

# HADOOP-PR000007<sup>Q&As</sup>

Hortonworks Certified Apache Hadoop 2.0 Developer (Pig and Hive Developer)

## Pass Hortonworks HADOOP-PR000007 Exam with 100% Guarantee

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

<https://www.pass2lead.com/hadoop-pr000007.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

Which describes how a client reads a file from HDFS?

- A. The client queries the NameNode for the block location(s). The NameNode returns the block location(s) to the client. The client reads the data directory off the DataNode(s).
- B. The client queries all DataNodes in parallel. The DataNode that contains the requested data responds directly to the client. The client reads the data directly off the DataNode.
- C. The client contacts the NameNode for the block location(s). The NameNode then queries the DataNodes for block locations. The DataNodes respond to the NameNode, and the NameNode redirects the client to the DataNode that holds the requested data block(s). The client then reads the data directly off the DataNode.
- D. The client contacts the NameNode for the block location(s). The NameNode contacts the DataNode that holds the requested data block. Data is transferred from the DataNode to the NameNode, and then from the NameNode to the client.

Correct Answer: A

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, How the Client communicates with HDFS?

---

### QUESTION 2

Given the following Hive command:

```
INSERT OVERWRITE TABLE mytable SELECT * FROM myothertable;
```

Which one of the following statements is true?

- A. The contents of myothertable are appended to mytable
- B. Any existing data in mytable will be overwritten
- C. A new table named mytable is created, and the contents of myothertable are copied into mytable
- D. The statement is not a valid Hive command

Correct Answer: B

---

### QUESTION 3

What data does a Reducer reduce method process?

- A. All the data in a single input file.
- B. All data produced by a single mapper.
- C. All data for a given key, regardless of which mapper(s) produced it.

D. All data for a given value, regardless of which mapper(s) produced it.

Correct Answer: C

Explanation: Reducing lets you aggregate values together. A reducer function receives an iterator of input values from an input list. It then combines these values together, returning a single output value.

All values with the same key are presented to a single reduce task.

Reference: Yahoo! Hadoop Tutorial, Module 4: MapReduce

---

#### QUESTION 4

A client application creates an HDFS file named foo.txt with a replication factor of 3. Identify which best describes the file access rules in HDFS if the file has a single block that is stored on data nodes A, B and C?

- A. The file will be marked as corrupted if data node B fails during the creation of the file.
- B. Each data node locks the local file to prohibit concurrent readers and writers of the file.
- C. Each data node stores a copy of the file in the local file system with the same name as the HDFS file.
- D. The file can be accessed if at least one of the data nodes storing the file is available.

Correct Answer: D

Explanation: HDFS keeps three copies of a block on three different datanodes to protect against true data corruption. HDFS also tries to distribute these three replicas on more than one rack to protect against data availability issues. The fact that HDFS actively monitors any failed datanode(s) and upon failure detection immediately schedules re-replication of blocks (if needed) implies that three copies of data on three different nodes is sufficient to avoid corrupted files. Note: HDFS is designed to reliably store very large files across machines in a large cluster. It stores each file as a sequence of blocks; all blocks in a file except the last block are the same size. The blocks of a file are replicated for fault tolerance. The block size and replication factor are configurable per file. An application can specify the number of replicas of a file. The replication factor can be specified at file creation time and can be changed later. Files in HDFS are write-once and have strictly one writer at any time. The NameNode makes all decisions regarding replication of blocks. HDFS uses rack-aware replica placement policy. In default configuration there are total 3 copies of a datablock on HDFS, 2 copies are stored on datanodes on same rack and 3rd copy on a different rack. Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers , How the HDFS Blocks are replicated?

---

#### QUESTION 5

To process input key-value pairs, your mapper needs to load a 512 MB data file in memory. What is the best way to accomplish this?

- A. Serialize the data file, insert in it the JobConf object, and read the data into memory in the configure method of the mapper.
- B. Place the data file in the DistributedCache and read the data into memory in the map method of the mapper.
- C. Place the data file in the DataCache and read the data into memory in the configure method of the mapper.
- D. Place the data file in the DistributedCache and read the data into memory in the configure method of the mapper.

Correct Answer: C

[HADOOP-PR000007 PDF Dumps](#)

[HADOOP-PR000007 Study Guide](#)

[HADOOP-PR000007 Exam Questions](#)