



Vendor: Cloudera

Exam Code: CCA-505

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

Version: DEMO

QUESTION 1

You are configuring your cluster to run HDFS and MapReduce v2 (MRv2) on YARN.
Which daemons need to be installed on your clusters master nodes?

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

Answer: ADE

QUESTION 2

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

Answer: FG

QUESTION 3

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: C

QUESTION 4

You have converted your Hadoop cluster from a MapReduce 1 (MRv1) architecture to a 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 developer to implement?

- A. Developers specify reduce tasks in the exact same way for both MapReduce version 1 (MRv1) and MapReduce version 2 (MRv2) on YARN. Thus, executing `-p mapreduce.job.reduce-2` will specify 2 reduce tasks.
- B. In YARN, the ApplicationMaster is responsible for requesting the resources required for a specific job. Thus, executing `-p yarn.applicationmaster.reduce.tasks-2` will specify that the ApplicationMaster launch two task containers on the worker nodes.
- C. In YARN, resource allocation is a function of megabytes of memory in multiple of 1024mb. Thus, they should specify the amount of memory resource they need by executing `-D mapreduce.reduce.memory-mp-2040`
- D. In YARN, resource allocation is a function of virtual cores specified by the ApplicationMaster making requests to the NodeManager where a reduce task is handled by a single container (and this 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`
- E. 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.

Answer: D

QUESTION 5

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.

Answer: C

QUESTION 6

Which two are Features of Hadoop's rack topology?

- A. Configuration of rack awareness is accomplished using a configuration file. You cannot use a rack topology script.
- B. Even for small clusters on a single rack, configuring rack awareness will improve performance.

- C. Rack location is considered in the HDFS block placement policy
- D. HDFS is rack aware but MapReduce daemons are not
- E. Hadoop gives preference to Intra rack data transfer in order to conserve bandwidth

Answer: BC

QUESTION 7

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: AB

QUESTION 8

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

Explanation:

failover - initiate a failover between two NameNodes This subcommand causes a failover from the first provided NameNode to the second. If the first NameNode is in the Standby state, this command simply transitions the second to the Active state without error. If the first NameNode is in the Active state, an attempt will be made to gracefully transition it to the Standby state. If this fails, the fencing methods (as configured by `dfs.ha.fencing.methods`) will be attempted in order until one of the methods succeeds. Only after this process will the second NameNode be transitioned to the Active state. If no fencing method succeeds, the second NameNode will not be transitioned to the Active state, and an error will be returned.

QUESTION 9

Your Hadoop cluster is configured with HDFS and MapReduce version 2 (MRv2) on YARN. Can you configure a worker node to run a NodeManager daemon but not a DataNode daemon and still have a functional cluster?

- A. Yes. The daemon will receive data from the NameNode to run Map tasks
- B. Yes. The daemon will get data from another (non-local) DataNode to run Map tasks
- C. Yes. The daemon will receive Reduce tasks only

Answer: A

QUESTION 10

Which YARN process runs as "controller O" of a submitted job and is responsible for resource requests?

- A. ResourceManager
- B. NodeManager
- C. JobHistoryServer
- D. ApplicationMaster
- E. JobTracker
- F. ApplicationManager

Answer: E

QUESTION 11

You have a cluster running with the Fair Scheduler enabled. There are currently no jobs running on the cluster, and you submit a job A, so that only job A is running on the cluster. A while later, you submit Job B. now job A and Job B are running on the cluster at the same time. How will the Fair Scheduler handle these two jobs?

- A. When job A gets submitted, it consumes all the tasks slots.
- B. When job A gets submitted, it doesn't consume all the task slots
- C. When job B gets submitted, Job A has to finish first, before job B can scheduled
- D. When job B gets submitted, it will get assigned tasks, while Job A continue to run with fewer tasks.

Answer: C

QUESTION 12

You observe 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 100 MB. How would you tune your io.sort.mb value to achieve maximum memory to disk I/O ratio?

- A. Decrease the io.sort.mb value to 0
- B. Increase the io.sort.mb to 1GB
- C. For 1GB child heap size an io.sort.mb of 128 MB will always maximize memory to disk I/O
- 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

Answer: D

Thank You for Trying Our Product

Lead2pass Certification Exam Features:

- ★ More than **99,900** Satisfied Customers Worldwide.
- ★ Average **99.9%** Success Rate.
- ★ **Free Update** to match latest and real exam scenarios.
- ★ **Instant Download** Access! No Setup required.
- ★ Questions & Answers are downloadable in **PDF** format and **VCE** test engine format.
- ★ Multi-Platform capabilities - **Windows, Laptop, Mac, Android, iPhone, iPod, iPad**.
- ★ **100%** Guaranteed Success or **100%** Money Back Guarantee.
- ★ **Fast**, helpful support **24x7**.



View list of all certification exams: <http://www.lead2pass.com/all-products.html>



Microsoft



ORACLE



CITRIX



JUNIPER
NETWORKS



EMC²
where information lives®

10% Discount Coupon Code: ASTR14