

# DUMPS ARENA

## Cloudera Certified Administrator for Apache Hadoop (CCA-H)

Cloudera CCA-500

Version Demo

Total Demo Questions: 10

Total Premium Questions: 60

Buy Premium PDF

<https://dumpsarena.com>

[sales@dumpsarena.com](mailto:sales@dumpsarena.com)

dumpsarena.com

**QUESTION NO: 1**

On a cluster running CDH 5.0 or above, you use the `hadoop fs -put` command to write a 300MB file into a previously empty directory using an HDFS block size of 64 MB. Just after this command has finished writing 200 MB of this file, what would another user see when they look in the directory?

- A. The directory will appear to be empty until the entire file write is completed on the cluster
- B. They will see the file with a `._COPYING_` extension on its name. If they view the file, they will see contents of the file up to the last completed block (as each 64MB block is written, that block becomes available)
- C. They will see the file with a `._COPYING_` extension on its name. If they attempt to view the file, they will get a `ConcurrentFileAccessException` until the entire file write is completed on the cluster
- D. They will see the file with its original name. If they attempt to view the file, they will get a `ConcurrentFileAccessException` until the entire file write is completed on the cluster

**ANSWER: B****QUESTION NO: 2**

A slave node in your cluster has 4 TB hard drives installed (4 x 2TB). The DataNode is configured to store HDFS blocks on all disks. You set the value of the `dfs.datanode.du.reserved` parameter to 100 GB. How does this alter HDFS block storage?

- A. 25GB on each hard drive may not be used to store HDFS blocks
- B. 100GB on each hard drive may not be used to store HDFS blocks
- C. All hard drives may be used to store HDFS blocks as long as at least 100 GB in total is available on the node
- D. A maximum of 100 GB on each hard drive may be used to store HDFS blocks

**ANSWER: C****QUESTION NO: 3**

You are configuring your cluster to run HDFS and MapReducer v2 (MRv2) on YARN. Which two daemons need to be installed on your cluster's master nodes? (Choose two)

- A. HMaster
- B. ResourceManager
- C. TaskManager
- D. JobTracker
- E. NameNode

F. DataNode

**ANSWER: B E**

#### QUESTION NO: 4

Which is the default scheduler in YARN?

- A. YARN doesn't configure a default scheduler, you must first assign an appropriate scheduler class in yarn-site.xml
- B. Capacity Scheduler
- C. Fair Scheduler
- D. FIFO Scheduler

**ANSWER: B**

#### QUESTION NO: 5

In CDH4 and later, which file contains a serialized form of all the directory and files inodes in the filesystem, giving the NameNode a persistent checkpoint of the filesystem metadata?

- A. fstime
- B. VERSION
- C. Fsimage\_N (where N reflects transactions up to transaction ID N)
- D. Edits\_N-M (where N-M transactions between transaction ID N and transaction ID N)

**ANSWER: C**

#### QUESTION NO: 6

During the execution of a MapReduce v2 (MRv2) job on YARN, where does the Mapper place the intermediate data of each Map Task?

- A. The Mapper stores the intermediate data on the node running the Job's ApplicationMaster so that it is available to YARN ShuffleService before the data is presented to the Reducer
- B. The Mapper stores the intermediate data in HDFS on the node where the Map tasks ran in the HDFS /usercache/&(user)/apache/application\_&(appid) directory for the user who ran the job
- C. The Mapper transfers the intermediate data immediately to the reducers as it is generated by the Map Task
- D. YARN holds the intermediate data in the NodeManager's memory (a container) until it is transferred to the Reducer

E. The Mapper stores the intermediate data on the underlying filesystem of the local disk in the directories `yarn.nodemanager.local-DIFS`

**ANSWER: E**

#### QUESTION NO: 7

Choose three reasons why should you run the HDFS balancer periodically? (Choose three)

- A. To ensure that there is capacity in HDFS for additional data
- B. To ensure that all blocks in the cluster are 128MB in size
- C. To help HDFS deliver consistent performance under heavy loads
- D. To ensure that there is consistent disk utilization across the DataNodes
- E. To improve data locality MapReduce

**ANSWER: C D E**

#### QUESTION NO: 8

What two processes must you do if you are running a Hadoop cluster with a single NameNode and six DataNodes, and you want to change a configuration parameter so that it affects all six DataNodes. (Choose two)

- A. You must modify the configuration files on the NameNode only. DataNodes read their configuration from the master nodes
- B. You must modify the configuration files on each of the DataNodes machines
- C. You don't need to restart any daemon, as they will pick up changes automatically
- D. You must restart the NameNode daemon to apply the changes to the cluster
- E. You must restart all six DataNode daemon to apply the changes to the cluster

**ANSWER: A D**

#### QUESTION NO: 9

Which two are features of Hadoop's rack topology? (Choose two)

- A. Configuration of rack awareness is accomplished using a configuration file. You cannot use a rack topology script.
- B. Hadoop gives preference to intra-rack data transfer in order to conserve bandwidth
- C. Rack location is considered in the HDFS block placement policy

D. HDFS is rack aware but MapReduce daemon are not

E. Even for small clusters on a single rack, configuring rack awareness will improve performance

**ANSWER: B C**

**QUESTION NO: 10**

You suspect that your NameNode is incorrectly configured, and is swapping memory to disk. Which Linux commands help you to identify whether swapping is occurring? (Select all that apply)

A. free

B. df

C. memcat

D. top

E. jps

F. vmstat

G. swapinfo

**ANSWER: A D F**