Vendor: Hortonworks

Exam Code:HDPCD

Exam Name: Hortonworks Data Platform Certified Developer

Version: Demo

QUESTION 1

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: A

Reference: How Map and Reduce operations are actually carried out

---

QUESTION 2

In Hadoop 2.2, which TWO of the following processes work together to provide automatic failover of the NameNode? Choose 2 answers

A. ZKFailoverController

B. ZooKeeper

C. QuorumManager

D. JournalNode

Correct Answer: AD

QUESTION 3

What are the TWO main components of the YARN ResourceManager process? Choose 2 answers

A. Job Tracker

B. Task Tracker

C. Scheduler

D. Applications Manager

Correct Answer: CD

QUESTION 4

MapReduce v2 (MRv2/YARN) splits which major functions of the JobTracker into separate daemons?

Select two.

A. Heath states checks (heartbeats)

B. Resource management

C. Job scheduling/monitoring

D. Job coordination between the ResourceManager and NodeManager

E. Launching tasks

F. Managing file system metadata

G. MapReduce metric reporting

H. Managing tasks

Correct Answer: BC

Explanation: The fundamental idea of MRv2 is to split up the two major functionalities of the JobTracker, resource management and job scheduling/monitoring, into separate daemons. The idea is to have a global ResourceManager (RM) and per-application ApplicationMaster (AM). An application is either a single job in the classical sense of Map- Reduce jobs or a DAG of jobs.

Note:

The central goal of YARN is to clearly separate two things that are unfortunately smushed together in current Hadoop, specifically in (mainly) JobTracker:

/ Monitoring the status of the cluster with respect to which nodes have which resources available. Under YARN, this will be global.

/ Managing the parallelization execution of any specific job. Under YARN, this will be done separately for each job.

Reference: Apache Hadoop YARN ?Conceptsand; Applications

QUESTION 5

In a MapReduce job with 500 map tasks, how many map task attempts will there be?

A. It depends on the number of reduces in the job.

B. Between 500 and 1000.

C. At most 500.

D. At least 500.

E. Exactly 500.

Correct Answer: D

From Cloudera Training Course: Task attempt is a particular instance of an attempt to execute a task ?There will be at least as many task attempts as there are tasks ?If a task attempt fails, another will be started by the JobTracker ?Speculative execution can also result in more task attempts than completed tasks

---

QUESTION 6

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: C

Reference: Hadoop Map-Reduce Tutorial; Yahoo! Hadoop Tutorial, Module 4: MapReduce

---

QUESTION 7

What does Pig provide to the overall Hadoop solution?

A. Legacy language Integration with MapReduce framework

B. Simple scripting language for writing MapReduce programs

C. Database table and storage management services

D. C++ interface to MapReduce and data warehouse infrastructure

Correct Answer: B

QUESTION 8

You want to perform analysis on a large collection of images. You want to store this data in HDFS and process it with MapReduce but you also want to give your data analysts and data scientists the ability to process the data directly from HDFS with an interpreted high- level programming language like Python. Which format should you use to store this data

in HDFS?

A. SequenceFiles

B. Avro

C. JSON

D. HTML

E. XML

F. CSV

Correct Answer: B

Reference: Hadoop binary files processing introduced by image duplicates finder

---

QUESTION 9

A NameNode in Hadoop 2.2 manages _____.

A. Two namespaces: an active namespace and a backup namespace

B. A single namespace

C. An arbitrary number of namespaces

D. No namespaces

Correct Answer: B

QUESTION 10

You have written a Mapper which invokes the following five calls to the OutputColletor.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

Explanation: 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 11

Review the following andapos;dataandapos; file and Pig code.

```
M,38,95111
M,62,95102

A = LOAD &apos;data&apos; USING PigStorage(&apos;,&apos;)
AS (gender:chararray, age:int, zip:chararray);
D = GROUP A BY gender;
DUMP D;
```

Which one of the following statements is true?

A. The Output Of the DUMP D command IS (M,{(M,62.95102),(M,38,95111)})

B. The output of the dump d command is (M, {(38,95in),(62,95i02)})

C. The code executes successfully but there is not output because the D relation is empty

D. The code does not execute successfully because D is not a valid relation

Correct Answer: A

QUESTION 12

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: B

Explanation: Only one key value pair will be passed from the two (the, 1) key value pairs.