CCD-410 Q&As

Cloudera Certified Developer for Apache Hadoop (CCDH)

Pass Cloudera CCD-410 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.geekcert.com/ccd-410.html

100% Passing Guarantee
100% Money Back Assurance

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

- **Instant Download** After Purchase
- **100% Money Back** Guarantee
- **365 Days** Free Update
- **800,000+** Satisfied Customers

QUESTION 1

You have user profile records in your OLPT database, that you want to join with web logs you have already ingested into the Hadoop file system. How will you obtain these user records?

A. HDFS command

B. Pig LOAD command

C. Sqoop import

D. Hive LOAD DATA command

E. Ingest with Flume agents F. Ingest with Hadoop Streaming

Correct Answer: C

Reference: Hadoop and Pig for Large-Scale Web Log Analysis

QUESTION 2

What is the disadvantage of using multiple reducers with the default HashPartitioner and distributing your workload across you cluster?

A. You will not be able to compress the intermediate data.

B. You will longer be able to take advantage of a Combiner.

C. By using multiple reducers with the default HashPartitioner, output files may not be in globally sorted order.

D. There are no concerns with this approach. It is always advisable to use multiple reduces.

Correct Answer: C

Multiple reducers and total ordering

If your sort job runs with multiple reducers (either because mapreduce.job.reduces in mapred- site.xml has been set to a number larger than 1, or because you\\'ve used the -r option to specify the number of reducers on the command-line), then by default Hadoop will use the HashPartitioner to distribute records across the reducers. Use of the HashPartitioner means that you can\\'t concatenate your output files to create a single sorted output file. To do this you\\'ll need total ordering,

Reference: Sorting text files with MapReduce

QUESTION 3

You have the following key-value pairs as output from your Map task:

(the, 1) (fox, 1) (faster, 1) (than, 1) (the, 1) (dog, 1)

How many keys will be passed to the Reducer\\'s reduce method?

A. Six

B. Five

C. Four

D. Two

E. One

F. Three

Correct Answer: B

Only one key value pair will be passed from the two (the, 1) key value pairs.

QUESTION 4

Given a directory of files with the following structure: line number, tab character, string:

Example: 1 abialkjfjkaoasdfjksdlkjhqweroij 2 kadfjhuwqounahagtnbvaswslmnbfgy 3 kjfteiomndscxeqalkzhtopedkfsikj

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 5

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 libjars 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

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 or

2.

Use the generic option -libjars.