

Top 13 Apache Storm Interview Questions & Answers

1) Explain what is Apache Storm? What are the components of Storm?

Apache storm is an open source distributed real-time computation system used for processing real time big data analytics. Unlike Hadoop batch processing, Apache storm does for real-time processing and can be used with any programming language.

Components of Apache Storm includes

- **Nimbus:** It works as a Hadoop’s Job Tracker. It distributes code across the cluster, uploads computation for execution, allocate workers across the cluster and monitors computation and reallocates workers as needed
- **Zookeeper:** It is used as a mediator for communication with the Storm Cluster
- **Supervisor:** Interacts with Nimbus through Zookeeper, depending on the signals received from the Nimbus, it executes the process.

2) Why Apache Storm is the first choice for Real Time Processing?

- **Easy to operate:** Operating storm is quiet easy
- **Real fast:** It can process 100 messages per second per node
- **Fault Tolerant:** It detects the fault automatically and re-starts the functional attributes
- **Reliable:** It guarantees that each unit of data will be executed at least once or exactly once
- **Scalable:** It runs across a cluster of machine

3) Explain how data is stream flow in Apache Storm?

In Apache storm, data is stream flow with three components **Spout, Bolt** and **Tuple**

- **Spout:** A spout is a source of data in Storm
- **Bolt:** A bolt processes these data’s
- **Tuple:** Data is passed as Tuple

4) Mention what is the difference between Apache Hbase and Storm?

Apache Storm

- It provides data processing in real-time
- It processes the data but not store
- You will streamline your data where data is processed in real time, so that alerts and actions can be raised if needed

Apache Hbase

- It offers you low-latency reads of processed data for querying later
- It stores the data but does not store
- _____

5) Explain how you can streamline log files using Apache storm?

To read from the log files you can configure your **spout** and emit per line as it read the log. The

output then can be assign to a bolt for analyzing.



Apache Storm

6) Explain what streams is and stream grouping in Apache storm?

In Apache Storm, stream is referred as a group or unbounded sequence of Tuples while stream grouping determines how stream should be partitioned among the bolt's tasks.

7) List out different stream grouping in Apache storm?

- Shuffle grouping
- Fields grouping
- Global grouping
- All grouping
- None grouping
- Direct grouping
- Local grouping

8) Mention how storm application can be beneficial in financial services?

In financial services, Storm can be helpful in preventing

- Securities fraud
- Order routing
- Pricing
- Compliance Violations

9) Explain what is `Topology_Message_Timeout_secs` in Apache Storm?

The maximum amount of time allotted to the topology to fully process a message released by a spout. If the message is not acknowledged in given time frame, Apache storm will fail the message on the spout.

10) Explain how message is fully processed in Apache Storm?

By calling the `nextTuple` procedure or method on the Spout, Storm requests a tuple from the Spout. The Spout avails the `SpoutOutputCollector` given in the `open` method to discharge a tuple to one of its output streams. While discharging a tuple, the `Spout` allocates a "message id" that will be used to recognize the tuple later.

After that, the tuple gets sent to consuming bolts, and storm takes charge of tracking the tree of messages that is produced. If the storm is confident that a tuple is processed thoroughly, then it can call the `ack` procedure on the originating `Spout` task with the message id that the Spout has given to the Storm.

11) Explain how to write the Output into a file using Storm?

In Spout, when you are reading file, make **FileReader object** in **Open()** method, as such that time it initializes the reader object for worker node. And use that object in nextTuple() method.

12) Mention what is the difference between Apache Kafka and Apache Storm?

- **Apache Kafka:** It is a distributed and robust messaging system that can handle huge amount of data and allows passage of messages from one end-point to another.
- **Apache Storm:** It is a real time message processing system, and you can edit or manipulate data in real time. Apache storm pulls the data from Kafka and applies some required manipulation.

13) Explain when using field grouping in storm, is there any time-out or limit to known field values?

Field grouping in storm uses a mod hash function to decide which task to send a tuple, ensuring which task will be processed in the correct order. For that, you don't require any cache. So, there is no time-out or limit to known field values.

Refer our [Apache Tutorials](#) for an extra edge in your interview.

[Guru99](#) Provides [FREE ONLINE TUTORIAL](#) on Various courses like

Java	MIS	MongoDB	BigData	Cassandra
Web Services	SQLite	JSP	Informatica	Accounting
SAP Training	Python	Excel	ASP Net	HBase
Project Management	Test Management	Business Analyst	Ethical Hacking	PMP
Live Project	SoapUI	Photoshop	Manual Testing	Mobile Testing
Data Warehouse	R Tutorial	Tableau	DevOps	AWS
Jenkins	Agile Testing	RPA	JUnit	Software Engineering
Selenium	CCNA	AngularJS	NodeJS	PLSQL

**Stay updated with new
courses at Guru99
Join our Newsletter**