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.pass2lead.com/cca-500.html

100% Passing Guarantee
100% Money Back Assurance

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

- **Instant Download** After Purchase
- **100% Money Back** Guarantee
- **365 Days** Free Update
- **800,000+** Satisfied Customers

QUESTION 1

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

QUESTION 2

You want to understand more about how users browse your public website. For example, you want to know which pages they visit prior to placing an order. You have a server farm of 200 web servers hosting your website. Which is the most efficient process to gather these web server across logs into your Hadoop cluster analysis?

A. Sample the web server logs web servers and copy them into HDFS using curl

B. Ingest the server web logs into HDFS using Flume

C. Channel these clickstreams into Hadoop using Hadoop Streaming

D. Import all user clicks from your OLTP databases into Hadoop using Sqoop

E. Write a MapReeeduce job with the web servers for mappers and the Hadoop cluster nodes for reducers

Correct Answer: B

QUESTION 3

Your Hadoop cluster is configuring with HDFS and MapReduce version 2 (MRv2) on YARN. Can you configure a worker node to run a NodeManager daemon but not a DataNode daemon and still have a functional cluster?

A. Yes. The daemon will receive data from the NameNode to run Map tasks

B. Yes. The daemon will get data from another (non-local) DataNode to run Map tasks

C. Yes. The daemon will receive Map tasks only

D. Yes. The daemon will receive Reducer tasks only

Correct Answer: B

QUESTION 4

You have a cluster running with the fair Scheduler enabled. There are currently no jobs running on the cluster, and you submit a job A, so that only job A is running on the cluster. A while later, you submit Job B. now Job A and Job B are running on the cluster at the same time. How will the Fair Scheduler handle these two jobs? (Choose two)

A. When Job B gets submitted, it will get assigned tasks, while job A continues to run with fewer tasks.

B. When Job B gets submitted, Job A has to finish first, before job B can gets scheduled.

C. When Job A gets submitted, it doesn\\'t consumes all the task slots.

D. When Job A gets submitted, it consumes all the task slots.

Correct Answer: B

QUESTION 5

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 6

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 7

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 1024mb. Thus, they should specify the amount of memory resource they need by executing D mapreducereduces.memory-mb-2048

C. In YARN, the ApplicationMaster is responsible for requesting the resource required for a specific launch. Thus, executing D yarn.applicationmaster.reduce.tasks=2 will specify that the ApplicationMaster launch two task contains on the worker nodes.

D. Developers specify reduce tasks in the exact same way for both MapReduce version 1 (MRv1) and MapReduce version 2 (MRv2) on YARN. Thus, executing D mapreduce.job.reduces-2 will specify reduce tasks.

E. 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

Correct Answer: D

QUESTION 8

Assuming you\\'re not running HDFS Federation, what is the maximum number of NameNode daemons you should run on your cluster in order to avoid a "split-brain" scenario with your NameNode when running HDFS High Availability (HA) using Quorum-based storage?

A. Two active NameNodes and two Standby NameNodes

B. One active NameNode and one Standby NameNode

C. Two active NameNodes and on Standby NameNode

D. Unlimited. HDFS High Availability (HA) is designed to overcome limitations on the number of NameNodes you can deploy

Correct Answer: B

QUESTION 9

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

Correct Answer: B

QUESTION 10

Your Hadoop cluster contains nodes in three racks. You have not configured the dfs.hosts property in the NameNode\\'s configuration file. What results?

A. The NameNode will update the dfs.hosts property to include machines running the DataNode daemon on the next NameNode reboot or with the command dfsadmin refreshNodes

B. No new nodes can be added to the cluster until you specify them in the dfs.hosts file

C. Any machine running the DataNode daemon can immediately join the cluster

D. Presented with a blank dfs.hosts property, the NameNode will permit DataNodes specified in mapred.hosts to join the cluster

Correct Answer: C