Top 100 Splunk Interview Questions & Answers

Here are important frequently asked Splunk interview questions for freshers as well as experienced candidates to get the right job.

1) Define Splunk

It is a software technology that is used for searching, visualizing, and monitoring machine-generated big data. It monitors and different types of log files and stores data in Indexers.

2) List out common ports used by Splunk.

Common ports used by Splunk are as follows:

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

3) Explain Splunk components

The fundamental components of Splunk are:

- Universal forward: It is a lightweight component which inserts data to Splunk forwarder.
- Heavy forward: It is a heavy component that allows you to filter the required data.
- Search head: This component is used to gain intelligence and perform reporting.
- License manager: The license is based on volume & usage. It allows you to use 50 GB per day. Splunk regular checks the licensing details.
- Load Balancer: In addition to the functionality of default Splunk loader, it also enables you to use your personalized load balancer.

4) What do you mean by Splunk indexer?

It is a component of Splunk Enterprise which creates and manages indexes. The primary functions of an indexer are 1) Indexing raw data into an index and 2) Search and manage Indexed data.

5) What are the disadvantages of using Splunk?

Some disadvantages of using Splunk tool are:

- Splunk can prove expensive for large data volumes.
- Dashboards are functional but not as effective as some other monitoring tools.
- Its learning curve is stiff, and you need Splunk training as it's a multi-tier architecture. So, you need to spend lots of time to learn this tool.
- Searches are difficult to understand, especially regular expressions and search syntax.

6) What are the pros of getting data into a Splunk instance using forwarders?

The advantages of getting data into Splunk via forwarders are TCP connection, bandwidth throttling, and secure SSL connection for transferring crucial data from a forwarder to an indexer.

7) What is the importance of license master in Splunk?

License master in Splunk ensures that the right amount of data gets indexed. It ensures that the environment remains within the limits of the purchased volume as Splunk license depends on the data volume, which comes to the platform within a 24-hour window.

8) Name some important configuration files of Splunk

Commonly used Splunk configuration files are:

- Inputs file

- Transforms file
- Server file
- Indexes file
- Props file

9) Explain license violation in Splunk.

It is a warning error that occurs when you exceed the data limit. This warning error will persist for 14 days. In a commercial license, you may have 5 warnings within a 1-month rolling window before which your Indexer search results and reports stop triggering.

However, in a free version, license violation warning shows only 3 counts of warning.

10) What is the use of Splunk alert?

Alerts can be used when you have to monitor for and respond to specific events. For example, sending an email notification to the user when there are more than three failed login attempts in a 24-hour period.

11) Explain map-reduce algorithm

Map-reduce algorithm is a technique used by Splunk to increase data searching speed. It is inspired by two functional programming functions 1) reduce () 2) map().

Here map() function is associated with Mapper class and reduce() function is associated with a Reducer class.

12) Explain different types of data inputs in Splunk?

Following are different types of data inputs in Splunk:

- Using files and directories as input
- Configuring Network ports to receive inputs automatically
- Add windows inputs. These windows inputs are of four types: 1) active directory monitor, 2) printer monitor, 3) network monitor, and 4) registry inputs monitor.

13) How Splunk avoids duplicate log indexing?

Splunk allows you to keeps track of indexed events in a fish buckets directory. It contains CRCs and seeks pointers for the files you are indexing, so Splunk can’t if it has read them already.

14) Explain pivot and data models.

Pivots are used to create the front views of your output and then choose the proper filter for a better view of this output. Both options are beneficial for the people from a semi-technical or non-technical background.

Data models are most commonly used for creating a hierarchical model of data. However, it can also be used when you have a large amount of unstructured data. It helps you make use of that information without using complicated search queries.

15) Explain search factor and replication factor?

Search factor determines the number of data maintained by the indexer cluster. It determines the number of searchable copies available in the bucket.

Replication factor determines the number of copies maintained by the cluster as well as the number of copies that each site maintains.

16) What is the use of lookup command?

Lookup command is generally used when you want to get some fields from an external file. It helps you to narrow the search results as it helps to reference fields in an external file that match fields in your event data.

17) Explain default fields for an event in Splunk

There are 5 default fields which are barcoded with every event into Splunk. They are: 1) host, 2) source, 3) source type, 4) index, and 5) timestamp.

18) How can you extract fields?

In order to extract fields from either sidebar, event lists or the settings menu using UI.

Another way to extract fields in Splunk is to write your regular expressions in a props configuration file.

19) What do you mean by summary index?

A summary index is a special index that stores that result calculated by Splunk. It is a fast and cheap way to run a query over a longer period of time.

20) How to prevent events from being indexed by Splunk?

You can prevent the event from being indexed by Splunk by excluding debug messages by putting them in the null queue. You have to keep the null queue in transforms.conf file at the forwarder level itself.

21) Define Splunk DB connect

It is a SQL database plugin which enables to import tables, rows, and columns from a database add the database. Splunk DB connect helps in providing reliable and scalable integration between databases and Splunk Enterprises.

22) Define Splunk buckets

It is the directory used by Splunk enterprise to store data and indexed files into the data. These index files contain various buckets managed by the age of the data.

23) What is the function of Alert Manager?

The alert manager adds workflow to Splunk. The purpose of alert manager o provides a common app with dashboards to search for alerts or events.

24) How can you troubleshoot Splunk performance issues?

Three ways to troubleshoot Splunk performance issue.

- See server performance issues.
- See for errors in splunkd.log.
- Install Splunk app and check for warnings and errors in the dashboard.

25) What is the difference between Index time and Search time?

Index time is a period when the data is consumed and the point when it is written to disk. Search time take place while the search is run as events are composed by the search.

26) How to reset the Splunk administrator password?

In order to reset the administrator password, perform the following steps:

1. Login into the server on which Splunk is installed
2. Rename the password file and then again start the Splunk.
3. After this, you can sign into the server by using username either administrator or admin with a password changeme.

27) Name the command which is used to the "filtering results" category

The command which is used to the "filtering results" category is: "where," "Sort," "rex," and "search."

28) List out different types of Splunk licenses

The types of Splunk licenses are as follows:

- Free license
- Beta license
- Search heads license
- Cluster members license
- Forwarder license

- Enterprise license

29) List out the number of categories of the SPL commands.

The SPL commands are classified into five categories:

1) Filtering Results, 2) Sorting Results, 3) Filtering Grouping Results, 4) Adding Fields, and 5) Reporting Results.

30) What is eval command?

This command is used to calculate an expression. Eval command evaluates boolean expressions, string, and mathematical articulations. You can use multiple eval expressions in a single search using a comma.

31) Name commands which are included in the reporting results category

Following are the commands which are included in the reporting results category:

- Rare
- Chart
- time chart
- Top
- Stats

32) What is SOS?

Splunk on Splunk or SOS is a Splunk app that helps you to analyze and troubleshoot Splunk environment performance and issues.

33) What is a replace command?

This command searches and replaces specified field values with replacement values.

34) Name features which are not available in Splunk free version?

Splunk free version lacks the following features:

- Distributed searching
- Forwarding in HTTP or TCP
- Agile statistics and reporting with Real-time architecture
- Offers analysis, search, and visualization capabilities to empower users of all types.
- Generate ROI faster

35) What is a null queue?

A null queue is an approach to filter out unwanted incoming events sent by Splunk enterprise.

36) Explain types of search modes in Splunk?

There are three types of search modules. They are:

- Fast mode: It increases the searching speed by limiting search data.
- Verbose mode: This mode returns all possible fields and event data.
- Smart mode: It is a default setting in a Splunk app. Smart mode toggles the search behavior based on transforming commands.

37) What is the main difference between source & source type

The source identifies as a source of the event which a particular event originates, while the sourcetype determines how Splunk processes the incoming data stream into events according to its nature.

38) What is a join command?

It is used to combine the results of a sub search with the results of the actual search. Here the fields must be common to each result set. You can also combine a search set of results to itself using the selfjoin command in Splunk.

39) How to start and stop Splunk service?

To start and stop Splunk serives use can use following commands:
[crayon-5f2d06aa14cf9548944296/]

40) Where to download Splunk Cloud?

Visit website: https://www.splunk.com/ to download a free trial of Splunk Cloud.

41) What is the difference between stats and timechart command?

Parameter	Stats	Timechart
Purpose	They are used to represent numerical data in tabular format.	Timechart is used to represent search result in a graphical view.
Fields usage	Stats can use more than one field.	It uses _time as default field in the graph.

42) Define deployment server

Deployment server is a Splunk instance that acts as a centralized configuration manager. It is used to deploy the configuration to other Splunk instances.

43) What is Time Zone property in Splunk?

Time zone property provides the output for a specific time zone. Splunk takes the default time zone from browser settings. The browser takes the current time zone from the computer system, which is currently in use. Splunk takes that time zone when users are searching and correlating bulk data coming from other sources.

44) What is Splunk sound unit connect?

Splunk sound unit is a plugin which allows adding info data with Splunk reports. It helps in providing reliable and ascendible integration between relative databases and Splunk enterprises.

45) How to install forwarder remotely?

You can make use of a bash script in order to install forwarder remotely.

46) What is the use of syslog server?

Syslog server is used to collect data from various devices like routers and switches and application logs from the web server. You can use R syslog or syslog NG command to configure a Syslog server.

47) How to monitor forwarders?

Use the forwarder tab available on the DMC (Distributed Management Console) to monitor the status of forwarders and the deployment server to manage them.

48) What is the use of Splunk btool?

It is a command-line tool that is designed to solve configuration related issues.

49) Name Splunk alternatives

Some Splunk alternatives are:

- Sumo logic
- Loglogic
- Loggy
- Logstash

50) What is KV store in Splunk?

Key Value(KV) allows to store and obtain data inside Splunk. KV also helps you to:

- Manage job queue
- Store metadata
- Examine the workflow

51) What do you mean by deployer in Splunk?

Deployer is a Splunk enterprise instant which is used to deploy apps to the cluster head. It can also be used to configure information for app and user.

52) When to use auto_high_volume in Splunk?

It is used when the indexes are of high volume, i.e., 10GB of data.

53) What is a stat command?

It is a Splunk command that is used to arrange report data in tabular format.

54) What is a regex command?

Regex command removes results which do not match with desired regular expression.

55) What is input lookup command?

This Splunk command returns lookup table in the search result.

56) What is the output lookup command?

Output lookup command searches the result for a lookup table on the hard disk.

57) List out various stages of bucket lifecycle

Stages of bucket lifecycle are as follows:

- Hot
- Warm
- Cold
- Frozen
- Thawed

58) Name stages of Splunk indexer

Stages of Splunk indexer are:

- Input
- Parsing
- Indexing
- Searching

59) Explain the distinction between Splunk and Spark

Parameter	Splunk	Spark
Purpose	Collect a large amount of computer-generated data.	Used for big data processing
Preference	Can be integrated easily with Hadoop	It is more preferred and can be used with apache projects.
Mode	Streaming mode	Streaming as well as batch mode

60) Explain how Splunk works?

There are three phases in which Splunk works:

- The first phase: It generates data and solves query from various sources.
- The second phase: It uses the data to solve the query.
- Third phase: it displays the answers via graph, report, or chart which is understood by audiences.

61) What are three versions if Splunk?

Splunk is available in three different versions. These versions are 1) Splunk enterprise, 2) Splunk light, 3) Splunk cloud.

- Splunk enterprise: Splunk Enterprise edition is used by many IT organizations. It helps you to analyze the data from various websites and applications.

- Splunk cloud: Splunk Cloud is a SaaS (Software as a Service) It offers almost similar features as the enterprise version, including APIs, SDKs, and apps.

- Splunk light: Splunk light is a free version which allows, to make a report, search and edit your log data. Splunk light version has limited functionalities and features compared to other versions.

62) Name companies which are using Splunk

Well known companies which are using Splunk tool are:

- Cisco
- Facebook
- Bosch
- Adobe
- IBM
- Walmart
- Salesforce

63) What is SLP?

Search Processing Language or SLP is a language which contains functions, commands, and arguments. It is used to get the desired output from the database.

64) Define monitoring in Splunk

Monitoring is a term related to reports you can visually monitor.

65) Name the domain in which knowledge objects can be used

Following are a few domains in which knowledge objects can be used:

- Application Monitoring
- Employee Management
- Physical Security
- Network Security

66) How many roles are there in Splunk?

There are three roles in Splunk: 1) Admin, 2) Power, and 3) User.

67) Are search terms in Splunk case sensitive?

No, Search terms in Splunk are not case sensitive.

68) Can search results be used to change the existing search?

Yes, the search result can be used to make changes in an existing search.

69) List out layout options for search results.

Following are a few layout options for search result:

- List
- Table
- Raw

70) What are the formats in which search result be exported?

The search result can be exported into JSON, CSV, XML, and PDF.

71) Explain types of Boolean operators in Splunk.

Splunk supports three types of Boolean operators; they are:

- AND: It is implied between two terms, so you do not need to write it.
- OR: It determines that either one of the two arguments should be true.

- NOT: used to filter out events having a specific word.

72) Explain the use of top command in Splunk

The top command is used to display the common values of a field, with their percentage and count.

73) What is the use of stats command?

It calculates aggregate statistics over a dataset, such as count, sum, and average.

74) What are the types of alerts in Splunk?

There are mainly three types of alerts available in Splunk:

- Scheduled alert: It is an alert that is based on a historical search. It runs periodically with a set schedule.
- Per result alert: This alert is based on a real time search which runs overall time.
- Rolling window alert: An alert that is based on real-time search. This search is set to run within a specific rolling time window that you define.

75) List various types of Splunk dashboards.

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

76) What is the use of tags in Splunk?

They are used to assign names to specific filed and value pairs. The filed can be event type, source, source type, and host.

77) How to increase the size of Splunk data storage?

In order to increase the size of data storage, you can either add more space to index or add more indexers.

78) Distinguish between Splunk apps and add-ons

There is only one difference between Splunk apps, and add-ons that is Splunk apps contains built-in reports, configurations, and dashboards. However, Splunk add-ons contain only built-in configurations they do not contain dashboards or reports.

79) Define dispatch directory in Splunk?

Dispatch directory stores status like running or completed.

80) What is the primary difference between stats and eventstats commands

Stats command provides summary statistics of existing fields available in search output, and then it stores them as values in new fields. On the other hand, in eventstats command aggregation results are added so that every event only if the aggregation applies to that particular event.

81) What do you mean by source type in Splunk?

Source field is a default field that finds the data structure of an event. It determines how Splunk formats the data while indexing.

82) Define calculated fields?

Calculated fields are the fields which perform the calculation which the values of two fields available in a specific event.

83) List out some Splunk search commands

Following are some search commands available in Splunk:

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

84) What does xyseries command do?

xyseries command converts the search results into a format that is suitable for graphing.

85) What is the use of spath command?

spath command is used to extract fields from structured data formats like JSON and XML.

86) How to adds summary statistics to all results in a streaming manner?

In order to add summary statistics in results, you can use streamstats.

87) Where to create knowledge objects, dashboards, and reports?

You can create knowledge, objects, reports, and dashboards in reporting and search app.

88) What is table command?

This command returns all fields of table in the argument list.

89) How to remove duplicate events having common values?

Use dedup command to remove duplicate events having common values.

90) What is the main difference between sort + and sort -?

- sort + displays search in ascending order
- sort - displays search in descending order.

91) Define reports in Splunk

They are results saved from a search action that shows the visualization and statistic of a particular event.

92) Define dashboard in Splunk

The dashboard is defined as a collection of views that are made of various panels.

93) What is the use of instant pivot in Splunk?

It is used to work with data without creating any data model.

Instant pivot is available to all users.

94) How is it possible to use the host value and not IP address or the DNS name for a TCP input?

Under stanza in the input configuration file, set the connection_host to none and mention the host value.

95) What is the full form of LDAP?

LDAP stands for Lightweight Directory Access Protocol

96) Define search head pooling

It is a group of servers connected with each other. These servers are used to share configuration, user data, and load.

97) Define search head clustering

It is a group of Splunk enterprise search heads that serves as a central resource for searching.

98) What is the full form of REST?

The abbreviation of REST is Representational State Transfer

99) Explain Splunk SDKs

Splunk SDKs are written on the base of Splunk REST APIs. Various languages supported by SDKs are: 1) Java, 2) Python, 3) JavaScript, and 4) C#.

100) Explain Splunk REST API

The Splunk REST API offers various processes for accessing every feature available in the product. Your program communicates to Splunk enterprise using HTTP or HTTPS. It uses the same protocols that any web browser uses to interact with web pages.

101) What is security accelerate data model in Splunk?

Splunk Enterprise Security accelerates data model provides a panel, dashboard, and correlation search results. It uses the indexers for processing and storage. The accelerated data is stored within each index by default.

102) Explain how indexer stores various indexes?

Indexers create various files which contain two types of data: 1) Raw data and 2) metadata index file. Both these files are used to constitute Splunk enterprise index.