**Name: Nikhil Jadhav**

**Student ID: 801075504**

**ITCS 3190 - 091: Cloud Computing for Data Analysis**

**VIDEO CASE – 1**

Watch following videos:

**Video 1**: https://www.youtube.com/watch?v=n64LnzXQXN0

**Video 2**: <https://www.youtube.com/watch?v=zgaGeyNgBYE>

Video 1 provide basic operations of MapReduce

Video 2 gives an example problem solved in MapReduce

**Video Case Questions:**

1. Give some advantages of using MapReduce

* Faster data processing – Using the MapReduce approach i.e. by using multiple server for data processing, large amount of data can be processed extremely fast compared to that of traditional method of data processing i.e. using single system to do all the data processing.
* Scalable – The MapReduce approach is highly scalable, as we can easily add or remove server used for data processing depending upon the size of data required for processing.
* Cost-effectiveness – Due to high scalability nature of MapReduce approach a lot of resources can be used in most efficient manner which can save time and resources for data processing which will later save money for the organization.
* Parallel processing – The primary approach the MapReduce takes for data processing is parallel execution of the data across different servers which saves a lot of execution time and resource utilization.

1. Describe briefly the steps that the MapReduce follow, after Mapper phases producing their intermediate outputs

* The mapper phase produces the chunk of data in the form of (key, value) pair.
* The key corresponds to the reducer to which the pair needs to go for further processing.
* Reducers then process the data and produce on final (key, value) pair for each key.

1. What happens in the Reducer phase?

* Each reducer is assigned a unique key.
* The reducer receives its respective key value pair from the mapper for the data processing.
* The reducer then performs the processing on all the (key, value) pair to produce one single (key, value) pair.

1. Mention two other applications that could use MapReduce to fasten their processing time

* **Data Analytics –** To calculate unique user visit to per page. Where user is the key and the page viewed, or page clicked on is the value.
* **Web Crawling –** A web crawling is a computer program that browses the World Wide Web in a methodical or automated manner to perform a particular operation i.e. searching a particular keyword etc. In this case MapReduce helps in duplicating the pages by URL and helps in identifying new pages to crawl.