



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

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: D

QUESTION 2

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

The input types fed to the mapper are controlled by the InputFormat used. The default input format, "TextInputFormat," will load data in as (LongWritable, Text) pairs. The long value is the byte offset of the line in the file. The Text object holds the string contents of the line of the file.

Note: The data types emitted by the reducer are identified by setOutputKeyClass() and setOutputValueClass(). The data types emitted by the reducer are identified by setOutputKeyClass() and setOutputValueClass().

By default, it is assumed that these are the output types of the mapper as well. If this is not the case, the methods setMapOutputKeyClass() and setMapOutputValueClass() methods of the JobConf class will override these.



Reference: Yahoo! Hadoop Tutorial, THE DRIVER METHOD

QUESTION 3

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 4

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 5

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?



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 VCE Dumps](#)

[CCD-410 Braindumps](#)