



# HADOOP-PR000007<sup>Q&As</sup>

Hortonworks Certified Apache Hadoop 2.0 Developer (Pig and Hive Developer)

**Pass Hortonworks HADOOP-PR000007 Exam with 100% Guarantee**

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.geekcert.com/hadoop-pr000007.html>

100% Passing Guarantee  
100% Money Back Assurance

Following Questions and Answers are all new published by Hortonworks Official Exam Center

-  **Instant Download** After Purchase
-  **100% Money Back** Guarantee
-  **365 Days** Free Update
-  **800,000+** Satisfied Customers





### QUESTION 1

Review the following andapos;dataandapos; file and Pig code.

```
M,38,95111
M,62,95102
```

```
A = LOAD andapos;dataandapos; USING PigStorage(andapos;,andapos;)
AS (gender:chararray, age:int, zip:chararray);
D = GROUP A BY gender;
DUMP D;
```

Which one of the following statements is true?

- A. The Output Of the DUMP D command IS (M,((M,62.95102),(M,38,95111)))
- B. The output of the dump d command is (M, {(38,95in),(62,95i02)})
- C. The code executes successfully but there is not output because the D relation is empty
- D. The code does not execute successfully because D is not a valid relation

Correct Answer: A

---

### QUESTION 2

Given the following Pig command:

```
logevents = LOAD andapos;input/my.logandapos; AS (date:chararray, levehstring, code:int, message:string);
```

Which one of the following statements is true?

- A. The logevents relation represents the data from the my.log file, using a comma as the parsing delimiter
- B. The logevents relation represents the data from the my.log file, using a tab as the parsing delimiter
- C. The first field of logevents must be a properly-formatted date string or table return an error
- D. The statement is not a valid Pig command

Correct Answer: B

---

### QUESTION 3

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

Correct 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 and Answers for Hadoop MapReduce developers, What are combiners? When should I use a combiner in my MapReduce Job?

---

#### QUESTION 4

Which best describes how TextInputFormat processes input files and line breaks?

- A. Input file splits may cross line breaks. A line that crosses file splits is read by the RecordReader of the split that contains the beginning of the broken line.
- B. Input file splits may cross line breaks. A line that crosses file splits is read by the RecordReaders of both splits containing the broken line.
- C. The input file is split exactly at the line breaks, so each RecordReader will read a series of complete lines.
- D. Input file splits may cross line breaks. A line that crosses file splits is ignored.
- E. Input file splits may cross line breaks. A line that crosses file splits is read by the RecordReader of the split that contains the end of the broken line.

Correct Answer: A

Reference: How Map and Reduce operations are actually carried out

---

#### QUESTION 5

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

[Latest HADOOP-PR000007 Dumps](#)

[HADOOP-PR000007 Practice Test](#)

[HADOOP-PR000007 Study Guide](#)