
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](#) | [Testing](#) | [Selenium](#) | [CCNA](#) | [NodeJS](#)

[TensorFlow](#) | [Data Warehouse](#) | [R Programming](#) | [Live Projects](#) | [DevOps](#)

Top 30 Hive Interview Questions & Answers

1) Explain what is Hive?

Hive is an ETL and Data warehousing tool developed on top of Hadoop Distributed File System (HDFS). It is a data warehouse framework for querying and analysis of data that is stored in HDFS. Hive is an open-source-software that lets programmers analyze large data sets on Hadoop.

2) When to use Hive?

- Hive is useful when making data warehouse applications
- When you are dealing with static data instead of dynamic data
- When application is on high latency (high response time)
- When a large data set is maintained
- When we are using queries instead of scripting

3) Mention what are the different modes of Hive?

Depending on the size of data nodes in Hadoop, Hive can operate in two modes.

These modes are,

- Local mode
- Map reduce mode

4) Mention when to use Map reduce mode?

Map reduce mode is used when,

- It will perform on large amount of data sets and query going to execute in a parallel way
- Hadoop has multiple data nodes, and data is distributed across different node we use

Hive in this mode

- Processing large data sets with better performance needs to be achieved

5) Mention key components of Hive Architecture?

Key components of Hive Architecture includes,

- User Interface
- Compiler
- Metastore
- Driver
- Execute Engine



6) Mention what are the different types of tables available in Hive?

There are two types of tables available in Hive.

- **Managed table:** In managed table, both the data and schema are under control of Hive
- **External table:** In the external table, only the schema is under the control of Hive.

7) Explain what is Metastore in Hive?

Metastore is a central repository in Hive. It is used for storing schema information or metadata in the external database.

8) Mention what Hive is composed of ?

Hive consists of 3 main parts,

1. Hive Clients
2. Hive Services
3. Hive Storage and Computing

9) Mention what are the type of database does Hive support ?

For single user metadata storage, Hive uses derby database and for multiple user Metadata or shared Metadata case Hive uses MYSQL.

10) Mention Hive default read and write classes?

Hive default read and write classes are

1. TextInputFormat/HiveIgnoreKeyTextOutputFormat
2. SequenceFileInputFormat/SequenceFileOutputFormat

11) Mention what are the different modes of Hive?

Different modes of Hive depends on the size of data nodes in Hadoop.

These modes are,

- Local mode
- Map reduce mode

12) Why is Hive not suitable for OLTP systems?

Hive is not suitable for OLTP systems because it does not provide insert and update function at the row level.

13) Mention what is the difference between Hbase and Hive?

Difference between Hbase and Hive is,

- Hive enables most of the SQL queries, but HBase does not allow SQL queries
- Hive does not support record level insert, update, and delete operations on table
- Hive is a data warehouse framework whereas HBase is NoSQL database
- Hive run on the top of MapReduce, HBase runs on the top of HDFS

14) Explain what is a Hive variable? What for we use it?

Hive variable is created in the Hive environment that can be referenced by Hive scripts. It is

used to pass some values to the hive queries when the query starts executing.

15) Mention what is ObjectInspector functionality in Hive?

ObjectInspector functionality in Hive is used to analyze the internal structure of the columns, rows, and complex objects. It allows to access the internal fields inside the objects.

16) Mention what is (HS2) HiveServer2?

It is a server interface that performs following functions.

- It allows remote clients to execute queries against Hive
- Retrieve the results of mentioned queries

Some advanced features Based on Thrift RPC in its latest version include

- Multi-client concurrency
- Authentication

17) Mention what Hive query processor does?

Hive query processor convert graph of MapReduce jobs with the execution time framework. So that the jobs can be executed in the order of dependencies.

18) Mention what are the components of a Hive query processor?

The components of a Hive query processor include,

- Logical Plan Generation
- Physical Plan Generation
- Execution Engine
- Operators
- UDF's and UDAF's
- Optimizer
- Parser
- Semantic Analyzer
- Type Checking

19) Mention what is Partitions in Hive?

Hive organizes tables into partitions.

- It is one of the ways of dividing tables into different parts based on partition keys.
- Partition is helpful when the table has one or more Partition keys.
- Partition keys are basic elements for determining how the data is stored in the table.

20) Mention when to choose “Internal Table” and “External Table” in Hive?

In Hive you can choose internal table,

- If the processing data available in local file system
- If we want Hive to manage the complete lifecycle of data including the deletion

You can choose External table,

- If processing data available in HDFS
- Useful when the files are being used outside of Hive

21) Mention if we can name view same as the name of a Hive table?

No. The name of a view must be unique compared to all other tables and as views present in the same database.

22) Mention what are views in Hive?

In Hive, Views are Similar to tables. They are generated based on the requirements.

- We can save any result set data as a view in Hive
- Usage is similar to as views used in SQL
- All type of DML operations can be performed on a view

23) Explain how Hive Deserialize and serialize the data?

Usually, while read/write the data, the user first communicate with inputformat. Then it connects with Record reader to read/write record. To serialize the data, the data goes to row. Here deserialized custom serde use object inspector to deserialize the data in fields.

24) What is Buckets in Hive?

- The data present in the partitions can be divided further into Buckets
- The division is performed based on Hash of particular columns that is selected in the table.

25) In Hive, how can you enable buckets?

In Hive, you can enable buckets by using the following command,

```
set.hive.enforce.bucketing=true;
```

26) In Hive, can you overwrite Hadoop MapReduce configuration in Hive?

Yes, you can overwrite Hadoop MapReduce configuration in Hive.

27) Explain how can you change a column data type in Hive?

You can change a column data type in Hive by using command,

```
ALTER TABLE table_name CHANGE column_name column_name new_datatype;
```

28) Mention what is the difference between order by and sort by in Hive?

- SORT BY will sort the data within each reducer. You can use any number of reducers for SORT BY operation.
- ORDER BY will sort all of the data together, which has to pass through one reducer. Thus, ORDER BY in hive uses a single

29) Explain when to use explode in Hive?

Hadoop developers sometimes take an array as input and convert into a separate table row. To convert complex data types into desired table formats, Hive use explode.

30) Mention how can you stop a partition from being queried?

You can stop a partition from being queried by using the ENABLE OFFLINE clause with ALTER TABLE statement.

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