



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.passapply.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

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 2

Identify the utility that allows you to create and run MapReduce jobs with any executable or script as the mapper and/or the reducer?

- A. Oozie
- B. Sqoop
- C. Flume
- D. Hadoop Streaming
- E. mapred

Correct Answer: D

Explanation: Hadoop streaming is a utility that comes with the Hadoop distribution. The utility allows you to create and run Map/Reduce jobs with any executable or script as the mapper and/or the reducer.

Reference: <http://hadoop.apache.org/common/docs/r0.20.1/streaming.html> (Hadoop Streaming, second sentence)

QUESTION 3

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 4

Which YARN component is responsible for monitoring the success or failure of a Container?

- A. ResourceManager
- B. ApplicationMaster
- C. NodeManager
- D. JobTracker

Correct Answer: A

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 Reduce 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 PDF Dumps](#)

[HADOOP-PR000007 Braindumps](#)