



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.geekcert.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

You write MapReduce job to process 100 files in HDFS. Your MapReduce algorithm uses TextInputFormat: the mapper applies a regular expression over input values and emits key- values pairs with the key consisting of the matching text, and the value containing the filename and byte offset. Determine the difference between setting the number of reduces to one and settings the number of reducers to zero.

- A. There is no difference in output between the two settings.
- B. With zero reducers, no reducer runs and the job throws an exception. With one reducer, instances of matching patterns are stored in a single file on HDFS.
- C. With zero reducers, all instances of matching patterns are gathered together in one file on HDFS. With one reducer, instances of matching patterns are stored in multiple files on HDFS.
- D. With zero reducers, instances of matching patterns are stored in multiple files on HDFS. With one reducer, all instances of matching patterns are gathered together in one file on HDFS.

Correct Answer: D

Explanation: * It is legal to set the number of reduce-tasks to zero if no reduction is desired.

In this case the outputs of the map-tasks go directly to the FileSystem, into the output path set by `setOutputPath(Path)`. The framework does not sort the map-outputs before writing them out to the FileSystem.

* Often, you may want to process input data using a map function only. To do this, simply set `mapreduce.job.reduces` to zero. The MapReduce framework will not create any reducer tasks. Rather, the outputs of the mapper tasks will be the final output of the job.

Note:

Reduce

In this phase the `reduce(WritableComparable, Iterator, OutputCollector, Reporter)` method is called for each pair in the grouped inputs.

The output of the reduce task is typically written to the FileSystem via `OutputCollector.collect(WritableComparable, Writable)`.

Applications can use the Reporter to report progress, set application-level status messages and update Counters, or just indicate that they are alive.

The output of the Reducer is not sorted.

QUESTION 2

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 3

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 4

Which two of the following statements are true about Pig's approach toward data? Choose 2 answers

- A. Accepts only data that has a key/value pair structure
- B. Accepts data whether it has metadata or not
- C. Accepts only data that is defined by metadata tables stored in a database
- D. Accepts tab-delimited text data only
- E. Accepts any data: structured or unstructured



Correct Answer: BE

QUESTION 5

Which one of the following statements regarding the components of YARN is FALSE?

- A. A Container executes a specific task as assigned by the ApplicationMaster
- B. The ResourceManager is responsible for scheduling and allocating resources
- C. A client application submits a YARW job to the ResourceManager
- D. The ResourceManager monitors and restarts any failed Containers

Correct Answer: D

[Latest HDPCD Dumps](#)

[HDPCD Study Guide](#)

[HDPCD Braindumps](#)