

# 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.pass2lead.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

What is 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 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

For each intermediate key, each reducer task can emit:

- A. As many final key-value pairs as desired. There are no restrictions on the types of those key- value pairs (i.e., they can be heterogeneous).
- B. As many final key-value pairs as desired, but they must have the same type as the intermediate key-value pairs.
- C. As many final key-value pairs as desired, as long as all the keys have the same type and all the values have the same type.
- D. One final key-value pair per value associated with the key; no restrictions on the type.
- E. One final key-value pair per key; no restrictions on the type.

Correct Answer: E

Reducer reduces a set of intermediate values which share a key to a smaller set of values.

Reducing lets you aggregate values together. A reducer function receives an iterator of input values from an input list. It then combines these values together, returning a single output value. Reference: Hadoop Map-Reduce Tutorial; Yahoo! Hadoop Tutorial, Module 4: MapReduce

---

### QUESTION 3

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 4

Identify the tool best suited to import a portion of a relational database every day as files into HDFS, and generate Java classes to interact with that imported data?

- A. Oozie
- B. Flume
- C. Pig
- D. Hue
- E. Hive
- F. Sqoop
- G. fuse-dfs

Correct Answer: F

Sqoop ("SQL-to-Hadoop") is a straightforward command-line tool with the following capabilities: Imports individual tables or entire databases to files in HDFS Generates Java classes to allow you to interact with your imported data Provides the ability to import from SQL databases straight into your Hive data warehouse Note: Data Movement Between Hadoop and Relational Databases Data can be moved between Hadoop and a relational database as a bulk data transfer, or relational tables

can be accessed from within a MapReduce map function. Note:

\* Cloudera's Distribution for Hadoop provides a bulk data transfer tool (i.e., Sqoop) that imports individual tables or entire databases into HDFS files. The tool also generates Java classes that support interaction with the imported data. Sqoop supports all relational databases over JDBC, and Quest Software provides a connector (i.e., OraOop) that has been optimized for access to data residing in Oracle databases.

Reference: <http://log.medcl.net/item/2011/08/hadoop-and-mapreduce-big-data-analytics-gartner/> (Data Movement between hadoop and relational databases, second paragraph)

---

**QUESTION 5**

Your cluster's HDFS block size is 64MB. You have a directory containing 100 plain text files, each of which is 100MB in size. The InputFormat for your job is TextInputFormat. Determine how many Mappers will run?

- A. 64
- B. 100
- C. 200
- D. 640

Correct Answer: C

Each file would be split into two as the block size (64 MB) is less than the file size (100 MB), so 200 mappers would be running.

Note:

If you're not compressing the files then Hadoop will process your large files (say 10G), with a number of mappers related to the block size of the file.

Say your block size is 64M, then you will have ~160 mappers processing this 10G file ( $160 * 64 \approx 10G$ ). Depending on how CPU intensive your mapper logic is, this might be an acceptable block size, but if you find that your mappers are executing in sub-minute times, then you might want to increase the work done by each mapper (by increasing the block size to 128, 256, 512M - the actual size depends on how you intend to process the data).

Reference: <http://stackoverflow.com/questions/11014493/hadoop-mapreduce-appropriate-input-files-size> (first answer, second paragraph)

[CCD-410 PDF Dumps](#)

[CCD-410 Practice Test](#)

[CCD-410 Study Guide](#)