Cloudera

Exam CCA-505

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

Version: 6.1

[Total Questions: 45]

Question No : 1

You have installed a cluster running HDFS and MapReduce version 2 (MRv2) on YARN. You have no afs.hosts entry()ies in your hdfs-alte.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 storing HDFS blocks?

A. Nothing; the worker node will automatically join the cluster when the DataNode daemon is started.
B. Without creating a dfs.hosts file or making any entries, run the command hadoop dfsadmin –refreshHadoop on the NameNode
C. Create a dfs.hosts file on the NameNode, add the worker node's name to it, then issue the command hadoop dfsadmin –refreshNodes on the NameNode
D. Restart the NameNode

Answer: B

Question No : 2

Given:

```
[user1@host1 ~] & yarn application –list
Total Applications: 3
```

Application-Id	Application-Name	Application-Type	User	Queue	State	Final State	Process	Tracking-URL
Application_1374638600275_0109	Sleep job	MAPREDUCE	user1	default	Killed	KILLED	100%	host1:54059
Application_1374638600275_0121	sleep job	MAPREDUCE	user1	default	SUCCEEDED	SUCCEEDED	100%	host1:19888/jobhistory/job/job_1374638600275_0121
Application_1374638600275_0020	sleep job	MAPREDUCE	user1	default	SUCCEEDED	SUCCEEDED	100%	host1:19888/jobhistory/job/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 –kill application_1374638600275_0109
B. Yarn rmadmin –refreshQueue
C. Yarn application –refreshJobHistory
D. Yarn rmadmin –kill application_1374638600275_0109

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

Question No : 3

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 am entry for the new node.
B. Restart the NameNode and ResourceManager deamons and resubmit any running jobs
C. Increase the value of dfs.number.of.needs in hdfs-site.xml
D. Add a new entry to /etc/nodes on the NameNode host.
E. Restart the NameNode daemon.

Answer: B

Question No : 4

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 you 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

Answer: C

Question No : 5

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

A. It manages the Edits file, which is a log changes to the HDFS filesystem.
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, which is a log of changes to the HDFS filesystem
D. It only keeps track of which NameNode is Active at any given time
E. Clients connect to ZoneKeeper to determine which NameNode is Active

Answer: D
Reference: http://www.cloudera.com/content/cloudera-content/cloudera-docs/CDH4/latest/PDF/CDH4-High-Availability-Guide.pdf (page 15)

Question No : 6

During the execution of a MapReduce v2 (MRv2) job on YARN, where does the Mapper place the intermediate data each Map task?

A. The Mapper stores the intermediate data on the mode running the job's ApplicationMaster so that is available to YARN's 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]sppcache/application_&(appid) directory for the user who ran the job
C. YARN holds the intermediate data in the NodeManager's memory (a container) until it is transferred to the Reducers
D. The Mapper stores the intermediate data on the underlying filesystem of the local disk in the directories yarn.nodemanager.local-dirs
E. The Mapper transfers the intermediate data immediately to the Reducers as it generated by the Map task

Answer: D

Question No : 7

Which Yarn daemon or service monitors a Container's per-application resource usage (e.g, memory, CPU)?

A. NodeManager
B. ApplicationMaster