Select the correct statements about the views in Hive



View's column types are not specified in CREATE VIEW, Hive will determine it from the view's defining SELECT expression

**This should be selected**



Changes to the underlying table’s schema will be reflected in the view's schema

**This should not be selected**

No, these changes don't affect the view’s schema, but if the changes invalidate the view an error appears only in response to the HiveQL query to this view



Views' metadata is read-only

**This should not be selected**

No, metadata can be changed with the ALTER VIEW statement



The views are read-only

**This should be selected**



CREATE VIEW will fail if the view's defining SELECT expression fails

**Un-selected is correct**

Question 2

Correct

1 / 1 points

**2. Question 2**

You have a bunch of data in your local filesystem and you need to load it in Hive. What would you do?



Put data in HDFS directory and create an external table on it or add this data to the directory of the existing table



Use a special form of the LOAD DATA statement



Both ways are allowed

**Correct**

Yes, you can put data in the right place manually or use the LOAD DATA statement

Question 3

Correct

1 / 1 points

**3. Question 3**

You have a Hive external table and want to store this data in more compact and efficient format. Your decision is:



Create a table in a new format with the CREATE TABLE statement and fill it from the external table with the INSERT INTO TABLE statement

**Correct**

Yes, that's the right way: create and fill a managed table



Create and fill a new table with the CREATE TABLE ... AS SELECT statement



CREATE VIEW over an external table

Question 4

Incorrect

0 / 1 points

**4. Question 4**

What table properties do you use to create the external table with Apache web server access logs?



ROW FORMAT DELIMITED FIELDS TERMINATED BY ... from CREATE TABLE

**This should not be selected**

No, even the ‘common’ format of the Apache access log is too complex to split fields with a separator



STORED AS from CREATE TABLE



ROW FORMAT SERDE and the a regular expression

Question 5

Incorrect

0 / 1 points

**5. Question 5**

What is the type of the following functions: count(), avg(), percentile()



UDF



UDAF



None of the above

**This should not be selected**

No, all these functions perform aggregation, so they all are UDAFs (User Defined Aggregate Functions)

Question 6

0.50 / 1 points

**6. Question 6**

Select the correct statements



In one SELECT statement multiple aggregations can be done

**This should be selected**



To register new UDF in Hive you need to execute ADD JAR statement and then CREATE FUNCTION

**Correct**

Yes, CREATE FUNCTION is required and if the function code was not loaded, it should be done with ADD JAR



If you need to make you own UDF, you can write it in Java, Python or C++

**This should not be selected**

No, Hive UDF can be written only in Java, see the "Hive Analytics: UDF, UDAF, UDTF" video



There can be columns and UDTF functions in one SELECT statement

**Un-selected is correct**

Question 7

Incorrect

0 / 1 points

**7. Question 7**

The table contains orders in an e-commerce shop in the following columns: date, customer\_id, order\_id, price. What type of function do you use to get top 10 most expensive orders for each month?



UDAF



Аnalytics function



UDTF



UDF

**This should not be selected**

No, it's necessary to make rating for each month, it can be done with the rank() or row\_number() analytics function

Question 8

Incorrect

0 / 1 points

**8. Question 8**

Is it possible to substitute the MAP and REDUCE keywords with TRANSFORM in the Hive streaming statement?



No



Yes, only with extra DISTRIBUTE BY or CLUSTER BY

**This should not be selected**

No, extra DISTRIBUTE BY or CLUSTER BY required anyway, they are used in the MAP … REDUCE statement



Yes

Question 9

Incorrect

0 / 1 points

**9. Question 9**

What role does DISTRIBUTE BY have in the Hive streaming statement?



Mapper



Combiner

**This should not be selected**

No, it is essentially an implementation of a "partitioner", see the "Hive Analytics with UDF and Streaming" video



Partitioner



Reducer

Question 10

0.50 / 1 points

**10. Question 10**

Select the correct statements about the user defined functions:



Partitioned table function can return less rows than receive as input

**Correct**

Yes, it's correct. PTF applies a condition to the given set of rows. So in most cases quantity of output rows is less than quantity of input ones.



UDF cannot change the type of fields.

**This should not be selected**

No, UDF can change the type by specifying the output type in the implementation of the UDF.



UDAF requires the Reduce phase

**This should be selected**



UDTF receives a set of rows as input but can return only a single row

**Un-selected is correct**

Is it possible to create a view on non-existent table?



No, because the view's defining SELECT expression is invalid

**Correct**

Yes, SELECT from the non-existent table is invalid, so it invalidates a CREATE VIEW statement



Yes, because the view's defining SELECT expression is executed only when you query the view

**This should not be selected**

No, because the SELECT expression is validated in CREATE VIEW. Moreover, Hive determines view's column types from the SELECT expression.



Create the view on an existing table, then you can delete the table

**Correct**

Yes, it's possible, because the views are not changed together with the changes in the underlying tables

Question 2

Correct

1 / 1 points

**2. Question 2**

You have some data in HDFS and you want to process it in Hive. How will you do it?



CREATE EXTERNAL TABLE

**Correct**

Yes, you can create an external table over your data



CREATE VIEW

Question 3

Correct

1 / 1 points

**3. Question 3**

You have a table and you want to transform it in more convenient form without extra space in HDFS. Your decision is:



Create the second table and fill it with the data from the first table



CREATE VIEW

**Correct**

Yes, views transform the table without extra space in HDFS

Question 4

Correct

1 / 1 points

**4. Question 4**

What table properties do you use to create the external table with Apache web server access logs?



STORED AS from CREATE TABLE



ROW FORMAT DELIMITED FIELDS TERMINATED BY ... from CREATE TABLE



ROW FORMAT SERDE and the a regular expression

**Correct**

Yes, it's possible to write an appropriate regular expression

Question 5

Correct

1 / 1 points

**5. Question 5**

What is the type of the following functions: count(), avg(), percentile()



UDF



UDAF

**Correct**

Yes, all these functions perform aggregation, so they all are UDAFs (User Defined Aggregate Functions)



None of the above

Question 6

Incorrect

0 / 1 points

**6. Question 6**

You want to process Hive table with your custom mapper written in Python. What ways are possible?



Pack your mapper as Hive UDF function

**This should not be selected**

No, it's impossible to write Hive UDF in Python. Rewrite your mapper in Java to make the Hive UDF or use the SELECT TRANSFORM statement



Use your mapper in the SELECT TRANSFORM statement as a streaming script

Question 7

Correct

1 / 1 points

**7. Question 7**

The table contains orders in an e-commerse shop in the following columns: date, customer\_id, order\_id, price. What type of function do you use to select top 10 customers by the total amount of money spent in the shop?



UDAF

**Correct**

Yes, use sum() to count the total amount and then sort by it



UDF



UDTF

Question 8

0.40 / 1 points

**8. Question 8**

How do you set up a reduce stage in Hive streaming statement?



By the DISTRIBUTE BY … TRANSFORM clause

**This should be selected**



By the DISTRIBUTE BY … REDUCE clause

**Correct**

Yes, the DISTRIBUTE BY clause defines a partitioner and the REDUCE clause defines a reducer script



By the CLUSTER BY … TRANSFORM clause

**This should be selected**



By the TRANSFORM clause

**This should not be selected**

The TRANSFORM clause can specify both mapper and reducer



By the REDUCE clause

**Un-selected is correct**

Question 9

Correct

1 / 1 points

**9. Question 9**

What role does DISTRIBUTE BY have in the Hive streaming statement?



Partitioner

**Correct**

Yes, DISTRIBUTE BY defines the key which is used to distribute between the reducers, that is what a partitioner does



Combiner



Reducer



Mapper

Question 10

Correct

1 / 1 points

**10. Question 10**

What type of Hive functions can increase rows quantity?



UDAF

**Un-selected is correct**



UDF

**Un-selected is correct**



UDTF

**Correct**

Yes. UDTF stands for "user-defined *table* function". It allows to return multiple rows and columns (i.e. a table) from one field given as input.



Partitioned table functions (windowing functions).

**Un-selected is correct**