Exam Number/Code:CCD-410

Exam Name: Cloudera Certified Developer for Apache Hadoop (CCDH)

Version: Demo

QUESTION NO: 1
What information is stored on disk on the NameNode? (Choose 4)

A. File permissions of the files in HDFS.
B. An edit log of changes that have been made since the last backup of the NameNode.
C. A catalog of DataNodes and the blocks that are stored on them.
D. Names of the files in HDFS.
E. The directory structure of the files in HDFS.
F. An edit log of changes that have been made since the last snapshot compaction by the Secondary NameNode.
G. The status of the heartbeats of each DataNode.

Answer: C,D,E,F

QUESTION NO: 2
Compare the hardware requirements of the NameNode with that of the DataNodes in a Hadoop cluster:

A. The NameNode requires more memory and no disk drives.
B. The NameNode requires more memory but less disk capacity.
C. The NameNode and DataNodes should have the same hardware configuration.
D. The NameNode requires less memory and fewer number of disk drives than the DataNodes.
E. The NameNode requires more memory and requires greater disk capacity than the DataNodes.

Answer: B

Reference:
http://www.atlantbh.com/how-to-build-optimal-hadoop-cluster/ (Hardware requirements, basic hardware recommendation)

QUESTION NO: 3
You have a cluster running with the Fair Scheduler enabled and configured. You submit multiple jobs to cluster. Each job is assigned to a pool. How are jobs scheduled? (Choose 2)

A. Each pool's share of task slots may change throughout the course of job execution.
B. Pools get a dynamically-allocated share of the available task slots (subject to additional constraints).

C. Each pool gets 1/M of the total available task slots, where M is the number of nodes in the cluster
D. Pools are assigned priorities. Pools with higher priorities are executed before pools with lower priorities.
E. Each pool gets 1/N of the total available task slots, where N is the number of jobs running on the cluster.
F. Each pool's share of task slots remains static within the execution of any individual job.

Answer: B,D

QUESTION NO: 4
How does the HDFS architecture provide redundancy?

A. Storing multiple replicas of data blocks on different DataNodes.
B. Reliance on RAID at each datanode.
C. Reliance on SAN devices as a DataNode interface.
D. DataNodes make copies of their data blocks, and put them on different local disks.

Answer: A

Reference:
http://bradhedlund.com/2011/09/10/understanding-hadoop-clusters-and-thenetwork/(
writing files to HDFS)

QUESTION NO: 5
Someone in your data center unplugs a slave node by accident. Users of the cluster notice via the Hadoop Web UI that the cluster size has shrunken and express concerns about data loss and HDFS performance. The replication factor of all the files in the cluster is unchanged from the default of 3. What can you tell the users?

A. The HDFS filesystem is corrupt until the administrator re adds the DataNode to the cluster. The warnings associated with the event should be reported.
B. After identifying the outage, the NameNode will naturally re-replicate the data and there will be no data loss. The administrator can re-add the DataNode at any time. The client can disregard warnings concerned with this event. Data will be under-replicated but will become properly replicated over time.
C. The NameNode will re replicate the data after the administrator issues a special command. The data is not lost but is underreplicated until the administrator issues this command.
D. The NameNode will identify the outage and re-replicate the data when the clients receive connection failures to the DataNode, so the end users can disregard such warnings.

Answer: B

QUESTION NO: 6
What is the Secondary NameNode?

A. An alternate data channel for clients to reach HDFS, should the NameNode become too busy.
B. A process that performs a checkpoint operation on the files produced by the NameNode.
C. A data channel between the primary name node and the tertiary NameNode.
D. A process purely intended to perform backups of the NameNode.
E. A standby NameNode, for high availability.

Answer: B

Reference:
http://wiki.apache.org/hadoop/FAQ#What_is_the_purpose_of_the_secondary_name-node.3F(3.2)