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.pass4itsure.com/hdpcd.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Hortonworks Official Exam Center

QUESTION 1

Which HDFS command uploads a local file X into an existing HDFS directory Y?

A. hadoop scp X Y

B. hadoop fs -localPut X Y

C. hadoop fs-put X Y

D. hadoop fs -get X Y

Correct Answer: C

QUESTION 2

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. (group,{(apple,peach,cherry,orange)})

B. {apple,peach,cherry,orange}

C. {1,4,2,3}

D. (apple,{(1,apple)})
(peach,{(4,peach)})
(cherry,{(2,cherry)})
(orange,{(3,orange)})

A. Option A

B. Option B

C. Option C

D. Option D

Correct Answer: D

QUESTION 3

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 4

Indentify 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 5

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/