Hadoop 2.0 Certification exam for Pig and Hive Developer

Hortonworks Apache-Hadoop-Developer

Version Demo

Total Demo Questions: 10

Total Premium Questions: 108

Buy Premium PDF

https://dumpsboss.com

support@dumpsboss.com

QUESTION NO: 1

In a MapReduce job, the reducer receives all values associated with same key. Which statement best describes the ordering of these values?

A. The values are in sorted order.

B. The values are arbitrarily ordered, and the ordering may vary from run to run of the same MapReduce job.

C. The values are arbitrary ordered, but multiple runs of the same MapReduce job will always have the same ordering.

D. Since the values come from mapper outputs, the reducers will receive contiguous sections of sorted values.

ANSWER: B

Explanation:

Note:

* Input to the Reducer is the sorted output of the mappers.

* The framework calls the application's Reduce function once for each unique key in the sorted order.

* Example:

For the given sample input the first map emits:

< Hello, 1>

< World, 1>

< Bye, 1>

< World, 1>

The second map emits:

< Hello, 1>

< Hadoop, 1>

< Goodbye, 1>

< Hadoop, 1>

QUESTION NO: 2

Which HDFS command displays the contents of the file x in the user's HDFS home directory?

A. hadoop fs -ls x

B. hdfs fs -get x

C. hadoop fs -cat x

D. hadoop fs -cp x

ANSWER: C

QUESTION NO: 3

Which one of the following is NOT a valid Oozie action?

A. mapreduce

B. pig

C. hive

D. mrunit

ANSWER: D

QUESTION NO: 4

Which Two of the following statements are true about hdfs? Choose 2 answers

A. An HDFS file that is larger than dfs.block.size is split into blocks

B. Blocks are replicated to multiple datanodes

C. HDFS works best when storing a large number of relatively small files

D. Block sizes for all files must be the same size

ANSWER: A B

QUESTION NO: 5

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: 6

In Hadoop 2.0, 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

ANSWER: A D

QUESTION NO: 7

Which two of the following statements are true about Pig's approach toward data? Choose 2 answers

A. Accepts only data that has a key/value pair structure

B. Accepts data whether it has metadata or not

C. Accepts only data that is defined by metadata tables stored in a database

D. Accepts tab-delimited text data only

E. Accepts any data: structured or unstructured

ANSWER: B E

QUESTION NO: 8

Which TWO of the following statements are true regarding Hive? Choose 2 answers

A. Useful for data analysts familiar with SQL who need to do ad-hoc queries

B. Offers real-time queries and row level updates

C. Allows you to define a structure for your unstructured Big Data

D. Is a relational database

ANSWER: A C

QUESTION NO: 9

You write MapReduce job to process 100 files in HDFS. Your MapReduce algorithm uses TextInputFormat: the mapper applies a regular expression over input values and emits key-values pairs with the key consisting of the matching text, and the value containing the filename and byte offset. Determine the difference between setting the number of reduces to one and settings the number of reducers to zero.

A. There is no difference in output between the two settings.

B. With zero reducers, no reducer runs and the job throws an exception. With one reducer, instances of matching patterns are stored in a single file on HDFS.

C. With zero reducers, all instances of matching patterns are gathered together in one file on HDFS. With one reducer, instances of matching patterns are stored in multiple files on HDFS.

D. With zero reducers, instances of matching patterns are stored in multiple files on HDFS. With one reducer, all instances of matching patterns are gathered together in one file on HDFS.

ANSWER: D

Explanation:

* It is legal to set the number of reduce-tasks to zero if no reduction is desired.

In this case the outputs of the map-tasks go directly to the FileSystem, into the output path set by setOutputPath(Path). The framework does not sort the map-outputs before writing them out to the FileSystem.

* Often, you may want to process input data using a map function only. To do this, simply set mapreduce.job.reduces to zero. The MapReduce framework will not create any reducer tasks. Rather, the outputs of the mapper tasks will be the final output of the job.

Note:

Reduce

In this phase the reduce(WritableComparable, Iterator, OutputCollector, Reporter) method is called for each pair in the grouped inputs.

The output of the reduce task is typically written to the FileSystem via OutputCollector.collect(WritableComparable, Writable).

Applications can use the Reporter to report progress, set application-level status messages and update Counters, or just indicate that they are alive.

The output of the Reducer is not sorted.

QUESTION NO: 10

MapReduce v2 (MRv2/YARN) is designed to address which two issues?

A. Single point of failure in the NameNode.

B. Resource pressure on the JobTracker.

C. HDFS latency.

D. Ability to run frameworks other than MapReduce, such as MPI.

E. Reduce complexity of the MapReduce APIs.

F. Standardize on a single MapReduce API.

ANSWER: A B

Explanation:

Reference: Apache Hadoop YARN – Concepts & Applications