Student ID: M13782945

Name: Jia-Bin, Chang

1. Explain the general concept in your own words

When computing on large data, using map function to encode the input data into key-value pairs first. Then implement the reduce function to merge all the key-value pairs into smaller sets.

1. Difference between Map and Reduce?

The map function is used when processing input data into key-value pairs. The reduce function is used to reduce the key-value pairs into a smaller set when getting the key-value pairs from map function.

1. How fault tolerance is handled?

The master keeps pinning all the worker periodically. If any worker has no response, the master will consider the worker failed and reassign the task to another worker.

1. What is the purpose of combiner?

The purpose of combiner is to let user to deal with repetition in their task. It can be design as an intermediate function before the reduce function, where it can merge some of the same key-value pairs before passing to reduce function.

1. Advantages and disadvantages of MapReduce versus other techniques- 2 each

The first advantage of map reduce is it can easily scale data over computing nodes, and the second is that the fault tolerance is higher. The first disadvantage is that it takes too much time handling iteration tasks, and the second is it is not real-time processing.