



CCD-410^{Q&As}

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.geekcert.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 write MapReduce job to process 100 files in HDFS. Your MapReduce algorithm uses TextInputFormat: the mapper applies a regular expression over input values and emits key- values pairs with the key consisting of the matching text, and the value containing the filename and byte offset. Determine the difference between setting the number of reduces to one and settings the number of reducers to zero.

- A. There is no difference in output between the two settings.
- B. With zero reducers, no reducer runs and the job throws an exception. With one reducer, instances of matching patterns are stored in a single file on HDFS.
- C. With zero reducers, all instances of matching patterns are gathered together in one file on HDFS. With one reducer, instances of matching patterns are stored in multiple files on HDFS.
- D. With zero reducers, instances of matching patterns are stored in multiple files on HDFS. With one reducer, all instances of matching patterns are gathered together in one file on HDFS.

Correct Answer: D

*

It is legal to set the number of reduce-tasks to zero if no reduction is desired.

In this case the outputs of the map-tasks go directly to the FileSystem, into the output path set by `setOutputPath(Path)`. The framework does not sort the map-outputs before writing them out to the FileSystem.

*

Often, you may want to process input data using a map function only. To do this, simply set `mapreduce.job.reduces` to zero. The MapReduce framework will not create any reducer tasks. Rather, the outputs of the mapper tasks will be the final output of the job.

Note:

Reduce

In this phase the `reduce(WritableComparable, Iterator, OutputCollector, Reporter)` method is called for each pair in the grouped inputs.

The output of the reduce task is typically written to the FileSystem via `OutputCollector.collect (WritableComparable, Writable)`.

Applications can use the Reporter to report progress, set application-level status messages and update Counters, or just indicate that they are alive.

The output of the Reducer is not sorted.

QUESTION 2

On a cluster running MapReduce v1 (MRv1), a TaskTracker heartbeats into the JobTracker on your cluster, and alerts the JobTracker it has an open map task slot. What determines how the JobTracker assigns each map task to a TaskTracker?



- A. The amount of RAM installed on the TaskTracker node.
- B. The amount of free disk space on the TaskTracker node.
- C. The number and speed of CPU cores on the TaskTracker node.
- D. The average system load on the TaskTracker node over the past fifteen (15) minutes.
- E. The location of the InputSplit to be processed in relation to the location of the node.

Correct Answer: E

The TaskTrackers send out heartbeat messages to the JobTracker, usually every few minutes, to reassure the JobTracker that it is still alive. These message also inform the JobTracker of the number of available slots, so the JobTracker can stay up to date with where in the cluster work can be delegated. When the JobTracker tries to find somewhere to schedule a task within the MapReduce operations, it first looks for an empty slot on the same server that hosts the DataNode containing the data, and if not, it looks for an empty slot on a machine in the same rack.

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, How JobTracker schedules a task?

QUESTION 3

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 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

You are developing a combiner that takes as input Text keys, IntWritable values, and emits Text keys, IntWritable values. Which interface should your class implement?

- A. Combiner
- B. Mapper
- C. Reducer
- D. Reducer
- E. Combiner

Correct Answer: D



VCE & PDF

GeekCert.com

<https://www.geekcert.com/ccd-410.html>

2024 Latest geekcert CCD-410 PDF and VCE dumps Download

[Latest CCD-410 Dumps](#)

[CCD-410 PDF Dumps](#)

[CCD-410 Exam Questions](#)