

Cloudera CCA-505 Exam

Volume: 45 Questions

Question No : 1

Your cluster implements HDFS High Availability (HA). Your two NameNodes are named nn01 and nn02. What occurs when you execute the command: `hdfs haadmin -failover nn01 nn02`

- A. nn02 becomes the standby NameNode and nn01 becomes the active NameNode
- B. nn02 is fenced, and nn01 becomes the active NameNode
- C. nn01 becomes the standby NameNode and nn02 becomes the active NameNode
- D. nn01 is fenced, and nn02 becomes the active NameNode

Answer: D

Question No : 2

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 3)

- A. free
- B. df
- C. memcat
- D. top
- E. vmstat
- F. swapinfo

Answer: A,D,E

Question No : 3

Your Hadoop cluster contains nodes in three racks. You have NOT configured the `dfs.hosts` property in the NameNode's configuration file. What results?

- A. No new nodes can be added to the cluster until you specify them in the `dfs.hosts` file

Cloudera CCA-505 Exam

- B. Presented with a blank `dfs.hosts` property, the NameNode will permit DatNode specified in `mapred.hosts` to join the cluster
- C. Any machine running the DataNode daemon can immediately join the cluster
- D. The NameNode will update the `dfs.hosts` property to include machine running DataNode daemon on the next NameNode reboot or with the command `dfsadmin -refreshNodes`

Answer: C

Question No : 4

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/&[user]spcache/application_&(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

Answer: D

Question No : 5

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`

Cloudera CCA-505 Exam

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

Answer: C

Question No : 6

You have installed a cluster running HDFS and MapReduce version 2 (MRv2) on YARN. You have no `dfs.hosts` entry()ies in your `hdfs-alte.xml` configuration file. You configure a new worker node by setting `fs.default.name` in its configuration files to point to the NameNode on your cluster, and you start the DataNode daemon on that worker node.

What do you have to do on the cluster to allow the worker node to join, and start storing HDFS blocks?

- A. Nothing; the worker node will automatically join the cluster when the DataNode daemon is started.
- B. Without creating a `dfs.hosts` file or making any entries, run the command `hadoop dfsadmin -refreshHadoop` on the NameNode
- C. Create a `dfs.hosts` file on the NameNode, add the worker node's name to it, then issue the command `hadoop dfsadmin -refreshNodes` on the NameNode
- D. Restart the NameNode

Answer: B

Question No : 7

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 of 64MB. Just after this command has finished writing 200MB of this file, what would another user see when they look in the directory?

- A. 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
- B. 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.
- C. 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)
- D. The directory will appear to be empty until the entire file write is completed on the cluster

Answer: C

Cloudera CCA-505 Exam

Question No : 8

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 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

Answer: A,B

Question No : 9

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 an 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

Answer: E

Cloudera CCA-505 Exam

Question No : 10

You are migrating a cluster from MapReduce version 1 (MRv1) to MapReduce version2 (MRv2) on YARN. To want to maintain your MRv1 TaskTracker slot capacities when you migrate. What should you do?

- A. Configure `yarn.applicationmaster.resource.memory-mb` and `yarn.applicationmaster.cpu-vcores` so that ApplicationMaster container allocations match the capacity you require.
- B. You don't need to configure or balance these properties in YARN as YARN dynamically balances resource management capabilities on your cluster
- C. Configure `yarn.nodemanager.resource.memory-mb` and `yarn.nodemanager.resource.cpu-vcores` to match the capacity you require under YARN for each NodeManager
- D. Configure `mapred.tasktracker.map.tasks.maximum` and `mapred.tasktracker.reduce.tasks.maximum` in `yarn.site.xml` to match your cluster's configured capacity set by `yarn.scheduler.minimum-allocation`

Answer: C

Question No : 11

You have a Hadoop cluster running HDFS, and a gateway machine external to the cluster from which clients submit jobs. What do you need to do in order to run on the cluster and submit jobs from the command line of the gateway machine?

- A. Install the `impslad` daemon, `statestored` daemon, and `catalogd` daemon on each machine in the cluster and on the gateway node
- B. Install the `impalad` daemon on each machine in the cluster, the `statestored` daemon and `catalogd` daemon on one machine in the cluster, and the `impala` shell on your gateway machine
- C. Install the `impalad` daemon and the `impala` shell on your gateway machine, and the `statestored` daemon and `catalog` daemon on one of the nodes in the cluster
- D. Install the `impalad` daemon, the `statestored` daemon, the `catalogd` daemon, and the `impala` shell on your gateway machine
- E. Install the `impalad` daemon, `statestored` daemon, and `catalogd` daemon on each machine in the cluster, and the `impala` shell on your gateway machine

Answer: B

Question No : 12