HADOOP MOCK TEST

This section presents you various set of Mock Tests related to **Hadoop Framework**. You can download these sample mock tests at your local machine and solve offline at your convenience. Every mock test is supplied with a mock test key to let you verify the final score and grade yourself.

HADOOP MOCK TEST IV

Q 1 - When a jobTracker schedules a task is first looks for

A - A node with empty slot in the same rack as datanode

B - Any node on the same rack as the datanode

C - Any node on the rack adjacent to rack of the datanode

D - Just any node in the cluster

Q 2 - The heartbeat signal are sent from

A - JObtracker to Tasktracker

B - Tasktracker to Job tracker

C - Jobtracker to namenode

D - Tasktracker to namenode

Q 3 - Job tracker runs on

A - Namenode

B - Datanode

C - Secondary namenode

D - Secondary datanode

Q 4 - Which of the following is not a scheduling option available in YARN

A - Balanced scheduler

B - Fair scheduler

C - Capacity scheduler

D - FiFO schesduler.

Q 5 - What is the default input format?

A - The default input format is xml. Developer can specify other input formats as appropriate if xml is not the correct input.

B - There is no default input format. The input format always should be specified.

C - The default input format is a sequence file format. The data needs to be preprocessed before using the default input format.

D - The default input format is TextInputFormat with byte offset as a key and entire line as a value.

Q 6 - Which one is not one of the big data feature?

A - Velocity

B - Veracity

C - volume

D - variety

Q 7 - Which technology is used to store data in Hadoop?

A - HBase

B - Avro

C - Sqoop

D - Zookeeper

Q 8 - Which technology is used to serialize the data in Hadoop?

A - HBase

B - Avro

C - Sqoop

D - Zookeeper

Q 9 - Which technology is used to import and export data in Hadoop?

A - HBase

B - Avro

C - Sqoop

D - Zookeeper

Q 10 - Which of the following technologies is a document store database?

A - HBase

B - Hive

C - Cassandra

D - CouchDB

Q 11 - Which one of the following is not true regarding to Hadoop?

A - It is a distributed framework.

B - The main algorithm used in it is Map Reduce

C - It runs with commodity hard ware

D - All are true

Q 12 - Which one of the following stores data?

A - Name node

B - Data node

C - Master node

D - None of these

Q 13 - Which one of the following nodes manages other nodes?

A - Name node

B - Data node

C - slave node

D - None of these

Q 14 - What is AVRO?

A - Avro is a java serialization library.

B - Avro is a java compression library.

C - Avro is a java library that create split table files.

D - None of these answers are correct.

Q 15 - Can you run Map - Reduce jobs directly on Avro data?

A - Yes, Avro was specifically designed for data processing via Map-Reduce.

B - Yes, but additional extensive coding is required.

C - No, Avro was specifically designed for data storage only.

D - Avro specifies metadata that allows easier data access. This data cannot be used as part of

map-reduce execution, rather input specification only.

Q 16 - What is distributed cache?

A - The distributed cache is special component on name node that will cache frequently used data for faster client response. It is used during reduce step.

B - The distributed cache is special component on data node that will cache frequently used data for faster client response. It is used during map step.

C - The distributed cache is a component that caches java objects.

D - The distributed cache is a component that allows developers to deploy jars for Map-Reduce processing.

Q 17 - What is writable?

A - Writable is a java interface that needs to be implemented for streaming data to remote servers.

B - Writable is a java interface that needs to be implemented for HDFS writes.

C - Writable is a java interface that needs to be implemented for MapReduce processing.

D - None of these answers are correct.

Q 18 - What is HBASE?

A - Hbase is separate set of the Java API for Hadoop cluster.

B - Hbase is a part of the Apache Hadoop project that provides interface for scanning large amount of data using Hadoop infrastructure.

D - HBase is a part of the Apache Hadoop project that provides a SQL like interface for data processing.

Q 19 - How does Hadoop process large volumes of data?

A - Hadoop uses a lot of machines in parallel. This optimizes data processing.

B - Hadoop was specifically designed to process large amount of data by taking advantage of MPP hardware.

C - Hadoop ships the code to the data instead of sending the data to the code.

D - Hadoop uses sophisticated caching techniques on name node to speed processing of data.

Q 20 - When using HDFS, what occurs when a file is deleted from the command line?

A - It is permanently deleted if trash is enabled.

B - It is placed into a trash directory common to all users for that cluster.

C - It is permanently deleted and the file attributes are recorded in a log file.

D - It is moved into the trash directory of the user who deleted it if trash is enabled.

Q 21 - When archiving Hadoop files, which of the following statements are true?
Choosetwoanswers

1. **Archived files will display with the extension .arc.**

2. **Many small files will become fewer large files.**

3. **MapReduce processes the original files names even after files are archived.**

4. **Archived files must be UN archived for HDFS and MapReduce to access the original, small files.**

5. **Archive is intended for files that need to be saved but no longer accessed by HDFS.**

A - 1 & 3

B - 2 & 3

C - 2 & 4

D - 3 & 4

Q 22 - When writing data to HDFS what is true if the replication factor is three?
Choose2answers

1. **Data is written to DataNodes on three separate racks *ifRackAware*.**

2. **The Data is stored on each DataNode with a separate file which contains a checksum value.**

3. **Data is written to blocks on three different DataNodes.**

4. **The Client is returned with a success upon the successful writing of the first block and checksum check.**

A - 1 & 3

B - 2 & 3

C - 3 & 4

D - 1 & 4

Q 23 - Which of the following are among the duties of the Data Nodes in HDFS?

A - Maintain the file system tree and metadata for all files and directories.

B - None of the options is correct.

C - Control the execution of an individual map task or a reduce task.

D - Store and retrieve blocks when told to by clients or the NameNode.

E - Manage the file system namespace.

Q 24 - Which of the following components retrieves the input splits directly from HDFS to determine the number of map tasks?

A - The NameNode.

B - The TaskTrackers.

C - The JobClient.

D - The JobTracker.

E - None of the options is correct.

Q 25 - The org.apache.hadoop.io.Writable interface declares which two methods? *Choose2answers.*

1. **public void readFields***DataInput***.**

2. **public void read***DataInput***.**

3. **public void writeFields***DataOutput***.**

4. **public void write***DataOutput***.**

A - 1 & 4

B - 2 & 3

C - 3 & 4

D - 2 & 4

Q 26 - Which one of the following statements is true regarding <key,value> pairs of a MapReduce job?

A - A key class must implement Writable.

B - A key class must implement WritableComparable.

C - A value class must implement WritableComparable.

D - A value class must extend WritableComparable.

Q 27 - Which one of the following statements is false regarding the Distributed Cache?

A - The Hadoop framework will ensure that any files in the Distributed Cache are distributed to all map and reduce tasks.

B - The files in the cache can be text files, or they can be archive files like zip and JAR files.

C - Disk I/O is avoided because data in the cache is stored in memory.

D - The Hadoop framework will copy the files in the Distributed Cache on to the slave node before any tasks for the job are executed on that node.

Q 28 - Which one of the following is not a main component of HBase?

A - Region Server.

B - Nagios.

C - ZooKeeper.

D - Master Server.

Q 29 - Which of the following is false about RawComparator ?

A - Compare the keys by byte.

B - Performance can be improved in sort and suffle phase by using RawComparator.

C - Intermediary keys are deserialized to perform a comparison.

Q 30 - Which demon is responsible for replication of data in Hadoop?

A - HDFS.

B - Task Tracker.

C - Job Tracker.

D - Name Node.

E - Data Node.

Q 31 - Keys from the output of shuffle and sort implement which of the following interface?

A - Writable.

B - WritableComparable.

C - Configurable.

D - ComparableWritable.

E - Comparable.

Q 32 - In order to apply a combiner, what is one property that has to be satisfied by the values emitted from the mapper?

A - Combiner can be applied always to any data

B - Output of the mapper and output of the combiner has to be same key value pair and they can be heterogeneous

C - Output of the mapper and output of the combiner has to be same key value pair. Only if the values satisfy associative and commutative property it can be done.

ANSWER SHEET

Question Number	Answer Key
1	A
2	B
3	A
4	A
5	D
6	B
7	A

8	B
9	C
10	D
11	D
12	B
13	A
14	A
15	A
16	B
17	C
18	B
19	C
20	C
21	B
22	C
23	D
24	D
25	A
26	B
27	C
28	B
29	C
30	D
31	B
32	C