



[www.HadoopExam.com](http://www.HadoopExam.com)

# Hive Interview Questions

---

[www.HadoopExam.com](http://www.HadoopExam.com)

Please visit [www.HadoopExam.com](http://www.HadoopExam.com) for various resources for BigData/Hadoop/Cassandra/MongoDB/Node.js/Scala etc.

# Professional Training for BigData and Apache Hadoop

**While watching we promise you will say WOW! At least once**

Accelerate Your and Organization Hadoop Education

Apache Hadoop is increasingly being adopted in a wide range of industries and as a result, Hadoop expertise is more valuable than ever for you and your organization. Using Hadoop, organizations can consolidate and analyze data in ways never before possible. Businesses can capture, manage and process information that they used to throw away. You can leverage years of Hadoop experience with training from [www.HadoopExam.com](http://www.HadoopExam.com)

The course/training is designed specifically for CEO, CTO to Managers, Software Architect to an Individual Developers and Testers to enhance their skills in BigData world. You will learn when the use of Hadoop is appropriate, what problems Hadoop addresses, how Hadoop fits into your existing environment, and what you need to know about deploying Hadoop.

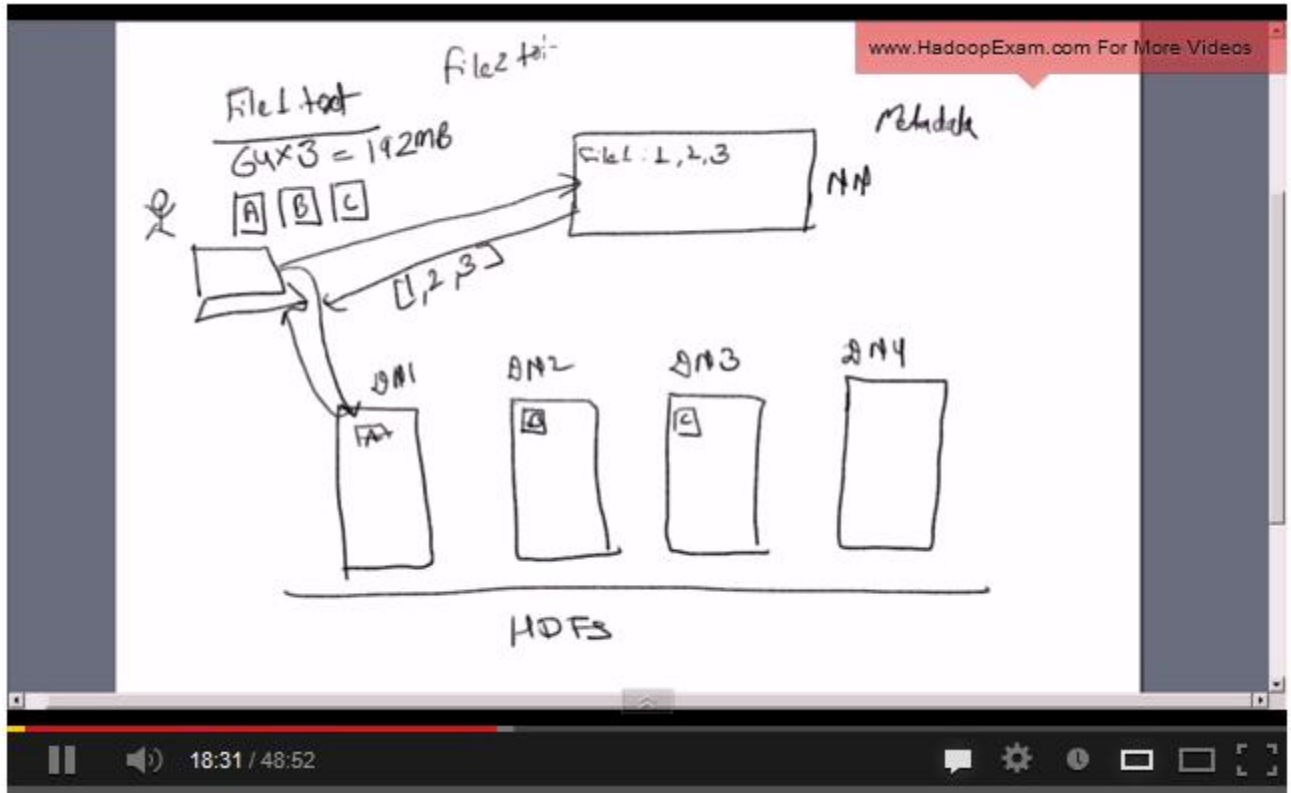
Learn the basics of the Hadoop Distributed File System (HDFS) and MapReduce framework and how to write programs against its API, as well as discuss design techniques for larger workflows. This training also covers advanced skills for debugging MapReduce programs and optimizing their performance, plus introduces participants to related projects in Distribution for Hadoop such as Hive, Pig, and Oozie. After completing the training, attendees can leverage our Hadoop Certification Exam Simulator for Developer as well as Administrator to clear the Hadoop Certification. Since launch 467+ attendees already cleared the exam with the help of our simulator.

**Industry Where Hadoop is Being Used:** Energy & Utilities, Financial Services ,Government, Healthcare & Life Sciences, Media & Entertainment, Retail, eCommerce Consumer, Product, Technology, Telecommunications, **start ups**(They are trying their each & every resource should have Hadoop knowledge)

## Features of HadoopExam Learning Resources Training

1. **Very Convenient (24/7 access)** : This training will be available online for three months and any time you can access as per your comfort, so no need to wait as soon as access provided to you, you can start watching the trainings.
2. **No to PPT at ALL, As effective as classroom trainings** : We have taken the advanced step for providing online training and avoided using PPTs and created recordings as white board sessions.
3. **No travel, No Leaves & No Weekend classes**: This training is created keeping in mind that professionals cannot take leave so easily to attend the trainings. Because these sessions are going to be used by CEO, CTO, Solution Architect, Managers, Individual Developer and Tester. So no to travel, no to leave and No to weekend classes.
4. **Immediate and Cost-effective (Equal to price of a Book)**: It is so cost effective that it cost you equal to buying a book on BigData and Hadoop.
5. **In Just two Weeks 3200+ views on Youtube** : We provide Module 2 as free of cost for the Demo and after launch in just two weeks we got **3200+** views and all positive appreciation you can see in below testimonial sections.

**Module 2 is freely available for Demo Watch Now (While watching you definitely says WOW!)**



### 1. What is Hive Metastore?

Ans : Hive metastore is a database that stores metadata about your Hive tables (eg. table name, column names and types, table location, storage handler being used, number of buckets in the table, sorting columns if any, partition columns if any, etc.). When you create a table, this metastore gets updated with the information related to the new table which gets queried when you issue queries on that table.

### 2. Wherever (Different Directory) I run hive query, it creates new metastore\_db, please explain the reason for it?

Ans: Whenever you run the hive in embedded mode, it creates the local metastore. And before creating the metastore it looks whether metastore already exist or not. This property is defined in configuration file hive-site.xml. Property is "javax.jdo.option.ConnectionURL" with default value "jdbc:derby:;databaseName=metastore\_db;create=true". So to change the behavior change the location to absolute path, so metastore will be used from that location.

### 3. Is it possible to use same metastore by multiple users, in case of embedded hive?

Ans: No, it is not possible to use metastore in sharing mode. It is recommended to use standalone "real" database like MySQL or PostgreSQL.

### 4. Is multiline comment supported in Hive Script ?

Ans: No.

### 5. If you run hive as a server, what are the available mechanism for connecting it from

## **application?**

Ans: There are following ways by which you can connect with the Hive Server:

1. Thrift Client: Using thrift you can call hive commands from a various programming languages e.g. C++, Java, PHP, Python and Ruby.
2. JDBC Driver : It supports the Type 4 (pure Java) JDBC Driver
3. ODBC Driver: It supports ODBC protocol.

## **6. What is SerDe in Apache Hive ?**

Ans : A SerDe is a short name for a Serializer Deserializer. Hive uses SerDe (and FileFormat) to read and write data from tables. An important concept behind Hive is that it DOES NOT own the Hadoop File System (HDFS) format that data is stored in. Users are able to write files to HDFS with whatever tools/mechanism takes their fancy("CREATE EXTERNAL TABLE" or "LOAD DATA INPATH," ) and use Hive to correctly "parse" that file format in a way that can be used by Hive. A SerDe is a powerful (and customizable) mechanism that Hive uses to "parse" data stored in HDFS to be used by Hive.

## **7. Which classes are used by the Hive to Read and Write HDFS Files**

Ans : Following classes are used by Hive to read and write HDFS files

- TextInputFormat/HiveIgnoreKeyTextOutputFormat: These 2 classes read/write data in plain text file format.
- SequenceFileInputFormat/SequenceFileOutputFormat: These 2 classes read/write data in hadoop SequenceFile format.

## **8. Give examples of the SerDe classes whihc hive uses to Serializa and Deserilize data ?**

Ans : Hive currently use these SerDe classes to serialize and deserialize data:

- MetadataTypedColumnsetSerDe: This SerDe is used to read/write delimited records like CSV, tab-separated control-A separated records (quote is not supported yet.)
- ThriftSerDe: This SerDe is used to read/write thrift serialized objects. The class file for the Thrift object must be loaded first.
- DynamicSerDe: This SerDe also read/write thrift serialized objects, but it understands thrift DDL so the schema of the object can be provided at runtime. Also it supports a lot of different protocols, including TBinaryProtocol, TJSONProtocol, TCTLSeparatedProtocol (which writes data in delimited records).

## **9. How do you write your own custom SerDe ?**

Ans :

- In most cases, users want to write a Deserializer instead of a SerDe, because users just want to read their own data format instead of writing to it.
- For example, the RegexDeserializer will deserialize the data using the configuration parameter 'regex', and possibly a list of column names
- If your SerDe supports DDL (basically, SerDe with parameterized columns and column types), you probably want to implement a Protocol based on DynamicSerDe, instead of writing a SerDe from scratch. The reason is that the framework passes DDL to SerDe through

"thrift DDL" format, and it's non-trivial to write a "thrift DDL" parser.

#### **10. What is ObjectInspector functionality ?**

Ans : Hive uses ObjectInspector to analyze the internal structure of the row object and also the structure of the individual columns.

ObjectInspector provides a uniform way to access complex objects that can be stored in multiple formats in the memory, including:

- Instance of a Java class (Thrift or native Java)
- A standard Java object (we use java.util.List to represent Struct and Array, and use java.util.Map to represent Map)
- A lazily-initialized object (For example, a Struct of string fields stored in a single Java string object with starting offset for each field)

A complex object can be represented by a pair of ObjectInspector and Java Object. The ObjectInspector not only tells us the structure of the Object, but also gives us ways to access the internal fields inside the Object.

#### **11. What is the functionality of Query Processor in Apached Hive ?**

Ans: This component implements the processing framework for converting SQL to a graph of map/reduce jobs and the execution time framework to run those jobs in the order of dependencies.



## Hadoop Certification Exam Simulator (Developer+Admin ) + Study Material

- Contains 4 practice Question Paper
- 200/238 (Developer/Admin) realistic Hadoop Certification Questions
- All Questions are on latest Pattern
- End time 15 Page revision notes for Developer (Save lot of time)
- Download from [www.HadoopExam.com](http://www.HadoopExam.com)

**Note: There is 50% talent gap in BigData domain, get Hadoop certification with the Hadoop Learning Resources Hadoop Exam Simulator.**

PappuPass Learning Resources

Exam PracticeExam Help

1 2 3 4 01:29:17

Question 15 :  
Which Daemon distributes individual task to machines

1. TaskTracker  
 2. JobTracker  
 3. MasterTracker  
 4. NameNode

Next Previous Finish Go To Question No. 1