

# DUMPS ARENA

## Hadoop 2.0 Certification exam for Pig and Hive Developer

Hortonworks Apache-Hadoop-Developer

Version Demo

Total Demo Questions: 10

Total Premium Questions: 108

Buy Premium PDF

<https://dumpsarena.co>

[sales@dumpsarena.co](mailto:sales@dumpsarena.co)

sales@dumpsarena.co  
dumpsarena.co

**QUESTION NO: 1**

Which two of the following statements are true about Pig's approach toward data? Choose 2 answers

- A. Accepts only data that has a key/value pair structure
- B. Accepts data whether it has metadata or not
- C. Accepts only data that is defined by metadata tables stored in a database
- D. Accepts tab-delimited text data only
- E. Accepts any data: structured or unstructured

**ANSWER: B E**

**QUESTION NO: 2**

Which two of the following are true about this trivial Pig program' (choose Two)

```
# pig
grunt> ABC = LOAD 'myfile';
grunt> DUMP ABC;
```

- A. The contents of myfile appear on stdout
- B. Pig assumes the contents of myfile are comma delimited
- C. ABC has a schema associated with it
- D. myfile is read from the user's home directory in HDFS

**ANSWER: A D**

**QUESTION NO: 3**

In Hadoop 2.0, which TWO of the following processes work together to provide automatic failover of the NameNode?  
Choose 2 answers

- A. ZKFailoverController
- B. ZooKeeper
- C. QuorumManager
- D. JournalNode

**ANSWER: A D**

**QUESTION NO: 4**

Which Two of the following statements are true about hdfs? Choose 2 answers

- A. An HDFS file that is larger than dfs.block.size is split into blocks
- B. Blocks are replicated to multiple datanodes
- C. HDFS works best when storing a large number of relatively small files
- D. Block sizes for all files must be the same size

**ANSWER: A B**

**QUESTION NO: 5**

What does the following WebHDFS command do?

Curl -1 -L "http://host:port/webhdfs/v1/foo/bar?op=OPEN"

- A. Make a directory /foo/bar
- B. Read a file /foo/bar
- C. List a directory /foo
- D. Delete a directory /foo/bar

**ANSWER: B**

**QUESTION NO: 6**

Examine the following Hive statements:

```
CREATE TABLE x (name STRING, age INT, zip INT, salary DOUBLE)
ROW FORMAT DELIMITED FIELDS TERMINATED BY
',';
LOAD DATA INPATH 'input/File1' OVERWRITE INTO TABLE x;
SELECT * FROM x SORT BY age;
```

Assuming the statements above execute successfully, which one of the following statements is true?

- A. Each reducer generates a file sorted by age

- B. The SORT BY command causes only one reducer to be used
- C. The output of each reducer is only the age column
- D. The output is guaranteed to be a single file with all the data sorted by age

**ANSWER: A**

### QUESTION NO: 7

Determine which best describes when the reduce method is first called in a MapReduce job?

- A. Reducers start copying intermediate key-value pairs from each Mapper as soon as it has completed. The programmer can configure in the job what percentage of the intermediate data should arrive before the reduce method begins.
- B. Reducers start copying intermediate key-value pairs from each Mapper as soon as it has completed. The reduce method is called only after all intermediate data has been copied and sorted.
- C. Reduce methods and map methods all start at the beginning of a job, in order to provide optimal performance for map-only or reduce-only jobs.
- D. Reducers start copying intermediate key-value pairs from each Mapper as soon as it has completed. The reduce method is called as soon as the intermediate key-value pairs start to arrive.

**ANSWER: B**

#### Explanation:

Reference: 24 Interview Questions & Answers for Hadoop MapReduce developers , When is the reducers are started in a MapReduce job?

### QUESTION NO: 8

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**

#### Explanation:

Reference: Apache Hadoop YARN – Concepts & Applications

**QUESTION NO: 9**

You've written a MapReduce job that will process 500 million input records and generated 500 million key-value pairs. The data is not uniformly distributed. Your MapReduce job will create a significant amount of intermediate data that it needs to transfer between mappers and reduces which is a potential bottleneck. A custom implementation of which interface is most likely to reduce the amount of intermediate data transferred across the network?

- A. Partitioner
- B. OutputFormat
- C. WritableComparable
- D. Writable
- E. InputFormat
- F. Combiner

**ANSWER: F****Explanation:**

Combiners are used to increase the efficiency of a MapReduce program. They are used to aggregate intermediate map output locally on individual mapper outputs. Combiners can help you reduce the amount of data that needs to be transferred across to the reducers. You can use your reducer code as a combiner if the operation performed is commutative and associative.

Reference: 24 Interview Questions & Answers for Hadoop MapReduce developers, What are combiners? When should I use a combiner in my MapReduce Job?

**QUESTION NO: 10**

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****Explanation:**

Reference: Apache Hadoop YARN – Concepts & Applications