

CCA-500 Dumps

Cloudera Certified Administrator for Apache Hadoop (CCA-H)

<https://www.certleader.com/CCA-500-dumps.html>



NEW QUESTION 1

Given:

```
[user1@host1 ~] yarn application -list
```

Total Applications: 3

Application ID	Application-Name	Application-Type	User	Queue	State	Final-State	Progress	Tracking
Application_1374638600275_0109	Sleep Job	MAPREDUCE	user1	KILLED	KILLED	KILLED	100%	host1:54059
Application_1374638600275_0121	Sleep Job	MAPREDUCE	user1	FINISHED	SUCCEEDED	SUCCEEDED	100%	host1:19888/Jobhistory/Job_1374638600275_0121
Application_1374638600275_0020	Sleep Job	MAPREDUCE	user1	FINISHED	SUCCEEDED	SUCCEEDED	100%	host1:19888/Jobhistory/Job_1374638600275_0020

You want to clean up this list by removing jobs where the State is KILLED. What command you enter?

- A. Yarn application --refreshJobHistory
- B. Yarn application --kill application_1374638600275_0109
- C. Yarn rmdadmin --refreshQueue
- D. Yarn rmdadmin --kill application_1374638600275_0109

Answer: B

Explanation: Reference:http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.1-latest/bk_using-apache-hadoop/content/common_mrv2_commands.html

NEW QUESTION 2

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

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

Answer: D

NEW QUESTION 3

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.

Answer: D

NEW QUESTION 4

A slave node in your cluster has 4 TB hard drives installed (4 x 2TB). The DataNode is configured to store HDFS blocks on all disks. You set the value of the dfs.datanode.du.reserved parameter to 100 GB. How does this alter HDFS block storage?

- A. 25GB on each hard drive may not be used to store HDFS blocks
- B. 100GB on each hard drive may not be used to store HDFS blocks
- C. All hard drives may be used to store HDFS blocks as long as at least 100 GB in total is available on the node
- D. A maximum of 100 GB on each hard drive may be used to store HDFS blocks

Answer: B

NEW QUESTION 5

You are running a Hadoop cluster with a NameNode on host mynamenode, a secondary NameNode on host mysecondarynamenode and several DataNodes. Which best describes how you determine when the last checkpoint happened?

- A. Execute hdfs namenode --report on the command line and look at the Last Checkpoint information
- B. Execute hdfs dfsadmin --saveNamespace on the command line which returns to you the last checkpoint value in fstime file
- C. Connect to the web UI of the Secondary NameNode (<http://mysecondary:50090/>) and look at the "Last Checkpoint" information
- D. Connect to the web UI of the NameNode (<http://mynamenode:50070/>) and look at the "Last Checkpoint" information

Answer: C

Explanation: Reference:<https://www.inkling.com/read/hadoop-definitive-guide-tom-white-3rd/chapter-10/hdfs>

NEW QUESTION 6

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

Answer: E

NEW QUESTION 7

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. The NameNode will update the dfs.hosts property to include machines running the DataNode daemon on the next NameNode reboot or with the command `dfsadmin–refreshNodes`
- B. No new nodes can be added to the cluster until you specify them in the dfs.hosts file
- C. Any machine running the DataNode daemon can immediately join the cluster
- D. Presented with a blank dfs.hosts property, the NameNode will permit DataNodes specified in `mapred.hosts` to join the cluster

Answer: C

NEW QUESTION 8

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

Answer: BE

NEW QUESTION 9

Which YARN daemon or service monitors a Controller's per-application resource using (e.g., memory CPU)?

- A. ApplicationMaster
- B. NodeManager
- C. ApplicationManagerService
- D. ResourceManager

Answer: A

NEW QUESTION 10

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 the third command?

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

Answer: CE

NEW QUESTION 10

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 use 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 nam
- C. 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. They will see the file with a `._COPYING_` extension on its nam
- E. If they attempt to view the file, they will get a `ConcurrentFileAccessException` until the entire file write is completed on the cluster
- F. They will see the file with its original nam
- G. If they attempt to view the file, they will get a `ConcurrentFileAccessException` until the entire file write is completed on the cluster

Answer: B

NEW QUESTION 11

Which scheduler would you deploy to ensure that your cluster allows short jobs to finish within a reasonable time without starting long-running jobs?

- A. Complexity Fair Scheduler (CFS)

- B. Capacity Scheduler
- C. Fair Scheduler
- D. FIFO Scheduler

Answer: C

Explanation: Reference:http://hadoop.apache.org/docs/r1.2.1/fair_scheduler.html

NEW QUESTION 16

For each YARN job, the Hadoop framework generates task log file. Where are Hadoop task log files stored?

- A. Cached by the NodeManager managing the job containers, then written to a log directory on the NameNode
- B. Cached in the YARN container running the task, then copied into HDFS on job completion
- C. In HDFS, in the directory of the user who generates the job
- D. On the local disk of the slave node running the task

Answer: D

NEW QUESTION 17

In CDH4 and later, which file contains a serialized form of all the directory and files inodes in the filesystem, giving the NameNode a persistent checkpoint of the filesystem metadata?

- A. fstime
- B. VERSION
- C. Fsimage_N (where N reflects transactions up to transaction ID N)
- D. Edits_N-M (where N-M transactions between transaction ID N and transaction ID N)

Answer: C

Explanation: Reference:<http://mikepluta.com/tag/namenode/>

NEW QUESTION 18

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 no need because three replicas are kept for each data block

Answer: B

Explanation: Reference:<https://twiki.grid.iu.edu/bin/view/Storage/HadoopRecovery>

NEW QUESTION 22

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?(Choose two)

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

Answer: B

NEW QUESTION 27

You are running a Hadoop cluster with a NameNode on host mynamenode. What are two ways to determine available HDFS space in your cluster?

- A. Run hdfs fs -du / and locate the DFS Remaining value
- B. Run hdfs dfsadmin -report and locate the DFS Remaining value
- C. Run hdfs dfs / and subtract NDFS Used from configured Capacity
- D. Connect to <http://mynamenode:50070/dfshealth.jsp> and locate the DFS remaining value

Answer: B

NEW QUESTION 30

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 .schedule.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

Answer: A

NEW QUESTION 34

Which YARN process run as “container 0” of a submitted job and is responsible for resource requests?

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

Answer: C

NEW QUESTION 35

You use 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 replicationfactor of 3). One of the nodes holding this file (a single block) fails. How will the cluster handle the replication of file in this situation?

- A. The file will remain under-replicated until the administrator brings that node back online
- B. The cluster will re-replicate the file the next time the system administrator reboots the NameNode daemon (as long as the file’s replication factor doesn’t fall below)
- C. This will be immediately re-replicated and all other HDFS operations on the cluster will halt until the cluster’s replication values are resorted
- 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 NameNodes

Answer: D

NEW QUESTION 36

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/&(user)/apache/application_&(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

Answer: E

NEW QUESTION 40

You are configuring a server running HDFS, MapReduce version 2 (MRv2) on YARN running Linux. How must you format underlying file system of each DataNode?

- A. They must be formatted as HDFS
- B. They must be formatted as either ext3 or ext4
- C. They may be formatted in any Linux file system
- D. They must not be formatted - - HDFS will format the file system automatically

Answer: B

NEW QUESTION 43

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

Answer: B

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.

NEW QUESTION 47

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

- A. Delete the /dev/vmswap file on the node
- B. Delete the /etc/swap file on the node

- C. Set the ram.swap parameter to 0 in core-site.xml
- D. Set vm.swapfile file on the node
- E. Delete the /swapfile file on the node

Answer: D

NEW QUESTION 52

You are running Hadoop cluster with all monitoring facilities properly configured. Which scenario will go undeselected?

- A. HDFS is almost full
- B. The NameNode goes down
- C. A DataNode is disconnected from the cluster
- D. Map or reduce tasks that are stuck in an infinite loop
- E. MapReduce jobs are causing excessive memory swaps

Answer: B

NEW QUESTION 56

You need to analyze 60,000,000 images stored in JPEG format, each of which is approximately 25 KB. Because you Hadoop cluster isn't optimized for storing and processing many small files, you decide to do the following actions:

1. Group the individual images into a set of larger files
2. Use the set of larger files as input for a MapReduce job that processes them directly with python using Hadoop streaming.

Which data serialization system gives the flexibility to do this?

- A. CSV
- B. XML
- C. HTML
- D. Avro
- E. SequenceFiles
- F. JSON

Answer: E

Explanation: Sequence files are block-compressed and provide direct serialization and deserialization of several arbitrary data types (not just text). Sequence files can be generated as the output of other MapReduce tasks and are an efficient intermediate representation for data that is passing from one MapReduce job to another.

NEW QUESTION 57

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

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

Answer: C

NEW QUESTION 62

Which two are features of Hadoop's rack topology?(Choose two)

- A. Configuration of rack awareness is accomplished using a configuration file
- B. You cannot use a rack topology script.
- C. Hadoop gives preference to intra-rack data transfer in order to conserve bandwidth
- D. Rack location is considered in the HDFS block placement policy
- E. HDFS is rack aware but MapReduce daemon are not
- F. Even for small clusters on a single rack, configuring rack awareness will improve performance

Answer: BC

NEW QUESTION 64

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 all that apply)

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

Answer: ADF

Explanation: Reference:<http://www.cyberciti.biz/faq/linux-check-swap-usage-command/>

NEW QUESTION 65

.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your CCA-500 Exam with Our Prep Materials Via below:

<https://www.certleader.com/CCA-500-dumps.html>