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

What does the following command do?

register andapos;/piggyban):/pig-files.jarandapos;;

A. Invokes the user-defined functions contained in the jar file

B. Assigns a name to a user-defined function or streaming command

C. Transforms Pig user-defined functions into a format that Hive can accept

D. Specifies the location of the JAR file containing the user-defined functions

Correct Answer: D

QUESTION 2

Your cluster\\'s HDFS block size in 64MB. You have directory containing 100 plain text files, each of which

is 100MB in size. The InputFormat for your job is TextInputFormat.

Determine how many Mappers will run?

A. 64

B. 100

C. 200

D. 640

Correct Answer: C

Explanation: Each file would be split into two as the block size (64 MB) is less than the file size (100 MB),

so 200 mappers would be running.

Note:

If you\\'re not compressing the files then hadoop will process your large files (say 10G), with a number of

mappers related to the block size of the file.

Say your block size is 64M, then you will have ~160 mappers processing this 10G file (160*64 ~= 10G).

Depending on how CPU intensive your mapper logic is, this might be an

acceptable blocks size, but if you find that your mappers are executing in sub minute times, then you might

want to increase the work done by each mapper (by increasing the block size to 128, 256, 512m - the

actual size depends on how you intend to process the data). Reference: http://stackoverflow.com/

QUESTION 3

Which one of the following statements is true regarding a MapReduce job?

A. The job\\'s Partitioner shuffles and sorts all (key.value) pairs and sends the output to all reducers

B. The default Hash Partitioner sends key value pairs with the same key to the same Reducer

C. The reduce method is invoked once for each unique value

D. The Mapper must sort its output of (key.value) pairs in descending order based on value

Correct Answer: A

QUESTION 4

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 5

Review the following andapos;dataandapos; file and Pig code.

```
M,38,95111
M,62,95102


A = LOAD &apos;data&apos; USING PigStorage(&apos;,&apos;)
AS (gender:chararray, age:int, zip:chararray);

D = GROUP A BY gender;
DUMP D;
```

Which one of the following statements is true?

A. The Output Of the DUMP D command IS (M,{(M,62.95102),(M,38,95111)})

B. The output of the dump d command is (M, {(38,95in),(62,95i02)})

C. The code executes successfully but there is not output because the D relation is empty

D. The code does not execute successfully because D is not a valid relation

Correct Answer: A

[Latest HDPCD Dumps](#)   [HDPCD PDF Dumps](#)   [HDPCD Study Guide](#)