

Cloud Computing Assignment 1

Name : Nikhil Raj Buduga

1. Explain the general concept in your own words (Word Limit : 50)

Map reduce is a software framework developed by google for processing large data sets in a distributed fashion. The main idea behind this concept is mapping our dataset into <key, value> collection pair, and then reducing over all pairs with the same key.

2. Difference between map and reduce? (Word Limit : 50)

- Mapping processes the input key value pairs into another set of key, value pairs in which all the values are grouped into a same immediate key.
- Reduce iterates over the values that are associated with a key and produces combined outputs.

3. How fault tolerance is handled? (Word Limit : 50)

The workers are constantly being checked by the master. If a worker does not respond, it is assumed to be dead. The checkpoints are responsible for handling the in progress tasks which are reset to idle to make the task eligible for rescheduling.

4. What is the purpose of combiner? (Word Limit : 50)

A combiner is used to minimize the network congestion between the mapper and the reducer. The data is partial merged by the combiner function before sending it over the network to enhance the overall performance.

5. Advantages and disadvantages - 2 each (Word Limit : 50)

Advantages:

- It has high scalability(due to simple design) and fault tolerance.
- It is cost effective solution for businesses.

Disadvantages:

- When the map generates enormous number of keys then the sorting process takes too much time.
- If the master fails, MapReduce will terminate the job.