

**100%** Money Back  
**Guarantee**

**Vendor:**Hortonworks

**Exam Code:**APACHE-HADOOP-DEVELOPER

**Exam Name:**Hadoop 2.0 Certification exam for Pig  
and Hive Developer

**Version:**Demo

### QUESTION 1

What does the following command do? `register andapos;:/piggyban):/pig-files.jarandapos;;`

- A. Invokes the user-defined functions contained in the jar file
- B. Assigns a name to a user-defined function or streaming command
- C. Transforms Pig user-defined functions into a format that Hive can accept
- D. Specifies the location of the JAR file containing the user-defined functions

Correct Answer: D

---

### QUESTION 2

To process input key-value pairs, your mapper needs to load a 512 MB data file in memory. What is the best way to accomplish this?

- A. Serialize the data file, insert in it the JobConf object, and read the data into memory in the configure method of the mapper.
- B. Place the data file in the DistributedCache and read the data into memory in the map method of the mapper.
- C. Place the data file in the DataCache and read the data into memory in the configure method of the mapper.
- D. Place the data file in the DistributedCache and read the data into memory in the configure method of the mapper.

Correct Answer: C

---

### QUESTION 3

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`

Correct Answer: B

---

### QUESTION 4

In a large MapReduce job with  $m$  mappers and  $n$  reducers, how many distinct copy operations will there be in the

sort/shuffle phase?

A.  $m \times n$  (i.e.,  $m$  multiplied by  $n$ )

B.  $n$

C.  $m$

D.  $m+n$  (i.e.,  $m$  plus  $n$ )

E.  $m^n$  (i.e.,  $m$  to the power of  $n$ )

Correct Answer: A

Explanation: A MapReduce job with  $m$  mappers and  $r$  reducers involves up to  $m * r$  distinct copy operations, since each mapper may have intermediate output going to every reducer.

---

#### QUESTION 5

A NameNode in Hadoop 2.0 manages \_\_\_\_\_.

A. Two namespaces: an active namespace and a backup namespace

B. A single namespace

C. An arbitrary number of namespaces

D. No namespaces

Correct Answer: B

---

#### QUESTION 6

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, zip:chararray);
```

What command to define B would produce the output (M,62,95102) when invoking the DUMP operator on B?

- A. B = FILTER A BY (zip = = '95102' AND gender = = M");
- B. B= FOREACH A BY (gender = = 'M' AND zip = = '95102');
- C. B = JOIN A BY (gender = = 'M' AND zip = = '95102');
- D. B= GROUP A BY (zip = = '95102' AND gender = = 'M');

Correct Answer: A

---

#### QUESTION 7

Which one of the following statements is false about HCatalog?

- A. Provides a shared schema mechanism
- B. Designed to be used by other programs such as Pig, Hive and MapReduce
- C. Stores HDFS data in a database for performing SQL-like ad-hoc queries
- D. Exists as a subproject of Hive

Correct Answer: C

---

#### QUESTION 8

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

Correct Answer: AD

---

#### QUESTION 9

Consider the following two relations, A and B.

```

A = LOAD 'data1' AS (a1:int,a2:chararray);
DUMP A;
(1,apple)
(3,orange)
(4,peach)
(2,cherry)

B = LOAD 'data2' AS (b1:chararray,b2:int);
DUMP B;
(Jim,2)
(Brian,4)
(Kim,0)
(Terry,3)
(Chris,2)

```

A Pig JOIN statement that combined relations A by its first field and B by its second field would produce what output?

- A. 2 Jim Chris 2 3 Terry 3 4 Brian 4
- B. 2 cherry 2 cherry 3 orange 4 peach
- C. 2 cherry Jim, Chris 3 orange Terry  
4 peach Brian
- D. 2 cherry Jim 2 2 cherry Chris 2 3 orange Terry 3 4 peach Brian 4

Correct Answer: D

---

#### QUESTION 10

Given the following Hive command:

```
INSERT OVERWRITE TABLE mytable SELECT * FROM myothertable;
```

Which one of the following statements is true?

- A. The contents of myothertable are appended to mytable
- B. Any existing data in mytable will be overwritten
- C. A new table named mytable is created, and the contents of myothertable are copied into mytable
- D. The statement is not a valid Hive command

Correct Answer: B

---

#### QUESTION 11

Which Hadoop component is responsible for managing the distributed file system metadata?

- A. NameNode

- B. Metanode
- C. DataNode
- D. NameSpaceManager

Correct Answer: A

---

## QUESTION 12

You have a directory named jobdata in HDFS that contains four files: \_first.txt, second.txt, .third.txt and #data.txt. How many files will be processed by the FileInputFormat.setInputPaths () command when it's given a path object representing this directory?

- A. Four, all files will be processed
- B. Three, the pound sign is an invalid character for HDFS file names
- C. Two, file names with a leading period or underscore are ignored
- D. None, the directory cannot be named jobdata
- E. One, no special characters can prefix the name of an input file

Correct Answer: C

Explanation: Files starting with \'\_\' are considered \'hidden\' like unix files starting with \'.\' . # characters are allowed in HDFS file names.