

## Exam Questions CCA-500

Cloudera Certified Administrator for Apache Hadoop (CCA-H)

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



#### NEW QUESTION 1

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 2

You have just run a MapReduce job to filter user messages to only those of a selected geographical region. The output for this job is in a directory named westUsers, located just below your home directory in HDFS. Which command gathers these into a single file on your local file system?

- A. Hadoop fs -getmerge -R westUsers.txt
- B. Hadoop fs -getemerge westUsers westUsers.txt
- C. Hadoop fs -cp westUsers/\* westUsers.txt
- D. Hadoop fs -get westUsers westUsers.txt

**Answer: B**

#### 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

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 6

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 7**

Which two features does Kerberos security add to a Hadoop cluster?(Choose two)

- A. User authentication on all remote procedure calls (RPCs)
- B. Encryption for data during transfer between the Mappers and Reducers
- C. Encryption for data on disk (“at rest”)
- D. Authentication for user access to the cluster against a central server
- E. Root access to the cluster for users hdfs and mapred but non-root access for clients

**Answer:** AD

**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**

Cluster Summary:

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

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

Refer to the above screenshot.

You configure a Hadoop cluster with seven DataNodes and on of your monitoring UIs displays the details shown in the exhibit.

What does the 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 bocks stored on the dead DatNode
- 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 10

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

**Answer:** A

**Explanation:** [http://wiki.apache.org/hadoop/FAQ#I\\_have\\_a\\_new\\_node\\_I\\_want\\_to\\_add\\_to\\_a\\_running\\_Hadoop\\_cluster.3B\\_how\\_do\\_I\\_start\\_services\\_on\\_just\\_one\\_node.3F](http://wiki.apache.org/hadoop/FAQ#I_have_a_new_node_I_want_to_add_to_a_running_Hadoop_cluster.3B_how_do_I_start_services_on_just_one_node.3F)

#### NEW QUESTION 14

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 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 19

Which is the default scheduler in YARN?

- A. YARN doesn't configure a default scheduler, you must first assign an appropriate scheduler class in yarn-site.xml
- B. Capacity Scheduler
- C. Fair Scheduler
- D. FIFO Scheduler

**Answer:** B

**Explanation:** Reference:<http://hadoop.apache.org/docs/r2.4.1/hadoop-yarn/hadoop-yarn-site/CapacityScheduler.html>

#### NEW QUESTION 20

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
- 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 22

Your company stores user profile records in an OLTP databases. You want to join these records with web server logs you have already ingested into the Hadoop file system. What is the best way to obtain and ingest these user records?

- A. Ingest with Hadoop streaming
- B. Ingest using Hive's IQAD DATA command
- C. Ingest with sqoop import
- D. Ingest with Pig's LOAD command
- E. Ingest using the HDFS put command

**Answer: C**

#### NEW QUESTION 26

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 29

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 1024m
- C. Thus, they should specify the amount of memory resource they need by executing `-D mapreduce-reduces.memory-mb-2048`
- D. In YARN, the ApplicationMaster is responsible for requesting the resource required for a specific launc
- E. Thus, executing `-D yarn.applicationmaster.reduce.tasks=2` will specify that the ApplicationMaster launch two task contains on the worker nodes.
- F. Developers specify reduce tasks in the exact same way for both MapReduce version 1 (MRv1) and MapReduce version 2 (MRv2) on YAR
- G. Thus, executing `-D mapreduce.job.reduces-2` will specify reduce tasks.
- H. In YARN, resource allocation is function of virtual cores specified by the ApplicationManager making requests to the NodeManager where a reduce task is handeled 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`

**Answer: D**

#### NEW QUESTION 30

Your cluster has the following characteristics:

? A rack aware topology is configured and on

? Replication is set to 3

? Cluster block size is set to 64MB

Which describes the file read process when a client application connects into the cluster and requests a 50MB file?

- A. The client queries the NameNode for the locations of the block, and reads all three copie
- B. The first copy to complete transfer to the client is the one the client reads as part of hadoop's speculative execution framework.
- C. The client queries the NameNode for the locations of the block, and reads from the first location in the list it receives.
- D. The client queries the NameNode for the locations of the block, and reads from a random location in the list it receives to eliminate network I/O loads by balancing which nodes it retrieves data from any given time.
- E. The client queries the NameNode which retrieves the block from the nearest DataNode to the client then passes that block back to the client.

**Answer: B**

#### NEW QUESTION 31

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 32

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 36

What does CDH packaging do on install to facilitate Kerberos security setup?

- A. Automatically configures permissions for log files at & MAPRED\_LOG\_DIR/userlogs
- B. Creates users for hdfs and mapreduce to facilitate role assignment
- C. Creates directories for temp, hdfs, and mapreduce with the correct permissions
- D. Creates a set of pre-configured Kerberos keytab files and their permissions
- E. Creates and configures your kdc with default cluster values

**Answer: B**

### NEW QUESTION 40

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 45

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 50

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 53

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 58

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 59

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 62

You have installed a cluster HDFS and MapReduce version 2 (MRv2) on YARN. You have no dfs.hosts entry(ies) in your hdfs-site.xml configuration file. You configure a new worker node by setting fs.default.name in its configuration files to point to the NameNode on your cluster, and you start the DataNode daemon on that worker node. What do you have to do on the cluster to allow the worker node to join, and start sorting HDFS blocks?

- A. Without creating a dfs.hosts file or making any entries, run the commands `hadoop.dfsadmin-refreshModes` on the NameNode
- B. Restart the NameNode
- C. Creating a dfs.hosts file on the NameNode, add the worker Node's name to it, then issue the command `hadoop dfsadmin -refresh Nodes =` on the Namenode
- D. Nothing; the worker node will automatically join the cluster when NameNode daemon is started

**Answer:** A

#### NEW QUESTION 64

Which process instantiates user code, and executes map and reduce tasks on a cluster running MapReduce v2 (MRv2) on YARN?

- A. NodeManager
- B. ApplicationMaster
- C. TaskTracker
- D. JobTracker
- E. NameNode
- F. DataNode
- G. ResourceManager

**Answer:** A

#### NEW QUESTION 66

.....

## 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 First Try
- \* CCA-500 Practice Test Questions in Multiple Choice Formats and Updates for 1 Year