DUMPSARENA

Cloudera Certified Developer for Apache Hadoop (CCDH)

Cloudera CCD-410

Version Demo

Total Demo Questions: 7

Total Premium Questions: 60

Buy Premium PDF

https://dumpsarena.co

sales@dumpsarena.co

sales@dumpsarena.co
dumpsarena.co

QUESTION NO: 1

In a MapReduce job, you want each of your input files processed by a single map task. How do you configure a MapReduce job so that a single map task processes each input file regardless of how many blocks the input file occupies?

A. Increase the parameter that controls minimum split size in the job configuration.

B. Write a custom MapRunner that iterates over all key-value pairs in the entire file.

C. Set the number of mappers equal to the number of input files you want to process.

D. Write a custom FileInputFormat and override the method isSplitable to always return false.

ANSWER: D

QUESTION NO: 2

What is the disadvantage of using multiple reducers with the default HashPartitioner and distributing your workload across you cluster?

A. You will not be able to compress the intermediate data.

B. You will longer be able to take advantage of a Combiner.

C. By using multiple reducers with the default HashPartitioner, output files may not be in globally sorted order.

D. There are no concerns with this approach. It is always advisable to use multiple reduces.

ANSWER: C

QUESTION NO: 3

You have the following key-value pairs as output from your Map task:

(the, 1)

(fox, 1)

(faster, 1)

(than, 1)

(the, 1)

(dog, 1)

How many keys will be passed to the Reducer's reduce method?

A. Six

B. Five

C. Four

D. Two

E. One

F. Three

ANSWER: B

QUESTION NO: 4

You want to perform analysis on a large collection of images. You want to store this data in HDFS and process it with MapReduce but you also want to give your data analysts and data scientists the ability to process the data directly from HDFS with an interpreted high-level programming language like Python. Which format should you use to store this data in HDFS?

A. SequenceFiles

B. Avro

C. JSON

D. HTML

E. XML

F. CSV

ANSWER: B

QUESTION NO: 5

MapReduce v2 (MRv2/YARN) splits which major functions of the JobTracker into separate daemons? Select two.

A. Heath states checks (heartbeats)

B. Resource management

C. Job scheduling/monitoring

D. Job coordination between the ResourceManager and NodeManager

E. Launching tasks

F. Managing file system metadata

G. MapReduce metric reporting

H. Managing tasks

ANSWER: B C

Explanation:

The fundamental idea of MRv2 is to split up the two major functionalities of the JobTracker, resource management and job scheduling/monitoring, into separate daemons. The idea is to have a global ResourceManager (RM) and per-application ApplicationMaster (AM). An application is either a single job in the classical sense of Map-Reduce jobs or a DAG of jobs.

Note:

The central goal of YARN is to clearly separate two things that are unfortunately smushed together in current Hadoop, specifically in (mainly) JobTracker:

/ Monitoring the status of the cluster with respect to which nodes have which resources available. Under YARN, this will be global.

/ Managing the parallelization execution of any specific job. Under YARN, this will be done separately for each job.

Reference: Apache Hadoop YARN – Concepts & Applications

QUESTION NO: 6

MapReduce v2 (MRv2/YARN) is designed to address which two issues?

A. Single point of failure in the NameNode.

B. Resource pressure on the JobTracker.

C. HDFS latency.

D. Ability to run frameworks other than MapReduce, such as MPI.

E. Reduce complexity of the MapReduce APIs.

F. Standardize on a single MapReduce API.

ANSWER: A B

QUESTION NO: 7

A client application creates an HDFS file named foo.txt with a replication factor of 3. Identify which best describes the file access rules in HDFS if the file has a single block that is stored on data nodes A, B and C?

A. The file will be marked as corrupted if data node B fails during the creation of the file.

B. Each data node locks the local file to prohibit concurrent readers and writers of the file.

C. Each data node stores a copy of the file in the local file system with the same name as the HDFS file.

D. The file can be accessed if at least one of the data nodes storing the file is available.

ANSWER: D