

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.leads4pass.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 have written a Mapper which invokes the following five calls to the `OutputCollector.collect` method:

```
output.collect (new Text ("Apple"), new Text ("Red") ) ;
```

```
output.collect (new Text ("Banana"), new Text ("Yellow") ) ; output.collect (new Text ("Apple"), new Text ("Yellow") ) ; output.collect (new Text ("Cherry"), new Text ("Red") ) ;
```

```
output.collect (new Text ("Apple"), new Text ("Green") ) ;
```

How many times will the Reducer's `reduce` method be invoked?

- A. 6
- B. 3
- C. 1
- D. 0
- E. 5

Correct Answer: B

`reduce()` gets called once for each [key, (list of values)] pair. To explain, let's say you called:

```
out.collect(new Text("Car"),new Text("Subaru");
```

```
out.collect(new Text("Car"),new Text("Honda");
```

```
out.collect(new Text("Car"),new Text("Ford");
```

```
out.collect(new Text("Truck"),new Text("Dodge");
```

```
out.collect(new Text("Truck"),new Text("Chevy");
```

Then `reduce()` would be called twice with the pairs

```
reduce(Car, )
```

```
reduce(Truck, )
```

Reference: Mapper `output.collect()`?

QUESTION 2

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 3

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. Iterative 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 4

Determine which best describes when the reduce method is first called in a MapReduce job?

- A. Reducers start copying intermediate key-value pairs from each Mapper as soon as it has completed. The programmer can configure in the job what percentage of the intermediate data should arrive before the reduce method begins.
- B. Reducers start copying intermediate key-value pairs from each Mapper as soon as it has completed. The reduce method is called only after all intermediate data has been copied and sorted.
- C. Reduce methods and map methods all start at the beginning of a job, in order to provide optimal performance for map-only or reduce-only jobs.
- D. Reducers start copying intermediate key-value pairs from each Mapper as soon as it has completed. The reduce method is called as soon as the intermediate key-value pairs start to arrive.

Correct Answer: B

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers , When is the reducers are started in a MapReduce job?

QUESTION 5

You need to perform statistical analysis in your MapReduce job and would like to call methods in the Apache Commons Math library, which is distributed as a 1.3 megabyte Java archive (JAR) file. Which is the best way to make this library available to your MapReducer job at runtime?

- A. Have your system administrator copy the JAR to all nodes in the cluster and set its location in the HADOOP_CLASSPATH environment variable before you submit your job.
- B. Have your system administrator place the JAR file on a Web server accessible to all cluster nodes and then set the HTTP_JAR_URL environment variable to its location.
- C. When submitting the job on the command line, specify the libjars option followed by the JAR file path.
- D. Package your code and the Apache Commons Math library into a zip file named JobJar.zip

Correct Answer: C

The usage of the jar command is like this,

Usage: `hadoop jar [mainClass] args...`

If you want the commons-math3.jar to be available for all the tasks you can do any one of these

1.

Copy the jar file in `$HADOOP_HOME/lib` dir or

2.

Use the generic option `-libjars`.

[CCD-410 PDF Dumps](#)

[CCD-410 Exam Questions](#)

[CCD-410 Braindumps](#)