Leads4Pass

HADOOP-PR000007$^{Q\&As}$

Hortonworks Certified Apache Hadoop 2.0 Developer (Pig and Hive Developer)

Pass Hortonworks HADOOP-PR000007 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.leads4pass.com/hadoop-pr000007.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Hortonworks Official Exam Center

⚙ **Instant Download** After Purchase

⚙ **100% Money Back** Guarantee

⚙ **365 Days** Free Update

⚙ **800,000+** Satisfied Customers

Leads4Pass

QUESTION 1

Which one of the following files is required in every Oozie Workflow application?

A. job.properties

B. Config-default.xml

C. Workflow.xml

D. Oozie.xml

Correct Answer: C

QUESTION 2

You want to Ingest log files Into HDFS, which tool would you use?

A. HCatalog

B. Flume

C. Sqoop

D. Ambari

Correct Answer: B

QUESTION 3

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

Correct Answer: D

QUESTION 4

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.

Correct Answer: C

QUESTION 5

You need to run the same job many times with minor variations. Rather than hardcoding all job configuration options in your drive code, you\\'ve decided to have your Driver subclass org.apache.hadoop.conf.Configured and implement the org.apache.hadoop.util.Tool interface.

Indentify 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 MyDrive ? 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

Explanation: Configure the property using the -D key=value notation:

-D mapred.job.name=\\'My Job\\'

---

https://www.leads4pass.com/hadoop-pr000007.html
2024 Latest leads4pass HADOOP-PR000007 PDF and VCE dumps Download

You can list a whole bunch of options by calling the streaming jar with just the -info argument

Reference: Python hadoop streaming : Setting a job name

QUESTION 6

Which two of the following are true about this trivial Pig program\\' (choose Two)

```
$ pig
grunt>  ABC  =  LOAD 'myfile';
grunt>  DUMP ABC;
```

A. The contents of myfile appear on stdout

B. Pig assumes the contents of myfile are comma delimited

C. ABC has a schema associated with it

D. myfile is read from the user\\'s home directory in HDFS

Correct Answer: AD

QUESTION 7

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 8

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 same type.

Correct Answer: D

Explanation: SequenceFile is a flat file consisting of binary key/value pairs.

There are 3 different SequenceFile formats:

Uncompressed key/value records.

Record compressed key/value records - only \'values\' are compressed here. Block compressed key/value

records - both keys and values are collected in \'blocks\' separately and compressed. The size of the \'block\'

is configurable.

Reference: http://wiki.apache.org/hadoop/SequenceFile

QUESTION 9

Analyze each scenario below and indentify which best describes the behavior of the default partitioner?

A. The default partitioner assigns key-values pairs to reduces based on an internal random number generator.

B. The default partitioner implements a round-robin strategy, shuffling the key-value pairs to each reducer in turn. This ensures an event partition of the key space.

C. The default partitioner computes the hash of the key. Hash values between specific ranges are associated with different buckets, and each bucket is assigned to a specific reducer.

D. The default partitioner computes the hash of the key and divides that valule modulo the number of reducers. The result determines the reducer assigned to process the key-value pair.

E. The default partitioner computes the hash of the value and takes the mod of that value with the number of reducers. The result determines the reducer assigned to process the key-value pair.

Correct Answer: D

Explanation: The default partitioner computes a hash value for the key and assigns the partition based on this result.

The default Partitioner implementation is called HashPartitioner. It uses the hashCode() method of the key objects modulo the number of partitions total to determine which partition to send a given (key, value) pair to.

In Hadoop, the default partitioner is HashPartitioner, which hashes a record\\'s key to determine which partition (and thus which reducer) the record belongs in.The number of partition is then equal to the number of reduce tasks for the job.

Reference: Getting Started With (Customized) Partitioning

QUESTION 10

Given the following Pig command:

logevents = LOAD andapos;input/my.logandapos; AS (date:chararray, levehstring, code:int, message:string);

Which one of the following statements is true?

A. The logevents relation represents the data from the my.log file, using a comma as the parsing delimiter

B. The logevents relation represents the data from the my.log file, using a tab as the parsing delimiter

C. The first field of logevents must be a properly-formatted date string or table return an error

D. The statement is not a valid Pig command

Correct Answer: B

QUESTION 11

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.

Correct 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 12

Which one of the following classes would a Pig command use to store data in a table defined in HCatalog?

A. org.apache.hcatalog.pig.HCatOutputFormat

B. org.apache.hcatalog.pig.HCatStorer

C. No special class is needed for a Pig script to store data in an HCatalog table

D. Pig scripts cannot use an HCatalog table

Correct Answer: B

QUESTION 13

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

A. You will not be able to compress the intermediate data.

B. You will 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 reduces.

Correct Answer: C

Explanation: Multiple reducers and total ordering

If your sort job runs with multiple reducers (either because mapreduce.job.reduces in mapred-site.xml has been set to a number larger than 1, or because you\\'ve used the -r option to specify the number of reducers on the command-line), then by default Hadoop will use the HashPartitioner to distribute records across the reducers. Use of the HashPartitioner means that you can\\'t concatenate your output files to create a single sorted output file. To do this you\\'ll need total ordering,

Reference: Sorting text files with MapReduce

QUESTION 14

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

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 15

To process input key-value pairs, your mapper needs to lead 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: C

[HADOOP-PR000007 Practice Test](#)

[HADOOP-PR000007 Exam Questions](#)

[HADOOP-PR000007 Braindumps](#)