# HADOOP-PR000007 Q&As

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.pass4itsure.com/hadoop-pr000007.html**

100% Passing Guarantee
100% Money Back Assurance

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



**QUESTION 1**

When can a reduce class also serve as a combiner without affecting the output of a MapReduce program?

A. When the types of the reduce operation\\'s input key and input value match the types of the reducer\\'s output key and output value and when the reduce operation is both communicative and associative.

B. When the signature of the reduce method matches the signature of the combine method.

C. Always. Code can be reused in Java since it is a polymorphic object-oriented programming language.

D. Always. The point of a combiner is to serve as a mini-reducer directly after the map phase to increase performance.

E. Never. Combiners and reducers must be implemented separately because they serve different purposes.

Correct Answer: A

Explanation: 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 2**

You have the following key-value pairs as output from your Map task:

(the, 1)

(fox, 1)

(faster, 1)

(than, 1)

(the, 1) (dog, 1)

How many keys will be passed to the Reducer\\'s reduce method?

A. Six

B. Five

C. Four

D. Two

E. One

F. Three

Correct Answer: B

Explanation: Only one key value pair will be passed from the two (the, 1) key value pairs.

**QUESTION 3**

Identify the tool best suited to import a portion of a relational database every day as files into HDFS, and generate Java classes to interact with that imported data?

A. Oozie

B. Flume

C. Pig

D. Hue

E. Hive

F. Sqoop

G. fuse-dfs

Correct Answer: F

Sqoop ("SQL-to-Hadoop") is a straightforward command-line tool with the following capabilities:

Imports individual tables or entire databases to files in HDFS Generates Java classes to allow you to

interact with your imported data Provides the ability to import from SQL databases straight into your Hive

data warehouse

Note:

Data Movement Between Hadoop and Relational Databases Data can be moved between Hadoop and a

relational database as a bulk data transfer, or relational tables can be accessed from within a MapReduce

map function.

Note:

* Cloudera\\'s Distribution for Hadoop provides a bulk data transfer tool (i.e., Sqoop) that imports individual tables or entire databases into HDFS files. The tool also generates Java classes that support interaction with the imported data. Sqoop supports all relational databases over JDBC, and Quest Software provides a connector (i.e., OraOop) that has been optimized for access to data residing in Oracle databases.

Reference: http://log.medcl.net/item/2011/08/hadoop-and-mapreduce-big-data-analytics- gartner/ (Data Movement between hadoop and relational databases, second paragraph)

**QUESTION 4**

What is the disadvantage of using multiple reducers with the default HashPartitioner and distributing your workload across you cluster?

A. You will not be able to compress the intermediate data.

B. You will longer be able to take advantage of a Combiner.

C. By using multiple reducers with the default HashPartitioner, output files may not be in globally sorted order.

D. There are no concerns with this approach. It is always advisable to use multiple reduces.

Correct Answer: C

Explanation: Multiple reducers and total ordering

If your sort job runs with multiple reducers (either because mapreduce.job.reduces in mapred-site.xml has been set to a number larger than 1, or because you\\'ve used the -r option to specify the number of reducers on the command-line), then by default Hadoop will use the HashPartitioner to distribute records across the reducers. Use of the HashPartitioner means that you can\\'t concatenate your output files to create a single sorted output file. To do this you\\'ll need total ordering,

Reference: Sorting text files with MapReduce

**QUESTION 5**

Examine the following Hive statements:

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

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

A. Hive reformats File1 into a structure that Hive can access and moves into to/user/joe/x/

B. The file named File1 is moved to to/user/joe/x/

C. The contents of File1 are parsed as comma-delimited rows and loaded into /user/joe/x/

D. The contents of File1 are parsed as comma-delimited rows and stored in a database

Correct Answer: B