



VCE & PDF

Pass4itSure.com

<https://www.pass4itsure.com/apache-hadoop-developer.html>  
2024 Latest pass4itsure 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.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

What does Pig provide to the overall Hadoop solution?

- A. Legacy language Integration with MapReduce framework
- B. Simple scripting language for writing MapReduce programs
- C. Database table and storage management services
- D. C++ interface to MapReduce and data warehouse infrastructure

Correct Answer: B

---

### QUESTION 2

Table metadata in Hive is:

- A. Stored as metadata on the NameNode.
- B. Stored along with the data in HDFS.
- C. Stored in the Metastore.
- D. Stored in ZooKeeper.

Correct Answer: C

Explanation: By default, hive use an embedded Derby database to store metadata information. The metastore is the "glue" between Hive and HDFS. It tells Hive where your data files live in HDFS, what type of data they contain, what tables they belong to, etc.

The Metastore is an application that runs on an RDBMS and uses an open source ORM layer called DataNucleus, to convert object representations into a relational schema and vice versa. They chose this approach as opposed to storing this information in hdfs as they need the Metastore to be very low latency. The DataNucleus layer allows them to plugin many different RDBMS technologies.

Note:

\*

By default, Hive stores metadata in an embedded Apache Derby database, and other client/server databases like MySQL can optionally be used.

\*

features of Hive include:

Metadata storage in an RDBMS, significantly reducing the time to perform semantic checks during query execution.



Reference: Store Hive Metadata into RDBMS

---

### 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

You need to perform statistical analysis in your MapReduce job and would like to call methods in the Apache Commons Math library, which is distributed as a 1.3 megabyte Java archive (JAR) file. Which is the best way to make this library available to your MapReducer job at runtime?

- A. Have your system administrator copy the JAR to all nodes in the cluster and set its location in the HADOOP\_CLASSPATH environment variable before you submit your job.
- B. Have your system administrator place the JAR file on a Web server accessible to all cluster nodes and then set the HTTP\_JAR\_URL environment variable to its location.
- C. When submitting the job on the command line, specify the ?ibjars option followed by the JAR file path.
- D. Package your code and the Apache Commons Math library into a zip file named JobJar.zip

Correct Answer: C

Explanation: The usage of the jar command is like this,

Usage: `hadoop jar [mainClass] args...`

If you want the commons-math3.jar to be available for all the tasks you can do any one of these

1. Copy the jar file in \$HADOOP\_HOME/lib dir

2.

or



Use the generic option -libjars.

---

### QUESTION 5

You wrote a map function that throws a runtime exception when it encounters a control character in input data. The input supplied to your mapper contains twelve such characters totals, spread across five file splits. The first four file splits each have two control characters and the last split has four control characters.

Identify the number of failed task attempts you can expect when you run the job with `mapred.max.map.attempts` set to 4:

- A. You will have forty-eight failed task attempts
- B. You will have seventeen failed task attempts
- C. You will have five failed task attempts
- D. You will have twelve failed task attempts
- E. You will have twenty failed task attempts

Correct Answer: E

Explanation: There will be four failed task attempts for each of the five file splits.

Note:

When the jobtracker is notified of a task attempt that has failed (by the tasktracker's heartbeat call), it will reschedule execution of the task. The jobtracker will try to avoid rescheduling the task on a tasktracker where it has previously failed. Furthermore, if a task fails four times (or more), it will not be retried further. This value is configurable: the maximum number of attempts to run a task is controlled by the `mapred.map.max.attempts` property for map tasks and `mapred.reduce.max.attempts` for reduce tasks. By default, if any task fails four times (or whatever the maximum number of attempts is configured to), the whole job fails.

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

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

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