By OnlineInterviewQuestions.com

Apache Storm Interview Questions

Apache Storm is distributed a real-time system of computation and is free and open-source. Using storm is very simple and has the ability to function with large data including data processing. It has the advantage of being quick and easy to use, having the potential to process a hundred messages in one second per node. A good benefit of the storm is that it is scalable and can work with any programming language. It can automatically detect the faults and it is a reliable choice that ensures the data is executed at least one time if not more. It has three components which are Nimbus, Zookeeper, and Supervisor. Apache storm fetches the data from Apache Kafka and adds the required manipulations. Apache storm can be used for streamlining log files. It consists of 8 stream groupings that are built in. Here you will find some of the most important Storm Interview Questions that are asked in the Storm Interview. These are a few hand-picked questions that will help you in cracking the Storm Interview. Read the questions carefully and get yourself prepared for the interview with Storm.

Here you will find some of the most important **Storm Interview Questions** that are asked in the Storm Interview. These questions are a few hand-picked questions that will help you in cracking the Storm Interview. Read the questions carefully and get yourself prepared for the interview with Storm. you can also download here the **Apache Storm Interview Questions PDF** for the best preparation for the Apache Storm Interview.

Q1. What is Apache Storm?

Apache Storm is a distributed real-time system of computation which is an open and free source. Apache storm is used for processing Big Data Analytics with being reliable. It is simple to use Storm. It is able to work with any programming language which gives it an advantage over Hadoop batch processing. It can function with huge quantities of data which also includes data processing.

Q2. What are the benefits of Apache Storm?

There are a bunch of benefits of using Apache Storm for real-time processing.

- First, it is very simple to operate Storm. Its standard configurations are helpful in deploying and using it easily.
- Second, its functioning is really quick and has the potential to process 100 messages in one second for each node.
- Third, Apache Storm is a scalable choice as it is capable of working with any programming language which makes it easy to run over a large number of Machines.
- Fourth, it is capable of automatically detecting faults and then the workers get restarted.
- Fifth, it is one of the most significant benefits of Apache Storm that it is reliable. It ensures the data to be executed once and in some cases more than once.

Q3. What are the types of Nodes in Apache Storm?

Nodes are classified into two types, namely **Master Node** and the **Worker Node**. The master node is responsible for the execution of daemon nimbus. Its duty is to assign tasks to the machines and also of monitoring their performances. The Worker node is responsible for running the daemon known as Supervisor. Its duty is to assign tasks and operate the other worker node according to the requirement.

Q4. Tell us about the main components of Apache Storm.

There are three main components of Apache Storm. These are Nimbus, Zookeeper, and Supervisor.

Nimbus	This works as a job tracker and its job is to distribute the code across the cluster and to monitor computation. It is responsible for computing and executing the code.
Zookeeper	Zookeeper works as an intermediate body for communicating with the Storm Cluster.
Supervisor	Supervisor contact Nimbus with the corporation of Zookeeper and is responsible for executing the process according to the directions of Nimbus.

Q5. What do you understand by Bolts in Storm?

Bolt is used for processing the data in topologies. Bolts have the duty of filtration, aggregating and communicating with the database. It is also capable of doing the simple stream transformation. It has the requirement of multiple bolts when it does the complex stream transformation. This is due to the reason that for complex stream transformation, it often needs multiple steps, so multiple bolts.

Q6. In Storm, tell us about the stream groups that are built-in?

Storms consist of 8 stream grouping that is built-in. These are :

- Shuffle Grouping
- Field Grouping
- Partial Key Grouping
- All Grouping
- Global Grouping
- None Grouping
- Direct Grouping
- Local Grouping

- In **Shuffle Grouping,** the same quantity of tuples is distributed to every bolt in a random manner.
- The stream is divided with the help of fields that are specified in the grouping, in **Field Grouping**.
- **Partial Grouping** is similar to field grouping as the stream is divided with the help of the field that is specified in the grouping but is different in the sense that it gives resources that are better utilized.
- **All Grouping** should be used carefully since the stream is duplicated across the task of all bolts.

- **Local Grouping** is also known as Shuffle Grouping.
- **Direct Grouping** is a different type of grouping than others. In this, it is decided by the producer that the task will be received by which task.
- In **Global Grouping**, the complete stream goes to each of the tasks of bolts.
- In **None Grouping**, the grouping of stream need not have cared.

Q7. How is Apache Storm different from Apache Kafka?

There is very little difference between **Apache Storm** and **Apache Kafka**.

Apache Kafka is a distributed and strong messaging system that has the potential to handle big data and is responsible for passing a message from one terminal to the other.

Apache Storm is a system for processing messages in real-time. Data is fetched by Apache storm from Apache Kafka and adds the required manipulations.

Q8. What is Real-time Analytics in Apache Storm?

The usage of all the enterprise data that is available, as per the need is Real-Time Analytics. It involves vigorous analysis and also involves reporting which is based on the data that is put in a system. Less then 60 seconds or a minute is taken before the real-time use. Real-time analytics is also known by other terms such as real-time data analytics and real-time intelligence.

Q9. What is the importance of real-time Analytics?

The **real-time analytics** is very important and the need is growing significantly. It is observed that the application provides fast solutions with real-time analytics. It has a wide range including the retail sector, telecommunication, and the banking sector. Many frauds are filed in the sector of banking. One of the frauds that are very often heard is fraud transactions. Such frauds are happening on a regular basis and real-time analytics helps in detecting and identifying the frauds. It also has its application in the circle of social networks such as Twitter. Twitter brings the most trending topics to show it to the users. Real-time analytics has the role to play in it which attracts traffic and generates revenue.

Q10. Mention the components that are used for streaming the flow of data.

There are three components that are used to streamflow of data.

1. Bolt	2. Spout	3.Tuple
In a storm, the processing logic unit is represented by the bolt.	In a storm, the source of the data is represented by the spout.	In a storm, a tuple is the major data structure.

Bolt is used for doing all types of processing which includes filtering and interaction to the data. Bolts also have the duty of acknowledging the processing done by tuples.

The spout is used for reading data through the data sources. Broadly, spouts are classified into two categories. These are called Reliable and Unreliable.

Tuples are helpful with its helper methods which help in getting field values without the need for casting the result.

Q11. What is the purpose of using Zookeeper in Storm?

Zookeeper is used in the storm for coordination of cluster. It's not the duty of the zookeeper to pass the messages, which makes the load on it lighter. Zookeeper clusters of a single node are good and can do most tasks. But for deploying storm clusters that are large, it might need larger zookeeper clusters. Zookeeper should be run very carefully for the reason being the process will get excited if the zookeeper encounters an error case.

Q12. What are the three distinct layers of the Codebase in Storm?

Storm's codebase consists of three distinct layers.

- First, a storm can run with the use of any language. This is due to the thrift structures. The storm is compatible to go with all the languages.
- Second, all the interfaces of the storm are specified as the interfaces of Java. All the users have to go with API of Java which implies that always all storm's features are accessible through Java.
- Third, Clojure has a large implementation of the storm. Half of the storm is Clojure code and the other half is Java code, with Clojure code being more expressive.

Q13. When Cleanup method is called in Apache Storm?

When a bolt is going to be shut down and that the opened resources need to be cleaned, then Cleanup is called. Its not sure for a Cleanup method to be called on the cluster.

Q14. What is the full form of SSL?

SSL is an abbreviation to the term ' Secure Socket Layer '.

Q15. Why SSL is not included in Apache?

SSL is not included in Apache due to some significant reasons. Some governments don't allow import, export and do not give permission for using the encryption technology which is required by SSL data transport. If SSL would be included in Apache then it won't be available freely due to various legal matters. Some technology of SSL which is used for talking to current clients is under the patent of RSA Data Security and it does not allow its usage without a license.

Q16. How can Apache storm be used for streamlining log files?

The spout can be configured and then by an emitting each line as the log is read, for reading the log files. Then, the bolt should be provided with the output for analyzing.

Q17. Explain the Server Type directive in the server of Apache?

In the server of Apache, the server type directive determine whether Apache should keep all the things in one process for it shall spawn as a child process. In Apache 2.0, the server type directive is not found because it is not available in it. It is however available in Apache 1.3 for the compatibility of background with Apache of version based on UNIX.

Q18. Is a search engine contained in Apache?

Yes. A search engine is included in Apache. It can be searched from 'Search Title' in Apache.

Q19. What do you understand by 'Streams'?

From the context of computer science, the stream is a series of elements of data that gets available over time. It is the central abstraction of the storm.

Please Visit OnlineInterviewquestions.com to download more pdfs