



APACHE-HADOOP-DEVELOPER^{Q&As}

Hadoop 2.0 Certification exam for Pig and Hive Developer

Pass Hortonworks APACHE-HADOOP-DEVELOPER Exam with 100% Guarantee

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

<https://www.pass4itsure.com/apache-hadoop-developer.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

When can a reduce class also serve as a combiner without affecting the output of a MapReduce program?

- A. When the types of the reduce operation's input key and input value match the types of the reducer's output key and output value and when the reduce operation is both communicative and associative.
- B. When the signature of the reduce method matches the signature of the combine method.
- C. Always. Code can be reused in Java since it is a polymorphic object-oriented programming language.
- D. Always. The point of a combiner is to serve as a mini-reducer directly after the map phase to increase performance.
- E. Never. Combiners and reducers must be implemented separately because they serve different purposes.

Correct Answer: A

Explanation: You can use your reducer code as a combiner if the operation performed is commutative and associative.

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, What are combiners? When should I use a combiner in my MapReduce Job?

QUESTION 2

You want to run Hadoop jobs on your development workstation for testing before you submit them to your production cluster. Which mode of operation in Hadoop allows you to most closely simulate a production cluster while using a single machine?

- A. Run all the nodes in your production cluster as virtual machines on your development workstation.
- B. Run the hadoop command with the `-jt local` and the `-fs file:///options`.
- C. Run the DataNode, TaskTracker, NameNode and JobTracker daemons on a single machine.
- D. Run simldoop, the Apache open-source software for simulating Hadoop clusters.

Correct Answer: C

QUESTION 3

Examine the following Hive statements:

```
CREATE TABLE x (name STRING, age INT, zip INT, salary DOUBLE)
ROW FORMAT DELIMITED FIELDS TERMINATED BY
',' LOCATION '/user/joe/x';
LOAD DATA INPATH 'input/FILE1' OVERWRITE INTO TABLE x;
```



Assuming the statements above execute successfully, which one of the following statements is true?

- A. Hive reformats File1 into a structure that Hive can access and moves into to/user/joe/x/
- B. The file named File1 is moved to to/user/joe/x/
- C. The contents of File1 are parsed as comma-delimited rows and loaded into /user/joe/x/
- D. The contents of File1 are parsed as comma-delimited rows and stored in a database

Correct Answer: B

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 'data' USING PigStorage(',')
AS (gender:chararray, age:int, zip:chararray);
```

What command to define B would produce the output (M,62,95102) when invoking the DUMP operator on B?

- A. B = FILTER A BY (zip == '95102' AND gender == 'M');
- B. B = FOREACH A BY (gender == 'M' AND zip == '95102');
- C. B = JOIN A BY (gender == 'M' AND zip == '95102');
- D. B = GROUP A BY (zip == '95102' AND gender == 'M');

Correct Answer: A

QUESTION 5

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 MapReduce is not suited and some papers that address the limitations of MapReduce.

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

[APACHE-HADOOP-DEVELOPER VCE Dumps](#)

[APACHE-HADOOP-DEVELOPER Study Guide](#)

[APACHE-HADOOP-DEVELOPER Exam Questions](#)