

100% Money Back
Guarantee

Vendor: Cloudera

Exam Code: CCD-470

Exam Name: Cloudera Certified Developer for Apache Hadoop CDH4 Upgrade Exam (CCDH)

Version: Demo

QUESTION 1

When is the earliest point at which the reduce method of a given Reducer can be called?

- A. As soon as at least one mapper has finished processing its input split.
- B. As soon as a mapper has emitted at least one record.
- C. Not until all mappers have finished processing all records.
- D. It depends on the InputFormat used for the job.

Correct Answer: C

QUESTION 2

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: C

QUESTION 3

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 <Text, IntWritable, Text, IntWritable>
- B. Mapper <Text, IntWritable, Text, IntWritable>
- C. Reducer <Text, Text, IntWritable, IntWritable>
- D. Reducer <Text, IntWritable, Text, IntWritable>
- E. Combiner <Text, Text, IntWritable, IntWritable>

Correct Answer: D

QUESTION 4

Identify the utility that allows you to create and run MapReduce jobs with any executable or script as the mapper and/or the reducer?

- A. Oozie
- B. Sqoop
- C. Flume
- D. Hadoop Streaming
- E. mapred

Correct Answer: D

QUESTION 5

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

QUESTION 6

Assuming default settings, which best describes the order of data provided to a reducer's reduce method:

- A. The keys given to a reducer aren't in a predictable order, but the values associated with those keys always are.
- B. Both the keys and values passed to a reducer always appear in sorted order.
- C. Neither keys nor values are in any predictable order.
- D. The keys given to a reducer are in sorted order but the values associated with each key are in no predictable order

Correct Answer: D

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

QUESTION 8

You want to populate an associative array in order to perform a map-side join. You've decided to put this information in a text file, place that file into the Distributed Cache and read it in your Mapper before any records are processed.

Identify which method in the Mapper you should use to implement code for reading the file and populating the associative array?

- A. combine
- B. map
- C. init
- D. configure

Correct Answer: D

QUESTION 9

You've written a MapReduce job that will process 500 million input records and generated 500 million key-value pairs. The data is not uniformly distributed. Your MapReduce job will create a significant amount of intermediate data that it needs to transfer between mappers and reducers which is a potential bottleneck. A custom implementation of which interface is most likely to reduce the amount of intermediate data transferred across the network?

- A. Partitioner
- B. OutputFormat

- C. WritableComparable
- D. Writable
- E. InputFormat
- F. Combiner

Correct Answer: F

QUESTION 10

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

QUESTION 11

You have just executed a MapReduce job. Where is intermediate data written to after being emitted from the Mapper's map method?

- A. Intermediate data is streamed across the network from Mapper to the Reducer and is never written to disk.
- B. Into in-memory buffers on the TaskTracker node running the Mapper that spill over and are written into HDFS.
- C. Into in-memory buffers that spill over to the local file system of the TaskTracker node running the Mapper.
- D. Into in-memory buffers that spill over to the local file system (outside HDFS) of the TaskTracker node running the Reducer
- E. Into in-memory buffers on the TaskTracker node running the Reducer that spill over and are written into HDFS.

Correct Answer: D

QUESTION 12

You want to understand more about how users browse your public website, such as which pages they visit prior to placing an order. You have a farm of 200 web servers hosting your website. How will you gather this data for your analysis?

- A. Ingest the server web logs into HDFS using Flume.
- B. Write a MapReduce job, with the web servers for mappers, and the Hadoop cluster nodes for reducers.
- C. Import all users clicks from your OLTP databases into Hadoop, using Sqoop.
- D. Channel these clickstreams into Hadoop using Hadoop Streaming.
- E. Sample the weblogs from the web servers, copying them into Hadoop using curl.

Correct Answer: B

QUESTION 13

MapReduce v2 (MRv2/YARN) is designed to address which two issues?

- A. Single point of failure in the NameNode.
- B. Resource pressure on the JobTracker.
- C. HDFS latency.
- D. Ability to run frameworks other than MapReduce, such as MPI.
- E. Reduce complexity of the MapReduce APIs.

F. Standardize on a single MapReduce API.

Correct Answer: BD

QUESTION 14

You need to run the same job many times with minor variations. Rather than hardcoding all job configuration options in your driver code, you've decided to have your Driver subclass `org.apache.hadoop.conf.Configured` and implement the `org.apache.hadoop.util.Tool` interface. Identify which invocation correctly passes `mapred.job.name` with a value of `Example` to Hadoop?

- A. `hadoop mapred.job.name=Example MyDriver input output`
- B. `hadoop MyDriver mapred.job.name=Example input output`
- C. `hadoop MyDriver mapred.job.name=Example input output`
- D. `hadoop setproperty mapred.job.name=Example MyDriver input output`
- E. `hadoop setproperty (mapred.job.name=Example) MyDriver input output`

Correct Answer: C

QUESTION 15

You are developing a MapReduce job for sales reporting. The mapper will process input keys representing the year (`IntWritable`) and input values representing product identifiers (`Text`). Identify what determines the data types used by the Mapper for a given job.

- A. The key and value types specified in the `JobConf.setMapInputKeyClass` and `JobConf.setMapInputValuesClass` methods
- B. The data types specified in `HADOOP_MAP_DATATYPES` environment variable
- C. The `mapper-specification.xml` file submitted with the job determine the mapper's input key and value types.
- D. The `InputFormat` used by the job determines the mapper's input key and value types.

Correct Answer: D

QUESTION 16

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: C

QUESTION 17

Which best describes how `TextInputFormat` processes input files and line breaks?

- A. Input file splits may cross line breaks. A line that crosses file splits is read by the `RecordReader` of the split that contains the beginning of the broken line.
- B. Input file splits may cross line breaks. A line that crosses file splits is read by the `RecordReaders` of both splits containing the broken line.
- C. The input file is split exactly at the line breaks, so each `RecordReader` will read a series of complete lines.
- D. Input file splits may cross line breaks. A line that crosses file splits is ignored.
- E. Input file splits may cross line breaks. A line that crosses file splits is read by the `RecordReader` of the split that contains the end of the broken line.

Correct Answer: E

QUESTION 18

For each input key-value pair, mappers can emit:

- A. As many intermediate key-value pairs as designed. There are no restrictions on the types of those key-value pairs (i.e., they can be heterogeneous).
- B. As many intermediate key-value pairs as designed, but they cannot be of the same type as the input key-value pair.
- C. One intermediate key-value pair, of a different type.
- D. One intermediate key-value pair, but of the same type.
- E. As many intermediate key-value pairs as designed, as long as all the keys have the same types and all the values have the same type.

Correct Answer: E

QUESTION 19

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: A

QUESTION 20

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: B

QUESTION 21

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

- A. You will not be able to compress the intermediate data.
- B. You will no 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 reducers.

Correct Answer: C

QUESTION 22

Given a directory of files with the following structure: line number, tab character, string:
Example:

- A.
- B.
- C.
- D.

Correct Answer:

QUESTION 23

abialkjfjkaoasdfjksdlkjhqwerioj

- A.
- B.
- C.
- D.

Correct Answer:

QUESTION 24

kadfjhuwqounahagtnbvaswslmnbfgy

- A.
- B.
- C.
- D.

Correct Answer:

QUESTION 25

kjfteiomndscxeqalkzhtopedkfsikj

You want to send each line as one record to your Mapper. Which InputFormat should you use to complete the line: `conf.setInputFormat (____.class) ; ?`

- A. SequenceFileAsTextInputFormat
- B. SequenceFileInputFormat
- C. KeyValueFileInputFormat
- D. BDBInputFormat

Correct Answer: B

QUESTION 26

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

QUESTION 27

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

QUESTION 28

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

QUESTION 29

What data does a Reducer reduce method process?

- A. All the data in a single input file.
- B. All data produced by a single mapper.
- C. All data for a given key, regardless of which mapper(s) produced it.
- D. All data for a given value, regardless of which mapper(s) produced it.

Correct Answer: C

QUESTION 30

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

QUESTION 31

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

QUESTION 32

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 the same type.

Correct Answer: D

QUESTION 33

A client application creates an HDFS file named foo.txt with a replication factor of 3. Identify which best describes the file access rules in HDFS if the file has a single block that is stored on data nodes A, B and C?

- A. The file will be marked as corrupted if data node B fails during the creation of the file.
- B. Each data node locks the local file to prohibit concurrent readers and writers of the file.
- C. Each data node stores a copy of the file in the local file system with the same name as the HDFS file.
- D. The file can be accessed if at least one of the data nodes storing the file is available.

Correct Answer: D

QUESTION 34

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

QUESTION 35

Which process describes the lifecycle of a Mapper?

- A. The JobTracker calls the TaskTracker's configure () method, then its map () method and finally its close () method.
- B. The TaskTracker spawns a new Mapper to process all records in a single input split.
- C. The TaskTracker spawns a new Mapper to process each key-value pair.
- D. The JobTracker spawns a new Mapper to process all records in a single file.

Correct Answer: C

QUESTION 36

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: D

QUESTION 37

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

QUESTION 38

To process input key-value pairs, your mapper needs to load a 512 MB data file in memory. What is the best way to accomplish this?

- A. Serialize the data file, insert in it the JobConf object, and read the data into memory in the configure method of the mapper.
- B. Place the data file in the DistributedCache and read the data into memory in the map method of the mapper.
- C. Place the data file in the DataCache and read the data into memory in the configure method of the mapper.
- D. Place the data file in the DistributedCache and read the data into memory in the configure method of the mapper.

Correct Answer: B

QUESTION 39

In a MapReduce job, the reducer receives all values associated with same key. Which statement best describes the ordering of these values?

- A. The values are in sorted order.
- B. The values are arbitrarily ordered, and the ordering may vary from run to run of the same MapReduce job.
- C. The values are arbitrary ordered, but multiple runs of the same MapReduce job will always have the same ordering.
- D. Since the values come from mapper outputs, the reducers will receive contiguous sections of sorted values.

Correct Answer: B

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

Trying 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 **69,000** Satisfied Customers Worldwide
- ★ Multi-Platform capabilities - **Windows, Mac, Android, iPhone, iPod, iPad, Kindle**

Need Help

Please provide as much detail as possible so we can best assist you.

To update a previously submitted ticket:



 One Year Free Update <p>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>	 Money Back Guarantee <p>To ensure that you are spending on quality products, we provide 100% money back guarantee for 30 days from the date of purchase.</p>	 Security & Privacy <p>We respect customer privacy. We use McAfee's security service to provide you with utmost security for your personal information & peace of mind.</p>
---	---	--

[Guarantee & Policy](#) | [Privacy & Policy](#) | [Terms & Conditions](#)

Any charges made through this site will appear as Global Simulators Limited.

All trademarks are the property of their respective owners.

Copyright © 2004-2015, All Rights Reserved.