



# 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 wrote a map function that throws a runtime exception when it encounters a control character in input data. The input supplied to your mapper contains twelve such characters total, spread across five file splits. The first four file splits each have two control characters and the last split has four control characters.

Identify the number of failed task attempts you can expect when you run the job with `mapred.map.max.attempts` set to 4:

- A. You will have forty-eight failed task attempts
- B. You will have seventeen failed task attempts
- C. You will have five failed task attempts
- D. You will have twelve failed task attempts
- E. You will have twenty failed task attempts

Correct Answer: E

There will be four failed task attempts for each of the five file splits.

Note:

When the jobtracker is notified of a task attempt that has failed (by the tasktracker's heartbeat call), it will reschedule execution of the task. The jobtracker will try to avoid rescheduling the task on a tasktracker where it has previously failed. Furthermore, if a task fails four times (or more), it will not be retried further. This value is configurable: the maximum number of attempts to run a task is controlled by the `mapred.map.max.attempts` property for map tasks and `mapred.reduce.max.attempts` for reduce tasks. By default, if any task fails four times (or whatever the maximum number of attempts is configured to), the whole job fails.

---

### QUESTION 2

All keys used for intermediate output from mappers must:

- A. Implement a splittable compression algorithm.
- B. Be a subclass of `FileInputFormat`.
- C. Implement `WritableComparable`.
- D. Override `isSplittable`.
- E. Implement a comparator for speedy sorting.

Correct Answer: C

The MapReduce framework operates exclusively on pairs, that is, the framework views the input to the job as a set of



pairs and produces a set of pairs as the output of the job, conceivably of different types.

The key and value classes have to be serializable by the framework and hence need to implement the Writable interface. Additionally, the key classes have to implement the WritableComparable interface to facilitate sorting by the framework.

Reference: MapReduce Tutorial

---

### QUESTION 3

In a MapReduce job with 500 map tasks, how many map task attempts will there be?

- A. It depends on the number of reduces in the job.
- B. Between 500 and 1000.
- C. At most 500.
- D. At least 500.
- E. Exactly 500.

Correct Answer: D

Explanation: From Cloudera Training Course: Task attempt is a particular instance of an attempt to execute a task. There will be at least as many task attempts as there are tasks. If a task attempt fails, another will be started by the JobTracker. Speculative execution can also result in more task attempts than completed tasks.

---

### QUESTION 4

Which describes how a client reads a file from HDFS?

- A. The client queries the NameNode for the block location(s). The NameNode returns the block location(s) to the client. The client reads the data directory off the DataNode(s).
- B. The client queries all DataNodes in parallel. The DataNode that contains the requested data responds directly to the client. The client reads the data directly off the DataNode.
- C. The client contacts the NameNode for the block location(s). The NameNode then queries the DataNodes for block locations. The DataNodes respond to the NameNode, and the NameNode redirects the client to the DataNode that holds the requested data block(s). The client then reads the data directly off the DataNode.
- D. The client contacts the NameNode for the block location(s). The NameNode contacts the DataNode that holds the requested data block. Data is transferred from the DataNode to the NameNode, and then from the NameNode to the client.

Correct Answer: A

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, How the Client communicates with HDFS?

---



### QUESTION 5

Which project gives you a distributed, Scalable, data store that allows you random, realtime read/write access to hundreds of terabytes of data?

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

Correct Answer: A

Use Apache HBase when you need random, realtime read/write access to your Big Data. Note: This project's goal is the hosting of very large tables -- billions of rows X millions of columns -- atop clusters of commodity hardware. Apache HBase is an open-source, distributed, versioned, column-oriented store modeled after Google's Bigtable: A Distributed Storage System for Structured Data by Chang et al. Just as Bigtable leverages the distributed data storage provided by the Google File System, Apache HBase provides Bigtable-like capabilities on top of Hadoop and HDFS.

#### Features

Linear and modular scalability.

Strictly consistent reads and writes.

Automatic and configurable sharding of tables

Automatic failover support between RegionServers.

Convenient base classes for backing Hadoop MapReduce jobs with Apache HBase tables.

Easy to use Java API for client access.

Block cache and Bloom Filters for real-time queries.

Query predicate push down via server side Filters

Thrift gateway and a REST-ful Web service that supports XML, Protobuf, and binary data encoding options

Extensible jruby-based (JIRB) shell

Support for exporting metrics via the Hadoop metrics subsystem to files or Ganglia; or via JMX

Reference: <http://hbase.apache.org/> (when would I use HBase? First sentence)

---

### QUESTION 6



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 7

Table metadata in Hive is:

- A. Stored as metadata on the NameNode.
- B. Stored along with the data in HDFS.
- C. Stored in the Metastore.
- D. Stored in ZooKeeper.

Correct Answer: C

By default, hive use an embedded Derby database to store metadata information. The metastore is the "glue" between Hive and HDFS. It tells Hive where your data files live in HDFS, what type of data they contain, what tables they belong to, etc.

The Metastore is an application that runs on an RDBMS and uses an open source ORM layer called DataNucleus, to convert object representations into a relational schema and vice versa. They chose this approach as opposed to storing this information in hdfs as they need the Metastore to be very low latency. The DataNucleus layer allows them to plugin many different RDBMS technologies.

Note:

\*

By default, Hive stores metadata in an embedded Apache Derby database, and other client/server databases like MySQL can optionally be used.

\*

features of Hive include:

Metadata storage in an RDBMS, significantly reducing the time to perform semantic checks during query execution.



Reference: Store Hive Metadata into RDBMS

---

### QUESTION 8

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 9

You want to perform analysis on a large collection of images. You want to store this data in HDFS and process it with MapReduce but you also want to give your data analysts and data scientists the ability to process the data directly from HDFS with an interpreted high-level programming language like Python. Which format should you use to store this data in HDFS?

- A. SequenceFiles
- B. Avro
- C. JSON
- D. HTML
- E. XML
- F. CSV

Correct Answer: B

Reference: Hadoop binary files processing introduced by image duplicates finder

---

### QUESTION 10



A combiner reduces:

- A. The number of values across different keys in the iterator supplied to a single reduce method call.
- B. The amount of intermediate data that must be transferred between the mapper and reducer.
- C. The number of input files a mapper must process.
- D. The number of output files a reducer must produce.

Correct Answer: B

Combiners are used to increase the efficiency of a MapReduce program. They are used to aggregate intermediate map output locally on individual mapper outputs. Combiners can help you reduce the amount of data that needs to be transferred across to the reducers. You can use your reducer code as a combiner if the operation performed is commutative and associative. The execution of combiner is not guaranteed, Hadoop may or may not execute a combiner. Also, if required it may execute it more than 1 times. Therefore your MapReduce jobs should not depend on the combiners execution.

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

[CCD-410 VCE Dumps](#)

[CCD-410 Exam Questions](#)

[CCD-410 Braindumps](#)