**Big Data Tech Questions**

**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

**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

**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

**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

**Question:**

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

Answer : a

**Question:**

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

Answer : a

**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

**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

**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

**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

**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

**Question**:

Where we store Bag on Pig

[A –](javascript:void(0);) {}

[B –](javascript:void(0);) [ ]

[C –](javascript:void(0);) ( )

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