

Exam Number/Code:CCA-332

Exam Name:Cloudera Certified
Administrator for Apache Hadoop

Version: Demo

<http://www.it-exams.com>

QUESTION NO: 1

Which of the following describe the functions of a scheduling algorithm? (Choose 4)

- A. Reduce the total amount of computation necessary to complete a job.
- B. Allow short Jobs to complete even when large, long jobs (consuming a lot of resources) are running.
- C. Support the implementation of service-level agreements for multiple cluster users.
- D. Allow multiple users to share clusters in a predictable, policy-guided manner.
- E. Run jobs at periodic times of the day.
- F. Reduce job latencies in an environment with multiple jobs of different sizes.

Answer: A,D,E,F

QUESTION NO: 2

You are running a Hadoop cluster with all monitoring facilities properly configured. Which scenario will go undetected?

- A. Map or reduce tasks that are stuck in an infinite loop.
- B. HDFS is almost full.
- C. The NameNode goes down.
- D. A DataNode is disconnected from the cluster.
- E. MapReduce jobs that are causing excessive memory swaps.

Answer: A

QUESTION NO: 3

Which of the following scenarios makes HDFS unavailable?

- A. JobTracker failure
- B. TaskTracker failure
- C. DataNode failure
- D. NameNode failure
- E. Secondary NameNode failure

Answer: C

Reference: <http://stackoverflow.com/questions/12362727/when-will-hdfs-be-unavailable>

QUESTION NO: 4

What's the relationship between JobTrackers and TaskTrackers?

- A. The JobTracker runs on a single master node and accepts MapReduce jobs from clients. A TaskTracker runs on every slave node and is responsible for managing actual map and reduce tasks.
- B. Every node in the cluster runs both a JobTracker and a TaskTracker. The JobTrackers manage jobs, and the TaskTrackers are responsible for managing actual map and reduce tasks.
- C. The TaskTrackers runs on a single master node and accepts MapReduce jobs from clients. A JobTracker runs on every slave node and is responsible for managing map and reduce tasks.
- D. The JobTracker runs on a single master node, but forks a separate instance of itself for every client MapReduce job. A TaskTracker runs on every slave node and is responsible for managing actual map and reduce tasks.

Answer: A

Reference:

http://hadoop.apache.org/mapreduce/docs/r0.22.0/mapred_tutorial.html (Overview, 4th paragraph)

QUESTION NO: 5

Assuming a large properly configured multi-rack Hadoop cluster, which scenario should not result in loss of HDFS data assuming the default replication factor settings?

- A. Ten percent of DataNodes simultaneously fail.
- B. All DataNodes simultaneously fail.
- C. An entire rack fails.
- D. Multiple racks simultaneously fail.
- E. Seventy percent of DataNodes simultaneously fail.

Answer: A

Reference:

<http://stackoverflow.com/questions/12399197/in-a-large-properly-configured-multirack-hadoop-cluster-which-scenarios-will-b>

QUESTION NO: 6

Which daemon spawns child JVMs to perform MapReduce processing?

- A. JobTracker
- B. NameNode
- C. DataNode
- D. TaskTracker

E. Secondary NameNode

Answer: D

Reference:

<http://www.mindmeister.com/75831919/hadoop-talk-nathan-milford-outbrain> (search Task Tracker)

QUESTION NO: 7

A client wants to read a file from HDFS. How does the data get from the DataNodes to the client?

- A. The NameNode reads the blocks from the DataNodes, and caches them. Then, the application reads the blocks from the NameNode.
- B. The application reads the blocks directly from the DataNodes.
- C. The blocks are sent to a single DataNode, then the application reads the blocks from that Data Node.

Answer: B

Reference:

<http://kazman.shidler.hawaii.edu/ArchDocOverview.html>

QUESTION NO: 8

What would be a reasonable configuration of disk drives in a Hadoop datanode?

- A. Four 1TB disk drives in a RAID configuration
- B. One 1TB disk drive
- C. Four 1TB disk drives in a JBOD configuration
- D. 48 1.5TB disk drives in a JBOD configuration
- E. 48 1.5 TB disk drives in a RAID configuration

Answer: C

Reference:

<http://www.cloudera.com/blog/2010/03/clouderas-support-team-shares-some-basichardware-recommendations/> (How to pick hardware for your hadoop cluster, see first bulleted point)

QUESTION NO: 9

You have a cluster running with the 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. Let's assume both jobs are running at the same priority.

How does the FIFO scheduler execute the jobs? (Choose 3)

- A. The order of execution of tasks within a job may vary.
- B. When a job is submitted, all tasks belonging to that job are scheduled.
- C. Given jobs A and B submitted in that order, all tasks from job A will be scheduled before all tasks from job B.
- D. Since job B needs only a few tasks, it might finish before job A completes.

Answer: A,B,C

Reference:

<http://seriss.com/rush-current/rush/rush-priority.html#FIFO%20Scheduling> (see fifo scheduling)

QUESTION NO: 10

When determining the most appropriate ratio of processor cores to disk drives for a Hadoop cluster, what are the considerations?

- A. Disk drives per machine, nothing else matters.
- B. Processors and disk drives per machine, the workload doesn't matter
- C. Processor core counts and workload type, the ratio doesn't matter
- D. Disk drives per core and workload types
- E. Disk drives per core without consideration to workload types

Answer: D