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.pass2lead.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

When is the earliest point at which the reduce method of a given Reducer can be called?

A. As soon as at least one mapper has finished processing its input split.

B. As soon as a mapper has emitted at least one record.

C. Not until all mappers have finished processing all records.

D. It depends on the InputFormat used for the job.

Correct Answer: C

Explanation: In a MapReduce job reducers do not start executing the reduce method until the all Map jobs have completed. Reducers start copying intermediate key-value pairs from the mappers as soon as they are available. The programmer defined reduce method is called only after all the mappers have finished.

Note: The reduce phase has 3 steps: shuffle, sort, reduce. Shuffle is where the data is collected by the reducer from each mapper. This can happen while mappers are generating data since it is only a data transfer. On the other hand, sort and reduce can only start once all the mappers are done.

Why is starting the reducers early a good thing? Because it spreads out the data transfer from the mappers to the reducers over time, which is a good thing if your network is the bottleneck.

Why is starting the reducers early a bad thing? Because they "hog up" reduce slots while only copying data. Another job that starts later that will actually use the reduce slots now can\\'t use them.

You can customize when the reducers startup by changing the default value of mapred.reduce.slowstart.completed.maps in mapred-site.xml. A value of 1.00 will wait for all the mappers to finish before starting the reducers. A value of 0.0 will start the reducers right away. A value of 0.5 will start the reducers when half of the mappers are complete. You can also change mapred.reduce.slowstart.completed.maps on a job-by-job basis.

Typically, keep mapred.reduce.slowstart.completed.maps above 0.9 if the system ever has multiple jobs running at once. This way the job doesn\\'t hog up reducers when they aren\\'t doing anything but copying data. If you only ever have one job running at a time, doing 0.1 would probably be appropriate.

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, When is the reducers are started in a MapReduce job?

QUESTION 2

What data does a Reducer reduce method process?

A. All the data in a single input file.

B. All data produced by a single mapper.

C. All data for a given key, regardless of which mapper(s) produced it.

D. All data for a given value, regardless of which mapper(s) produced it.

Correct Answer: C

Explanation: Reducing lets you aggregate values together. A reducer function receives an iterator of input values from an input list. It then combines these values together, returning a single output value.

All values with the same key are presented to a single reduce task.

Reference: Yahoo! Hadoop Tutorial, Module 4: MapReduce

QUESTION 3

A combiner reduces:

A. The number of values across different keys in the iterator supplied to a single reduce method call.

B. The amount of intermediate data that must be transferred between the mapper and reducer.

C. The number of input files a mapper must process.

D. The number of output files a reducer must produce.

Correct Answer: B

Explanation: Combiners are used to increase the efficiency of a MapReduce program. They are used to aggregate intermediate map output locally on individual mapper outputs. Combiners can help you reduce the amount of data that needs to be transferred across to the reducers. You can use your reducer code as a combiner if the operation performed is commutative and associative. The execution of combiner is not guaranteed, Hadoop may or may not execute a combiner. Also, if required it may execute it more then 1 times. Therefore your MapReduce jobs should not depend on the combiners execution.

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

Review the following data and Pig code.

M,38,95111

F,29,95060

F,45,95192

M,62,95102

F,56,95102

A = LOAD andapos;dataandapos; USING PigStorage(andapos;.andapos;) as (gender:Chararray, age:int,

zlp:chararray);

B = FOREACH A GENERATE age;

Which one of the following commands would save the results of B to a folder in hdfs named myoutput?

A. STORE A INTO andapos;myoutputandapos; USING PigStorage(andapos;,andapos;);

B. DUMP B using PigStorage(andapos;myoutputandapos;);

C. STORE B INTO andapos;myoutputandapos;;

D. DUMP B INTO andapos;myoutputandapos;;

Correct Answer: C

QUESTION 5

A NameNode in Hadoop 2.2 manages _____.

A. Two namespaces: an active namespace and a backup namespace

B. A single namespace

C. An arbitrary number of namespaces

D. No namespaces

Correct Answer: B