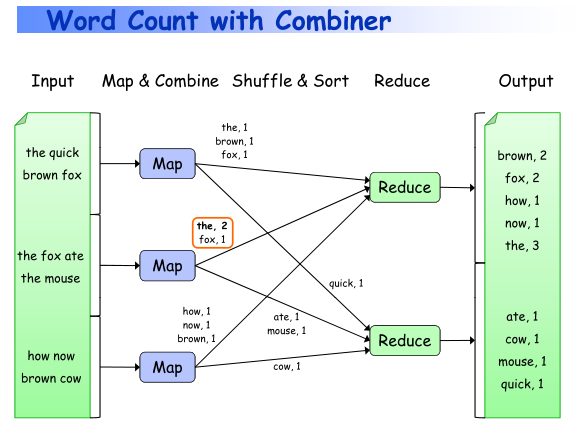
**Cloud Computing**

**Exercise 7 – MapReduce II**

1. What is distributed file systems?
2. Why we needed HDFS in mapReduce? What are the main components in a HDFS system.
3. Discuss the fault tolerance in mapReduce when tasks or nodes crash.
4. What is the straggler, and how to handle the straggler.
5. In the following application, please write the codes using combiner.



1. Please write codes for numerical integration using mapReduce.
2. Using both pairs and stripes to write the codes for conditional probabilities.