

## Exam Questions CCA-500

Cloudera Certified Administrator for Apache Hadoop (CCA-H)

<https://www.2passeasy.com/dumps/CCA-500/>



### 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](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

Which three basic configuration parameters must you set to migrate your cluster from MapReduce 1 (MRv1) to MapReduce V2 (MRv2)?(Choose three)

- A. Configure the NodeManager to enable MapReduce services on YARN by setting the following property in yarn-site.xml:<name>yarn.nodemanager.hostname</name><value>your\_nodeManager\_shuffle</value>
- B. Configure the NodeManager hostname and enable node services on YARN by setting the following property in yarn-site.xml:<name>yarn.nodemanager.hostname</name><value>your\_nodeManager\_hostname</value>
- C. Configure a default scheduler to run on YARN by setting the following property in mapred-site.xml:<name>mapreduce.jobtracker.taskScheduler</name><Value>org.apache.hadoop.mapred.JobQueueTaskScheduler</value>
- D. Configure the number of map tasks per jon YARN by setting the following property in mapred:<name>mapreduce.job.maps</name><value>2</value>
- E. Configure the ResourceManager hostname and enable node services on YARN by setting the following property in yarn-site.xml:<name>yarn.resourcemanager.hostname</name><value>your\_resourceManager\_hostname</value>
- F. Configure MapReduce as a Framework running on YARN by setting the following property in mapred-site.xml:<name>mapreduce.framework.name</name><value>yarn</value>

**Answer:** AEF

### 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 if 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

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. Set an HDFS replication factor that provides data redundancy, protecting against node failure
- C. Run a Secondary NameNode on a different master from the NameNode in order to provide automatic recovery from a NameNode failure.
- D. Run the ResourceManager on a different master from the NameNode in order to load- share HDFS metadata processing
- E. Configure the cluster's disk drives with an appropriate fault tolerant RAID level

**Answer:** D

#### 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

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 9

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 access 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

**Answer:** B

**Explanation:** Apache Flume is a service for streaming logs into Hadoop.

Apache Flume is a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of streaming data into the Hadoop Distributed File System (HDFS). It has a simple and flexible architecture based on streaming data flows; and is robust and fault tolerant with tunable reliability mechanisms for failover and recovery.

#### NEW QUESTION 10

Cluster Summary:

45 files and directories, 12 blocks = 57 total. Heap size is 15.31 MB/193.38MB(7%)

Configured capacity	:	17.33GB
DFS Used	:	144KB
Non DFS Used	:	5.49GB
DFS Remaining	:	11.84GB
DFS Used %	:	0%
DFS Remaining %	:	68.32GB
Live Nodes	:	6
Dead Nodes	:	1
Decommissioning Nodes	:	0
Number of Under-Replicated Blocks	:	6

Refer to the above screenshot.

You configure a Hadoop cluster with seven DataNodes and one of your monitoring UIs displays the details shown in the exhibit. What does this tell you?

- A. The DataNode JVM on one host is not active
- B. Because your under-replicated blocks count matches the Live Nodes, one node is dead, and your DFS Used % equals 0%, you can't be certain that your cluster has all the data you've written it.
- C. Your cluster has lost all HDFS data which had blocks stored on the dead DataNode
- D. The HDFS cluster is in safe mode

**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 11

You are working on a project where you need to chain together MapReduce, Pig jobs. You also need the ability to use forks, decision points, and path joins. Which ecosystem project should you use to perform these actions?

- A. Oozie
- B. ZooKeeper
- C. HBase
- D. Sqoop
- E. HUE

**Answer:** A

#### NEW QUESTION 12

You decide to create a cluster which runs HDFS in High Availability mode with automatic failover, using Quorum Storage. What is the purpose of ZooKeeper in such a configuration?

- A. It only keeps track of which NameNode is Active at any given time
- B. It monitors an NFS mount point and reports if the mount point disappears
- C. It both keeps track of which NameNode is Active at any given time, and manages the Edits file
- D. Which is a log of changes to the HDFS filesystem
- E. It only manages the Edits file, which is log of changes to the HDFS filesystem
- F. Clients connect to ZooKeeper to determine which NameNode is Active



**Answer:** A

**Explanation:** Reference: Reference:<http://www.cloudera.com/content/cloudera-content/cloudera-docs/CDH4/latest/PDF/CDH4-High-Availability-Guide.pdf>(page 15)

#### NEW QUESTION 16

Each node in your Hadoop cluster, running YARN, has 64GB memory and 24 cores. Your yarn.site.xml has the following configuration:

```
<property>
<name>yarn.nodemanager.resource.memory-mb</name>
<value>32768</value>
</property>
<property>
<name>yarn.nodemanager.resource.cpu-vcores</name>
<value>12</value>
</property>
```

You want YARN to launch no more than 16 containers per node. What should you do?

- A. Modify yarn-site.xml with the following property:<name>yarn.scheduler.minimum-allocation-mb</name><value>2048</value>
- B. Modify yarn-sites.xml with the following property:<name>yarn.scheduler.minimum-allocation-mb</name><value>4096</value>
- C. Modify yarn-site.xml with the following property:<name>yarn.nodemanager.resource.cpu-vcores</name>
- D. No action is needed: YARN's dynamic resource allocation automatically optimizes the node memory and cores

**Answer:** A

#### NEW QUESTION 18

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 the 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
- 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 name
- 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 name
- 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 23

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](http://hadoop.apache.org/docs/r1.2.1/fair_scheduler.html)

#### NEW QUESTION 27

For each YARN job, the Hadoop framework generates task log files. 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 28

Your cluster's `mapred-start.xml` includes the following parameters

```
<name>mapreduce.map.memory.mb</name>
<value>4096</value>
<name>mapreduce.reduce.memory.mb</name>
<value>8192</value>
```

And any cluster's `yarn-site.xml` includes the following parameters

```
<name>yarn.nodemanager.vmem-pmem-ratio</name>
<value>2.1</value>
```

What is the maximum amount of virtual memory allocated for each map task before YARN will kill its Container?

- A. 4 GB
- B. 17.2 GB
- C. 8.9 GB
- D. 8.2 GB
- E. 24.6 GB

**Answer:** D

### NEW QUESTION 33

You have a cluster running with a FIFO scheduler enabled. You submit a large job A to the cluster, which you expect to run for one hour. Then, you submit job B to the cluster, which you expect to run a couple of minutes only.

You submit both jobs with the same priority.

Which two best describes how FIFO Scheduler arbitrates the cluster resources for job and its tasks?(Choose two)

- A. Because there is a more than a single job on the cluster, the FIFO Scheduler will enforce a limit on the percentage of resources allocated to a particular job at any given time
- B. Tasks are scheduled on the order of their job submission
- C. The order of execution of job may vary
- D. Given job A and submitted in that order, all tasks from job A are guaranteed to finish before all tasks from job B
- E. The FIFO Scheduler will give, on average, and equal share of the cluster resources over the job lifecycle
- F. The FIFO Scheduler will pass an exception back to the client when Job B is submitted, since all slots on the cluster are use

**Answer:** AD

### NEW QUESTION 36

Assuming you're not running HDFS Federation, what is the maximum number of NameNode daemons you should run on your cluster in order to avoid a "split-brain" scenario with your NameNode when running HDFS High Availability (HA) using Quorum- based storage?

- A. Two active NameNodes and two Standby NameNodes
- B. One active NameNode and one Standby NameNode
- C. Two active NameNodes and on Standby NameNode
- D. Unlimite
- E. HDFS High Availability (HA) is designed to overcome limitations on the number of NameNodes you can deploy

**Answer:** B

### NEW QUESTION 40

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

**Answer:** B

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

### NEW QUESTION 42

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 gets scheduled.
- C. When Job A gets submitted, it doesn't consumes all the task slots.
- D. When Job A gets submitted, it consumes all the task slots.

**Answer:** B

### NEW QUESTION 47

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 50

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 55

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 60

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 65

Choose three reasons why should you run the HDFS balancer periodically?(Choose three)

- A. To ensure that there is capacity in HDFS for additional data
- B. To ensure that all blocks in the cluster are 128MB in size
- C. To help HDFS deliver consistent performance under heavy loads
- D. To ensure that there is consistent disk utilization across the DataNodes
- E. To improve data locality MapReduce

**Answer:** CDE

**Explanation:** <http://www.quora.com/Apache-Hadoop/It-is-recommended-that-you-run-the-HDFS-balancer-periodically-Why-Choose-3>

#### NEW QUESTION 68

Your Hadoop cluster is configuring 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. Ye
- B. The daemon will receive data from the NameNode to run Map tasks
- C. Ye
- D. The daemon will get data from another (non-local) DataNode to run Map tasks
- E. Ye
- F. The daemon will receive Map tasks only
- G. Ye
- H. The daemon will receive Reducer tasks only

**Answer:** B

#### NEW QUESTION 71

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 72

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 73

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 77

A user comes to you, complaining that when she attempts to submit a Hadoop job, it fails. There is a Directory in HDFS named /data/input. The Jar is named j.jar, and the driver class is named DriverClass.

She runs the command:

Hadoop jar j.jar DriverClass /data/input/data/output The error message returned includes the line:

PrivilegedActionException as:training (auth:SIMPLE) cause:org.apache.hadoop.mapreduce.lib.input.InvalidInputException:

Input path does not exist: file:/data/input What is the cause of the error?

- A. The user is not authorized to run the job on the cluster
- B. The output directory already exists
- C. The name of the driver has been spelled incorrectly on the command line
- D. The directory name is misspelled in HDFS
- E. The Hadoop configuration files on the client do not point to the cluster

**Answer:** A

#### NEW QUESTION 78

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 80

.....



## THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual CCA-500 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the CCA-500 Product From:

<https://www.2passeasy.com/dumps/CCA-500/>

## Money Back Guarantee

### CCA-500 Practice Exam Features:

- \* CCA-500 Questions and Answers Updated Frequently
- \* CCA-500 Practice Questions Verified by Expert Senior Certified Staff
- \* CCA-500 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- \* CCA-500 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year