

1. Map/Reduce Library partitions the input data into M pieces of typically 16-64 MB.

True.

2. If there are M partitions of the input, there are M map workers running simultaneously.

False. There are generally less worker nodes than partitions.

3. Reduce function accepts intermediate key and a set of values for that key and merges together these values to form possibly a smaller set of values.

True.

4. Map and Reduce are user defined functions

True.

5. MapReduce is a restricted programming model?

True.

Following are multiple Choice Questions.

Question 1: Which of the following functions of cloud computing is optimized by MapReduce?

e) All of the Above

Question 2: The value of R, the number of Reduce workers, is determined by

b) master program - number of partitions are defined by user though

Question 3: In Map Reduce function, what can be said about the input key, output key and the intermediate key values?

C. The input keys are drawn from a different domain as output keys and the intermediate keys are drawn from the same domain as output keys.

Question 4: What is the relationship between intermediate keys and intermediate value?

A 1:N

Question 5: Which of the following Key/Value pair is used by Map function in Map Reduce?

b) $(k_1, v_1) \rightarrow \text{List}(k_2, v_2)$

Question 6: How is the intermediate Key/Value pair arranged and processed during partition of tasks?

a) Key/value pairs are processed in increasing key order

Question 7: Which is the reduce function in the example of Distributed Grep:

- c) Copy the intermediate data to output

Question 8: Which is the reduce function in the example of Text indexing?

- d) none of the above - count document IDs for each word

Question 9: Which of the following properties of Worker Machine is stored by MapReduce?

- a) State
- b) Identity

Question 10: Map is phased in M pieces and Reduce is phased into R pieces. What is the maximum number of scheduling operations required to be performed?

- b) $O(M+R)$

Question 11: How is the network bandwidth maintained by Map Reduce in cloud computing?

- a) Data is stored locally on disks on different machines with 64 MB blocks

Question 12: How MapReduce environment handle failures?

- b) Master worker pings the client worker regularly to check the status

Question 13: MapReduce provides the following benefits:

- a) Reducing the amount of data sent across the network
- b) Optimization the data by storing it locally
- c) Redundant execution to handle machine failures and data loss

Question 14: What is the difference between the Combiner function and Reduce function?

- b) Output of combiner function is written to intermediate file and Reduce function to output file