NAME: SAI KRISHNA MANNAVA

STUDENT ID: 801136361

**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

**Answer:**

* Scalability: The Hadoop platform is highly flexible as it has the ability to store and restore large data sets across many of the servers.
* Fast: Hadoop platform uses distributed file system which uses the mapping system to locate the data in the clusters. The tools used for the data processing are located on the same server which allows fast processing of data.
* Security: Map Reducer works with HDFS which allows more security and allows only approved users to work with the data stored in the system.
* Parallel processing: Map Reducer allows the divided tasks to execute parallelly so that the entire program is completed in less time.
* Simple programming: It allows the users to develop the MapReduce that can handle with more ease and efficiency using the languages such as JAVA etc.

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

**Answer:**

* After the mapper phase producing the key value sets, these sets are passed to the reducer based on the shuffle and sorting.
* Then the reducer performs its function on every key value set, the data is aggregated, filtered and passes zero or more key-valued pairs to the final phase.
* Then the output formatter that translates the final key-value pairs from the Reducer function and writes them onto a file using a record writer.

1. What happens in the Reducer phase?

**Answer:**

* After the shuffle and sorting phase, the Reducer takes the grouped key-value paired data as input and runs a Reducer function on each one of the key pair values.
* Here, the data can be aggregated, filtered, and combined in a number of ways, and it requires a wide range of processing.
* After the execution is over, it gives zero or more key-value pairs to the final phase.

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

**Answer:**

* Fraud detection
* social networking
* page rank
* analysis of biosensors application
* Streaming data