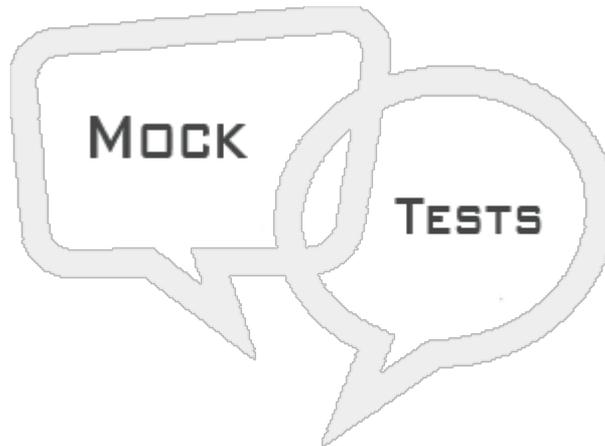


HADOOP MOCK TEST

<http://www.tutorialspoint.com>

Copyright © tutorialspoint.com

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 I

Q 1 - The concept using multiple machines to process data stored in distributed system is not new.

The High-performance computing *HPC* uses many computing machines to process large volume of data stored in a storage area network *SAN*. As compared to *HPC*, Hadoop

- A - Can process a larger volume of data.
- B - Can run on a larger number of machines than HPC cluster.
- C - Can process data faster under the same network bandwidth as compared to HPC.
- D - Cannot run compute intensive jobs.

Q 2 - Hadoop differs from volunteer computing in

- A - Volunteers donating CPU time and not network bandwidth.
- B - Volunteers donating network bandwidth and not CPU time.
- C - Hadoop cannot search for large prime numbers.
- D - Only Hadoop can use mapreduce.

Q 3 - As compared to RDBMS, Hadoop

- A - Has higher data Integrity.
- B - Does ACID transactions
- C - IS suitable for read and write many times
- D - Works better on unstructured and semi-structured data.

Q 4 - What is the main problem faced while reading and writing data in parallel from multiple disks?

- A - Processing high volume of data faster.
- B - Combining data from multiple disks.
- C - The software required to do this task is extremely costly.
- D - The hardware required to do this task is extremely costly.

Q 5 - Which of the following is true for disk drives over a period of time?

- A - Data Seek time is improving faster than data transfer rate.
- B - Data Seek time is improving more slowly than data transfer rate.
- C - Data Seek time and data transfer rate are both increasing proportionately.
- D - Only the storage capacity is increasing without increase in data transfer rate.

Q 6 - Data locality feature in Hadoop means

- A - store the same data across multiple nodes.
- B - relocate the data from one node to another.
- C - co-locate the data with the computing nodes.
- D - Distribute the data across multiple nodes.

Q 7 - Which of these provides a Stream processing system used in Hadoop ecosystem?

- A - Solr
- B - Tez
- C - Spark
- D - Hive

Q 8 - HDFS files are designed for

- A - Multiple writers and modifications at arbitrary offsets.
- B - Only append at the end of file
- C - Writing into a file only once.
- D - Low latency data access.

Q 9 - A file in HDFS that is smaller than a single block size

- A - Cannot be stored in HDFS.
- B - Occupies the full block's size.
- C - Occupies only the size it needs and not the full block.

D - Can span over multiple blocks.

Q 10 - HDFS block size is larger as compared to the size of the disk blocks so that

A - Only HDFS files can be stored in the disk used.

B - The seek time is maximum

C - Transfer of a large files made of multiple disk blocks is not possible.

D - A single file larger than the disk size can be stored across many disks in the cluster.

Q 11 - In a Hadoop cluster, what is true for a HDFS block that is no longer available due to disk corruption or machine failure?

A - It is lost for ever

B - It can be replicated form its alternative locations to other live machines.

C - The namenode allows new client request to keep trying to read it.

D - The Mapreduce job process runs ignoring the block and the data stored in it.

Q 12 - Which utility is used for checking the health of a HDFS file system?

A - fchk

B - fsck

C - fsch

D - fcks

Q 13 - Which command lists the blocks that make up each file in the filesystem.

A - hdfs fsck / -files -blocks

B - hdfs fsck / -blocks -files

C - hdfs fchk / -blocks -files

D - hdfs fchk / -files -blocks

Q 14 - The datanode and namenode are respectivley

A - Master and worker nodes

B - Worker and Master nodes

C - Both are worker nodes

D - None

Q 15 - In the local disk of the namenode the files which are stored persistently are –

A - namespace image and edit log

- B - block locations and namespace image
- C - edit log and block locations
- D - Namespace image, edit log and block locations.

Q 16 - When a client communicates with the HDFS file system, it needs to communicate with

- A - only the namenode
- B - only the data node
- C - both the namenode and datanode
- D - None of these

Q 17 - What mechanisms Hadoop uses to make namenode resilient to failure.

- A - Take backup of filesystem metadata to a local disk and a remote NFS mount.
- B - Store the filesystem metadata in cloud.
- C - Use a machine with at least 12 CPUs
- D - Using expensive and reliable hardware.

Q 18 - The main role of the secondary namenode is to

- A - Copy the filesystem metadata from primary namenode.
- B - Copy the filesystem metadata from NFS stored by primary namenode
- C - Monitor if the primary namenode is up and running.
- D - Periodically merge the namespace image with the edit log.

Q 19 - For the frequently accessed HDFS files the blocks are cached in

- A - the memory of the datanode
- B - in the memory of the namenode
- C - Both A&B
- D - In the memory of the client application which requested the access to these files.

Q 20 - User applications can instruct the namenode to cache the files by

- A - adding cache file names to cache pool
- B - adding cache config to cache pool
- C - adding cache directive to cache pool
- D - passing the file names as parameters to the cache pool

Q 21 - In Hadoop 2.x release HDFS federation means

- A - Allowing namenodes to communicate with each other.
- B - Allow a cluster to scale by adding more datanodes under one namenode.
- C - Allow a cluster to scale by adding more namenodes.
- D - Adding more physical memory to both namenode and datanode.

Q 22 - Under HDFS federation

- A - Each namenode manages metadata of the entire filesystem.
- B - Each namenode manages metadata of a portion of the filesystem.
- C - Failure of one namenode causes loss of some metadata availability from the entire filesystem.
- D - Each datanode registers with each namenode.

Q 23 - The main goal of HDFS High availability is

- A - Faster creation of the replicas of primary namenode.
- B - To reduce the cycle time required to bring back a new primary namenode after existing primary fails.
- C - Prevent data loss due to failure of primary namenode.
- D - Prevent the primary namenode from becoming single point of failure.

Q 24 - As part of the HDFS high availability a pair of primary namenodes are configured. What is true for them?

- A - When a client request comes, one of them chosen at random serves the request.
- B - One of them is active while the other one remains powered off.
- C - Datanodes send block reports to only one of the namenodes.
- D - The standby node takes periodic checkpoints of active namenode's namespace.

Q 25 - Zookeeper ensures that

- A - All the namenodes are actively serving the client requests
- B - Only one namenode is actively serving the client requests
- C - A failover is triggered when any of the datanode fails.
- D - A failover can not be started by hadoop administrator.

Q 26 - Under Hadoop High Availability, Fencing means

- A - Preventing a previously active namenode from start running again.
- B - Preventing the start of a failover in the event of network failure with the active namenode.

C - Preventing the power down to the previously active namenode.

D - Preventing a previously active namenode from writing to the edit log.

Q 27 - Which of the following is not a fencing mechanism for a previously active namenode?

A - Disabling its network port via a remote management command.

B - Revoking its access to shared storage directory.

C - Formatting its disk drive.

D - STONITH

Q 28 - The property used to set the default filesystem for Hadoop in core-site.xml is-

A - filesystem.default

B - fs.default

C - fs.defaultFS

D - hdfs.default

Q 29 - The default replication factor for HDFS file system in hadoop is

A - 1

B - 2

C - 3

D - 4

Q 30 - When running on a pseudo distributed mode the replication factor is set to

A - 2

B - 1

C - 0

D - 3

Q 31 - For a HDFS directory the replication factor RF is

A - same as the RF of the files in that directory

B - Zero

C - 3

D - Does not apply.

Q 32 - The following is not permitted on HDFS files

A - Deleting

B - Renaming

C - Moving

D - Executing.

ANSWER SHEET

Question Number	Answer Key
-----------------	------------

1	C
---	---

2	A
---	---

3	D
---	---

4	B
---	---

5	B
---	---

6	C
---	---

7	C
---	---

8	B
---	---

9	C
---	---

10	D
----	---

11	B
----	---

12	B
----	---

13	A
----	---

14	B
----	---

15	A
----	---

16	C
----	---

17	A
----	---

18	D
----	---

19	A
----	---

20	C
----	---

21	C
----	---

22	B
----	---

23	B
----	---

24	D
----	---

25	B
----	---

26	D
----	---

27	C
----	---

28	B
29	C
30	B
31	D
32	D

Loading [Mathjax]/jax/output/HTML-CSS/jax.js