

TestHorse

Certified IT practice exam authority

Accurate study guides, High passing rate!
Testhorse provides update free of charge in one year!



Exam : **CCD-333**

Title : Cloudera Certified
Developer for Apache
Hadoop

Version : Demo

1.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.

Answer: D

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

Example:

```
abialkijfkaoasdfjksdlkjhqwerioj
```

```
kadf jhuwqounahagtnbvaswslmnbfgy
```

```
kjfteiomndscxeqalkzhtopedkfskj
```

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

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

Answer: C

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.

Answer: B

4.Which of the following best describes the workings of TextInputFormat?

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

Answer: D

5.Which of the following statements most accurately describes the relationship between MapReduce and

Pig?

- A. Pig provides additional capabilities that allow certain types of data manipulation not possible with MapReduce.
- B. Pig provides no additional capabilities to MapReduce. Pig programs are executed as MapReduce jobs via the Pig interpreter.
- C. Pig programs rely on MapReduce but are extensible, allowing developers to do special-purpose processing not provided by MapReduce.
- D. Pig provides the additional capability of allowing you to control the flow of multiple MapReduce jobs.

Answer: D

6.You need to import a portion of a relational database every day as files to HDFS, and generate Java classes to Interact with your imported data. Which of the following tools should you use to accomplish this?

- A. Pig
- B. Hue
- C. Hive
- D. Flume
- E. Sqoop
- F. Oozie
- G. fuse-dfs

Answer: C,E

7.You have an employee who is a Data Analyst and is very comfortable with SQL. He would like to run ad-hoc analysis on data in your HDFS duster. Which of the following is a data warehousing software built on top of Apache Hadoop that defines a simple SQL-like query language well-suited for this kind of user?

- A. Pig
- B. Hue
- C. Hive
- D. Sqoop
- E. Oozie
- F. Flume
- G. Hadoop Streaming

Answer: C

8.Workflows expressed in Oozie can contain:

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

Answer: D

9.You need a distributed, scalable, data Store that allows you random, realtime read/write access to hundreds of terabytes of data. Which of the following would you use?

- A. Hue
- B. Pig
- C. Hive
- D. Oozie
- E. HBase
- F. Flume
- G. Sqoop

Answer: E

10.Which of the following utilities 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

Answer: D