APACHE-HADOOP-DEVELOPER^{Q&As}

Hadoop 2.0 Certification exam for Pig and Hive Developer

Pass Hortonworks APACHE-HADOOP-DEVELOPER Exam with 100% Guarantee

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

https://www.geekcert.com/apache-hadoop-developer.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

QUESTION 1

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.

Correct 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 2

You have just executed a MapReduce job. Where is intermediate data written to after being emitted from the Mapper\\'s map method?

A. Intermediate data in streamed across the network from Mapper to the Reduce and is never written to disk.

B. Into in-memory buffers on the TaskTracker node running the Mapper that spill over and are written into HDFS.

C. Into in-memory buffers that spill over to the local file system of the TaskTracker node running the Mapper.

D. Into in-memory buffers that spill over to the local file system (outside HDFS) of the TaskTracker node running the Reducer

E. Into in-memory buffers on the TaskTracker node running the Reducer that spill over and are written into HDFS.

Correct Answer: C

Explanation: The mapper output (intermediate data) is stored on the Local file system (NOT HDFS) of each individual mapper nodes. This is typically a temporary directory location which can be setup in config by the hadoop administrator. The intermediate data is cleaned up after the Hadoop Job completes.

Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers, Where is the Mapper Output (intermediate kay-value data) stored ?

QUESTION 3

A client application creates an HDFS file named foo.txt with a replication factor of 3. Identify which best describes the file access rules in HDFS if the file has a single block that is stored on data nodes A, B and C?

A. The file will be marked as corrupted if data node B fails during the creation of the file.

B. Each data node locks the local file to prohibit concurrent readers and writers of the file.

C. Each data node stores a copy of the file in the local file system with the same name as the HDFS file.

D. The file can be accessed if at least one of the data nodes storing the file is available.

Correct Answer: D

Explanation: HDFS keeps three copies of a block on three different datanodes to protect against true data corruption. HDFS also tries to distribute these three replicas on more than one rack to protect against data availability issues. The fact that HDFS actively monitors any failed datanode(s) and upon failure detection immediately schedules re-replication of blocks (if needed) implies that three copies of data on three different nodes is sufficient to avoid corrupted files. Note: HDFS is designed to reliably store very large files across machines in a large cluster. It stores each file as a sequence of blocks; all blocks in a file except the last block are the same size. The blocks of a file are replicated for fault tolerance. The block size and replication factor are configurable per file. An application can specify the number of replicas of a file. The replication factor can be specified at file creation time and can be changed later. Files in HDFS are write-once and have strictly one writer at any time. The NameNode makes all decisions regarding replication of blocks. HDFS uses rack-aware replica placement policy. In default configuration there are total 3 copies of a datablock on HDFS, 2 copies are stored on datanodes on same rack and 3rd copy on a different rack. Reference: 24 Interview Questions and Answers for Hadoop MapReduce developers , How the HDFS Blocks are replicated?

QUESTION 4

Which one of the following statements is true about a Hive-managed table?

A. Records can only be added to the table using the Hive INSERT command.

B. When the table is dropped, the underlying folder in HDFS is deleted.

C. Hive dynamically defines the schema of the table based on the FROM clause of a SELECT query.

D. Hive dynamically defines the schema of the table based on the format of the underlying data.

Correct Answer: B

QUESTION 5

Which HDFS command uploads a local file X into an existing HDFS directory Y?

A. hadoop scp X Y

B. hadoop fs -localPut X Y

C. hadoop fs-put X Y

D. hadoop fs -get X Y

Correct Answer: C

APACHE-HADOOP-
DEVELOPER PDF Dumps

APACHE-HADOOP-
DEVELOPER Study Guide

APACHE-HADOOP-
DEVELOPER Exam
Questions