



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.pass4itsure.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**

Can you use MapReduce to perform a relational join on two large tables sharing a key? Assume that the two tables are formatted as comma-separated files in HDFS.

- A. Yes.
- B. Yes, but only if one of the tables fits into memory
- C. Yes, so long as both tables fit into memory.
- D. No, MapReduce cannot perform relational operations.
- E. No, but it can be done with either Pig or Hive.

Correct Answer: A

Explanation: Note:

*

Join Algorithms in MapReduce A) Reduce-side join B) Map-side join

C) In-memory join / Striped Striped variant variant / Memcached variant

*

Which join to use? / In-memory join > map-side join > reduce-side join / Limitations of each? In-memory join: memory
Map-side join: sort order and partitioning Reduce-side join: general purpose

QUESTION 2

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 3



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 Commands 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 4

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 5

MapReduce v2 (MRv2/YARN) splits which major functions of the JobTracker into separate daemons? Select two.

- A. Health states checks (heartbeats)
- B. Resource management



- C. Job scheduling/monitoring
- D. Job coordination between the ResourceManager and NodeManager
- E. Launching tasks
- F. Managing file system metadata
- G. MapReduce metric reporting H. Managing tasks

Correct Answer: BC

Explanation: The fundamental idea of MRv2 is to split up the two major functionalities of the JobTracker, resource management and job scheduling/monitoring, into separate daemons. The idea is to have a global ResourceManager (RM) and per-application ApplicationMaster (AM). An application is either a single job in the classical sense of Map- Reduce jobs or a DAG of jobs.

Note:

The central goal of YARN is to clearly separate two things that are unfortunately smushed together in current Hadoop, specifically in (mainly) JobTracker:

/ Monitoring the status of the cluster with respect to which nodes have which resources available. Under YARN, this will be global.

/ Managing the parallelization execution of any specific job. Under YARN, this will be done separately for each job.

Reference: Apache Hadoop YARN ?Conceptsand; Applications

[Latest HADOOP-PR000007 Dumps](#)

[HADOOP-PR000007 VCE Dumps](#)

[HADOOP-PR000007 Exam Questions](#)