**Cloud Computing Homework - 1**

**Q1.**

* Google created the software framework Map Reduce to handle massive data volumes in a distributed manner. The fundamental idea behind this notion is to reduce overall pairings that have the same key after mapping our dataset into collection pairs.

**Q2.**

* Mapping converts the input key-value pairs into a new set of key-value pairs where each immediate key contains a group of values.
* Reduce generates combined outputs by iterating over the values linked to a key.

**Q3**

* The boss continually inspects the employees. It is presumed that a worker will die if it doesn't react. The in-progress tasks that are reset to idle to make the task suitable for rescheduling are handled by the checkpoints.

**Q4**

* To alleviate network congestion between the mapper and the reducer, a combiner is utilized. To improve overall performance, the combiner function partially merges the data before sending it over the network.

**Q5.**

**Advantages:**

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

**Disadvantages:**

* The sorting procedure takes too long when the map creates many keys.
* If the master fails, MapReduce will terminate the job.