

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.leads4pass.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

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 2

You are developing a MapReduce job for sales reporting. The mapper will process input keys representing the year (IntWritable) and input values representing product identifiers (Text).

Identify what determines the data types used by the Mapper for a given job.

- A. The key and value types specified in the `JobConf.setMapInputKeyClass` and `JobConf.setMapInputValuesClass` methods
- B. The data types specified in `HADOOP_MAP_DATATYPES` environment variable
- C. The `mapper-specification.xml` file submitted with the job determine the mapper's input key and value types.
- D. The `InputFormat` used by the job determines the mapper's input key and value types.

Correct Answer: D

Explanation: The input types fed to the mapper are controlled by the `InputFormat` used. The default input format, `TextInputFormat`, will load data in as `(LongWritable, Text)` pairs. The long value is the byte offset of the line in the file. The `Text` object holds the string contents of the line of the file.

Note: The data types emitted by the reducer are identified by `setOutputKeyClass()`

and `setOutputValueClass()`. The data types emitted by the reducer are identified by `setOutputKeyClass()` and `setOutputValueClass()`.

By default, it is assumed that these are the output types of the mapper as well. If this is not the case, the methods `setMapOutputKeyClass()` and `setMapOutputValueClass()` methods of the `JobConf` class will override these.

Reference: Yahoo! Hadoop Tutorial, THE DRIVER METHOD

QUESTION 3

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 4

Which of the following tool was designed to import data from a relational database into HDFS?

- A. HCatalog
- B. Sqoop
- C. Flume
- D. Ambari

Correct Answer: B

QUESTION 5

Given the following Pig commands:

```
logevents = LOAD 'input/my.log';
severe = FILTER logevents BY ($1 == 'severe' AND $2 >= 500);
grouped = GROUP severe BY $2;
DUMP grouped;
```

Which one of the following statements is true?

- A. The \$1 variable represents the first column of data in `\\'my.log\\'`
- B. The \$1 variable represents the second column of data in `\\'my.log\\'`
- C. The severe relation is not valid
- D. The grouped relation is not valid

Correct Answer: B

[HDPCD Practice Test](#)

[HDPCD Study Guide](#)

[HDPCD Braindumps](#)