Review Test Submission: Hadoop MapReduce Quiz

User	Venkata Kowsik Temididapathi
Course	CS 6350.002 - Big Data Management and Analytics - S22
Test	Hadoop MapReduce Quiz
Started	2/12/22 12:40 PM
Submitted	2/12/22 1:16 PM
Due Date	2/13/22 11:59 PM
Status	Completed
Attempt Score	100 out of 100 points
Time Elapsed	35 minutes
Results Displayed	All Answers, Submitted Answers, Correct Answers

Question 1

10 out of 10 points

In Hadoop MapReduce output of Mapper is stored on:

Selected Answer: ✓ Local Disk

Answers:
- Memory
- HDFS
- ✓ Local Disk
- Remote Location

Question 2

10 out of 10 points

Which of the following is the correct order of operations for MapReduce in Hadoop

Answers	Selected Answer
✓ 1. Reading input data from HDFS	✓ 1. Reading input data from HDFS

2.
Map operation performed is performed so that useful (key, value) pairs can be identified from input data

Intermediate outputs are stored on local disk

3.

Data is sorted and shuffled to group by key values

4.

5. Reduce operation is performed

6. Results are written back to HDFS

2.
Map operation performed is performed so that useful (key, value) pairs can be identified from input data

Intermediate outputs are stored on local disk

3.

Data is sorted and shuffled to group by key values

4.

5. Reduce operation is performed

6. Results are written back to HDFS

Question 3

10 out of 10 points

What is meant by locality of computation in Hadoop?

Selected Answer: Data storage and processing can be co-located on the same node to optimize overall performance.

Answers:

Data storage and processing can be co-located on the same node to optimize overall performance.

Computation should be performed at a remote location

Data storage should be as distributed as possible

All data processing should happen at only a single node

Question 4

10 out of 10 points

What is the input to the reduce function?

Selected Answer: One key and list of all associated values

Answers:

One key and a list of some (partial) values associated with the key

One key and list of all associated values

One key and one value

One value and a list of all associated keys

Question 5

10 out of 10 points

The number of map tasks is dependent on:

Selected Answer: size of input data

Answers:
- none of the above
- size of input data
- number of useful values in the data
- number of useful keys in the data

Question 6 10 out of 10 points

The number of reduce tasks is dependent on:

Selected Answer: number of useful keys in the data

Answers:
- number of useful keys in the data
- number of useful values in the data
- none of the above
- size of input data

Question 7 10 out of 10 points

You have a Hadoop cluster with 20 machines each having 250 GB of HDFS disk space. The system settings are: block size = 128 MB, replication factor = 3. Assuming the cluster is totally free i.e. no stored data and no jobs. You want to upload 10 text files, each of size 250 GB and then perform a Wordcount job i.e. count the frequency of words. What is going to happen?

Selected Answer: The data upload fails for an intermediate file

Answers:
- The data upload fails for the first file
- The data upload fails for an intermediate file
- WordCount will run successfully
- Map step fails as there are too many inputs

Question 8 10 out of 10 points

If a node running map tasks fails, how will the map tasks be recovered?

Selected Answer: The application master will re-run the completed map tasks on another node, since their results will be lost when node crashes.

Answers:
- The application master will re-run the completed map tasks on another node, since their results will be lost when node crashes.
- The completed map tasks' results can be recovered because their results will be stored on HDFS
- Entire Mapreduce job will have to be re-run
- A node can never crash

Question 9 10 out of 10 points

In Hadoop MapReduce, reduce tasks takes input from:

Selected Answer: multiple map tasks

Answers:
- multiple map tasks
- single map task
- all map tasks that ran on the same machine as the reducer
- reduce tasks doesn't need any input.

Question 10 10 out of 10 points

Suppose I create the following function in Scala:

```
def myFunction(x: Int): Int = {
    if (x % 2 == 0)    2*x
    else   3*x
}
```

I also define a list as:

```
val list = List(1, 2, 3, 4, 5)
```

Now, I want to run a MapReduce job that will apply the function myFunction to every element of the list and then reduce it using the max operator. Which of the following accomplishes this?

Selected Answer:
```
import scala.math._
list.map(myFunction).reduce(max)
```

Answers:
- None of the above
-
```
import scala.math._
list.reduce(max.map(myFunction))
```
-
```
import scala.math._
list.reduce(max).map(myFunction)
```

```
import scala.math._
list.map(myFunction).reduce(max)
```

Saturday, March 12, 2022 9:19:20 AM CST

← OK