

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.pass2lead.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

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)
```

What is the output of the following Pig commands?

```
X = GROUP A BY S1;
```

```
DUMP X;
```

- A. $\{(\text{group},\{\{\text{apple,peach,cherry,orange}\}\})\}$
- B. $\{\text{apple,peach,cherry,orange}\}$
- C. $\{1,4,2,3\}$
- D. $\{(\text{apple},\{\{1,\text{apple}\}\})\}$
 $\{(\text{peach},\{\{4,\text{peach}\}\})\}$
 $\{(\text{cherry},\{\{2,\text{cherry}\}\})\}$
 $\{(\text{orange},\{\{3,\text{orange}\}\})\}$

- A. Option A
- B. Option B
- C. Option C
- D. Option D

Correct Answer: D

QUESTION 2

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 3

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

Correct Answer: AC

QUESTION 4

Examine the following Pig commands:

```
customers = LOAD 'data' USING PigStorage(',')
AS (name:chararray,age:int,zip:int,receipts:double);
s = SAMPLE customers 0.2;
DUMP s;
```

Which one of the following statements is true?

- A. The SAMPLE command generates an "unexpected symbol" error
- B. Each MapReduce task will terminate after executing for 0.2 minutes
- C. The reducers will only output the first 20% of the data passed from the mappers
- D. A random sample of approximately 20% of the data will be output

Correct Answer: D

QUESTION 5

You have just executed a MapReduce job. Where is intermediate data written to after being emitted from the Mapper's map method?

- A. Intermediate data is streamed across the network from Mapper to the Reducer and is never written to disk.
- B. Into in-memory buffers on the TaskTracker node running the Mapper that spill over and are written into HDFS.
- C. Into in-memory buffers that spill over to the local file system of the TaskTracker node running the Mapper.
- D. Into in-memory buffers that spill over to the local file system (outside HDFS) of the TaskTracker node running the Reducer.
- E. Into in-memory buffers on the TaskTracker node running the Reducer that spill over and are written into HDFS.

Correct Answer: C

Explanation: The mapper output (intermediate data) is stored on the Local file system (NOT HDFS) of each individual mapper nodes. This is typically a temporary directory location which can be setup in config by the hadoop administrator. The intermediate data is cleaned up after the Hadoop Job completes.

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, Where is the Mapper Output (intermediate key-value data) stored ?

[Latest HADOOP-PR000007 Dumps](#)

[HADOOP-PR000007 VCE Dumps](#)

[HADOOP-PR000007 Exam Questions](#)