SIEM

SPLUNK

INTERVIEW

QUESTIONS

&

ANSWERS

1. Define Splunk

Splunk is a software platform that allows users to analyze machine-generated data (from hardware devices, networks, servers, IoT devices, etc.). Splunk is widely used for searching, visualizing, monitoring, and reporting enterprise data. It processes and analyzes machine data and converts it into powerful operational intelligence by offering real-time insights into the data through accurate visualizations.

Splunk is used for analyzing machine data because:

- It offers business insights – Splunk understands the patterns hidden within the data and turns it into real-time business insights that can be used to make informed business decisions.
- It provides operational visibility – Splunk leverages machine data to get end-to-end visibility into company operations and then breaks it down across the infrastructure.
- It facilitates proactive monitoring – Splunk uses machine data to monitor systems in real-time to identify system issues and vulnerabilities (external/internal breaches and attacks).

2. Name the common port numbers used by Splunk.

The common port numbers for Splunk are:

- Splunk Web Port: 8000
- Splunk Management Port: 8089
- Splunk Network port: 514
- Splunk Index Replication Port: 8080
- Splunk Indexing Port: 9997
- KV store: 8191

3. Name the components of Splunk architecture.

The Splunk architecture is made of the following components:

- Search Head – It provides GUI for searching
- Indexer – It indexes the machine data
- Forwarder – It forwards logs to the Indexer

Deployment server – It manages the Splunk components in a distributed environment and distributes configuration apps.

4. What are the different types of Splunk dashboards?

There are three different kinds of Splunk dashboards:

- Real-time dashboards
- Dynamic form-based dashboards
- Dashboards for scheduled reports

5. Name the types of search modes supported in Splunk.

Splunk supports three types of dashboards, namely:

- Fast mode
- Smart mode
- Verbose mode

6. Name the different kinds of Splunk Forwarders.

There are two types of Splunk Forwarders:

- **Universal Forwarder (UF)** – It is a lightweight Splunk agent installed on a non-Splunk system to gather data locally. UF cannot parse or index data.
- **Heavyweight Forwarder (HWF)** – It is a heavyweight Splunk agent with advanced functionalities, including parsing and indexing capabilities. It is used for filtering data.

7. What are the benefits of feeding data into a Splunk instance through Splunk Forwarders?

If you feed the data into a Splunk instance via Splunk Forwarders, you can reap three significant benefits – TCP connection, bandwidth throttling, and an encrypted SSL connection to transfer data from a Forwarder to an Indexer. Splunk's architecture is such that the data forwarded to the Indexer is load-balanced by default.

So, even if one Indexer goes down due to some reason, the data can re-route itself via another Indexer instance quickly. Furthermore, Splunk Forwarders cache the events locally before forwarding it, thereby creating a temporary backup of the data.

8. What is the "Summary Index" in Splunk?

In Splunk, the Summary Index refers to the default Splunk index that stores data resulting from scheduled searches over time. Essentially, it is the index that Splunk Enterprise uses if a user does not specify or indicate another one.

The most significant advantage of the Summary Index is that it allows you to retain the analytics and reports even after your data has aged.

9. What is the purpose of Splunk DB Connect?

Splunk DB Connect is a generic SQL database plugin designed for Splunk. It enables users to integrate database information with Splunk queries and reports seamlessly.

10. What is the function of the Splunk Indexer?

As the name suggests, the Splunk Indexer creates and manages indexes. It has two core functions – to index raw data into an index and to search and manage the indexed data.

11. Name a few important Splunk search commands.

Some of the important search commands in Splunk are:

- Abstract
- Erex
- Addtotals
- Accum
- Filldown
- Typer
- Rename
- Anomalies

12. What are some of the most important configuration files in Splunk?

The most crucial configuration files in Splunk are:

- props.conf
- indexes.conf
- inputs.conf
- transforms.conf
- server.conf

13. What is the importance of the License Master in Splunk? What happens if the License Master is unreachable?

In Splunk, the License Master ensures that the right amount of data gets indexed. Since the Splunk license is based on the data volume that reaches the platform within a 24hr-window, the License Master ensures that your Splunk environment stays within the constraints of the purchased volume.

If ever the License Master is unreachable, a user cannot search the data. However, this will not affect the data flowing into the Indexer – data will continue to flow in the Splunk deployment, and the Indexers will index the data. But the top of the Search Head will display a warning message that the user has exceeded the indexing volume. In this case, they must either reduce the amount of data flowing in or must purchase additional capacity of the Splunk license.

14. Explain 'license violation' in the Splunk perspective.

Anytime you exceed the data limit, the 'license violation' error will show on the dashboard. This warning will remain for 14 days. For a commercial Splunk license, users can have five warnings in a 30-day window before which Indexer's search results and reports will not trigger. However, for the free version, users get only three warning counts.

15. What is the general expression for extracting IP address from logs?

Although you can extract the IP address from logs in many ways, the regular experssion for it would be:

```
rex field=_raw “(?<ip_address>\d+\.\d+\.\d+\.\d+)”
```

OR

```
rex field=_raw “(?<ip_address>([0-9]{1,3}[\.]){3}[0-9]{1,3})”
```

16. How can you troubleshoot Splunk performance issues?

To troubleshoot Splunk performance issues, perform the following steps:

- Check splunkd.log to find any errors
- Check server performance issues (CPU/memory usage, disk i/o, etc.)
- Check the number of saved searches that are running at present and also their system resources consumption.
- Install the SOS (Splunk on Splunk) app and see if the dashboard displays any warning or errors.
- Install Firebug (a Firefox extension) and enable it in your system. After that, you have to log into Splunk using Firefox, open Firebug's panels, and go to the 'Net' panel to enable it). The Net panel displays the HTTP requests and responses, along with the time spent in each. This will allow you to see which requests are slowing down Splunk and affecting the overall performance.

17. What are Buckets? Explain Splunk Bucket Lifecycle.

Buckets are directories that store the indexed data in Splunk. So, it is a physical directory that chronicles the events of a specific period. A bucket undergoes several stages of transformation over time. They are:

- Hot – A hot bucket comprises of the newly indexed data, and hence, it is open for writing and new additions. An index can have one or more hot buckets.
- Warm – A warm bucket contains the data that is rolled out from a hot bucket.
- Cold – A cold bucket has data that is rolled out from a warm bucket.
- Frozen – A frozen bucket contains the data rolled out from a cold bucket. The Splunk Indexer deletes the frozen data by default. However, there's an option to archive it. An important thing to remember here is that frozen data is not searchable.

18. What purpose does the Time Zone property serve in Splunk?

In Splunk, Time Zone is crucial for searching for events from a security or fraud perspective. Splunk sets the default Time Zone for you from your browser settings. The browser further picks up the current Time Zone from the machine you are using. So, if you search for any event with the wrong Time Zone, you will not find anything relevant for that search.

The Time Zone becomes extremely important when you are searching and correlating data pouring in from different and multiple sources.

19. Define Sourcetype in Splunk.

In Splunk, Sourcetype refers to the default field that is used to identify the data structure of an incoming event. Sourcetype should be set at the forwarder level for indexer extraction to help identify different data formats. It determines how Splunk Enterprise formats the data during the indexing process. This being the case, you must ensure to assign the correct Sourcetype to your data. To make data searching even easier, you should provide accurate timestamps, and event breaks to the indexed data (the event data).

20. Explain the difference between Stats and Eventstats commands.

In Splunk, the Stats command is used to generate the summary statistics of all the existing fields in the search results and save them as values in newly created fields. Although the Eventstats command is pretty similar to the Stats command, it adds the aggregation results inline to each event (if only the aggregation is pertinent to that particular event). So, while both the commands compute the requested statistics, the Eventstats command aggregates the statistics into the original raw data.

21. Differentiate between Splunk App and Add-on.

Splunk Apps refer to the complete collection of reports, dashboards, alerts, field extractions, and lookups. However, Splunk Add-ons only contain built-in configurations – they do not have dashboards or reports.

22. What is the command to stop and start Splunk service?

The command to start Splunk service is: ./splunk start

The command to stop Splunk service is: ./splunk stop

23. How can you clear the Splunk search history?

To clear the Splunk search history, you need to delete the following file from Splunk server: $splunk_home/var/log/splunk/searches.log

24. What is Btool in Splunk?

Btool in Splunk is a command-line tool that is used for troubleshooting configuration file issues. It also helps check what values are being used by a user's Splunk Enterprise installation in the existing environment.

25. What is the need for Splunk Alert? Specify the type of options you get while setting up Splunk Alerts.

Splunk Alerts help notify users of any erroneous condition in their systems. For instance, a user can set up Alerts for email notification to be sent to the admin in case there are more than three failed login attempts within 24 hours.

The different options you get while setting up Alerts include:

- You can create a webhook. This will allow you to write to HipChat or GitHub – you can write an email to a group of machines containing your subject, priorities, and the body of your email.
- You can add results in CSV or pdf formats or in line with the body of the message to help the recipient understand the location and conditions of the alert that has been triggered and what actions have been taken for the same.
- You can create tickets and throttle alerts based on specific conditions such as the machine name or IP address. These alerts can be controlled from the alert window.

26. What is a Fishbucket and what is the Index for it?

Fishbucket is an index directory resting at the default location, that is:

```
/opt/splunk/var/lib/splunk
```

Fishbucket includes seek pointers and CRCs for the indexed files. To access the Fishbucket, you can use the GUI for searching:

```
index=_thefishbucket
```

27. How to know when Splunk has completed indexing a log file?

You can figure out whether or not **Splunk has completed indexing a log file in two ways:**

1. By monitoring the data from Splunk's metrics log in real-time:

```
 index=”_internal” source=”*metrics.log” group=”per_sourcetype_thruput”
series=”&lt;your_sourcetype_here&gt;” |
eval MB=kb/1024 | chart sum(MB)
```

2. By monitoring all the metrics split by source type:

```
index=”_internal” source=”*metrics.log” group=”per_sourcetype_thruput” | eval MB=kb/1024 |
chart sum(MB) avg(eps) over series
```

28. What is the Dispatch Directory?

The Dispatch Directory includes a directory for individual searches that are either running or have completed. The configuration for the Dispatch Directory is as follows:

```
$SPLUNK_HOME/var/run/splunk/dispatch
```

Let's assume, there is a directory named 1434308943.358. This directory will contain a CSV file of all the search results, a search.log containing the details about the search execution, and other relevant information. By using the default configuration, you can delete this directory within 10 minutes after the search completes. If you save the search results, they will be deleted after seven days.

29. How can you add folder access logs from a Windows machine to Splunk?

To add folder access logs from a Windows machines to Splunk, you must follow the steps listed below:

- Go to Group Policy and enable Object Access Audit on the Windows machine where the folder is located.

- Now you have to enable auditing on the specific folder for which you want to monitor access logs.
- Install Splunk Universal Forwarder on the Windows machine.
- Configure the Universal Forwarder to send security logs to the Splunk Indexer.

30. How does Splunk avoid duplicate indexing of logs?

Among many, one of the common Splunk interview questions and answers is this. The Splunk Indexer keeps track of all the indexed events in a directory – the Fishbuckets directory that contains seek pointers and CRCs for all the files being indexed presently. So, if there's any seek pointer or CRC that has been already read, splunkd will point it out.

31. What is the configuration files precedence in Splunk?

The precedence of configuration files in Splunk is as follows:

- System Local Directory (highest priority)
- App Local Directories
- App Default Directories
- System Default Directory (lowest priority)

32. Define "Search Factor" and "Replication Factor."

Both Search Factor (SF) and Replication Factor (RF) are clustering terminologies in Splunk. While the SF (with a default value of 2) determines the number of searchable copies of data maintained by the Indexer cluster, the RF represents the number of copies of data maintained by the Indexer cluster. An important thing to remember is that SF must always be less than or equal to the replication factor. Also, the Search Head cluster only has a Search Factor, whereas an Indexer cluster has both SF and RF.

33. Why is the lookup command used? Differentiate between inputlookup & outputlookup commands.

In Splunk, lookup commands are used when you want to receive specific fields from an external file (for example, a Python-based script, or a CSV file) to obtain a value of an event. It helps narrow the search results by referencing the fields in an external CSV file that matches fields in the event data.

The inputlookup command is used when you want to take an input. For instance, the command can take the product price or product name as input and then match it with an internal field such as a product ID. On the contrary, the outputlookup command is used to produce an output from an existing field list.

34. Differentiate between Splunk SDK and Splunk Framework.

Splunk SDKs are primarily designed to help users develop applications from scratch. They do not require Splunk Web or any other component from the Splunk App Framework to function. Splunk SDKs are separately licensed from Splunk. As opposed to this, the Splunk App Framework rests within the Splunk Web Server. It allows users to customize the Splunk Web UI that accompanies the product. Although it lets you develop Splunk apps, you have to do so by using the Splunk Web Server.

35. What are the pros of getting data into a Splunk instance using forwarders?

The benefits of using forwarders to enter data into Splunk include secure SSL connections, bandwidth throttling, and TCP connections for sending valuable data from a forwarder to an indexer.

36. In which form does Splunk stores its data?

When asked in which form does Splunk stores its data, you can answer by mentioning Splunk stores data in a flat-file format. Based on the amount and age of the data, Splunk stores all data in an index and in hot, warm, and cold buckets.

37. Explain the map-reduce algorithm?

To speed up data searches, Splunk employs the map-reduce method. It takes its cues from two functional programming constructs: 1) reduce () 2) map ().

Here, the reduce () function is connected to a Reducer class, while the map () function is connected to a Mapper class.

38. Explain various types of data inputs in Splunk?

The following list includes various Splunk data inputs:

- Using files and directories as sources
- Setting up network ports to start receiving inputs
- Include Windows inputs. There are four different kinds of windows inputs: the active directory monitor, printer monitor, network monitor, and registry inputs monitor.

39. What are Pivot and Data Models?

Pivots are used to build the front views of your output and choose the proper filter for a better view of this output.

When processing enormous amounts of unstructured data in Splunk, data models are utilized to build a hierarchical model without running complicated search queries on the data. Data models are frequently used to build authentication structures for multiple applications, add access levels, and create sales reports.

On the contrary, Pivots allow you to design numerous views and view the outcomes as you see fit. Even managers of stakeholders with no technical background can construct views and learn more about their departments with pivots.

40. What are Workflow Actions?

One of the commonly asked **Splunk interview questions and answers** is this. In Splunk, "workflow actions" are knowledge objects with a high degree of configuration that let you interact with other areas and websites. Splunk workflow actions can be used to construct HTML links and utilize them to search field values, send HTTP post requests to particular URLs, and carry out secondary searches on particular events.

41. Which component of a bucket stores raw event data?

'Which component of a bucket stores raw event data?' can most likely be asked by interviewers to test your in-depth knowledge. You can answer this in the following manner.

Each bucket has a compressed journal in time-series index files. Splunk stores our unprocessed event data in the journal. It is made up of numerous smaller compressed slices, each measuring roughly 128kB. The index keys to our journal file are the time-series index files or TSIDX files.

42. Specify the command that is used for the "Filtering results" category.

The following commands are used for the "filtering results" category: "where," "Sort," "rex," and "search."

43. List the various Splunk licensing types.

Splunk licenses come in the following types:

- Free license

- Beta license
- Search head license
- Cluster member license
- Forwarder license
- Enterprise license

44. Who are the largest competitors to Splunk?

The largest competitors of Splunk are logstash, Loggly, LogLogic, sumo logic, etc.

45. What do Splunk licenses specify?

They specify how much data you can index per calendar day.

46. How does Splunk determine 1 day from the licensing point of view?

Splunk determines the time from Midnight to midnight on the clock of the license master.

47. How are forwarder licenses purchased?

You need not purchase forwarder licenses separately, as they are included with Splunk.

48. What is the command to restart only the Splunk web server?

The command is – Splunk start Splunk web.

49. What is the command to restart only the Splunk daemon?

The command is – Splunk start Splunk.

50. What are the three Splunk versions?

There are three different versions of Splunk available. Splunk enterprise, Splunk light, and Splunk cloud are the three available versions.

- Splunk Enterprise: Many IT firms utilize the Splunk Enterprise edition. You can use it to examine data from numerous applications and websites.
- Splunk Cloud: Splunk Cloud is a SaaS (Software as a Service) that includes features like APIs, SDKs, and apps that are similar to those of the business edition.
- Splunk lite: Splunk light is a free version that lets you search, edit, and create reports using your log data. Splunk Light has limited features as compared to its other versions.

EXTRA

51. What is Splunk?

Splunk is a machine data analytics platform that lets you monitor, track, and troubleshoot your customers' businesses in real time. This allows you to deliver insights that drive business value and gain actionable intelligence. You can also monitor a diverse set of inputs such as network and system logs, databases, websites, mobile applications, and operational technology OT system. Using this information to take immediate action and increase customer value is called real-time intelligence. Real-time intelligence lets you identify customer issues and take action before they impact customer satisfaction and business. The platform provides a searchable store of all customer generated data, including app usage, customer behaviour and business performance.

52. What is the Splunk architecture?

Splunk is a server-based software application that collects, indexes and stores machine-generated data for trending, analysing and reporting. Splunk software is installed on a dedicated forwarder device or computer. This device is then connected to the target machine or server that it is collecting data from. The data is then sent to Splunk over the network and is stored in a database. Once the data is stored it can be used to perform various types of monitoring and reporting.

53. What are search modes in Splunk?

You can gather data from a variety of different sources to analyse using Splunk. Splunk supports several search modes, representing different ways of interacting with data in Splunk. You can use the search bar to query the Splunk index and to generate search results without defining any search modes in advance. The 3 types of search modes in Splunk are – Fast, Smart, and Verbose modes.