



HADOOP-PR000007^{Q&As}

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.passapply.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

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 2

A combiner reduces:

- A. The number of values across different keys in the iterator supplied to a single reduce method call.
- B. The amount of intermediate data that must be transferred between the mapper and reducer.
- C. The number of input files a mapper must process.
- D. The number of output files a reducer must produce.

Correct Answer: B

Explanation: Combiners are used to increase the efficiency of a MapReduce program. They are used to aggregate intermediate map output locally on individual mapper outputs. Combiners can help you reduce the amount of data that needs to be transferred across to the reducers. You can use your reducer code as a combiner if the operation performed is commutative and associative. The execution of combiner is not guaranteed, Hadoop may or may not execute a combiner. Also, if required it may execute it more than 1 times. Therefore your MapReduce jobs should not depend on the combiners execution.

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, What are combiners? When should I use a combiner in my MapReduce Job?

QUESTION 3

What is the term for the process of moving map outputs to the reducers?

- A. Reducing



- B. Combining
- C. Partitioning
- D. Shuffling and sorting

Correct Answer: D

QUESTION 4

Which one of the following statements is true about a Hive-managed table?

- A. Records can only be added to the table using the Hive INSERT command.
- B. When the table is dropped, the underlying folder in HDFS is deleted.
- C. Hive dynamically defines the schema of the table based on the FROM clause of a SELECT query.
- D. Hive dynamically defines the schema of the table based on the format of the underlying data.

Correct Answer: B

QUESTION 5

Which best describes how TextInputFormat processes input files and line breaks?

- A. Input file splits may cross line breaks. A line that crosses file splits is read by the RecordReader of the split that contains the beginning of the broken line.
- B. Input file splits may cross line breaks. A line that crosses file splits is read by the RecordReaders of both splits containing the broken line.
- C. The input file is split exactly at the line breaks, so each RecordReader will read a series of complete lines.
- D. Input file splits may cross line breaks. A line that crosses file splits is ignored.
- E. Input file splits may cross line breaks. A line that crosses file splits is read by the RecordReader of the split that contains the end of the broken line.

Correct Answer: A

Reference: How Map and Reduce operations are actually carried out

QUESTION 6

Which one of the following statements describes a Pig bag, tuple, and map, respectively?

- A. Unordered collection of maps, ordered collection of tuples, ordered set of key/value pairs
- B. Unordered collection of tuples, ordered set of fields, set of key value pairs



- C. Ordered set of fields, ordered collection of tuples, ordered collection of maps
- D. Ordered collection of maps, ordered collection of bags, and unordered set of key/value pairs

Correct Answer: B

QUESTION 7

What are the TWO main components of the YARN ResourceManager process? Choose 2 answers

- A. Job Tracker
- B. Task Tracker
- C. Scheduler
- D. Applications Manager

Correct Answer: CD

QUESTION 8

Given a directory of files with the following structure: line number, tab character, string: Example: 1abialkjjkaoasdfjksdlkjhweroij 2kadjhuwqounahagtnbvaswslmnbfgy 3kjfteiomndscxeqalkzhtopedkfsikj You want to send each line as one record to your Mapper. Which InputFormat should you use to complete

the line: `conf.setInputFormat (____.class) ; ?`

- A. SequenceFileAsTextInputFormat
- B. SequenceFileInputFormat
- C. KeyValueFileInputFormat
- D. BDBInputFormat

Correct Answer: C

<http://stackoverflow.com/questions/9721754/how-to-parse-customwritable-from-text-in-hadoop>

QUESTION 9

In a MapReduce job, the reducer receives all values associated with same key. Which statement best describes the ordering of these values?

- A. The values are in sorted order.
- B. The values are arbitrarily ordered, and the ordering may vary from run to run of the same MapReduce job.
- C. The values are arbitrary ordered, but multiple runs of the same MapReduce job will always have the same ordering.



D. Since the values come from mapper outputs, the reducers will receive contiguous sections of sorted values.

Correct Answer: B

Note:

*

Input to the Reducer is the sorted output of the mappers.

*

The framework calls the application's Reduce function once for each unique key in the sorted order.

*

Example:

For the given sample input the first map emits:

The second map emits:

QUESTION 10

Your cluster's HDFS block size is 64MB. You have a directory containing 100 plain text files, each of which is 100MB in size. The InputFormat for your job is TextInputFormat.

Determine how many Mappers will run?

A. 64

B. 100

C. 200

D. 640

Correct Answer: C



Explanation: Each file would be split into two as the block size (64 MB) is less than the file size (100 MB), so 200 mappers would be running.

Note:

If you're not compressing the files then hadoop will process your large files (say 10G), with a number of mappers related to the block size of the file.

Say your block size is 64M, then you will have ~160 mappers processing this 10G file ($160 \times 64 \approx 10G$).

Depending on how CPU intensive your mapper logic is, this might be an acceptable blocks size, but if you find that your mappers are executing in sub minute times, then you might want to increase the work done by each mapper (by increasing the block size to 128, 256, 512m - the actual size depends on how you intend to process the data). Reference: <http://stackoverflow.com/>

QUESTION 11

Which best describes what the map method accepts and emits?

- A. It accepts a single key-value pair as input and emits a single key and list of corresponding values as output.
- B. It accepts a single key-value pairs as input and can emit only one key-value pair as output.
- C. It accepts a list key-value pairs as input and can emit only one key-value pair as output.
- D. It accepts a single key-value pairs as input and can emit any number of key-value pair as output, including zero.

Correct Answer: D

Explanation: public class Mapper extends Object Maps input key/value pairs to a set of intermediate key/value pairs.

Maps are the individual tasks which transform input records into a intermediate records. The transformed intermediate records need not be of the same type as the input records. A given input pair may map to zero or many output pairs. Reference: org.apache.hadoop.mapreduce

Class Mapper

QUESTION 12

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 13

Which one of the following statements describes a Hive user-defined aggregate function?

- A. Operates on multiple input rows and creates a single row as output
- B. Operates on a single input row and produces a single row as output
- C. Operates on a single input row and produces a table as output
- D. Operates on multiple input rows and produces a table as output

Correct Answer: A

QUESTION 14

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 Reducer 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 15

Given the following Pig commands: Which one of the following statements is true?



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

- 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

[HADOOP-PR000007 Study Guide](#)

[HADOOP-PR000007 Exam Questions](#)

[HADOOP-PR000007 Braindumps](#)