



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.geekcert.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

Identify which best defines a SequenceFile?

- A. A SequenceFile contains a binary encoding of an arbitrary number of homogeneous Writable objects
- B. A SequenceFile contains a binary encoding of an arbitrary number of heterogeneous Writable objects
- C. A SequenceFile contains a binary encoding of an arbitrary number of WritableComparable objects, in sorted order.
- D. A SequenceFile contains a binary encoding of an arbitrary number key-value pairs. Each key must be the same type. Each value must be the same type.

Correct Answer: D

Explanation: SequenceFile is a flat file consisting of binary key/value pairs.

There are 3 different SequenceFile formats:

Uncompressed key/value records.

Record compressed key/value records - only `values` are compressed here. Block compressed key/value records - both keys and values are collected in `blocks` separately and compressed. The size of the `block` is configurable.

Reference: <http://wiki.apache.org/hadoop/SequenceFile>

QUESTION 2

Determine which best describes when the reduce method is first called in a MapReduce job?

- A. Reducers start copying intermediate key-value pairs from each Mapper as soon as it has completed. The programmer can configure in the job what percentage of the intermediate data should arrive before the reduce method begins.
- B. Reducers start copying intermediate key-value pairs from each Mapper as soon as it has completed. The reduce method is called only after all intermediate data has been copied and sorted.
- C. Reduce methods and map methods all start at the beginning of a job, in order to provide optimal performance for map-only or reduce-only jobs.
- D. Reducers start copying intermediate key-value pairs from each Mapper as soon as it has completed. The reduce method is called as soon as the intermediate key-value pairs start to arrive.

Correct Answer: B

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers , When is the reducers are started in a MapReduce job?



QUESTION 3

For each input key-value pair, mappers can emit:

- A. As many intermediate key-value pairs as designed. There are no restrictions on the types of those key-value pairs (i.e., they can be heterogeneous).
- B. As many intermediate key-value pairs as designed, but they cannot be of the same type as the input key-value pair.
- C. One intermediate key-value pair, of a different type.
- D. One intermediate key-value pair, but of the same type.
- E. As many intermediate key-value pairs as designed, as long as all the keys have the same types and all the values have the same type.

Correct Answer: E

Explanation: Mapper maps input key/value pairs to a set of intermediate key/value pairs.

Maps are the individual tasks that transform input records into intermediate records. The transformed intermediate records do not need to be of the same type as the input records. A given input pair may map to zero or many output pairs.

Reference: Hadoop Map-Reduce Tutorial

QUESTION 4

Your client application submits a MapReduce job to your Hadoop cluster. Identify the Hadoop daemon on which the Hadoop framework will look for an available slot schedule a MapReduce operation.

- A. TaskTracker
- B. NameNode
- C. DataNode
- D. JobTracker
- E. Secondary NameNode

Correct Answer: D

Explanation: JobTracker is the daemon service for submitting and tracking MapReduce jobs in Hadoop. There is only One Job Tracker process run on any hadoop cluster. Job Tracker runs on its own JVM process. In a typical production cluster its run on a separate machine. Each slave node is configured with job tracker node location. The JobTracker is single point of failure for the Hadoop MapReduce service. If it goes down, all running jobs are halted. JobTracker in Hadoop performs following actions(from Hadoop Wiki:)

Client applications submit jobs to the Job tracker. The JobTracker talks to the NameNode to determine the location of the data The JobTracker locates TaskTracker nodes with available slots at or near the data The JobTracker submits the work to the chosen TaskTracker nodes. The TaskTracker nodes are monitored. If they do not submit heartbeat signals often enough, they are deemed to have failed and the work is scheduled on a different TaskTracker. A TaskTracker will notify the JobTracker when a task fails. The JobTracker decides what to do then: it may resubmit the job elsewhere, it may mark that specific record as something to avoid, and it may may even blacklist the TaskTracker as unreliable.



When the work is completed, the JobTracker updates its status.

Client applications can poll the JobTracker for information.

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, What is a JobTracker in Hadoop?
How many instances of JobTracker run on a Hadoop Cluster?

QUESTION 5

Examine the following Hive statements:

```
CREATE TABLE x (name STRING, age INT, zip INT, salary DOUBLE)
ROW FORMAT DELIMITED FIELDS TERMINATED BY
',';
LOAD DATA INPATH 'input/File1' OVERWRITE INTO TABLE x;
SELECT * FROM x SORT BY age;
```

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

- A. Each reducer generates a file sorted by age
- B. The SORT BY command causes only one reducer to be used
- C. The output of each reducer is only the age column
- D. The output is guaranteed to be a single file with all the data sorted by age

Correct Answer: A

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

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

[APACHE-HADOOP-DEVELOPER Braindumps](#)