



# 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.pass4itsure.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**

Identify which best defines a SequenceFile?

- A. A SequenceFile contains a binary encoding of an arbitrary number of homogeneous Writable objects
- B. A SequenceFile contains a binary encoding of an arbitrary number of heterogeneous Writable objects
- C. A SequenceFile contains a binary encoding of an arbitrary number of WritableComparable objects, in sorted order.
- D. A SequenceFile contains a binary encoding of an arbitrary number key-value pairs. Each key must be the same type. Each value must be the same type.

Correct Answer: D

SequenceFile is a flat file consisting of binary key/value pairs.

There are 3 different SequenceFile formats:

Uncompressed key/value records.

Record compressed key/value records - only `\\values\\` are compressed here.

Block compressed key/value records - both keys and values are collected in `\\blocks\\` separately and

compressed. The size of the `\\block\\` is configurable.

Reference: <http://wiki.apache.org/hadoop/SequenceFile>

---

**QUESTION 2**

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

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 3

When can a reduce class also serve as a combiner without affecting the output of a MapReduce program?

- A. When the types of the reduce operation's input key and input value match the types of the reducer's output key and output value and when the reduce operation is both communicative and associative.
- B. When the signature of the reduce method matches the signature of the combine method.
- C. Always. Code can be reused in Java since it is a polymorphic object-oriented programming language.
- D. Always. The point of a combiner is to serve as a mini-reducer directly after the map phase to increase performance.
- E. Never. Combiners and reducers must be implemented separately because they serve different purposes.

Correct Answer: A

You can use your reducer code as a combiner if the operation performed is commutative and associative.

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, What are combiners? When should I use a combiner in my MapReduce Job?

---

### QUESTION 4

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

---

### QUESTION 5

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



- A. Heath 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

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 Concepts and Applications

[Latest CCD-410 Dumps](#)

[CCD-410 Practice Test](#)

[CCD-410 Exam Questions](#)