



HDPCD^{Q&As}

Hortonworks Data Platform Certified Developer

Pass Hortonworks HDPCD Exam with 100% Guarantee

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

<https://www.passapply.com/hdpcd.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

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 2

Which project gives you a distributed, Scalable, data store that allows you random, realtime read/write access to hundreds of terabytes of data?

- A. HBase
- B. Hue
- C. Pig
- D. Hive
- E. Oozie
- F. Flume
- G. Sqoop

Correct Answer: A

Explanation: Use Apache HBase when you need random, realtime read/write access to your Big Data.

Note: This project's goal is the hosting of very large tables -- billions of rows X millions of columns -- atop clusters of commodity hardware. Apache HBase is an open-source, distributed, versioned, column-oriented store modeled after Google's Bigtable: A Distributed Storage System for Structured Data by Chang et al. Just as Bigtable leverages the distributed data storage provided by the Google File System, Apache HBase provides Bigtable-like capabilities on top of Hadoop and HDFS.

Features



Linear and modular scalability.

Strictly consistent reads and writes.

Automatic and configurable sharding of tables

Automatic failover support between RegionServers.

Convenient base classes for backing Hadoop MapReduce jobs with Apache HBase tables.

Easy to use Java API for client access.

Block cache and Bloom Filters for real-time queries.

Query predicate push down via server side Filters

Thrift gateway and a REST-ful Web service that supports XML, Protobuf, and binary data encoding options

Extensible jruby-based (JIRB) shell

Support for exporting metrics via the Hadoop metrics subsystem to files or Ganglia; or via JMX

Reference: <http://hbase.apache.org/> (when would I use HBase? First sentence)

QUESTION 3

Examine the following Pig commands: Which one of the following statements is true?

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

- 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 4

You have written a Mapper which invokes the following five calls to the OutputCollector.collect method:

```
output.collect (new Text ("Apple"), new Text ("Red")) ;
```

```
output.collect (new Text ("Banana"), new Text ("Yellow")) ;
```

```
output.collect (new Text ("Apple"), new Text ("Yellow")) ;
```



```
output.collect (new Text ("Cherry"), new Text ("Red") ) ;
```

```
output.collect (new Text ("Apple"), new Text ("Green") ) ;
```

How many times will the Reducer's reduce method be invoked?

- A. 6
- B. 3
- C. 1
- D. 0
- E. 5

Correct Answer: B

Explanation: reduce() gets called once for each [key, (list of values)] pair. To explain, let's say you called:

```
out.collect(new Text("Car"),new Text("Subaru");
```

```
out.collect(new Text("Car"),new Text("Honda");
```

```
out.collect(new Text("Car"),new Text("Ford");
```

```
out.collect(new Text("Truck"),new Text("Dodge");
```

```
out.collect(new Text("Truck"),new Text("Chevy");
```

Then reduce() would be called twice with the pairs reduce(Car,)

reduce(Truck,)

Reference: Mapper output.collect()?

QUESTION 5

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

[HDPCD VCE Dumps](#)

[HDPCD Exam Questions](#)

[HDPCD Braindumps](#)