



Hortonworks Data Platform Certified Developer

Pass Hortonworks HDPCD Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.passapply.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

Given the following Hive command:

CREATE EXTERNAL TABLE mytable (name string, age int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/home/user/mydata/';

Which one of the following statements is true?

- A. The files in the mydata folder are copied to a subfolder of /apps/hlve/warehouse
- B. The files in the mydata folder are moved to a subfolder of /apps/hive/wa re house
- C. The files in the mydata folder are copied into Hive\\'s underlying relational database
- D. The files in the mydata folder do not move from their current location In HDFS

Correct Answer: D

QUESTION 2

What types of algorithms are difficult to express in MapReduce v1 (MRv1)?

- A. Algorithms that require applying the same mathematical function to large numbers of individual binary records.
- B. Relational operations on large amounts of structured and semi-structured data.
- C. Algorithms that require global, sharing states.
- D. Large-scale graph algorithms that require one-step link traversal.
- E. Text analysis algorithms on large collections of unstructured text (e.g, Web crawls).

Correct Answer: C

Explanation: See 3) below.

Limitations of Mapreduce ?where not to use Mapreduce

While very powerful and applicable to a wide variety of problems, MapReduce is not the answer to every

problem. Here are some problems I found where MapReudce is not suited and some papers that address

the limitations of MapReuce.

1.

Computation depends on previously computed values If the computation of a value depends on previously computed values, then MapReduce cannot be used. One good example is the Fibonacci series where each value is summation of the previous two values. i.e., f(k+2) = f(k+1) + f(k). Also, if the data set is small enough to be computed on a single machine, then it is better to do it as a single reduce(map(data)) operation rather than going through the entire map reduce process.



2.

Full-text indexing or ad hoc searching The index generated in the Map step is one dimensional, and the Reduce step must not generate a large amount of data or there will be a serious performance degradation. For example, CouchDB\\'s MapReduce may not be a good fit for full-text indexing or ad hoc searching. This is a problem better suited for a tool such as Lucene.

3.

Algorithms depend on shared global state Solutions to many interesting problems in text processing do not require global synchronization. As a result, they can be expressed naturally in MapReduce, since map and reduce tasks run independently and in isolation. However, there are many examples of algorithms that depend crucially on the existence of shared global state during processing, making them difficult to implement in MapReduce (since the single opportunity for global synchronization in MapReduce is the barrier between the map and reduce phases of processing)

Reference: Limitations of Mapreduce ?where not to use Mapreduce

QUESTION 3

Identify the MapReduce v2 (MRv2 / YARN) daemon responsible for launching application containers and monitoring application resource usage?

- A. ResourceManager
- B. NodeManager
- C. ApplicationMaster
- D. ApplicationMasterService
- E. TaskTracker
- F. JobTracker

Correct Answer: B

Reference: Apache Hadoop YARN ?Conceptsand; Applications

QUESTION 4

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

QUESTION 5

Which one of the following files is required in every Oozie Workflow application?

- A. job.properties
- B. Config-default.xml
- C. Workflow.xml
- D. Oozie.xml
- Correct Answer: C

HDPCD VCE Dumps

HDPCD Study Guide

HDPCD Braindumps