**Hadoop Questions and Answers – Introduction to Hive**

This set of Hadoop Multiple Choice Questions & Answers (MCQs) focuses on “Introduction to Hive”.

1. Which of the following command sets the value of a particular configuration variable (key) ?  
a) set -v  
**b) set <key>=<value>**  
c) set  
d) reset  
**View Answer**

**Answer: b  
Explanation: If you misspell the variable name, the CLI will not show an error.**

2. Point out the correct statement :  
**a) Hive Commands are non-SQL statement such as setting a property or adding a resource**  
b) Set -v prints a list of configuration variables that are overridden by the user or Hive  
c) Set sets a list of variables that are overridden by the user or Hive  
d) None of the mentioned  
**View Answer**

**Answer: a  
Explanation: Commands can be used in HiveQL scripts or directly in the CLI or Beeline.**

3. Which of the following operator executes a shell command from the Hive shell ?  
a) |  
**b) !**  
c) ^  
d) +  
**View Answer**

**Answer: b  
Explanation: Excalamation operator is for execution of command.**

4. Which of the following will remove the resource(s) from the distributed cache ?  
a) delete FILE[S] <filepath>\*  
b) delete JAR[S] <filepath>\*  
c) delete ARCHIVE[S] <filepath>\*  
d) all of the mentioned  
**View Answer**

**Answer: d  
Explanation: Delete command is used to remove existing resource.**

5. Point out the wrong statement :  
a) source FILE executes a script file inside the CLI  
**b) bfs executes a dfs command from the Hive shell**c) hive is Query lanaguage similar to SQL  
d) none of the mentioned

**Answer: b  
Explanation: dfs executes a dfs command View Answer from the Hive shell.**

6. \_\_\_\_\_\_\_\_\_ is a shell utility which can be used to run Hive queries in either interactive or batch mode.  
a) $HIVE/bin/hive  
b) $HIVE\_HOME/hive  
**c) $HIVE\_HOME/bin/hive**  
d) All of the mentioned  
**View Answer**

**Answer: c  
Explanation: Various types of command line operations are available in the shell utilty.**

7. Which of the following is a command line option ?  
**a) -d,–define <key=value>**  
b) -e,–define <key=value>  
c) -f,–define <key=value>  
d) None of the mentioned  
View Answer

Answer: a  
Explanation: Variable substitution to apply to hive commands. e.g. -d A=B or –define A=B.

8. Which is the additional command line option is available in Hive 0.10.0 ?  
**a) –database <dbname>**  
b) –db <dbname>  
c) –dbase <<dbname> d) All of the mentioned

[expand title="View Answer"]Answer: a Explanation: Database is specified which is to be used.[/expand]

9. The CLI when invoked without the -i option will attempt to load $HIVE\_HOME/bin/.hiverc

and $HOME/.hiverc as \_\_\_\_\_\_\_ files

a) processing

b) termination

**c) initialization**

d) none of the mentioned

[expand title="View Answer"]

Answer: c Explanation: Hiverc file is loaded as per options selected.[/expand]

10. When $HIVE\_HOME/bin/hive is run without either the -e or -f option, it enters \_\_\_\_\_\_\_ mode.

a) Batch

**b) Interactive shell**

c) Multiple

d) None of the mentioned

[expand title="View Answer"]Answer: b Explanation: Use ";" (semicolon) to terminate commands for multiple options available.[/expand]

**Hadoop Questions and Answers – HiveQL-1**

**This set of MCQs focuses on “Hive”.**

1. Hive uses \_\_\_\_\_\_\_\_\_ for logging.  
a) logj4  
b) log4l  
c) log4i  
**d) log4j**  
View Answer

Answer: d  
Explanation: By default Hive will use hive-log4j.default in the conf/ directory of the Hive installation.

2. Point out the correct statement :  
a) list FILE[S] <filepath>\* executes a Hive query and prints results to standard output  
**b) <query string> executes a Hive query and prints results to standard output**c) <query> executes a Hive query and prints results to standard output  
d) All of the mentioned  
**View Answer**

**Answer: b  
Explanation: list FILE[S] <filepath>\* checks whether the given resources are already added to the distributed cache or not. See Hive Resources below for more information.**

3. What does the hive.rrot.logger specifies in the following statement ?

$HIVE\_HOME/bin/hive --hiveconf hive.root.logger=INFO,console

**a) Log level**  
b) Log modes  
c) Log source  
d) All of the mentioned  
View Answer

Answer: a  
Explanation: hive.root.logger specifies the logging level as well as the log destination. Specifying console as the target sends the logs to the standard error.

4. HiveServer2 introduced in Hive 0.11 has a new CLI called \_\_\_\_\_\_\_\_\_\_  
**a) BeeLine**  
b) SqlLine  
c) HiveLine  
d) CLilLine  
View Answer

**Answer: a  
Explanation: Beeline is a JDBC client based on SQLLine.**

5. Point out the wrong statement :  
**a) There are four namespaces for variables in Hive**  
b) Custom variables can be created in a separate namespace with the define  
c) Custom variables can also be created in a separate namespace with hivevar  
d) None of the mentioned  
**View Answer**

**Answer: a  
Explanation: Three namespaces for variables are hiveconf, system, and env.**

6. HCatalog is installed with Hive, starting with Hive release  
a) 0.10.0  
b) 0.9.0  
**c) 0.11.0**  
d) 0.12.0  
View Answer

Answer: c  
Explanation: hcat commands can be issued as hive commands, and vice versa.

7. hiveconf variables are set as normal by using the following statement :  
a) set -v x=myvalue  
b) set x=myvalue  
c) reset x=myvalue  
**d) none of the mentioned**  
View Answer

Answer: d  
Explanation: The hiveconf variables are set as normalby set x=myvalue.

8. Variable Substitution is disabled by using :  
**a) set hive.variable.substitute=false;**  
b) set hive.variable.substitutevalues=false;  
c) set hive.variable.substitute=true;  
d) all of the mentioned  
View Answer

**Answer: a  
Explanation: Variable substitution is on by default (hive.variable.substitute=true)**

9. \_\_\_\_\_\_\_ supports a new command shell Beeline that works with HiveServer2.  
**a) HiveServer2**  
b) HiveServer3  
c) HiveServer4  
d) None of the mentioned  
View Answer

**Answer: a  
Explanation: The Beeline shell works in both embedded mode as well as remote mode.**

10. In \_\_\_\_\_\_ mode HiveServer2 only accepts valid Thrift calls.  
**a) Remote**  
b) HTTP  
c) Embedded  
d) Interactive  
View Answer

Answer: a  
Explanation: In HTTP mode, the message body contains Thrift payloads.

**Hadoop Questions and Answers – Querying data with HiveQL-2**

**This set of Hadoop MCQs focuses on “Querying data with HiveQL”.**

1. Serialization of string columns uses a \_\_\_\_\_\_\_\_ to form unique column values.  
a) Footer  
b) STRIPES  
**c) Dictionary**  
d) Index  
**View Answer**

**Answer: c  
Explanation: The dictionary is sorted to speed up predicate filtering and improve compression ratios.**

2. Point out the correct statement :  
a) The Avro file dump utility analyzes ORC files  
**b) Streams are compressed using a codec, which is specified as a table property for all streams in that table**  
c) The ODC file dump utility analyzes ORC files  
d) All of the mentioned  
View Answer

Answer: b  
Explanation: The codec can be Snappy, Zlib, or none.

3. \_\_\_\_\_\_\_ is a lossless data compression library that favors speed over compression ratio.  
**a) LOZ**  
b) LZO  
c) OLZ  
d) All of the mentioned  
View Answer

**Answer: a  
Explanation: lzo and lzop need to be installed on every node in the Hadoop cluster.**

4. Which of the following will prefix the query string with parameters:  
**a) SET hive.exec.compress.output=false**  
b) SET hive.compress.output=false  
c) SET hive.exec.compress.output=true  
d) All of the mentioned  
**View Answer**

**Answer: a  
Explanation: Use lzop command utility or your custom Java to generate .lzo.index for the .lzo files.**

5. Point out the wrong statement :  
a) TIMESTAMP is Only available starting with Hive 0.10.0  
**b) DECIMAL introduced in Hive 0.11.0 with a precision of 38 digits**c) Hive 0.13.0 introduced user definable precision and scale  
d) All of the mentioned  
**View Answer**

**Answer: b  
Explanation: TIMESTAMP is available starting with Hive 0.8.0**

6. Integral literals are assumed to be \_\_\_\_\_\_\_\_\_ by default.  
a) SMALL INT  
**b) INT**  
c) BIG INT  
d) TINY INT  
View Answer

Answer: b  
Explanation: Integral literals are assumed to be INT by default, unless the number exceeds the range of INT in which case it is interpreted as a BIGINT, or if one of the following postfixes is present on the number.

7. Hive uses \_\_\_\_\_-style escaping within the strings.  
**a) C**  
b) Java  
c) Python  
d) Scala  
**View Answer**

**Answer: a  
Explanation: String literals can be expressed with either single quotes (‘) or double quotes (“).**

8. Which of the following statement will create column with varchar datatype ?  
a) CREATE TABLE foo (bar CHAR(10))  
**b) CREATE TABLE foo (bar VARCHAR(10))**  
c) CREATE TABLE foo (bar CHARVARYING(10))  
d) All of the mentioned  
**View Answer**

**Answer: b  
Explanation: Varchar datatype was introduced in Hive 0.12.0**

9. \_\_\_\_\_\_\_\_\_ will overwrite any existing data in the table or partition  
a) INSERT WRITE  
b) INSERT OVERWRITE  
**c) INSERT INTO**  
d) None of the mentioned  
View Answer

Answer: c  
Explanation: INSERT INTO will append to the table or partition, keeping the existing data intact.

10. Hive does not support literals for \_\_\_\_\_\_ types.  
a) Scalar  
**b) Complex**  
c) INT  
d) CHAR  
View Answer

Answer: b  
Explanation: It is not possible to use them in INSERT INTO…VALUES clauses.

**Hadoop Questions and Answers – HiveQL-2**

**This set of Hadoop Multiple Choice Questions & Answers (MCQs) for quizzes focuses on “HiveQL”.**

1. Hive specific commands can be run from Beeline, when the Hive \_\_\_\_\_\_\_ driver is used.  
a) ODBC  
**b) JDBC**  
c) ODBC-JDBC  
d) All of the Mentioned  
View Answer

Answer: b  
Explanation: Hive specific commands are same as Hive CLI commands.

2. Point out the correct statement :  
a) –helpusage display a usage message  
b) The JDBC connection URL format has the prefix jdbc:hive:  
**c) Starting with Hive 0.14, there are improved SV output formats**  
d) None of the mentioned  
View Answer

Answer: c  
Explanation: Output formats available are namely DSV, CSV2 and TSV2.

3. \_\_\_\_\_\_\_\_\_ reduce the amount of informational messages displayed (true) or not (false).  
**a) –silent=[true/false].**  
b) –autosave=[true/false].  
c) –force=[true/false].  
d) All of the mentioned  
View Answer

Answer: a  
Explanation: It also stops displaying the log messages for the query from HiveServer2.

4. Which of the following is used to set transaction isolation level ?  
a) –incremental=[true/false].  
**b) –isolation=LEVEL**  
c) –force=[true/false].  
d) –truncateTable=[true/false].  
View Answer

Answer: b  
Explanation: Set the transaction isolation level to TRANSACTION\_READ\_COMMITTED or TRANSACTION\_SERIALIZABLE.

5. Point out the wrong statement :  
a) HiveServer2 has a new JDBC driver  
**b) CSV and TSV output formats are maintained for forward compatibility**  
c) HiveServer2 supports both embedded and remote access to HiveServer2  
d) None of the mentioned  
View Answer

Answer: b  
Explanation: CSV and TSV output formats are maintained for backward compatibility.

6. The \_\_\_\_\_\_\_\_ allows users to read or write Avro data as Hive tables.  
**a) AvroSerde**  
b) HiveSerde  
c) SqlSerde  
d) None of the mentioned  
View Answer

Answer: a  
Explanation: AvroSerde understands compressed Avro files.

7. Starting in Hive \_\_\_\_\_\_\_, the Avro schema can be inferred from the Hive table schema.  
**a) 0.14**  
b) 0.12  
c) 0.13  
d) 0.11  
View Answer

Answer: a  
Explanation: Starting in Hive 0.14, columns can be added to an Avro backed Hive table using the Alter Table statement.

8. The AvroSerde has been built and tested against Hive 0.9.1 and later, and uses Avro \_\_\_\_\_\_\_ as of Hive 0.13 and 0.14.  
a) 1.7.4  
b) 1.7.2  
**c) 1.7.3**  
d) None of the mentioned  
View Answer

Answer: d  
Explanation: The AvroSerde uses Avro 1.7.5.

9. Which of the following data type is supported by Hive ?  
a) map  
b) record  
c) string  
**d) enum**  
View Answer

Answer: d  
Explanation: Hive has no concept of enums.

10. Which of the following data type is converted to Array prior to Hive 0.12.0 ?  
a) map  
b) long  
c) float  
**d) bytes**  
View Answer

Answer: d  
Explanation: Bytes are converted to Array[smallint] prior to Hive 0.12.0.

**Hadoop Questions and Answers – Querying data with HiveQL-1**

**This set of Interview Questions & Answers focuses on “Hive”.**

1. Avro-backed tables can simply be created by using \_\_\_\_\_\_\_\_\_ in a DDL statement.  
**a) “STORED AS AVRO”**  
b) “STORED AS HIVE”  
c) “STORED AS AVROHIVE”  
d) “STORED AS SERDE”  
View Answer

Answer: a  
Explanation: AvroSerDe takes care of creating the appropriate Avro schema from the Hive table schema.

2. Point out the correct statement :  
a) Avro Fixed type should be defined in Hive as lists of tiny ints  
**b) Avro Bytes type should be defined in Hive as lists of tiny ints**  
c) Avro Enum type should be defined in Hive as strings  
d) All of the mentioned  
**View Answer**

**Answer: b  
Explanation: The AvroSerde will convert these to Bytes during the saving process.**

3. Types that may be null must be defined as a \_\_\_\_\_\_ of that type and Null within Avro.  
**a) Union**  
b) Intersection  
c) Set  
d) All of the mentioned  
View Answer

Answer: a  
Explanation: A null in a field that is not so defined will result in an exception during the save. No changes need be made to the Hive schema to support this, as all fields in Hive can be null.

4. The files that are written by the \_\_\_\_\_\_\_ job are valid Avro files.  
a) Avro  
b) Map Reduce  
**c) Hive**  
d) All of the mentioned  
View Answer

Answer: c  
Explanation: If you copy these files out, you’ll likely want to rename them with .avro.

5. Point out the wrong statement :  
a) To create an Avro-backed table, specify the serde as org.apache.hadoop.hive.serde2.avro.AvroSerDe  
b) Avro-backed tables can be created in Hive using AvroSerDe  
**c) The AvroSerde cannot serialize any Hive table to Avro files**  
d) None of the mentioned  
View Answer

Answer: c  
Explanation: The AvroSerde can serialize any Hive table to Avro files.

6. Use \_\_\_\_\_\_\_\_ and embed the schema in the create statement.  
**a) schema.literal**  
b) schema.lit  
c) row.literal  
d) all of the mentioned  
**View Answer**

**Answer: a  
Explanation: You can embed the schema directly into the create statement.**

7. \_\_\_\_\_\_\_ is interpolated into the quotes to correctly handle spaces within the schema.  
**a) $SCHEMA**  
b) $ROW  
c) $SCHEMASPACES  
d) $NAMESPACES  
View Answer

Answer: a  
Explanation: Use none to ignore either avro.schema.literal or avro.schema.url.

8. To force Hive to be more verbose, it can be started with :  
**a) \*hive –hiveconf hive.root.logger=INFO,console\***  
b) \*hive –hiveconf hive.subroot.logger=INFO,console\*  
c) \*hive –hiveconf hive.root.logger=INFOVALUE,console\*  
d) All of the mentioned  
**View Answer**

**Answer: a  
Explanation: This Statement will spit orders of magnitude more information to the console and will likely include any information the AvroSerde is trying to get you about what went wrong.**

9. \_\_\_\_\_\_\_\_ was designed to overcome limitations of the other Hive file formats.  
**a) ORC**  
b) OPC  
c) ODC  
d) None of the mentioned  
View Answer

Answer: a  
Explanation: The Optimized Row Columnar (ORC) file format provides a highly efficient way to store Hive data.

10. An ORC file contains groups of row data called :  
a) postscript  
**b) stripes**  
c) script  
d) none of the mentioned  
View Answer

**Answer: b  
Explanation: The default stripe size is 250 MB. Large stripe sizes enable large, efficient reads from HDFS.**