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.geekcert.com/hadoop-pr000007.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Hortonworks Official Exam Center

QUESTION 1

In a MapReduce job with 500 map tasks, how many map task attempts will there be?

A. It depends on the number of reduces in the job.

B. Between 500 and 1000.

C. At most 500.

D. At least 500.

E. Exactly 500.

Correct Answer: D

From Cloudera Training Course: Task attempt is a particular instance of an attempt to execute a task ?There will be at least as many task attempts as there are tasks ?If a task attempt fails, another will be started by the JobTracker ?Speculative execution can also result in more task attempts than completed tasks

---

QUESTION 2

Given the following Pig command:

logevents = LOAD andapos;input/my.logandapos; AS (date:chararray, levehstring, code:int, message:string);

Which one of the following statements is true?

A. The logevents relation represents the data from the my.log file, using a comma as the parsing delimiter

B. The logevents relation represents the data from the my.log file, using a tab as the parsing delimiter

C. The first field of logevents must be a properly-formatted date string or table return an error

D. The statement is not a valid Pig command

Correct Answer: B

QUESTION 3

Which YARN component is responsible for monitoring the success or failure of a Container?

A. ResourceManager

B. ApplicationMaster

C. NodeManager

D. JobTracker

Correct Answer: A

QUESTION 4

Assuming the following Hive query executes successfully:

```
from inputdata select context_ngrams(sentences(lines),
array("you", "are", null), 80);
```

Which one of the following statements describes the result set?

A. A bigram of the top 80 sentences that contain the substring "you are" in the lines column of the input data A1 table.

B. An 80-value ngram of sentences that contain the words "you" or "are" in the lines column of the inputdata table.

C. A trigram of the top 80 sentences that contain "you are" followed by a null space in the lines column of the inputdata table.

D. A frequency distribution of the top 80 words that follow the subsequence "you are" in the lines column of the inputdata table.

Correct Answer: D

QUESTION 5

The Hadoop framework provides a mechanism for coping with machine issues such as faulty configuration or impending hardware failure. MapReduce detects that one or a number of machines are performing poorly and starts more copies of a map or reduce task. All the tasks run simultaneously and the task finish first are used. This is called:

A. Combine

B. IdentityMapper

C. IdentityReducer

D. Default Partitioner

E. Speculative Execution

Correct Answer: E

Explanation: Speculative execution: One problem with the Hadoop system is that by dividing the tasks across many nodes, it is possible for a few slow nodes to rate-limit the rest of the program. For example if one node has a slow disk controller, then it may be reading its input at only 10% the speed of all the other nodes. So when 99 map tasks are already complete, the system is still waiting for the final map task to check in, which takes much longer than all the other nodes. By forcing tasks to run in isolation from one another, individual tasks do not know where their inputs come from. Tasks trust the Hadoop platform to just deliver the appropriate input. Therefore, the same input can be processed multiple times in parallel, to exploit differences in machine capabilities. As most of the tasks in a job are coming to a close, the Hadoop platform will schedule redundant copies of the remaining tasks across several nodes which do not have other work to perform. This process is known as speculative execution. When tasks complete, they announce this fact to the JobTracker. Whichever copy of a task finishes first becomes the definitive copy. If other copies were executing speculatively, Hadoop tells the TaskTrackers to abandon the tasks and discard their outputs. The Reducers then receive their inputs from whichever Mapper completed successfully, first.

Reference: Apache Hadoop, Module 4: MapReduce

Note:

*

Hadoop uses "speculative execution." The same task may be started on multiple boxes. The first one to finish wins, and the other copies are killed.

*

There are a few reasons Hadoop can kill tasks by his own decisions:

Failed tasks are tasks that error out.

a) Task does not report progress during timeout (default is 10 minutes)

b) FairScheduler or CapacityScheduler needs the slot for some other pool (FairScheduler) or queue (CapacityScheduler).

c) Speculative execution causes results of task not to be needed since it has completed on other place.

Reference: Difference failed tasks vs killed tasks