Hortonworks

Exam Hortonworks-Certified-Apache-Hadoop-2.0-Developer

Hadoop 2.0 Certification exam for Pig and Hive Developer

Version: 7.0

[Total Questions: 108]

Question No : 1

What does the following command do?

register '/piggyban):/pig-files.jar';

A. Invokes the user-defined functions contained in the jar file
B. Assigns a name to a user-defined function or streaming command
C. Transforms Pig user-defined functions into a format that Hive can accept
D. Specifies the location of the JAR file containing the user-defined functions

Answer: D

Question No : 2

Consider the following two relations, A and B.

```
A = LOAD 'data1' AS (a1:int,a2:chararray);
DUMP A;
(1,apple)
(3,orange)
(4,peach)
(2,cherry)

B = LOAD 'data2' AS (b1:chararray,b2:int);
DUMP B;
(Jim,2)
(Brian,4)
(Kim,0)
(Terry,3)
(Chris,2)
```

A Pig JOIN statement that combined relations A by its first field and B by its second field would produce what output?

A. 2 Jim Chris 2
3 Terry 3
4 Brian 4
B. 2 cherry
2 cherry
3 orange

4 peach
C. 2 cherry Jim, Chris
3 orange Terry
4 peach Brian
D. 2 cherry Jim 2
2 cherry Chris 2
3 orange Terry 3
4 peach Brian 4

Answer: D

Question No : 3

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

Answer: C
Reference: Hadoop and Pig for Large-Scale Web Log Analysis

Question No : 4

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

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, <Subaru, Honda, Ford>)
reduce(Truck, <Dodge, Chevy>)

Reference: Mapper output.collect()?

Question No : 5

Given the following Pig commands:

```
logevents = LOAD 'input/my.log';
severe = FILTER logevents BY ($1 =='severe' AND $2 >= 500);
grouped = GROUP severe BY $2;
DUMP grouped;
```

Which one of the following statements is true?

A. The $1 variable represents the first column of data in 'my.log'
B. The $1 variable represents the second column of data in 'my.log'
C. The severe relation is not valid
D. The grouped relation is not valid

Answer: B

Question No : 6

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.

Answer: C

Explanation: Reducing lets you aggregate values together. A reducer function receives an iterator of input values from an input list. It then combines these values together, returning a single output value.

All values with the same key are presented to a single reduce task.

Reference: Yahoo! Hadoop Tutorial, Module 4: MapReduce

Question No : 7

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 isSplitable.
E. Implement a comparator for speedy sorting.

Answer: C

Explanation: The MapReduce framework operates exclusively on <key, value> pairs, that

5

is, the framework views the input to the job as a set of <key, value> pairs and produces a set of <key, value> pairs as the output of the job, conceivably of different types.

The key and value classes have to be serializable by the framework and hence need to implement the Writable interface. Additionally, the key classes have to implement the WritableComparable interface to facilitate sorting by the framework.

Reference: MapReduce Tutorial

Question No : 8

Which Hadoop component is responsible for managing the distributed file system metadata?

A. NameNode
B. Metanode
C. DataNode
D. NameSpaceManager

Answer: A

Question No : 9

You need to move a file titled "weblogs" into HDFS. When you try to copy the file, you can't. You know you have ample space on your DataNodes. Which action should you take to relieve this situation and store more files in HDFS?

A. Increase the block size on all current files in HDFS.
B. Increase the block size on your remaining files.
C. Decrease the block size on your remaining files.
D. Increase the amount of memory for the NameNode.
E. Increase the number of disks (or size) for the NameNode.
F. Decrease the block size on all current files in HDFS.

Answer: C

Question No : 10

In the reducer, the MapReduce API provides you with an iterator over Writable values. What does calling the next () method return?

A. It returns a reference to a different Writable object time.
B. It returns a reference to a Writable object from an object pool.
C. It returns a reference to the same Writable object each time, but populated with different data.
D. It returns a reference to a Writable object. The API leaves unspecified whether this is a reused object or a new object.
E. It returns a reference to the same Writable object if the next value is the same as the previous value, or a new Writable object otherwise.

Answer: C

Explanation: Calling Iterator.next() will always return the SAME EXACT instance of IntWritable, with the contents of that instance replaced with the next value.

Reference: manupulating iterator in mapreduce

Question No : 11

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

Answer: B,C

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 – Concepts & Applications

Question No : 12

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.

Answer: E

Explanation: Mapper maps input key/value pairs to a set of intermediate key/value pairs.

Maps are the individual tasks that transform input records into intermediate records. The transformed intermediate records do not need to be of the same type as the input records. A given input pair may map to zero or many output pairs.

Reference: Hadoop Map-Reduce Tutorial

Question No : 13

Which one of the following statements describes the relationship between the ResourceManager and the ApplicationMaster?

A. The ApplicationMaster requests resources from the ResourceManager
B. The ApplicationMaster starts a single instance of the ResourceManager
C. The ResourceManager monitors and restarts any failed Containers of the ApplicationMaster
D. The ApplicationMaster starts an instance of the ResourceManager within each Container

Answer: A

Question No : 14

Given the following Hive commands:

```
CREATE TABLE mytable (name chararray, age int) ROW FORMAT
DELIMITED FIELDS TERMINATED BY ',' STORED AS
TEXTFILE;
LOAD DATA INPATH '/home/user/mydata.txt' INTO TABLE mytable;
```

Which one of the following statements Is true?

A. The file mydata.txt is copied to a subfolder of /apps/hive/warehouse
B. The file mydata.txt is moved to a subfolder of /apps/hive/warehouse
C. The file mydata.txt is copied into Hive's underlying relational database 0.
D. The file mydata.txt does not move from Its current location in HDFS

Answer: A

Question No : 15

Which YARN component is responsible for monitoring the success or failure of a Container?

A. ResourceManager
B. ApplicationMaster
C. NodeManager
D. JobTracker

Answer: A

Question No : 16

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.

Answer: A
Explanation: You can use your reducer code as a combiner if the operation performed is commutative and associative.

Reference: 24 Interview Questions & Answers for Hadoop MapReduce developers, What are combiners? When should I use a combiner in my MapReduce Job?

Question No : 17

Review the following data and Pig code: