# Hadoop Hive Interview Questions

# Hive Interview Questions

### **What is Hive ?**

Hive is a data warehouse software which is used for facilitates querying and managing large data sets residing in distributed storage.Hive language almost look like SQL language called HiveQL.Hive also allows traditional map reduce programs to customize mappers and reducers when it is inconvenient or inefficient to execute the logic in HiveQL (User Defined Functions UDFS)

**What is Hive Metastore ?**

Hive Meta store is a database that stores metadata of your hive tables like table name,columnname,datatypes,tablelocation,number of buckets in the table etc.

Hadoop Hive Interview Questions

**What is Hive Present Version ?**

hive-0.13.1

**What is the stable version of Hive ?**

hive-0.12.0

**Hive new version supported Hadoop Versions ?**

This release works with Hadoop 0.20.x, 0.23.x.y, 1.x.y, 2.x.y

**Where we have to set the Hive Installation Path ?**

we can set hive path in ~/.bashrc file or hadoop-env.sh file

**~/.bashrc file or hadoop-env.sh file which one is better to set the path ?**

~/.bashrc.sh

**Why ~/.bashrc.sh is better then hadoop-env.sh ?**

~/.bashrc.sh starts the work when system login but hadoop-env.sh starts the work only when hadoop starts in system

**What is Hive Installation Path ?**

export HIVE\_HOME=/home/hadoop/work/hive-x.y.z

export PATH=$PATH:$HIVE\_HOME/bin

**How to Install Hive ?**

check the answer above menu-hive installation tab

**Which companies are mostly using Hive ?**

Facebook,Netflix

**Which company initially developed Hive ?**

Facebook

**How Facebook Uses Hadoop,Hive and Hbase ?**

Facebook data stored on HDFS,everyday millions of photos uploaded into facebook with the help of Hadoop

1. Facebook Messages,Likes and statues updates running on top of Hbase
2. Hive to generate reports for third-party developers and advertisers who need to track the success of their applications or campaigns.

**What is Apache Hcatalog ?**

HCatalog is built on top of the Hive metastore and incorporates Hive’s DDL.ApacheHcatalog is a table and data management layer for hadoop,we can process the data on Hcatalog by using APachepig,ApacheMapreduce and Apache Hive.There is no need to worry in Hcatalog where data is stored and which format of data generated.HCatalog displays data from RCFile format, text files, or sequence files in a tabular view. It also provides REST APIs so that external systems can access these tables’ metadata.

**What is the work of Hive/Hcatalog ?**

Hive/HCatalog also enables sharing of data structure with external systems including traditional data management tools.

**What is WebHCatServer ?**

The WebHcatServer provides a REST – like web API for Hcatalog.Applications make HTTP requests to run Pig, Hive, and HCatalog DDL from within applications.

**What is SerDe in Apache Hive ?**

SerDe full form is  SerializerDeserializer.Hive uses  SerializerDeserializer to read and write the data from hive table.Theimportent one behind hive is hive does not have own Hadoop distributed file system(HDFS) format that data is stored in.Users have to write store the hive data on HDFS by using (“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.Hive uses to “parse” data stored in HDFS to be used by Hive

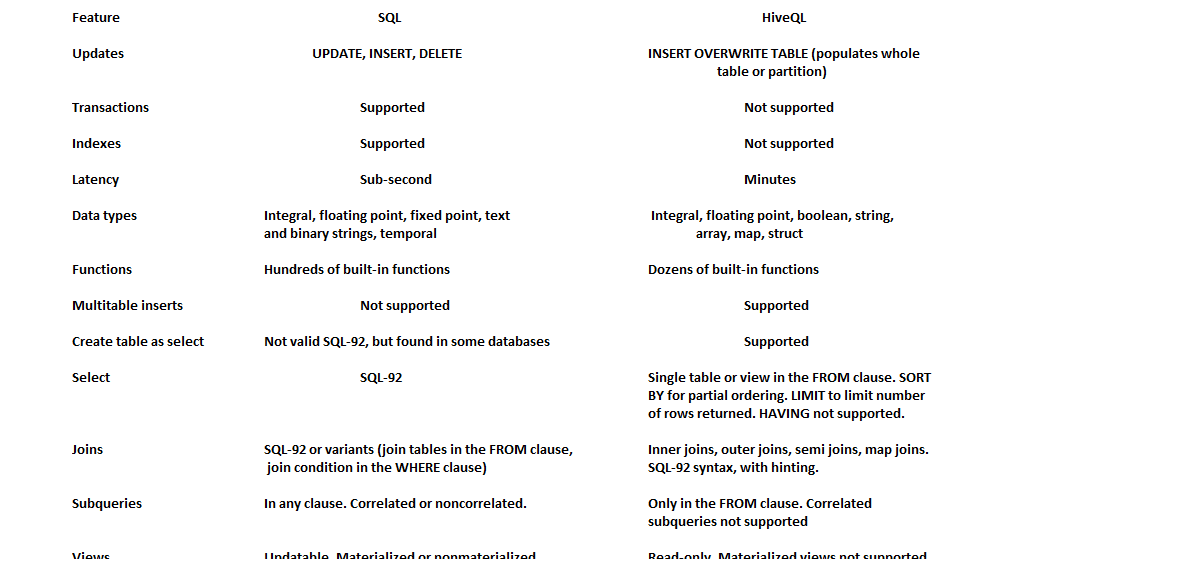
**Is it possible to use same metastore by multiple users, in case of embedded hive?**

No, it is not possible to use metastore in sharing mode. It is recommended to use standalone “real” database like MySQL or PostGresSQL.

**Is multiline comment supported in Hive Script ?**

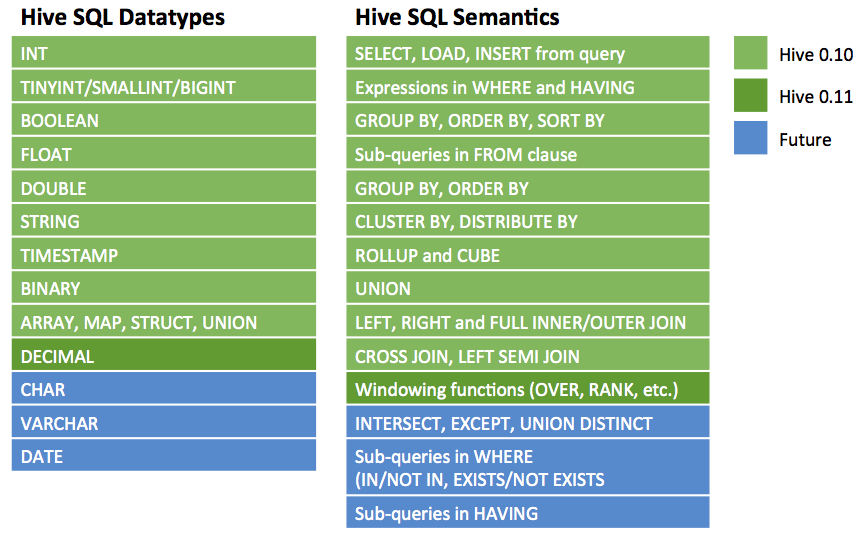
NO

**Difference between SQL and HiveQL ?**

[](http://www.hadooptpoint.com/wp-content/uploads/2014/11/Difference-between-SQL-HiveQL.png)

Hive Interview QuestionsHadoop Hive Interview Questions

**Hive Data types ?**

[](http://www.hadooptpoint.com/wp-content/uploads/2014/11/Hive-Datatypes.png)Hadoop Hive Interview Questions

Hive Interview Questions Part-2

Hive Interview Questions

**What is Hive Shell ?**

The shell is the primary way that we will interact with hive by using hiveqlcommands.In other words shell is nothing but a prompt which is used to enter the Hiveql commands for interacting the Hive shell

**How we can enter into Hive shell from normal terminal ?**

Hive Interview Questions

just by entering the hive command like ‘hive’

**How we can check the hive shell is working or not ?**

After entered into hive shell just enter another Hiveql command like ‘show databases;’

**Is it necessary to add semi colon (;) at end of the Hiveqlcommands ?**

Yes,We have to add semicolon (;) at end of the Hiveql every  command

**What is the default database of Hive ?**

Derby database

**Which Databases are supported for Hive ?**

Derby database is default one,mysql,postgresql database

**What is the meaning of Hiveql ?**

Hiveqlmeaing is Hive + SQL is called Hiveql,thehiveql commands ae almost same as SQL commands.

**What is Hive Table ?**

A Hive table is logically made up of the data being stored and the associated metadata describing the layout of the data int the table.The data typically associated resides in HDFS,although it may reside in any hadoop file system,including the local file system or s3.Hive stores the metadata in a relational databases and not in HDFS.

**How many types we can create tables in hive ?**

Two types

**What are that two types of tables ?**

i) Managed tables

ii) External tables

**Where Hive stores Metadata ?**

Hive stores metadata in relational databases and not in HDFS.

**What is Managed Table in Hive ?**

Managed system is nothing but when you create the table in hive,by default hive will manage the data,which means that hive moves the data into its warehouse directory.

**What is External Table in Hive ?**

External table is nothing but when you create the table in hive,the data stored at an existing location outside the warehouse directory

**What is the main differences between an external and a managed table in Hive ?**

One of the main differences between an external and a managed table in Hive is that when an external table is dropped, the data associated with it  doesn’t get deleted, only the metadata (number of columns, type of columns, terminators,etc.) gets dropped from the Hive meta store. When a managed table gets dropped, both the metadata and data get dropped.

**Which type  tables is always preferred making tables in hive ?**

External Table,because external table data stored in HDFS

**What is the use of SET command in Hive ?**

SET command is useful for changing  Hive or Map Reduce job settings for a particular query.

**Example for SET command ?**

hive> SET hive.enforce.bucketing;

hive.enforce.bucketing=true

The above command shows value of that property.

**Property setting files in HIVE ?**

i) The Hive SET command

ii) The command line -hiveconf option

iii) hive-site.xml

iv) hive-default.xml

v) hadoop-site.xml (or,equivalently , core-site.xml,hdfs-site.xml, and mapred-site.xml)

vi) hadoop-default.xml (or,equivalently, core-default.xml,hdfs-default.xml and mapred-default.xml)

**What is the command for Hive Services ?**

hive –service help to get a list of available service names

**what are the Hive services ?**

i)cli

ii) hiveserver

iii) hwi

iv) jar

v)metastore

vi)hive client

vii)Thrift client

viii)JDBC Driver

ix)ODBC Driver

**What is CLI service ?**

The command Line Interface to Hive (Shell).This is the default service

**What is HiveServer ?**

Runs hive as a server exposing a thrift service,enabling access from a range of clients written in different languages.Applications using the thrift,JDBC,ODBC connectors need to run hive server to communicate with Hive.

**What is hwi service in hive ?**

The Hive Web Interface

**what is Jar service in hive ?**

The hive jar is equivalent to hadoopjar,a convenient way to run java applications that includes both Hadoop and Hive classes on the class path.

**what is meta store service in hive ?**

It is possible to run the meta store as a standalone(remote) process.

**what is hive clients ?**

If you run Hive as a server (hive –server hiveserver),then there are a number of different mechanisms for connecting to it from applications.The relationship between hive client and Hive services.

**what is Thrift client ?**

The main aim of Hive thrift client is easy to run hive hive commands from other programming languages.Thrift client supported languages are c++,java,php,python, and ruby.They can be found in the src/service/src sub directory in the hive distribution

**What is JDBC Driver service ?**

Hive provides type 4 pure java server.the defined in class org.apache.hadoop.hive.jdbc.HiveDriver.When it configured with JDBC URI of the form of  **jdbc:hive://host:port/dbname**

**What is ODBC Driver service ?**

That supports ODBC protocol to connect to hive.(Like the JDBC Driver,the ODBC driver Thrift to communicate with the hive server)