



VCE & PDF

PassApply.com

<https://www.passapply.com/apache-hadoop-developer.html>

2024 Latest passapply APACHE-HADOOP-DEVELOPER PDF and VCE dumps  
Download

# APACHE-HADOOP-DEVELOPER<sup>Q&As</sup>

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.passapply.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

Review the following andapos;dataandapos; file and Pig code.

```
M,38,95111  
M,62,95102
```

```
A = LOAD &apos;data&apos; USING PigStorage(&apos;,&apos;)  
AS (gender:chararray, age:int, zip:chararray);  
D = GROUP A BY gender;  
DUMP D;
```

Which one of the following statements is true?

- A. The Output Of the DUMP D command IS (M,{{M,62.95102),(M,38,95111}})
- B. The output of the dump d command is (M, {(38,95in),(62,95i02}})
- C. The code executes successfully but there is not output because the D relation is empty
- D. The code does not execute successfully because D is not a valid relation

Correct Answer: A

---

### 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 simldooop, the Apache open-source software for simulating Hadoop clusters.

Correct Answer: C

---

### QUESTION 3

You have just executed a MapReduce job. Where is intermediate data written to after being emitted from the Mapper's map method?



- A. Intermediate data is streamed across the network from Mapper to the Reduce and is never written to disk.
- B. Into in-memory buffers on the TaskTracker node running the Mapper that spill over and are written into HDFS.
- C. Into in-memory buffers that spill over to the local file system of the TaskTracker node running the Mapper.
- D. Into in-memory buffers that spill over to the local file system (outside HDFS) of the TaskTracker node running the Reducer
- E. Into in-memory buffers on the TaskTracker node running the Reducer that spill over and are written into HDFS.

Correct Answer: C

Explanation: The mapper output (intermediate data) is stored on the Local file system (NOT HDFS) of each individual mapper nodes. This is typically a temporary directory location which can be setup in config by the hadoop administrator. The intermediate data is cleaned up after the Hadoop Job completes.

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, Where is the Mapper Output (intermediate key-value data) stored ?

#### QUESTION 4

Which process describes the lifecycle of a Mapper?

- A. The JobTracker calls the TaskTracker's configure () method, then its map () method and finally its close () method.
- B. The TaskTracker spawns a new Mapper to process all records in a single input split.
- C. The TaskTracker spawns a new Mapper to process each key-value pair.
- D. The JobTracker spawns a new Mapper to process all records in a single file.

Correct Answer: B

Explanation: For each map instance that runs, the TaskTracker creates a new instance of your mapper.

Note:

\*

The Mapper is responsible for processing Key/Value pairs obtained from the InputFormat. The mapper may perform a number of Extraction and Transformation functions on the Key/Value pair before ultimately outputting none, one or many Key/Value pairs of the same, or different Key/Value type.

\*

With the new Hadoop API, mappers extend the org.apache.hadoop.mapreduce.Mapper class. This class defines an 'Identity' map function by default - every input Key/Value pair obtained from the InputFormat is written out.

Examining the run() method, we can see the lifecycle of the mapper:



```
/**
 *
 * Expert users can override this method for more complete control over the
 *
 * execution of the Mapper.
 *
 * @param context
 *
 * @throws IOException
 */
public void run(Context context) throws IOException, InterruptedException { setup(context);
while (context.nextKeyValue()) {
map(context.getCurrentKey(), context.getCurrentValue(), context); }
cleanup(context);
}
setup(Context) - Perform any setup for the mapper. The default implementation is a no-op method. map(Key, Value,
Context) - Perform a map operation in the given Key / Value pair. The default implementation calls Context.write(Key,
Value) cleanup(Context) - Perform any cleanup for the mapper. The default implementation is a no-op method.
```

Reference: Hadoop/MapReduce/Mapper

---

## QUESTION 5

Can you use MapReduce to perform a relational join on two large tables sharing a key? Assume that the two tables are formatted as comma-separated files in HDFS.

- A. Yes.
- B. Yes, but only if one of the tables fits into memory
- C. Yes, so long as both tables fit into memory.
- D. No, MapReduce cannot perform relational operations.
- E. No, but it can be done with either Pig or Hive.

Correct Answer: A

Explanation: Note:



\*

Join Algorithms in MapReduce A) Reduce-side join B) Map-side join

C) In-memory join / Striped Striped variant variant / Memcached variant

\*

Which join to use? / In-memory join > map-side join > reduce-side join / Limitations of each? In-memory join: memory  
Map-side join: sort order and partitioning Reduce-side join: general purpose

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

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

[APACHE-HADOOP-DEVELOPER Practice Test](#)