



# CCD-410<sup>Q&As</sup>

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.passapply.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 are developing a MapReduce job for sales reporting. The mapper will process input keys representing the year (IntWritable) and input values representing product identifiers (Text). Identify what determines the data types used by the Mapper for a given job.

- A. The key and value types specified in the JobConf.setMapInputKeyClass and JobConf.setMapInputValuesClass methods
- B. The data types specified in HADOOP\_MAP\_DATATYPES environment variable
- C. The mapper-specification.xml file submitted with the job determine the mapper's input key and value types.
- D. The InputFormat used by the job determines the mapper's input key and value types.

Correct Answer: D

The input types fed to the mapper are controlled by the InputFormat used. The default input format, "TextInputFormat," will load data in as (LongWritable, Text) pairs. The long value is the byte offset of the line in the file. The Text object holds the string contents of the line of the file.

Note: The data types emitted by the reducer are identified by setOutputKeyClass() and setOutputValueClass(). The data types emitted by the reducer are identified by setOutputKeyClass() and setOutputValueClass().

By default, it is assumed that these are the output types of the mapper as well. If this is not the case, the methods setMapOutputKeyClass() and setMapOutputValueClass() methods of the JobConf class will



override these.

Reference: Yahoo! Hadoop Tutorial, THE DRIVER METHOD

---

### QUESTION 3

In a MapReduce job, the reducer receives all values associated with same key. Which statement best describes the ordering of these values?

- A. The values are in sorted order.
- B. The values are arbitrarily ordered, and the ordering may vary from run to run of the same MapReduce job.
- C. The values are arbitrary ordered, but multiple runs of the same MapReduce job will always have the same ordering.
- D. Since the values come from mapper outputs, the reducers will receive contiguous sections of sorted values.

Correct Answer: B

Note:

\*

Input to the Reducer is the sorted output of the mappers.

\*

The framework calls the application's Reduce function once for each unique key in the sorted order.

\*

Example:

For the given sample input the first map emits:

The second map emits:



#### QUESTION 4

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

Example: 1 abialkjjkaoasdfjksdlkjhqwer0ij 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 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

[CCD-410 PDF Dumps](#)

[CCD-410 Practice Test](#)

[CCD-410 Braindumps](#)