DUMPSARENA

# Hortonworks Data Platform Certified Developer

## Hortonworks HDPCD

## Version Demo

**Total Demo Questions: 10**

**Total Premium Questions: 108**

**Buy Premium PDF**

https://dumpsarena.com

sales@dumpsarena.com

## QUESTION NO: 1

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: 2

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: 3

Your cluster's HDFS block size in 64MB. You have directory containing 100 plain text files, each of which is 100MB in size. The InputFormat for your job is TextInputFormat. Determine how many Mappers will run?

**A.** 64

**B.** 100

**C.** 200

**D.** 640

## ANSWER: C

**Explanation:**

Each file would be split into two as the block size (64 MB) is less than the file size (100 MB), so 200 mappers would be running.

Note:

If you're not compressing the files then hadoop will process your large files (say 10G), with a number of mappers related to the block size of the file.

Say your block size is 64M, then you will have ~160 mappers processing this 10G file (160*64 ~= 10G). Depending on how CPU intensive your mapper logic is, this might be an

acceptable blocks size, but if you find that your mappers are executing in sub minute times, then you might want to increase the work done by each mapper (by increasing the block size to 128, 256, 512m - the actual size depends on how you intend to process the data).

Reference: http://stackoverflow.com/questions/11014493/hadoop-mapreduce-appropriate-input-files-size (first answer, second paragraph)

## QUESTION NO: 4

Which HDFS command displays the contents of the file x in the user's HDFS home directory?

**A.** hadoop fs -ls x

**B.** hdfs fs -get x

**C.** hadoop fs -cat x

**D.** hadoop fs -cp x

**ANSWER: C**

## QUESTION NO: 5

Review the following data and Pig code.

M,38,95111

F,29,95060

F,45,95192

M,62,95102

F,56,95102

A = LOAD 'data' USING PigStorage('.') as (gender:Chararray, age:int, zlp:chararray);

B = FOREACH A GENERATE age;

Which one of the following commands would save the results of B to a folder in hdfs named myoutput?

**A.** STORE A INTO 'myoutput' USING PigStorage(',');

**B.** DUMP B using PigStorage('myoutput');

**C.** STORE B INTO 'myoutput';

**D.** DUMP B INTO 'myoutput';

**ANSWER: C**

## QUESTION NO: 6

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: 7

What does Pig provide to the overall Hadoop solution?

**A.** Legacy language Integration with MapReduce framework

**B.** Simple scripting language for writing MapReduce programs

**C.** Database table and storage management services

**D.** C++ interface to MapReduce and data warehouse infrastructure

**ANSWER: B**

## QUESTION NO: 8

Which TWO of the following statements are true regarding Hive? Choose 2 answers

**A.** Useful for data analysts familiar with SQL who need to do ad-hoc queries

**B.** Offers real-time queries and row level updates

**C.** Allows you to define a structure for your unstructured Big Data

**D.** Is a relational database

**ANSWER: A C**

## QUESTION NO: 9

Given the following Hive command:

```
CREATE EXTERNAL TABLE mytable (name string, age int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION '/home/user/mydata/';
```

Which one of the following statements is true?

**A.** The files in the mydata folder are copied to a subfolder of /apps/hlve/warehouse

**B.** The files in the mydata folder are moved to a subfolder of /apps/hive/wa re house

**C.** The files in the mydata folder are copied into Hive's underlying relational database

**D.** The files in the mydata folder do not move from their current location In HDFS

**ANSWER: D**

## 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

DUMPSARENA