



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 want to populate an associative array in order to perform a map-side join. You've decided to put this information in a text file, place that file into the DistributedCache and read it in your Mapper before any records are processed.

Identify which method in the Mapper you should use to implement code for reading the file and populating the associative array?

- A. combine
- B. map
- C. init
- D. configure

Correct Answer: B

Reference: `org.apache.hadoop.filecache` , Class `DistributedCache`

QUESTION 2

You have a directory named `jobdata` in HDFS that contains four files: `_first.txt`, `second.txt`, `.third.txt` and `#data.txt`. How many files will be processed by the `FileInputFormat.setInputPaths ()` command when it's given a path object representing this directory?

- A. Four, all files will be processed
- B. Three, the pound sign is an invalid character for HDFS file names
- C. Two, file names with a leading period or underscore are ignored
- D. None, the directory cannot be named `jobdata`
- E. One, no special characters can prefix the name of an input file

Correct Answer: C

Files starting with `_'_` are considered 'hidden' like unix files starting with `.'.`

`#` characters are allowed in HDFS file names.

QUESTION 3

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: E

As the Map operation is parallelized the input file set is first split to several pieces called FileSplits. If an individual file is so large that it will affect seek time it will be split to several Splits. The splitting does not know anything about the input file's internal logical structure, for example line-oriented text files are split on arbitrary byte boundaries. Then a new map task is created per FileSplit.

When an individual map task starts it will open a new output writer per configured reduce task. It will then proceed to read its FileSplit using the RecordReader it gets from the specified InputFormat. InputFormat parses the input and generates key-value pairs. InputFormat must also handle records that may be split on the FileSplit boundary. For example TextInputFormat will read the last line of the FileSplit past the split boundary and, when reading other than the first FileSplit, TextInputFormat ignores the content up to the first newline.

Reference: How Map and Reduce operations are actually carried out

QUESTION 4

You want to run Hadoop jobs on your development workstation for testing before you submit them to your production cluster. Which mode of operation in Hadoop allows you to most closely simulate a production cluster while using a single machine?

- A. Run all the nodes in your production cluster as virtual machines on your development workstation.
- B. Run the hadoop command with the jt local and the fs file:///options.
- C. Run the DataNode, TaskTracker, NameNode and JobTracker daemons on a single machine.
- D. Run simldoop, the Apache open-source software for simulating Hadoop clusters.

Correct Answer: C

QUESTION 5

In a large MapReduce job with m mappers and n reducers, how many distinct copy operations will there be in the sort/shuffle phase?

- A. $m \times n$ (i.e., m multiplied by n)
- B. n
- C. m
- D. $m+n$ (i.e., m plus n)



E. mn (i.e., m to the power of n)

Correct Answer: A

A MapReduce job with m mappers and r reducers involves up to $m * r$ distinct copy operations, since each mapper may have intermediate output going to every reducer.

[CCD-410 PDF Dumps](#)

[CCD-410 VCE Dumps](#)

[CCD-410 Study Guide](#)