**HIVE**

What is the Language used by HIVE?

SQL

Oracle SQL

HIVEQL

HIVE SQL

Which of the following is the collection data types in hive?

Arrays

Maps

Structs

All the above

What is the Default engine for running hive queries?

MapReduce

Tez

Spark

None of these

HIVEQL is based on

SQL

Oracle SQL

HIVEQL

HIVE SQL

Logging library used by hive is

log4p

log4l

log4j

log4k

Default database used by hive is

derby

mysql

Cassandra

Mongodb

Hive queries are converted to

Shell Scripts

HDFS commands

SQL queries

MR jobs

Metadata of hive tables are stored in

Hive Metastore

Hive database

HDFS

Local disk

By default hive table data are stored in

Hive Metastore

Hive database

HDFS

Local disk

Hive metastore can be run using the command

hive metastore start

hive --service metastore

metastore service start

hive metastore --service start

command to load data in a local file into hive table

load data 'path of the file' into table\_name

load data local inpath 'path of the file' into table\_name

load data in local 'path of the file' into table\_name

load data in 'path of the file' into table\_name

Function to repeat a string 'n' number of times in hive

REPEAT('string','n')

REPEAT('string',n)

REPEAT(string,'n')

REPEAT(string,n)

Command to create database in hive

create database dbname

create db dbname

create hivedb dbname

creata dbname

Function to convert text into lower case in hive

LOWER

LCASE

both a & b

none of these

Function to remove the spaces associated with a string

trim

remove\_spaces

remove char where c==' '

None of these

Function to remove the spaces associated with a string

trim

remove\_spaces

remove char where c==' '

None of these

How to select the records from a table whose name starts with 'a'

select \* from table\_name whose name='s'

select \* from table\_name where name like 's.\*'

select \* from table\_name where name like 's\*'

select \* from table\_name where name ='s\*'

Which command enables to check all the available functions in hive

get functions

show functions

which functions

all functions

Command used to set a database as your current working database, analogous to changing working directories in a file system.

Create

Select

Use

Nine of the above

If we drop a managed table

only schema is dropped

only data is dropped

Both schema and data are dropped

Only table structure is altered

Can Hive be used in real time

True

False

True for some distributions

True if data set is small

How to arrange the values in ascending order in hive

sort by

order by

Both of these

None of these

Which of the below sorts records as per each reducer

sort by

order by

Both of these

None of these

Which of the below sorts records as per each reducer

sort by

order by

Both of these

None of these

Group by clause groups all the records with a particular

collection column

collection row

collection value

collection field

Partially ordered results are given by

sort by

order by

Both of these

None of these

If there are more than 1 reducer then which of the following is better for sorting

sort by

order by

Both of these

None of these

Which of the following syntax is correct for order by clause

select \* from emp order by sal

select \* from emp orderby sal

select \* from emp ordered by sal

None of these

Hive uses the columns in Distribute By to distribute the

columns among mappers

columns among reducers

rows among mappers

rows among reducers

How can you differentiate the partition columns and sort columns

Using cluster by and sort by

Using distribute by and sort by

Both of these

None of these

Which of the following is a short cut for both Distribute By and Sort By.

Cluster by

Order by

Both of these

None of these

Distribute By does not guarantee clustering or sorting properties on the

distributed columns

distributed rows

distributed keys

None of these

Partitioning is used to

distrubute load horizantally

distrubute load vertically

distribute load orthogonally

All the above

The way in which tasks are encoded permits the system to optimize their execution automatically, allowing the user to focus on

semantics rather than efficiency

time give to system for execution

efficiency rather than semantics

map-reduce programs

\_\_\_\_\_\_\_\_Clause is used to divide the table into buckets

Select

Clustered by

Distribute By

None of the above

How to view the databases which start with letter 'a'

SHOW DATABASE 'a'

SHOW DATABASE LIKE 'a'

SHOW DATABASE LIKE 'a.\*'

SHOW DATABASE LIKE 'a.'

Partitioning data will distribute the

Load vertically

Columns vertically

Load Horizantally

All of the above

Partitioning organizes data in

Physical fashion

Logical fashion

Remote location

All of the above

Which of the following runs a MR job

Static partition

Dynamic partition

Both a & b

None of these

Which of the following decomposes data into more manageable parts

Static partition

Dynamic partition

Bucketing

None of these

The Bucketing concept is based on

hash column

hash function

hash key

None of these

In hive a partition is --------- and a bucket is ------ respectively

Directory, File

File, Directory

Directory, Directory

File, File

Which of the following is a type of join in hive?

Left outer join

Right outer join

Full outer join

All the above

Hive property to set the maximum number of dynamic partitions per node

set.hive.max.dynamic.partition.pernode

set hive.exec.max.dynamic.partition.pernode

set hive.exec.max.dynamic.partition.pernode

c.set hive.exec.max.dynamic.partition.pernode set.hive.max.dynamic.partition.pernode

A table may be joined to itself.

True

False

None of the above

An outer join requires each record in the two joined tables to have a matching record.

True

False

When do we use a HAVING clause?

To limit the output of a query

To limit the output of a query using an aggregate function only

When GROUP by is used

both b and c

Table “Customer” contains customer KYC Data with customer ID and “Transaction” tables contains all the transaction done with customer ID and with Date column. I want know all the customers who has not done any purchase in month of FEB 2016. Which join would you use for this requirement?

Customer inner join Transaction

Transaction left join Customer

Customer right join Transaction

Customer left join Transaction

Which of the following is true about Cartesian Products?

A Cartesian product is formed when a join condition is omitted.

A Cartesian product is formed when a join condition is valid.

Some rows in the first table are joined to all rows in the second table.

All rows in the first table are joined to some rows in the second table.

To perform a bucket join, if the number of buckets in one table is 5 then the number of buckets in the other table can be

6

20

10

both b and c

To perform a SMBJ, if the number of buckets in one table is 5 then the number of buckets in other table can be

6

5

10

both b and c

If you have two datasets one is over 1 billion rows, one is about 100 million rows and both and bucketed into same number of buckets then which of the following join is better

Left semi join

Bucket-Map join

SMBJ

both b and c

File formats available in hive

a.Avro file

sequence file

Orc file

All the above

How to disable variable substitution in hive

set.hive.variable.substitute=flase

set.hive.variable.substitute=true

set.hive.substitute.variable=true

set.hive.substitute.variable=false

Hive Script supports multi line commenting?

True

False

Select the complex dataype in hive which is an ordered collection of fields which are indexable using zero-based integers.

MAP

STRUCT

ARRAY

Only (a) and (b)

\_\_\_\_\_\_\_ complex data type in hive can have collection of elements of different types

STRUCT

ARRAY

Both (a) and (b)

\_\_\_\_\_\_\_\_\_splits data horizontally into row groups

RCFile

Sequence

Both (a) and (b)

None of the above

internal Storage order in RCFile is

Row-based

Column-based

Both (a) and (b)

None of the above

I

nternal Storage order in SequenceFile is

Row-based

Column-based

Both (a) and (b)

None of the above

\_\_\_\_\_\_\_\_\_\_introduces a lightweight indexing that enables skipping of

complete blocks of rows that do not match a query.

RC File format

ORC File format

Both (a) and (b)

None of the above

\_\_\_\_\_\_\_\_\_\_\_files are flat files consisting of binary key-value pairs.

Text

Sequence

RC

ORC

Facebook uses \_\_\_\_\_\_\_\_\_\_FILE as its default file format for storing of data in their data warehouse as they perform different types of analytics using Hive.

Text

Sequence

RC

ORC

\_\_\_\_\_\_\_\_\_\_/\_\_\_\_\_\_\_\_\_\_\_ These 2 classes read/write data in hadoop SequenceFile format.

SequenceFileInputFormat/SequenceFileOutputFormat

BinaryFileInputFormat/SequenceFileOutputFormat

TextInputFormat/SequenceFileOutputFormat

None of the above

\_\_\_\_\_\_\_\_\_\_\_\_: These 2 classes read/write data in plain text file format.

TextInputFormat/TextOutputFormat

BinaryFileInputFormat/TextOutputFormat

TextInputFormat/SequenceFileOutputFormat

TextInputFormat/HiveIgnoreKeyTextOutputFormat

While writing hive udf the function need to be used is

Evaluate

Perform

extern

None of these

Which of the following perform row wise operations

udf

udaf

udtf

Both a &c

How to write UDF's in hive

extending 'hiveudf' clas

extending 'udf' class

extending 'hive\_udf' class

extending 'udf\_hive' class

Which of the following method should be implemented in Hive UDF

init

iterate

evaluate

merge

Hive udf can be added into the hive using the command

USE

cp

REGISTER

ADD jar

explode function is an example of

udf

udaf

udtf

Both a &c

Which is not a valid method while writing custom UDF in hive

Initialize()

Execute()

Evaluate()

getDisplayString()

Who enables the hive to access the range of clients written in different languages?

Hive server

Hive server

Thrift server

JDBC/ODBC serve

Hive server2 accepts valid thrift calls in

sql mode

http mode

https mode

None of these

The CLI present in hive server 2 is

hiveline

Beeline

Sqlline

All the above

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ is a table and storage management layer for Hadoop that enables users with different data processing tools - Pig, MapReduce - to more easily read and write data on the grid.

HCatalog

HiveServer

Hive Metastore

None of the above

HCatalog is built on top of the Hive metastore

True

False

HiveServer2 supports a command shell \_\_\_\_\_\_\_\_\_\_ that works with HiveServer2.

Hive CLI

Beeline

None of the above

The Beeline shell works in both embedded mode as well as remote

mode.

True

False

\_\_\_\_\_\_\_\_\_\_\_\_\_ HiveServer2 mode is recommended for production use, as it is more secure and doesn't require direct HDFS/metastore access to be granted for users.

Embedded

Remote

Both (a) and (b)

None of the above

\_\_\_\_\_\_\_\_ is a powerful mechanism to convert hive queries

SerDe

ThriftServer

HiveServer

None of the above

Which of the SeDe class is used to serialize and serialize thrift serialized object.

MetadataTypedColumnsetSerDe

ThriftSerDe

ThriftDynamicSerDe

None of the above