

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

Examine the following Hive statements:

```
CREATE TABLE x (name STRING, age INT, zip INT, şalary 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

QUESTION 2

You need to create a job that does frequency analysis on input data. You will do this by writing a Mapper that uses TextInputFormat and splits each value (a line of text from an input file) into individual characters. For each one of these characters, you will emit the character as a key and an InputWritable as the value. As this will produce proportionally more intermediate data than input data, which two resources should you expect to be bottlenecks?

A. Processor and network I/O

- B. Disk I/O and network I/O
- C. Processor and RAM
- D. Processor and disk I/O

Correct Answer: B

QUESTION 3

You want to count the number of occurrences for each unique word in the supplied input data. You\\'ve decided to implement this by having your mapper tokenize each word and emit a literal value 1, and then have your reducer increment a counter for each literal 1 it receives. After successful implementing this, it occurs to you that you could optimize this by specifying a combiner. Will you be able to reuse your existing Reduces as your combiner in this case and why or why not?

A. Yes, because the sum operation is both associative and commutative and the input and output types to the reduce method match.



B. No, because the sum operation in the reducer is incompatible with the operation of a Combiner.

C. No, because the Reducer and Combiner are separate interfaces.

D. No, because the Combiner is incompatible with a mapper which doesn\\'t use the same data type for both the key and value.

E. Yes, because Java is a polymorphic object-oriented language and thus reducer code can be reused as a combiner.

Correct Answer: A

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 then 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 4

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.

Indentify 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.

QUESTION 5

Which Two of the following statements are true about hdfs? Choose 2 answers

- A. An HDFS file that is larger than dfs.block.size is split into blocks
- B. Blocks are replicated to multiple datanodes
- C. HDFS works best when storing a large number of relatively small files
- D. Block sizes for all files must be the same size

Correct Answer: AB

HADOOP-PR000007 PDF Dumps HADOOP-PR000007 Exam Questions HADOOP-PR000007 Braindumps