

**100%** Money Back  
**Guarantee**

**Vendor:**Cloudera

**Exam Code:**CCA-505

**Exam Name:**Cloudera Certified Administrator for  
Apache Hadoop (CCAH) CDH5 Upgrade Exam

**Version:**Demo

### QUESTION 1

You want to understand more about how users browse your public website. For example, you want to know which pages they visit prior to placing an order. You have a server farm of 200 web servers hosting your website. Which is the most efficient process to gather these web server logs into your Hadoop cluster for analysis?

- A. Sample the web server logs from web servers and copy them into HDFS using curl
- B. Ingest the server web logs into HDFS using Flume
- C. Import all users clicks from your OLTP databases into Hadoop using Sqoop
- D. Write a MapReduce job with the web servers from mappers and the Hadoop cluster nodes reducers
- E. Channel these clickstream into Hadoop using Hadoop Streaming

Correct Answer: AB

---

### QUESTION 2

Your cluster's `mapred-site.xml` includes the following parameters

`mapreduce.map.memory.mb`

4096

`mapreduce.reduce.memory.mb`

8192

And your cluster's `yarn-site.xml` includes the following parameters

`yarn.nodemanager/vmem-pmem-ratio`

2.1

What is the maximum amount of virtual memory allocated for each map before YARN will kill its Container?

- A. 4 GB
- B. 17.2 GB
- C. 24.6 GB
- D. 8.2 GB

Correct Answer: D

---

### QUESTION 3

Your cluster has the following characteristics:

A rack aware topology is configured and on

Replication is not set to 3

Cluster block size is set to 64 MB

Which describes the file read process when a client application connects into the cluster and requests a 50MB file?

A. The client queries the NameNode which retrieves the block from the nearest DataNode to the client and then passes that block back to the client.

B. The client queries the NameNode for the locations of the block, and reads from a random location in the list it retrieves to eliminate network I/O leads by balancing which nodes it retrieves data from at any given time.

C. The client queries the NameNode for the locations of the block, and reads all three copies. The first copy to complete transfer to the client is the one the client reads as part of Hadoop's

speculative execution framework.

D. The client queries the NameNode for the locations of the block, and reads from the first location in the list it receives.

Correct Answer: A

---

#### QUESTION 4

You are running a Hadoop cluster with MapReduce version 2 (MRv2) on YARN. You consistently see that MapReduce map tasks on your cluster are running slowly because of excessive garbage collection of JVM, how do you increase JVM heap property to 3GB to optimize performance?

A. `Yarn.application.child.java.opts=-Xax3072m`

B. `Yarn.application.child.java.opts=-3072m`

C. `Mapreduce.map.java.opts=-Xmx3072m`

D. `Mapreduce.map.java.opts=-Xms3072m`

Correct Answer: C

---

#### QUESTION 5

You decide to create a cluster which runs HDFS in High Availability mode with automatic failover, using Quorum-based Storage. What is the purpose of ZooKeeper in such a configuration?

A. It manages the Edits file, which is a log changes to the HDFS filesystem.

B. It monitors an NFS mount point and reports if the mount point disappears

C. It both keeps track of which NameNode is Active at any given time, and manages the Edits file, which is a log of changes to the HDFS filesystem

D. It only keeps track of which NameNode is Active at any given time

E. Clients connect to ZoneKeeper to determine which NameNode is Active

Correct Answer: D

---

### QUESTION 6

You are upgrading a Hadoop cluster from HDFS and MapReduce version 1 (MRv1) to one running HDFS and MapReduce version 2 (MRv2) on YARN. You want to set and enforce a block of 128MB for all new files written to the cluster after the upgrade. What should you do?

- A. Set `dfs.block.size` to 128M on all the worker nodes, on all client machines, and on the NameNode, and set the parameter to final.
- B. Set `dfs.block.size` to 134217728 on all the worker nodes, on all client machines, and on the NameNode, and set the parameter to final.
- C. Set `dfs.block.size` to 134217728 on all the worker nodes and client machines, and set the parameter to final. You do need to set this value on the NameNode.
- D. Set `dfs.block.size` to 128M on all the worker nodes and client machines, and set the parameter to final. You do need to set this value on the NameNode.
- E. You cannot enforce this, since client code can always override this value.

Correct Answer: C

---

### QUESTION 7

Assuming a cluster running HDFS, MapReduce version 2 (MRv2) on YARN with all settings at their default, what do you need to do when adding a new slave node to a cluster?

- A. Nothing, other than ensuring that DNS (or `/etc/hosts` files on all machines) contains an entry for the new node.
- B. Restart the NameNode and ResourceManager daemons and resubmit any running jobs
- C. Increase the value of `dfs.number.of.replicas` in `hdfs-site.xml`
- D. Add a new entry to `/etc/nodes` on the NameNode host.
- E. Restart the NameNode daemon.

Correct Answer: B

---

### QUESTION 8

Assume you have a file named `foo.txt` in your local directory. You issue the following three commands:

```
Hadoop fs mkdir input Hadoop fs put foo.txt input/foo.txt Hadoop fs put foo.txt input
```

What happens when you issue that third command?

- A. The write succeeds, overwriting foo.txt in HDFS with no warning
- B. The write silently fails
- C. The file is uploaded and stored as a plain named input
- D. You get an error message telling you that input is not a directory
- E. You get a error message telling you that foo.txt already exists. The file is not written to HDFS
- F. You get an error message telling you that foo.txt already exists, and asking you if you would like to overwrite
- G. You get a warning that foo.txt is being overwritten

Correct Answer: E

---

### QUESTION 9

On a cluster running MapReduce v2 (MRv2) on YARN, a MapReduce job is given a directory of 10 plain text as its input directory. Each file is made up of 3 HDFS blocks. How many Mappers will run?

- A. We cannot say; the number of Mappers is determined by the ResourceManager
- B. We cannot say; the number of Mappers is determined by the ApplicationManager
- C. We cannot say; the number of Mappers is determined by the developer
- D. 30
- E. 3
- F. 10

Correct Answer: E

---

### QUESTION 10

Which three basic configuration parameters must you set to migrate your cluster from MapReduce1 (MRv1) to MapReduce v2 (MRv2)?

- A. Configure the NodeManager hostname and enable services on YARN by setting the following property in yarn-site.xml: `yarn.nodemanager.hostname your_nodeManager_hostname`
- B. Configure the number of map tasks per job on YARN by setting the following property in mapredsite.xml: `mapreduce.job.maps 2`
- C. Configure MapReduce as a framework running on YARN by setting the following property in mapredsite.xml: `mapreduce.framework.name yarn`
- D. Configure the ResourceManager hostname and enable node services on YARN by setting the following property in yarn-site.xml: `yarn.resourcemanager.hostname your_responseManager_hostname`
- E. Configure a default scheduler to run on YARN by setting the following property in sapred-site.xml: `mapreduce.jobtracker.taskScheduler org.apache.hadoop.mapred.JobQueueTaskScheduler`

F. Configure the NodeManager to enable MapReduce services on YARN by adding following property in yarn-site.xml:  
yarn.nodemanager.aux-services mapreduce\_shuffle

Correct Answer: ABD

---

### QUESTION 11

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

- A. The Mapper stores the intermediate data on the node running the job's ApplicationMaster so that is available to YARN's 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/and[user]sppcache/application\_and(appid) directory for the user who ran the job
- C. YARN holds the intermediate data in the NodeManager's memory (a container) until it is transferred to the Reducers
- D. The Mapper stores the intermediate data on the underlying filesystem of the local disk in the directories yarn.nodemanager.local-dirs
- E. The Mapper transfers the intermediate data immediately to the Reducers as it generated by the Map task

Correct Answer: D

---

### QUESTION 12

You are the hadoop fs put command to add a file "sales.txt" to HDFS. This file is small enough that it fits into a single block, which is replicated to three nodes in your cluster (with a replication factor of 3). One of the nodes holding this file (a single block) fails. How will the cluster handle the replication of this file in this situation/

- A. The cluster will re-replicate the file the next time the system administrator reboots the NameNode daemon (as long as the file's replication doesn't fall two)
- B. This file will be immediately re-replicated and all other HDFS operations on the cluster will halt until the cluster's replication values are restored
- C. The file will remain under-replicated until the administrator brings that nodes back online
- D. The file will be re-replicated automatically after the NameNode determines it is under replicated based on the block reports it receives from the DataNodes

Correct Answer: B