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

Which one of the following statements is false about HCatalog?

A. Provides a shared schema mechanism

B. Designed to be used by other programs such as Pig, Hive and MapReduce

C. Stores HDFS data in a database for performing SQL-like ad-hoc queries

D. Exists as a subproject of Hive

Correct Answer: C

QUESTION 2

Which Hadoop component is responsible for managing the distributed file system metadata?

A. NameNode

B. Metanode

C. DataNode

D. NameSpaceManager

Correct Answer: A

QUESTION 3

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 4

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 5

In the reducer, the MapReduce API provides you with an iterator over Writable values. What does calling the next () method return?

A. It returns a reference to a different Writable object time.

B. It returns a reference to a Writable object from an object pool.

C. It returns a reference to the same Writable object each time, but populated with different data.

D. It returns a reference to a Writable object. The API leaves unspecified whether this is a reused object or a new object.

E. It returns a reference to the same Writable object if the next value is the same as the previous value, or a new Writable object otherwise.

Correct Answer: C

Explanation: Calling Iterator.next() will always return the SAME EXACT instance of IntWritable, with the contents of that instance replaced with the next value.

Reference: manupulating iterator in mapreduce