

CCA-505^{Q&As}

Cloudera Certified Administrator for Apache Hadoop (CCA) CDH5
Upgrade Exam

Pass Cloudera CCA-505 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.leads4pass.com/cca-505.html>

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Cloudera
Official Exam Center

- ⚙️ **Instant Download** After Purchase
- ⚙️ **100% Money Back** Guarantee
- ⚙️ **365 Days** Free Update
- ⚙️ **800,000+** Satisfied Customers



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

You want a node to only swap Hadoop daemon data from RAM to disk when absolutely necessary. What should you do?

- A. Delete the /swapfile file on the node
- B. Set vm.swappiness to 0 in /etc/sysctl.conf
- C. Set the ram.swap parameter to 0 in core-site.xml
- D. Delete the /etc/swap file on the node
- E. Delete the /dev/vmswap file on the node

Correct Answer: B

QUESTION 3

You have a 20 node Hadoop cluster, with 18 slave nodes and 2 master nodes running HDFS High Availability (HA). You want to minimize the chance of data loss in your cluster. What should you do?

- A. Add another master node to increase the number of nodes running the JournalNode which increases the number of machines available to HA to create a quorum
- B. Configure the cluster's disk drives with an appropriate fault tolerant RAID level
- C. Run the ResourceManager on a different master from the NameNode in the order to load share HDFS metadata processing
- D. Run a Secondary NameNode on a different master from the NameNode in order to load provide automatic recovery from a NameNode failure

E. Set an HDFS replication factor that provides data redundancy, protecting against failure

Correct Answer: C

QUESTION 4

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 5

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

Correct Answer: D

[CCA-505 PDF Dumps](#)

[CCA-505 Study Guide](#)

[CCA-505 Exam Questions](#)