

100% Money Back
Guarantee

Vendor:Cloudera

Exam Code:CCA-500

Exam Name:Cloudera Certified Administrator for
Apache Hadoop (CCAHA)

Version:Demo

QUESTION 1

Your cluster is running MapReduce version 2 (MRv2) on YARN. Your ResourceManager is configured to use the FairScheduler. Now you want to configure your scheduler such that a new user on the cluster can submit jobs into their own queue application submission. Which configuration should you set?

- A. You can specify new queue name when user submits a job and new queue can be created dynamically if the property `yarn.scheduler.fair.allow-undecleared-pools = true`
- B. `Yarn.scheduler.fair.user.fair-as-default-queue = false` and `yarn.scheduler.fair.allow-undecleared-pools = true`
- C. You can specify new queue name when user submits a job and new queue can be created dynamically if `yarn.scheduler.fair.user-as-default-queue = false`
- D. You can specify new queue name per application in `allocations.xml` file and have new jobs automatically assigned to the application queue

Correct Answer: A

QUESTION 2

Which command does Hadoop offer to discover missing or corrupt HDFS data?

- A. `Hdfs fs du`
- B. `Hdfs fsck`
- C. `Dskchk`
- D. The map-only checksum
- E. Hadoop does not provide any tools to discover missing or corrupt data; there is not need because three replicas are kept for each data block

Correct Answer: B

QUESTION 3

You are planning a Hadoop cluster and considering implementing 10 Gigabit Ethernet as the network fabric. Which workloads benefit the most from faster network fabric?

- A. When your workload generates a large amount of output data, significantly larger than the amount of intermediate data
- B. When your workload consumes a large amount of input data, relative to the entire capacity of HDFS
- C. When your workload consists of processor-intensive tasks
- D. When your workload generates a large amount of intermediate data, on the order of the input data itself

Correct Answer: A

QUESTION 4

You have recently converted your Hadoop cluster from a MapReduce 1 (MRv1) architecture to MapReduce 2 (MRv2) on YARN architecture. Your developers are accustomed to specifying map and reduce tasks (resource allocation) tasks when they run jobs: A developer wants to know how specify to reduce tasks when a specific job runs. Which method should you tell that developers to implement?

- A. MapReduce version 2 (MRv2) on YARN abstracts resource allocation away from the idea of "tasks" into memory and virtual cores, thus eliminating the need for a developer to specify the number of reduce tasks, and indeed preventing the developer from specifying the number of reduce tasks.
- B. In YARN, resource allocations is a function of megabytes of memory in multiples of 1024mb. Thus, they should specify the amount of memory resource they need by executing `D mapreducereduces.memory-mb-2048`
- C. In YARN, the ApplicationMaster is responsible for requesting the resource required for a specific launch. Thus, executing `D yarn.applicationmaster.reduce.tasks=2` will specify that the ApplicationMaster launch two task contains on the worker nodes.
- D. Developers specify reduce tasks in the exact same way for both MapReduce version 1 (MRv1) and MapReduce version 2 (MRv2) on YARN. Thus, executing `D mapreduce.job.reduces-2` will specify reduce tasks.
- E. In YARN, resource allocation is function of virtual cores specified by the ApplicationManager making requests to the NodeManager where a reduce task is handled by a single container (and thus a single virtual core). Thus, the developer needs to specify the number of virtual cores to the NodeManager by executing `p yarn.nodemanager.cpu-vcores=2`

Correct Answer: D

QUESTION 5

You observed that the number of spilled records from Map tasks far exceeds the number of map output records. Your child heap size is 1GB and your `io.sort.mb` value is set to 1000MB. How would you tune your `io.sort.mb` value to achieve maximum memory to disk I/O ratio?

- A. For a 1GB child heap size an `io.sort.mb` of 128 MB will always maximize memory to disk I/O
- B. Increase the `io.sort.mb` to 1GB
- C. Decrease the `io.sort.mb` value to 0
- D. Tune the `io.sort.mb` value until you observe that the number of spilled records equals (or is as close to equals) the number of map output records.

Correct Answer: D

QUESTION 6

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 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

Correct Answer: B

QUESTION 7

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

Correct Answer: B

QUESTION 8

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/and(user)/apache/application_and(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`

Correct Answer: E

QUESTION 9

You are configuring your cluster to run HDFS and MapReducer v2 (MRv2) on YARN. Which two daemons needs to be

installed on your cluster's master nodes? (Choose two)

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

Correct Answer: BE

QUESTION 10

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 across logs into your Hadoop cluster 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. Channel these clickstreams into Hadoop using Hadoop Streaming
- D. Import all user clicks from your OLTP databases into Hadoop using Sqoop
- E. Write a MapReduce job with the web servers for mappers and the Hadoop cluster nodes for reducers

Correct Answer: B

QUESTION 11

On a cluster running MapReduce v2 (MRv2) on YARN, a MapReduce job is given a directory of 10 plain text files 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 developer
- C. 30
- D. 3
- E. 10
- F. We cannot say; the number of mappers is determined by the ApplicationMaster

Correct Answer: E

QUESTION 12

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 cluster?

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

Correct Answer: A