh result every time it runs.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

HTTP event Collector:

We can push data without universal forwarder with HEC.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

bucket --> JOURNAl.GZ and .tsidx (time series Index Files)

JOURNAL -- \_raw\_event data it contain slices of 128kb each files

tsidx -- are index keys to our journal files it tells splunk which slice of to openup inorder find the result looking for the search

tsidx files are made - splunk take raw and chooses common unique term in data an tokenize them to lexicon in .tsidx

Each lexicon has unique posting list array it will give the location of the data from the journal array.

Bloomfilter: it will be created when the data move from hot to warm

the unqiue terms go through certain hashing functions and set the 0 and 1 in bloomfilters simillarlly search key words go through same hashing and set 0 and 1 and it will compare for the match bloomfilter in all bucket . Comparing bloomfilter is much efficient way than searching on tsidx

Transforming command: stats, timechart, chart, top, rare. it is executed in search Head

Streaming commands: centralized - transaction, streamstats and Distributable - eval, rename, feild, regex. it is executed in the Indexer.

Segmatation:major and minor breakerse

lisby expression: splunk uses lisby expression to build bloomfilters and to locate term in tsidx files.

Troubleshoot command:

- | makeresults used to generate temporary event in memory.

it is used when one or more eval commands used

must be the first command in search

if we need to add any static data at search time it is not in your index

- | fieldsummary calculates a variety of summary statistics.

Host=mycomputer | fieldsummary

majo

IF CONDITION:

Index=web | **eval IsproductNull = if (isnull(productid),”yes”,”no”)** | table IsproductNull, productID

TYPEof:

| Eval field = typeof(X)

creating feild extraction: (manual)

feilds are searchable key value pair

they may be extracted in index time or search time or may be persistant or created in temporary memeory

The feild extrator:

two methods for feild extractor - Regular exp and delimiters

regular exp - for unstructured data.

delimiter - is used when the data are seperated same character like , | \ ....

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

| chart, timechart, top, rare, stats

chart - | chart count over id by host limit=10 useother=f .....

timechart - it will count all data based on time

top -- show the highest value

rare -- written the least count

stats -- it is similar to table

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

functions- sum

| stats min(price) as cost ,

max(id) as ID, avg(price) as price , sum(price) as price by component

| stats values(user) as "user" -- it is display unique users

|stats list(user) as "user" -- it display with duplicate values

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

|eval -- it is for operations (arithmetic, concatenation, boolean, comparison)

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

Data model

3 types of dataset - events, searches, transaction

event command doesn't contain transforming com

search com contain pipe

tansaction com contain transaction com

INPUT.CONF:

It is used to specify the source file path and the and the index name

*[monitor://$SPLUNK\_HOME\var\log\splunk]*

*index = \_internal*

default file --- $SPLUNK\_HOME/etc/system/default/

custom file ----$SPLUNK\_HOME/etc/system/local/

# You must restart the Splunk platform instance to enable configuration

changes.

index --- to store event from this input