KE KILL EXAMS

HDPCD Dumps
HDPCD Braindumps
HDPCD Real Questions
HDPCD Practice Test
HDPCD Actual Questions

PASS

killexams.com

Hortonworks

HDPCD

Hortonworks Data Platform Certified Developer

ORDER FULL VERSION

QUESTION: 97

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 <key, (list of values)> 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: 98

Indentify the utility that allows you to create and run MapReduce jobs with any executable or script as the mapper and/or the reducer?

A. Oozie
B. Sqoop
C. Flume
D. Hadoop Streaming
E. mapred

Answer: D

Explanation:
Hadoop streaming is a utility that comes with the Hadoop distribution. The utility allows you to create and run Map/Reduce jobs with any executable or script as the mapper and/or the reducer.

Reference:
http://hadoop.apache.org/common/docs/r0.20.1/streaming.html (Hadoop Streaming, second sentence)

QUESTION: 99
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.

Answer: B

QUESTION: 100
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.

QUESTION: 101
Which process describes the lifecycle of a Mapper?

A. The JobTracker calls the TaskTracker's configure () method, then its map () method and finally its close () method.
B. The TaskTracker spawns a new Mapper to process all records in a single input split.
C. The TaskTracker spawns a new Mapper to process each key-value pair.
D. The JobTracker spawns a new Mapper to process all records in a single file.

Explanation:
For each map instance that runs, the TaskTracker creates a new instance of your mapper.
Note:
* The Mapper is responsible for processing Key/Value pairs obtained from the InputFormat. The mapper may perform a number of Extraction and Transformation functions on the Key/Value pair before ultimately outputting none, one or many Key/Value pairs of the same, or different Key/Value type.
* With the new Hadoop API, mappers extend the org.apache.hadoop.mapreduce.Mapper class. This class defines an 'Identity' map function by default - every input Key/Value pair obtained from the InputFormat is written out.
Examining the run() method, we can see the lifecycle of the mapper:
/**
* Expert users can override this method for more complete control over the
* execution of the Mapper.
* @param context
* @throws IOException
*/
public void run(Context context) throws IOException, InterruptedException {
setup(context);
while (context.nextKeyValue()) {
map(context.getCurrentKey(), context.getCurrentValue(), context);
}
cleanup(context);

}

setup(Context) - Perform any setup for the mapper. The default implementation is a no-op method.
map(Key, Value, Context) - Perform a map operation in the given Key / Value pair. The default implementation calls Context.write(Key, Value)
cleanup(Context) - Perform any cleanup for the mapper. The default implementation is a no-op method.

Reference:
Hadoop/MapReduce/Mapper

QUESTION: 102
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

Answer: C

QUESTION: 103
Which one of the following statements is FALSE regarding the communication between DataNodes and a federation of NameNodes in Hadoop 2.2?

A. Each DataNode receives commands from one designated master NameNode.
B. DataNodes send periodic heartbeats to all the NameNodes.
C. Each DataNode registers with all the NameNodes.
D. DataNodes send periodic block reports to all the NameNodes.

Answer: A

QUESTION: 104
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.

Answer: D

Explanation:
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: 105
Review the following 'data' 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

Answer: A

QUESTION: 106
Which one of the following is NOT a valid Oozie action?

A. mapreduce
B. pig
C. hive
D. mrunit

Answer: D

QUESTION: 107
Examine the following Hive statements:

```
CREATE TABLE x (name STRING, age INT, zip INT, salary DOUBLE)
ROW FORMAT DELIMITED FIELDS TERMINATED BY
',';
LOAD DATA INPATH 'input/File1' OVERWRITE INTO TABLE x;
SELECT * FROM x SORT BY age;
```

Assuming the statements above execute successfully, which one of the following statements is true?

A. Each reducer generates a file sorted by age
B. The SORT BY command causes only one reducer to be used
C. The output of each reducer is only the age column
D. The output is guaranteed to be a single file with all the data sorted by age

Answer: A

QUESTION: 108
Your client application submits a MapReduce job to your Hadoop cluster. Identify the Hadoop daemon on which the Hadoop framework will look for an available slot schedule a MapReduce operation.

A. TaskTracker
B. NameNode
C. DataNode
D. JobTracker
E. Secondary NameNode

Answer: D

Explanation:
JobTracker is the daemon service for submitting and tracking MapReduce jobs in Hadoop. There is only One Job Tracker process run on any hadoop cluster. Job Tracker runs on its own JVM process. In a typical production cluster its run on a separate machine. Each slave node is configured with job tracker node location. The JobTracker is single point of failure for the Hadoop MapReduce service. If it goes down, all running jobs are halted. JobTracker in Hadoop performs following actions(from Hadoop Wiki:)
Client applications submit jobs to the Job tracker.
The JobTracker talks to the NameNode to determine the location of the data
The JobTracker locates TaskTracker nodes with available slots at or near the data The JobTracker submits the work to the chosen TaskTracker nodes.
The TaskTracker nodes are monitored. If they do not submit heartbeat signals often enough, they are deemed to have failed and the work is scheduled on a different TaskTracker.
A TaskTracker will notify the JobTracker when a task fails. The JobTracker decides what to do then: it may resubmit the job elsewhere, it may mark that specific record as something to avoid, and it may may even blacklist the TaskTracker as unreliable.
When the work is completed, the JobTracker updates its status. Client applications can poll the JobTracker for information.

Reference:
24 Interview Questions & Answers for Hadoop MapReduce developers, What is a JobTracker in Hadoop? How many instances of JobTracker run on a Hadoop Cluster?

SAMPLE QUESTIONS

KILL EXAMS

*These questions are for demo purpose only. **Full version** is up to date and contains actual questions and answers.*

Killexams.com is an online platform that offers a wide range of services related to certification exam preparation. The platform provides actual questions, exam dumps, and practice tests to help individuals prepare for various certification exams with confidence. Here are some key features and services offered by Killexams.com:

ORDER FULL VERSION
Special Discount Coupon

Actual Exam Questions: Killexams.com provides actual exam questions that are experienced in test centers. These questions are updated regularly to ensure they are up-to-date and relevant to the latest exam syllabus. By studying these actual questions, candidates can familiarize themselves with the content and format of the real exam.

Exam Dumps: Killexams.com offers exam dumps in PDF format. These dumps contain a comprehensive collection of questions and answers that cover the exam topics. By using these dumps, candidates can enhance their knowledge and improve their chances of success in the certification exam.

Practice Tests: Killexams.com provides practice tests through their desktop VCE exam simulator and online test engine. These practice tests simulate the real exam environment and help candidates assess their readiness for the actual exam. The practice tests cover a wide range of questions and enable candidates to identify their strengths and weaknesses.

Guaranteed Success: Killexams.com offers a success guarantee with their exam dumps. They claim that by using their materials, candidates will pass their exams on the first attempt or they will refund the purchase price. This guarantee provides assurance and confidence to individuals preparing for certification exams.

Updated Content: Killexams.com regularly updates its question bank and exam dumps to ensure that they are current and reflect the latest changes in the exam syllabus. This helps candidates stay up-to-date with the exam content and increases their chances of success.

Technical Support: Killexams.com provides free 24x7 technical support to assist candidates with any queries or issues they may encounter while using their services. Their certified experts are available to provide guidance and help candidates throughout their exam preparation journey.