



# 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**

Your client application submits a MapReduce job to your Hadoop cluster. Identify the Hadoop daemon on which the Hadoop framework will look for an available slot schedule a MapReduce operation.

- A. TaskTracker
- B. NameNode
- C. DataNode
- D. JobTracker
- E. Secondary NameNode

Correct Answer: D

JobTracker is the daemon service for submitting and tracking MapReduce jobs in Hadoop. There is only One Job Tracker process run on any hadoop cluster. Job Tracker runs on its own JVM process. In a typical production cluster its run on a separate machine. Each slave node is configured with job tracker node location. The JobTracker is single point of failure for the Hadoop MapReduce service. If it goes down, all running jobs are halted. JobTracker in Hadoop performs following actions(from Hadoop Wiki:)

Client applications submit jobs to the Job tracker.

The JobTracker talks to the NameNode to determine the location of the data

The JobTracker locates TaskTracker nodes with available slots at or near the data

The JobTracker submits the work to the chosen TaskTracker nodes.

The TaskTracker nodes are monitored. If they do not submit heartbeat signals often enough, they are deemed to have failed and the work is scheduled on a different TaskTracker.

A TaskTracker will notify the JobTracker when a task fails. The JobTracker decides what to do then: it may resubmit the job elsewhere, it may mark that specific record as something to avoid, and it may even blacklist the TaskTracker as unreliable.

When the work is completed, the JobTracker updates its status.

Client applications can poll the JobTracker for information.

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, What is a JobTracker in Hadoop? How many instances of JobTracker run on a Hadoop Cluster?

---

**QUESTION 2**

What is the disadvantage of using multiple reducers with the default HashPartitioner and distributing your workload across you cluster?

- A. You will not be able to compress the intermediate data.
- B. You will longer be able to take advantage of a Combiner.



- C. By using multiple reducers with the default HashPartitioner, output files may not be in globally sorted order.
- D. There are no concerns with this approach. It is always advisable to use multiple reduces.

Correct Answer: C

Multiple reducers and total ordering

If your sort job runs with multiple reducers (either because `mapreduce.job.reduces` in `mapred-site.xml` has been set to a number larger than 1, or because you've used the `-r` option to specify the number of reducers on the command-line), then by default Hadoop will use the HashPartitioner to distribute records across the reducers. Use of the HashPartitioner means that you can't concatenate your output files to create a single sorted output file. To do this you'll need total ordering,

Reference: [Sorting text files with MapReduce](#)

---

### QUESTION 3

In a MapReduce job, you want each of your input files processed by a single map task. How do you configure a MapReduce job so that a single map task processes each input file regardless of how many blocks the input file occupies?

- A. Increase the parameter that controls minimum split size in the job configuration.
- B. Write a custom MapRunner that iterates over all key-value pairs in the entire file.
- C. Set the number of mappers equal to the number of input files you want to process.
- D. Write a custom FileInputFormat and override the method `isSplittable` to always return false.

Correct Answer: D

FileInputFormat is the base class for all file-based InputFormats. This provides a generic implementation of `getSplits(JobContext)`. Subclasses of FileInputFormat can also override the `isSplittable(JobContext, Path)` method to ensure input-files are not split-up and are processed as a whole by Mappers.

Reference: [org.apache.hadoop.mapreduce.lib.input, Class FileInputFormat](#)

---

### QUESTION 4

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 5

Workflows expressed in Oozie can contain:

- A. Sequences of MapReduce and Pig. These sequences can be combined with other actions including forks, decision points, and path joins.
- B. Sequences of MapReduce job only; on Pig on Hive tasks or jobs. These MapReduce sequences can be combined with forks and path joins.
- C. Sequences of MapReduce and Pig jobs. These are limited to linear sequences of actions with exception handlers but no forks.
- D. Iterntive repetition of MapReduce jobs until a desired answer or state is reached.

Correct Answer: A

Oozie workflow is a collection of actions (i.e. Hadoop Map/Reduce jobs, Pig jobs) arranged in a control dependency DAG (Direct Acyclic Graph), specifying a sequence of actions execution. This graph is specified in hPDL (a XML Process

Definition Language).

hPDL is a fairly compact language, using a limited amount of flow control and action nodes. Control nodes define the flow of execution and include beginning and end of a workflow (start, end and fail nodes) and mechanisms to control the

workflow execution path ( decision, fork and join nodes).

Workflow definitions

Currently running workflow instances, including instance states and variables

Reference: Introduction to Oozie

Note: Oozie is a Java Web-Application that runs in a Java servlet-container - Tomcat and uses a database to store:

---

## QUESTION 6

You have user profile records in your OLPT database, that you want to join with web logs you have already ingested



into the Hadoop file system. How will you obtain these user records?

- A. HDFS command
- B. Pig LOAD command
- C. Sqoop import
- D. Hive LOAD DATA command
- E. Ingest with Flume agents
- F. Ingest with Hadoop Streaming

Correct Answer: C

Reference: Hadoop and Pig for Large-Scale Web Log Analysis

#### QUESTION 7

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.max.map.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 8



The Hadoop framework provides a mechanism for coping with machine issues such as faulty configuration or impending hardware failure. MapReduce detects that one or a number of machines are performing poorly and starts more copies of a map or reduce task. All the tasks run simultaneously and the task finish first are used. This is called:

- A. Combine
- B. IdentityMapper
- C. IdentityReducer
- D. Default Partitioner
- E. Speculative Execution

Correct Answer: E

Speculative execution: One problem with the Hadoop system is that by dividing the tasks across many nodes, it is possible for a few slow nodes to rate-limit the rest of the program. For example if one node has a slow disk controller, then it may be reading its input at only 10% the speed of all the other nodes. So when 99 map tasks are already complete, the system is still waiting for the final map task to check in, which takes much longer than all the other nodes.

By forcing tasks to run in isolation from one another, individual tasks do not know where their inputs come from. Tasks trust the Hadoop platform to just deliver the appropriate input. Therefore, the same input can be processed multiple times in parallel, to exploit differences in machine capabilities. As most of the tasks in a job are coming to a close, the Hadoop platform will schedule redundant copies of the remaining tasks across several nodes which do not have other work to perform. This process is known as speculative execution. When tasks complete, they announce this fact to the JobTracker. Whichever copy of a task finishes first becomes the definitive copy. If other copies were executing speculatively, Hadoop tells the TaskTrackers to abandon the tasks and discard their outputs. The Reducers then receive their inputs from whichever Mapper completed successfully, first.

Reference: Apache Hadoop, Module 4: MapReduce

Note:

\*

Hadoop uses "speculative execution." The same task may be started on multiple boxes. The first one to finish wins, and the other copies are killed. Failed tasks are tasks that error out.

\*

There are a few reasons Hadoop can kill tasks by his own decisions:

- a) Task does not report progress during timeout (default is 10 minutes)
- b) FairScheduler or CapacityScheduler needs the slot for some other pool (FairScheduler) or queue (CapacityScheduler).
- c) Speculative execution causes results of task not to be needed since it has completed on other place.

Reference: Difference failed tasks vs killed tasks

---

## QUESTION 9

Identify the MapReduce v2 (MRv2 / YARN) daemon responsible for launching application containers and monitoring



application resource usage?

- A. ResourceManager
- B. NodeManager
- C. ApplicationMaster
- D. ApplicationMasterService
- E. TaskTracker
- F. JobTracker

Correct Answer: B

Reference: Apache Hadoop YARN Concepts and Applications

---

#### QUESTION 10

How are keys and values presented and passed to the reducers during a standard sort and shuffle phase of MapReduce?

- A. Keys are presented to reducer in sorted order; values for a given key are not sorted.
- B. Keys are presented to reducer in sorted order; values for a given key are sorted in ascending order.
- C. Keys are presented to a reducer in random order; values for a given key are not sorted.
- D. Keys are presented to a reducer in random order; values for a given key are sorted in ascending order.

Correct Answer: A

Reducer has 3 primary phases:

1.

Shuffle

The Reducer copies the sorted output from each Mapper using HTTP across the network.

2.

Sort

The framework merge sorts Reducer inputs by keys (since different Mappers may have output the same key).

The shuffle and sort phases occur simultaneously i.e. while outputs are being fetched they are merged.

SecondarySort

To achieve a secondary sort on the values returned by the value iterator, the application should extend the key with the secondary key and define a grouping comparator. The keys will be sorted using the entire key, but will be grouped using



the grouping comparator to decide which keys and values are sent in the same call to reduce.

3.

Reduce

In this phase the reduce(Object, Iterable, Context) method is called for each in the sorted inputs.

The output of the reduce task is typically written to a RecordWriter via TaskInputOutputContext.write(Object, Object).

The output of the Reducer is not re-sorted.

Reference: org.apache.hadoop.mapreduce, Class

Reducer

---

### QUESTION 11

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 12

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?

[Latest CCD-410 Dumps](#)

[CCD-410 Study Guide](#)

[CCD-410 Exam Questions](#)



To Read the [Whole Q&As](#), please purchase the [Complete Version](#) from [Our website](#).

## Try our product !

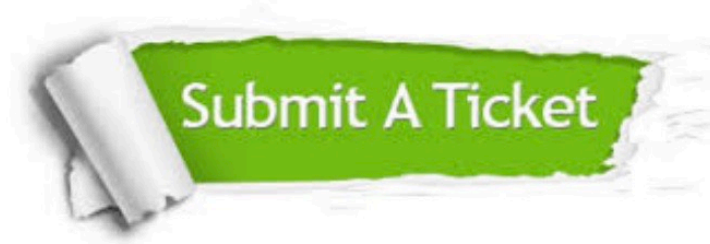
100% Guaranteed Success  
100% Money Back Guarantee  
365 Days Free Update  
Instant Download After Purchase  
24x7 Customer Support  
Average 99.9% Success Rate  
More than 800,000 Satisfied Customers Worldwide  
Multi-Platform capabilities - [Windows](#), [Mac](#), [Android](#), [iPhone](#), [iPod](#), [iPad](#), [Kindle](#)

We provide exam PDF and VCE of Cisco, Microsoft, IBM, CompTIA, Oracle and other IT Certifications. You can view Vendor list of All Certification Exams offered:

<https://www.pass4itsure.com/allproducts>

## Need Help

Please provide as much detail as possible so we can best assist you.  
To update a previously submitted ticket:



 <p><b>One Year Free Update</b> Free update is available within One Year after your purchase. After One Year, you will get 50% discounts for updating. And we are proud to boast a 24/7 efficient Customer Support system via Email.</p>	 <p><b>Money Back Guarantee</b> To ensure that you are spending on quality products, we provide 100% money back guarantee for 30 days from the date of purchase.</p>	 <p><b>Security &amp; Privacy</b> We respect customer privacy. We use McAfee's security service to provide you with utmost security for your personal information &amp; peace of mind.</p>
---	---	--

Any charges made through this site will appear as Global Simulators Limited.  
All trademarks are the property of their respective owners.  
Copyright © pass4itsure, All Rights Reserved.