

DUMPS ARENA

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.com>

sales@dumpsarena.com

dumpsarena.com

QUESTION NO: 1

You need to perform statistical analysis in your MapReduce job and would like to call methods in the Apache Commons Math library, which is distributed as a 1.3 megabyte Java archive (JAR) file. Which is the best way to make this library available to your MapReducer job at runtime?

- A. Have your system administrator copy the JAR to all nodes in the cluster and set its location in the HADOOP_CLASSPATH environment variable before you submit your job.
- B. Have your system administrator place the JAR file on a Web server accessible to all cluster nodes and then set the HTTP_JAR_URL environment variable to its location.
- C. When submitting the job on the command line, specify the `-libjars` option followed by the JAR file path.
- D. Package your code and the Apache Commons Math library into a zip file named JobJar.zip

ANSWER: C**QUESTION NO: 2**

Which process describes the lifecycle of a Mapper?

- A. The JobTracker calls the TaskTracker's `configure ()` method, then its `map ()` method and finally its `close ()` method.
- B. The TaskTracker spawns a new Mapper to process all records in a single input split.
- C. The TaskTracker spawns a new Mapper to process each key-value pair.
- D. The JobTracker spawns a new Mapper to process all records in a single file.

ANSWER: B**QUESTION NO: 3**

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: 4

You want to understand more about how users browse your public website, such as which pages they visit prior to placing an order. You have a farm of 200 web servers hosting your website. How will you gather this data for your analysis?

- A. Ingest the server web logs into HDFS using Flume.
- B. Write a MapReduce job, with the web servers for mappers, and the Hadoop cluster nodes for reduces.
- C. Import all users' clicks from your OLTP databases into Hadoop, using Sqoop.
- D. Channel these clickstreams into Hadoop using Hadoop Streaming.
- E. Sample the weblogs from the web servers, copying them into Hadoop using curl.

ANSWER: A

QUESTION NO: 5

Assuming default settings, which best describes the order of data provided to a reducer's reduce method:

- A. The keys given to a reducer aren't in a predictable order, but the values associated with those keys always are.
- B. Both the keys and values passed to a reducer always appear in sorted order.
- C. Neither keys nor values are in any predictable order.
- D. The keys given to a reducer are in sorted order but the values associated with each key are in no predictable order

ANSWER: D

QUESTION NO: 6

Identify the MapReduce v2 (MRv2 / YARN) daemon responsible for launching application containers and monitoring application resource usage?

- A. ResourceManager
- B. NodeManager
- C. ApplicationMaster
- D. ApplicationMasterService
- E. TaskTracker
- F. JobTracker

ANSWER: B**QUESTION NO: 7**

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

- A. Health 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