CCA-500 Q&As

Cloudera Certified Administrator for Apache Hadoop (CCAH)

Pass Cloudera CCA-500 Exam with 100% Guarantee

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

https://www.certbus.com/cca-500.html

100% Passing Guarantee
100% Money Back Assurance

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

QUESTION 1

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 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

Correct Answer: A

QUESTION 2

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.dfsadminrefreshModes 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

Correct Answer: A

QUESTION 3

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

A. Configuration of rack awareness is accomplished using a configuration file. You cannot use a rack topology script.

B. Hadoop gives preference to intra-rack data transfer in order to conserve bandwidth

C. Rack location is considered in the HDFS block placement policy

D. HDFS is rack aware but MapReduce daemon are not

E. Even for small clusters on a single rack, configuring rack awareness will improve performance

Correct Answer: BC

QUESTION 4

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

Correct Answer: E

QUESTION 5

You are running a Hadoop cluster with MapReduce version 2 (MRv2) on YARN. You consistently see that MapReduce map tasks on your cluster are running slowly because of excessive garbage collection of JVM, how do you increase JVM heap size property to 3GB to optimize performance?

A. yarn.application.child.java.opts=-Xsx3072m

B. yarn.application.child.java.opts=-Xmx3072m

C. mapreduce.map.java.opts=-Xms3072m

D. mapreduce.map.java.opts=-Xmx3072m

Correct Answer: C

QUESTION 6

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 exists. The file is not written to HDFS

F. You get an error message telling you that input is not a directory

G. The write silently fails

Correct Answer: CE

QUESTION 7

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

Correct Answer: A

QUESTION 8

Identify two features/issues that YARN is designated to address: (Choose two)

A. Standardize on a single MapReduce API

B. Single point of failure in the NameNode

C. Reduce complexity of the MapReduce APIs

D. Resource pressure on the JobTracker

E. Ability to run framework other than MapReduce, such as MPI

F. HDFS latency

Correct Answer: DE

QUESTION 9

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

Correct Answer: B

QUESTION 10

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

Correct Answer: C