Splunk Interview Questions

1. **What is Splunk?**

   Splunk is an observability and Security application. It helps capture, index and correlate real-time data in a searchable repository, from which it can generate graphs, reports, alerts, dashboards and visualisations. Splunk uses machine data to identify patterns, provide metrics, diagnose problems, and provide intelligence for business operations.

2. **What are the standard ports for Splunk**

Service	Port Number Used
Splunk Web port	8000
Splunk Management port	8089
Splunk Indexing port	9997

Splunk Index Replication port	8080
Splunk Network port	514 (Used to get data from the Network port, i.e., UDP data)
KV Store	8191

3. **What are the components of Splunk? Explain Splunk architecture.?**

   Splunk is a powerful platform for searching, monitoring, and analysing machine-generated data. Its basic components include:

   Forwarders: These are agents or daemons installed on the machines generating data. Forwarders collect data from various sources and send it to the Splunk indexer for further processing and storage.

   Indexers: Indexers are responsible for receiving, indexing, and storing the data sent by forwarders. They organise the data into events and create an index for faster searching. Indexers can also perform searches locally.

   Search Head: The Search Head is the user interface of Splunk. It provides a web-based GUI for searching, analysing, and visualising the data stored in the indexers. The Search Head is where users interact with Splunk to run searches, create dashboards, and generate reports.

4. **Which is the latest Splunk version in use?**

   9.1. 1 was released on August 30, 2023

5. **What is a Splunk forwarder? What are the types of Splunk forwarders?**

   Heavy Forwarder: The Heavy Forwarder was an advanced version of the Universal Forwarder with additional data processing capabilities. It could perform data transformations, filtering, and more before forwarding data to the indexer.

   Universal Forwarder: This is the most commonly used type of forwarder. It is a lightweight, dedicated agent explicitly designed for forwarding data to a Splunk indexer. Universal forwarders are suitable for environments where resource usage needs to be minimised. They have a smaller footprint compared to heavy forwarders.

6. **Types of Splunk Licenses**

   Splunk Enterprise License: This is the standard license for Splunk's flagship product, Splunk Enterprise. It provides access to various features and capabilities, including data indexing, searching, alerting, reporting, and visualisation.

Splunk Free License: Splunk offers a free version that allows users to index a limited amount of daily data. It provides basic functionality for searching and analysing data. This license is often used for smaller environments or evaluation purposes.

Splunk Trial License: Splunk offers trial licenses that provide access to the complete set of features in Splunk Enterprise for a limited period. These licenses are typically used for evaluation purposes.

Splunk License Types

Mr.Varma
VarmaTrainer@gmail.com
Splunk,SIEM,SOAR,CEH,CISSP,CCSPTrainings

- **Enterprise trial license**
  - Downloads with product
  - Features same as Enterprise except for 500mb per day limit
  - Only valid for 60 days, after which one of the other 3 license types must be activated
  - ***Sales trial license*** is a trial Enterprise license of varying size and duration
- **Enterprise license**
  - Purchased from Splunk
  - Full functionality for indexing, search head, deployment server, etc.
  - Sets the daily indexing volume
  - No-enforcement license, allows users to keep searching even if you are in a license violation period

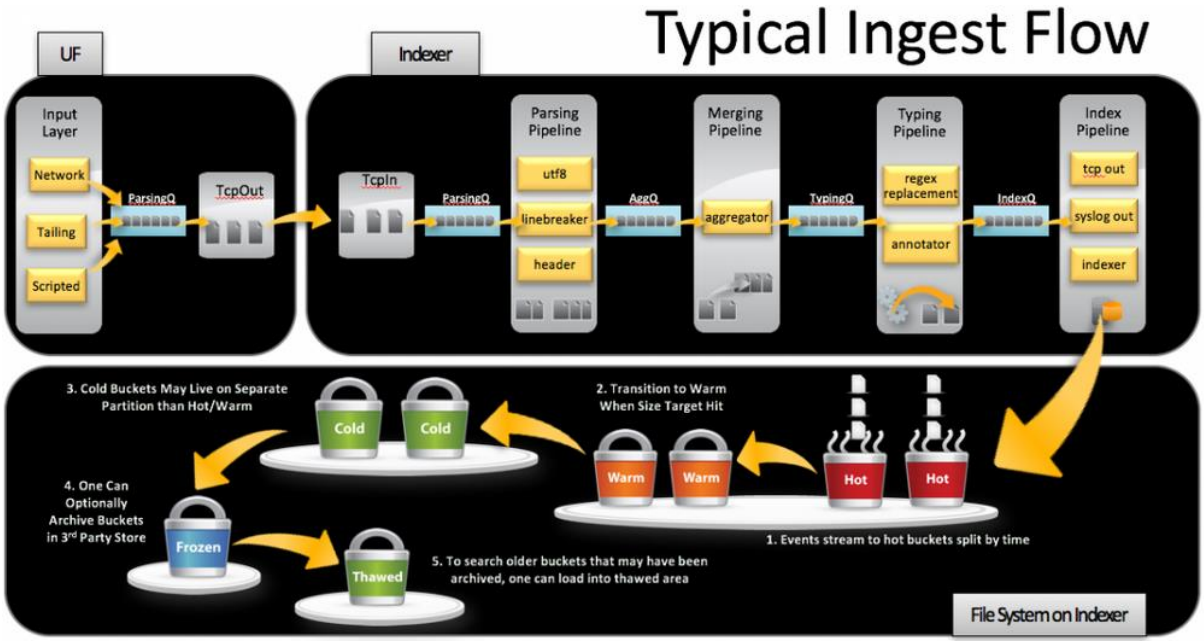
Splunk License Types (cont.)

- **Free license**
  - Disables alerts, authentication, clustering, distributed search, summarization, and forwarding to non-Splunk servers
  - Allows 500mb/day of indexing and forwarding to other Splunk instances
- **Forwarder license**
  - Sets the server up as a heavy forwarder
  - Applies to non-indexing forwarders
  - Allows authentication, but no indexing
- Splunk license comparison:

https://www.splunk.com/en_us/products/splunk-enterprise/free-vs-enterprise.html

Mr.Varma
VarmaTrainer@gmail.com
Splunk,SIEM,SOAR,CEH,CISSP,CCSP Trainings

7. **What is a Splunk indexer? What are the stages of Splunk indexing?**



using the 2 different stanzas, we can send data to the Hot bucket to Cold bucket.

- frozenTimePeriodInSecs
- maxTotalDataSizeMB

    - **Input:** Inputs data from source. Source-wide keys, such as source/sourcetypes/hosts, are annotated here. The output of these pipelines is sent to the parsingQueue.
    - **Parsing:** Parsing of UTF8 decoding, Line Breaking, and header is done here. This is the first place to split the data stream into a single-line event. Note that in a Universal Forwarder (UF), this parsing pipeline does "NOT" do parsing jobs.
    - **Merging:** Line Merging for multi-line events and Time Extraction for each event are done here.
    - **Typing:** Regex Replacement, Punct. Extractions are done here.
    - **IndexPipe:** Tcpout to another Splunk, syslog output, and indexing are done here. In addition, this pipeline is responsible for byte quota, block signing, and indexing metrics such as thruput, etc.

8. Where is the Splunk default configuration stored?

```
splunkhome/etc/system/default
```

9. What happens if the license master is unreachable?
    - Splunk will continue to function normally for a period of time based on its license quota. The license master communicates with the indexers to enforce license usage.

- If the license master remains unreachable for an extended period, Splunk may start issuing warnings about the impending license expiration.

10. **What is a summary index in Splunk?**
    To store aggregated or summarised data. It allows users to pre-calculate and store summary statistics, metrics, or key performance indicators (KPIs) based on existing indexed data. This can significantly improve the performance of searches and reports that require the analysis of large datasets or complex calculations.

11. **What is Splunk DB Connect?**
    Splunk DB Connect is particularly valuable for organisations that rely on databases as a critical part of their IT infrastructure and want to integrate that data into their overall Splunk-based monitoring and analytics platform. It allows for a more comprehensive view of an organisation's data by incorporating structured, relational data alongside machine-generated logs and events.

12. **Stats vs. Transaction Command**

    Use stats to perform summary statistics and aggregations on individual events or groups of events based on specified fields.
    Use transaction when you want to group events based on specific criteria (e.g., time, shared field value) to understand the relationship between events in a transaction.

    In some cases, you might even use both commands together, using transaction to group events and then applying stats to analyse the transactions.

13. **Troubleshoot Splunk performance issues.**

    Check splunkd.log for errors, warnings, and other indicators of issues.

    Check server performance issues and monitor CPU, memory, and disk usage on all components (indexers, search heads, forwarders, etc.). Identify any resource spikes or consistently high usage.

    Install the SOS (Splunk on Splunk) app and check for warnings and errors in its dashboard. Check the number of saved searches currently running and their consumption of system resources.

    Additional Details:

    Indexer Health: Verify the health of the indexers. Ensure they are processing data efficiently, have enough disk space, and are not encountering errors in indexing.

    Search Head Performance: Analyze the performance of search heads. Check if there are any long-running searches, resource-intensive dashboards, or misconfigured settings.

    Review Search Queries: Look for complex or inefficient search queries. Evaluate whether they can be optimised or if they are overloading the system.

Check Forwarders: Verify that forwarders are functioning correctly and not overwhelming the indexers with data. Ensure they are sending data at an appropriate rate.

Review Indexing Pipeline: Understand the flow of data through the indexing pipeline. Identify any bottlenecks, such as heavy transformations or congested queues.

Review Data Volume and Retention: Ensure data volume is within expected limits. Consider adjusting data retention policies to manage storage requirements.

Check for Data Model Acceleration: Evaluate if data models are being used and if they are accelerating searches. Improperly configured or overly complex data models can impact performance.

Check Summary Indexes: Evaluate the size and performance of summary indexes. Ensure they are not consuming excessive resources or causing performance degradation.

Review Hardware and Configuration: Verify that hardware meets Splunk's recommended specifications. Ensure configurations align with best practices.

Review Search Head Clustering (if applicable): If using a search head cluster, ensure it is balanced and all members are healthy. Monitor search head captain elections.

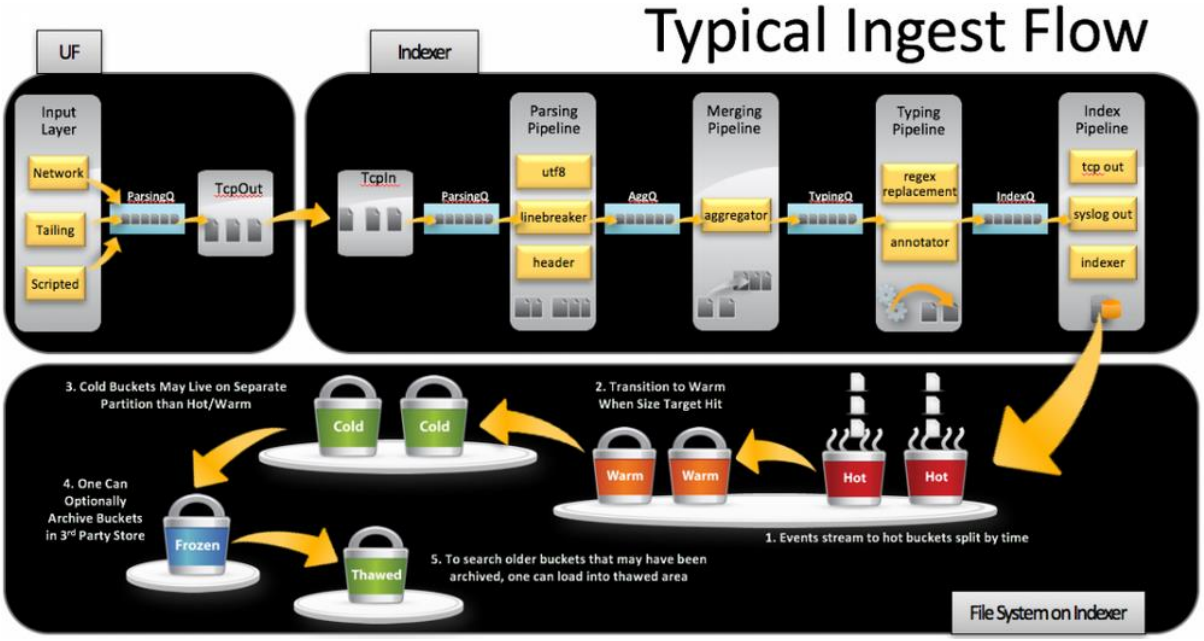
Consider Distributed Deployment: If the environment is large, consider distributing components to spread the load and improve performance.

Engage Splunk Support (if necessary): If you cannot resolve the issue, consider contacting Splunk Support for further assistance. Please provide them with relevant logs, metrics, and details about the environment.

Implement Performance Enhancements: Based on your findings, implement necessary changes such as query optimisation, index tuning, or hardware upgrades.

Monitor and Iterate: Continuously monitor the environment and make adjustments as needed. Performance tuning is an ongoing process.

14. **What are buckets? Explain the Splunk bucket lifecycle.**
    How Splunk Indexing works



Hot: A hot bucket contains newly indexed data. It is open for writing. There can be one or more hot buckets for each index.

Warm: A warm bucket consists of data rolled out from a hot bucket. There are many warm buckets.

Cold: A cold bucket has data rolled out from a warm bucket. There are many cold buckets.

Frozen: A frozen bucket comprises data rolled out from a cold bucket. The indexer deletes frozen data by default, but we can archive it. Archived data can later be thawed (data in a frozen bucket is not searchable).

15. **What command checks the running Splunk processes on Unix/Linux?**

```
ps aux | grep splunk
```

16. **What is a sourcetype in Splunk?**
    Sourcetype is a metadata attribute defining the format or type of data being indexed. It helps Splunk understand how to interpret and process the raw data it receives. Source types are crucial for parsing, field extraction, and applying the appropriate knowledge objects (such as event types and tags) to the data.

17. **What is Btool? How will you troubleshoot Splunk configuration files?**

    btool can be used to check the syntax and validity of configuration files. This helps identify any errors or misconfigurations that might be causing issues.

    It provides insights into how configuration files are layered and applied in Splunk. This is important for understanding precedence and inheritance rules.

18. **What is the '.conf' file's precedence in Splunk?**
    - System local directory — highest priority
    - App local directories
    - App default directories
    - System default directory — lowest priority

19. **What is a fishbucket? What is a fishbucket index?**

    Fishbucket is an index containing seek pointers and CRCs for the files we are indexing so 'splunkd' can tell us if it has read them already.

    We can access it through the GUI by searching for index=_thefishbucket

20. **How do I exclude some events from being indexed by Splunk?**

    This can be done by defining a regex to match the necessary event(s) and sending everything else to NullQueue. Here is a basic example that will drop everything except events that contain the string login:

    In **props.conf:**

```
[example_sourcetype]
TRANSFORMS-nullQueue = setnull
```

    In **transforms.conf**

```
[setnull]
REGEX = .
```

```
DEST_KEY = queue
FORMAT = nullQueue
```

21. **How do you set the default search time in Splunk?**

```
$SPLUNK_HOME/etc/system/local/ui-prefs.conf file
```

includes:

```
[search]
dispatch.earliest_time = @d
dispatch.latest_time = now
```

22. **What is a dispatch directory?**
    It contains a directory for each search that is running or has been completed.

    For example, a directory named 1434308943.358 will contain a CSV file of its search results, a search.log with details about the search execution, etc.

23. **Define the terms 'search factor' and 'replication factor.'**
    Search factor: The search factor (SF) decides the number of searchable copies an indexer cluster can maintain of the data/bucket. For example, the search factor value 3 shows that the cluster can keep up to 3 copies of each bucket.

    Replication factor: The replication factor (RF) determines how many users receive copies of your data/buckets. However, the search factor should not be more significant than the replication factor.

24. How to stop/start the Splunk service?

Goto $splunk-Home/bin/

```
#splunk stop
./splunk stop

#splunk start
./splunk start

#splunk restart
./splunk restart
```

25. Important Top 25 Splunk Commands

- **Search** The most fundamental command used to search for events in Splunk.
- **table**: Formats search results into a table for easy visualisation.
- **stats**: Performs statistical operations like count, sum, average, etc., on search results.
- **timechart**: Generates time-based visualisations like line charts, bar charts, and area charts.
- **eval**: Allows the creation of calculated fields based on existing fields or functions.
- **where**: Filters events based on specified conditions.
- **rename**: Renames fields for better readability.
- **dedup**: Removes duplicate events based on specified fields.
- **rex**: Performs regular expression extractions to extract specific information from events.
- **transaction**: Groups events together based on defined criteria (e.g., time, common field value).
- **join**: Joins results from multiple sub-searches based on common fields.
- **lookup**: Enriches events with additional information from external lookup tables.
- **append**: Combines results from two or more searches.
- **Top**: Displays the top values for a specified field based on a defined statistic (e.g., count).
- **chart**: Create various charts and graphs based on search results.
- **time**: Sets the time range for a search.
- **filldown**: Copies field values from one event to another based on specified conditions.
- **mvexpand**: Expands multivalue fields into separate events.
- **rename**: Renames fields for better readability.
- **inputlookup**: Reads data from external lookup tables.
- **outputlookup**: Writes search results to an external lookup table.
- **delete**: Deletes events or fields from search results. (you need can_delete role added to user role)
- **savedsearch**: Executes a saved search within a search.

26. **Search Modes in Splunk**

Fast mode: speeds up your search result by limiting the types of data.

Verbose mode: Slower compared to the fast mode, but returns the information for as many events as possible.

Smart mode: It toggles between different modes and search behaviours to provide maximum results in the shortest period of time.

27. **Mention some critical configuration files in Slunk.**

props.conf: Used for configuring event processing, such as field extractions, timestamps, and event line merging.

transforms.conf: Defines field extractions, aliasing, masking, and other data transformations.

inputs.conf: Specifies the inputs that Splunk monitors, such as files, directories, network ports, and scripts.

outputs.conf: Configures where events are forwarded to, such as indexers, other Splunk instances, or external systems.

indexes.conf: Contains settings for defining and managing indexes, including storage locations, retention policies, and access permissions.

server.conf: Contains various server-wide settings, including settings for web server, license management, and deployment configurations.

limits.conf: Controls system-wide limits and settings for various aspects of Splunk, such as concurrency, search limits, and resource allocation.

authentication.conf: Contains settings for user authentication methods, including local authentication, LDAP, and SAML.

authorize.conf: Defines user roles, capabilities, and role-based access control (RBAC) settings.

savedsearches.conf: Contains saved searches and pre-configured search queries that can be reused or scheduled.

28. stats vs eventstats command

Stats: The Stats command in Splunk calculates statistics for every field in your events (search results) and stores these values in newly created fields.

Eventstats: Similar to the stats command, this calculates a statistical result. While the Eventstats command is similar to the Stats command, it adds the aggregate results inline to each event (if only the aggregate is relevant to that event).

29. What is the Lookup command? State the difference between Inputlookup and Outputlookup commands.

Splunk lookup commands can retrieve specific fields from an external file (e.g., Python script, CSV file, etc.) to get the value of an event.

Inputlookup: Inputlookup can be used to search the contents of a lookup table (CSV lookup or a KV store lookup). It is used to take input. This command, for instance, could take the product price or product name as input and match it with an internal field like the product ID.

Outputlookup: Conversely, the outputlookup command outputs search results to a specified lookup table, i.e., it places a search result into a specific lookup table. 6. Explain what is Splunk Btool.

30. What do you mean by SF (Search Factor) and RF (Replication Factor)?

SF (Search Factor) & RF (Replication Factor) are terms associated with Clustering techniques, i.e., Search head clustering & Indexer clustering.

Search Factor: It is only associated with indexer clustering. It determines how many searchable copies of data the indexing cluster maintains. By default, the value of the search factor is 2.

Replication Factor: It is associated with both Search head clustering & Indexer clustering. In the case of the indexer cluster, the replication factor determines the number of copies of the data that an indexer cluster maintains. In contrast, in the case of the search head cluster, the replication factor determines the minimum number of copies of the search artefacts that a search head cluster maintains. For the replication factor, the default value is 3

31. Splunk universal forwarder – Troubleshooting steps

- The Splunk forwarder is an agent for log collection from remote machines.
- The role of the Splunk forwarder is to collect the logs from remote machines and forward them to the indexer for further processing and storage.
- Splunk universal Forwarders provide reliable, secure data collection from remote sources and forward that data into Splunk Enterprise for indexing and consolidation.
- Below are the few most common checks to help identify and resolve the problem efficiently.

Step 1: Check if the Splunk process is running on the Splunk forwarder

For Windows check services | for Linux use the below command

ps -ef |grep splunkd
Or
cd $SPLUNK HOME/bin
./splunk status

Step 2: Check if the Splunk forwarder forwarding port is open by using the below command

netstat -an | grep 9997
If the output of the above command is blank, then your port is not open. You need to open it.

Step 3: Check on the indexer if receiving is enabled on port 9997 and if port 9997 is open on the indexer

Step 4: Check if receiving is configured: on the indexer, go to setting>>forwarding and receiving >> check if receiving is enabled on port 9997. If not, enable it.

Step 5: Check if you can ping the indexer from the forwarder host

ping indexer name
If you are not able to ping the server, then check the network issue

Step 6: Confirm on the indexer if your file is already indexed or not by using the below search query

In the Splunk UI, run the following search - index=_internal "FileInputTracker" **

You will get a list of log files indexed as the output of the search query.

Step 7: Check if the forwarder has completed processing the log file (i.e. the tailing process by using the below URL)

https://splunk forwarder server name:8089/services/admin/inputstatus/TailingProcessor:FileStatus
In tailing process output, you can check if the forwarder is having an issue with processing the file

Step 8: Check out the log file permissions you send to Splunk. Verify if the Splunk user has access to the log file

Step 9: Checkout filesystem for the last modification and verify if the forwarder is monitoring it

Verify inputs.conf and outputs.conf for proper configuration

Below are sample configuration files for comparison:

Inputs.conf example:
[monitor:///var/log/secure]
disabled = false
sourcetype = linux_secure

[monitor:///var/log/messages]
disabled = false

sourcetype = syslog

outputs.conf example:
[tcpout:imp_A]
server = impAserver01.domain:9997,impAserver02.domain:9997
autoLB = true
[tcpout]
defaultGroup = imp_B
[tcpout:imp_B]
server = impBserver01.domain:9997,impBserver02.domain:9997

autoLB = true

Checkout disk space availability on the indexer

Step 10:

**Check splunkd.log on forwarder at location $SPLUNK_HOME/var/log/splunk for any errors. Like for messages that are from 'TcpOutputProc', they should give you an indication as to what is occurring when the forwarder tries to connect to the indexer

tcpdump port 9997 data for any errors

tcpdump -i eth0 port 9997

Step 11: Check out ulimit if you have installed Forwarder on linux. And set it to limites or max (65535 -Splunk recommended)

- ulimit is the limit set by default in linux for the number of files opened by a process
- check ulimit command: ulimit -n
- set ulimit command: ulimit -n expected size

Step 12: Finally, try restarting Splunk on the forwarder

32. What is the use of Syslog server?

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

33. How many default roles are there in Splunk?

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

34. Explain the 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.

35. In what scenario do we implement Heavy Forwarder?

1. IF my servers are geographically far, then I will keep HF as an intermediary to get data, parse and send to the Indexer.
2. If My indexer cluster requires any helping hand to reduce load, at the same time not want to scale horizonally, then I will use Heavy Forwarders as helping brothers.

36. Difference between props.conf and transforms.conf

- So in props.conf, you say "events with the sourcetype XXX has the extraction YYY applied to it at parse time" or "events from host HHH has lookup JJJ applied at search time".

- transforms.conf would specify exactly how extraction XXX worked, or where lookup JJJ comes from.

- some of the rules are specified directly in props.conf. Some of these (e.g., rules for parsing timestamps or line breaks) are only specified in props.conf, while others (search time field extractions) can be either directly defined in props.conf, or referenced back to transforms.conf

- regex based field extraction can be specified:
  (1) inline in props.conf via the EXTRACT- field or
  (2) in transforms.conf and referenced from props.conf via REPORT-

  If you are reusing the extraction rule on other sources/sourcetypes/hosts it is recommended that you use (2), for extractions that you know are not going to be reused use (1)

- Example

In **props.conf:**

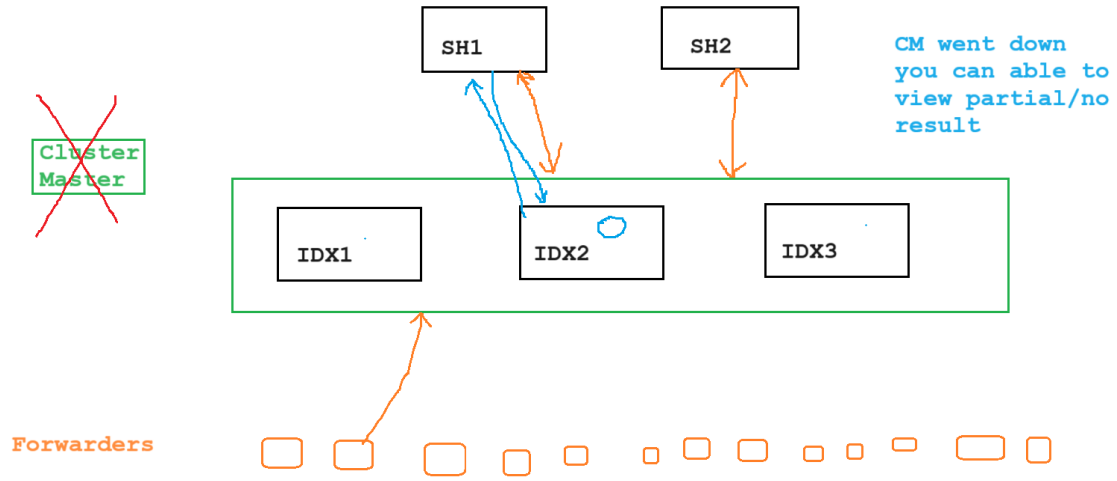
```
[example_sourcetype]
TRANSFORMS-nullQueue = setnull
```

In **transforms.conf**

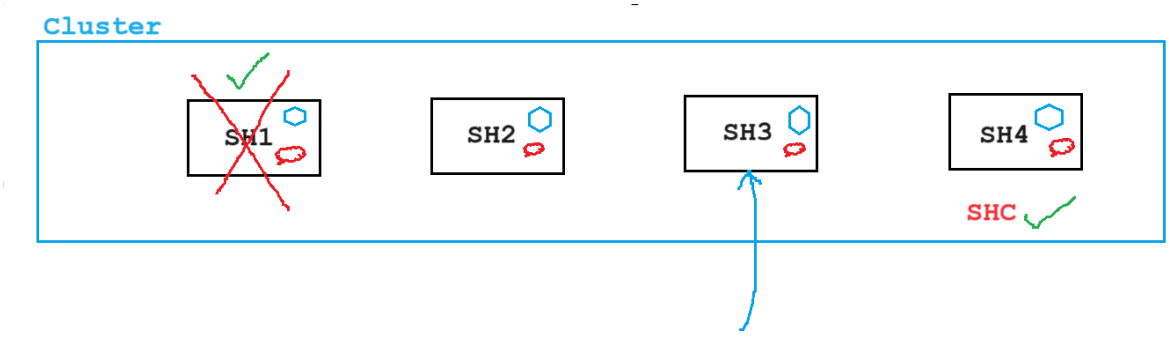
```
[setnull]
REGEX = .
DEST_KEY = queue
FORMAT = nullQueue
```

Scenario Based Questions

1. If the Cluster master goes down? How do the indexers behave?



2. how are alerts distributed among the remaining search heads if the search head captain goes down?



3. What is the recent issue you have faced? How did you solve it?