Topic / Module: Big Data overview

Q. No. 1

**Question:**

What is not true about Big Data

Answer Choices

A: Hadoop ecosystem handles Big Data

B: It is represented by 4 V's

C: It references OLTP system

D: It references OLAP system.

**Answer**:C

Q. No. 2

**Question:**

What is not true about Hadoop

Answer Choices

A: It is a distributed parallel processing ecosystem.

B: It is ideally a Datawarehouse solution

C: It can replace RDBMS systems completely

D: It is a file system

**Answer**:C

Q. No. 3

**Question:**

Which one of the following is not among 4V's of Big Data

Answer Choices

A) Volume –Scale of data  
B) Velocity –Different forms of data  
C) Variety –Analysis of streaming data  
D) Volatile –Synchronzation of data

**Answer**:D

Q. No. 4

**Question:**

Which one of the following is not Hadoop's Distributiion

Answer Choices

A) MapR  
B) Cloudera  
C) Hortonworks  
D) MapReduce

**Answer**:D

Q. No. 5

**Question:**

Which one of the following is not a part of Hadoop's Ecosystem

Answer Choices

A) HDFS  
B) MapReduce  
C) Hbase  
D) MongoDB

**Answer**:D

Q. No. 6

**Question:**

Hadoop is a framework that works with a variety of related tools. Common cohorts include:  
A) MapReduce, Hive and HBase  
B) MapReduce, MySQL and Google Apps  
C) MapReduce, Hummer and Iguana  
D) MapReduce, Heron and Trumpet

**Answer:**A

Q. No. 7

**Question:**

\_\_\_\_\_\_\_\_\_\_ can best be described as a programming model used to develop Hadoop-based applications that can process massive amounts of data.  
a) MapReduce  
b) Mahout  
c) Oozie  
d) All of the mentioned

Answer:a

Q. No. 8

**Question:**

\_\_\_\_\_\_\_\_\_\_ can best be described as a programming model used to develop Hadoop-based applications that can process massive amounts of data.  
a) MapReduce  
b) Mahout  
c) Oozie  
d) All of the mentioned

Answer:a

Q. No. 9

**Question:**

Point out the correct statement :  
a) Hive is not a relational database, but a query engine that supports the parts of SQL specific to querying data  
b) Hive is a relational database with SQL support  
c) Pig is a relational database with SQL support  
d) All of the mentioned

Answer : a

Q. No. 10

**Question:**

The Pig Latin scripting language is not only a higher-level data flow language but also has operators similar to :  
a) SQL  
b) JSON  
c) XML  
d) All of the mentioned

Answer : a

Q. No. 11

**Question:**

A \_\_\_\_\_\_\_\_ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.  
a) MapReduce  
b) Mapper  
c) TaskTracker  
d) JobTracker

Answer : c

Q. No. 12

**Question:**

Point out the correct statement :  
a) MapReduce tries to place the data and the compute as close as possible  
b) Map Task in MapReduce is performed using the Mapper() function  
c) Reduce Task in MapReduce is performed using the Map() function  
d) All of the mentioned

Answer : a

Q. No. 13

**Question:**

 \_\_\_\_\_\_\_\_\_ function is responsible for consolidating the results produced by each of the Map() functions/tasks.  
a) Reduce  
b) Map  
c) Reducer  
d) All of the mentioned

Answer : a

Q. No. 14

**Question:**

\_\_\_\_\_\_\_\_\_ is the default Partitioner for partitioning key space.  
a) HashPar  
b) Partitioner  
c) HashPartitioner  
d) None of the mentioned

Answer : a

Q. No. 15

**Question:**

Input to the \_\_\_\_\_\_\_ is the sorted output of the mappers.  
a) Reducer  
b) Mapper  
c) Shuffle  
d) All of the mentioned

Answer : a

Q. No. 16

**Question:**

Point out the wrong statement :  
a) Reducer has 2 primary phases  
b) Increasing the number of reduces increases the framework overhead, but increases load balancing and lowers the cost of failures  
c) It is legal to set the number of reduce-tasks to zero if no reduction is desired  
d) The framework groups Reducer inputs by keys (since different mappers may have output the same key) in sort stage

Answer : a

Q. No. 17

**Question:**

Which of the following phases occur simultaneously ?  
a) Shuffle and Sort  
b) Reduce and Sort  
c) Shuffle and Map  
d) All of the mentioned

Answer : a

Q. No. 18

**Question:**

 \_\_\_\_\_\_\_\_\_ is the primary interface for a user to describe a MapReduce job to the Hadoop framework for execution.  
a) Map Parameters  
b) JobConf  
c) MemoryConf  
d) None of the mentioned

Answer : b

Q. No. 19

**Question**:

Which of the following phases occur simultaneously ?  
a) Shuffle and Sort  
b) Reduce and Sort  
c) Shuffle and Map  
d) All of the mentioned

Answer: a

Q. No. 20

**Question**:

The need for data replication can arise in various scenarios like :  
a) Replication Factor is changed  
b) DataNode goes down  
c) Data Blocks get corrupted  
d) All of the mentioned

Answer :d

Q. No. 21

**Question**:

\_\_\_\_\_\_\_\_ is the slave/worker node and holds the user data in the form of Data Blocks.  
a) DataNode  
b) NameNode  
c) Data block  
d) Replication

Answer :a

Q. No. 22

**Question**:

The daemons associated with the MapReduce phase are \_\_\_\_\_\_\_\_ and task-trackers.  
a) job-tracker  
b) map-tracker  
c) reduce-tracker  
d) All of the mentioned

Answer :a

Q. No. 23

**Question**:

The JobTracker pushes work out to available \_\_\_\_\_\_\_ nodes in the cluster, striving to keep the work as close to the data as possible  
a) DataNodes  
b) TaskTracker  
c) ActionNodes  
d) All of the mentioned

Answer :a

Q. No. 24

**Question**:

InputFormat class calls the \_\_\_\_\_\_\_\_ function and computes splits for each file and then sends them to the jobtracker.  
a) puts  
b) gets  
c) getSplits  
d) All of the mentioned

Answer :a

Q. No. 25

**Question**:

InputFormat class calls the \_\_\_\_\_\_\_\_ function and computes splits for each file and then sends them to the jobtracker.  
a) puts  
b) gets  
c) getSplits  
d) All of the mentioned

Answer :c

Q. No. 26

**Question**:

On a tasktracker, the map task passes the split to the createRecordReader() method on InputFormat to obtain a \_\_\_\_\_\_\_\_\_ for that split.  
a) InputReader  
b) RecordReader  
c) OutputReader  
d) None of the mentioned

Answer :b

Q. No. 27

**Question**:

The default InputFormat is \_\_\_\_\_\_\_\_\_\_ which treats each value of input a new value and the associated key is byte offset.  
a) TextFormat  
b) TextInputFormat  
c) InputFormat  
d) All of the mentioned

Answer :b

Q. No. 28

**Question**:

\_\_\_\_\_\_\_\_\_\_ controls the partitioning of the keys of the intermediate map-outputs.  
a) Collector  
b) Partitioner  
c) InputFormat  
d) None of the mentioned

Answer :b

Q. No. 29

**Question**:

Output of the mapper is first written on the local disk for sorting and \_\_\_\_\_\_\_\_\_ process.  
a) shuffling  
b) secondary sorting  
c) forking  
d) reducing

Answer :a

Q. No. 30

**Question**:

The \_\_\_\_\_\_\_\_\_\_ is a framework-specific entity that negotiates resources from the ResourceManager  
a) NodeManager  
b) ResourceManager  
c) ApplicationMaster  
d) All of the mentioned

Answer :c

Q. No. 31

**Question**:

Apache Hadoop YARN stands for :  
a) Yet Another Reserve Negotiator  
b) Yet Another Resource Network  
c) Yet Another Resource Negotiator  
d) All of the mentioned

Answer :c

Q. No. 32

**Question**:

The \_\_\_\_\_\_\_\_\_\_\_\_ is the ultimate authority that arbitrates resources among all the applications in the system.  
a) NodeManager  
b) ResourceManager  
c) ApplicationMaster  
d) All of the mentioned

Answer :b

Q. No. 33

**Question**:

The \_\_\_\_\_\_\_\_\_\_ is responsible for allocating resources to the various running applications subject to familiar constraints of capacities, queues etc.  
a) Manager  
b) Master  
c) Scheduler  
d) None of the mentioned

Answer :b

Q. No. 34

**Question**:

ZooKeeper allows distributed processes to coordinate with each other through registers, known as :  
a) znodes  
b) hnodes  
c) vnodes  
d) rnodes

Answer :a

Q. No. 35

**Question**:

ZooKeeper allows distributed processes to coordinate with each other through registers, known as :  
a) znodes  
b) hnodes  
c) vnodes  
d) rnodes

Answer :a

Q. No. 36

**Question**:

In Hive SerDe stands for

[A - serialize and Desrialize](javascript:void(0);)

[B - serializer and Deserializer](javascript:void(0);)

[C - Serialize and Destruct](javascript:void(0);)

[D - serve and destruct](javascript:void(0);)

Answer :B

Q. No. 37

**Question**:

To select all columns starting with the word 'Sell' form the table GROSS\_SELL the query is

[A - select '$Sell\*'  from GROSS\_SELL](javascript:void(0);)

[B - select 'Sell\*' from GROSS\_SELL](javascript:void(0);)

[C - select 'sell.\*' from GROSS\_SELL](javascript:void(0);)

[D - select  'sell[\*]' from GROSS\_SELL](javascript:void(0);)

Answer :C

Q. No. 38

**Question**:

Which of the following hint is used to optimize the join queries

[A - /\* joinlast(table\_name) \*/](javascript:void(0);)

[B - /\* joinfirst(table\_name) \*/](javascript:void(0);)

[C - /\* streamtable(table\_name) \*/](javascript:void(0);)

[D - /\* cacheable(table\_name) \*/](javascript:void(0);)

Answer :C

Q. No. 39

**Question**:

The drawback of managed tables in hive is

[A - they are always stored under default directory](javascript:void(0);)

[B - They cannot grow bigger than a fixed size of 100GB](javascript:void(0);)

[C - They can never be dropped](javascript:void(0);)

[D - They cannot be shared with other applications](javascript:void(0);)

Answer:D

Q. No. 40

**Question**:

In case of one large table and 2 small tables, for an optimized query performance

[A - The largest one should be cached to memory and small ones should be streamed](javascript:void(0);)

[B - The small Ones should be cached and large one should be streamed](javascript:void(0);)

[C - All of the table should be cached](javascript:void(0);)

[D - All the tables should be streamed.](javascript:void(0);)

Answer:B

Q. No. 41

**Question**:

What are collection data types in Pig

[A -](javascript:void(0);) Tuple

[B -](javascript:void(0);) Bag

[C -](javascript:void(0);) Map

[D - All](javascript:void(0);)

Answer:D

Q. No. 42

**Question**:

What are collection data types in Pig

[A -](javascript:void(0);) Tuple

[B -](javascript:void(0);) Bag

[C -](javascript:void(0);) Map

[D - All](javascript:void(0);)

Answer:D

Q. No. 43

**Question**:

How to refer fields in Pig

[A –](javascript:void(0);) By Names

[B –](javascript:void(0);) By Positional Notation

[C -](javascript:void(0);) Both

[D - None](javascript:void(0);)

Answer:C

Q. No. 44

**Question**:

Where we store Bag on Pig

[A –](javascript:void(0);) {}

[B –](javascript:void(0);) [ ]

[C –](javascript:void(0);) ( )

[D - < >](javascript:void(0);)

Answer: A

Total Number of Questions Generated: \_\_\_\_\_\_44\_\_\_\_\_\_\_\_\_