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

**VIDEO CASE – 1**

|  |
| --- |
| **Submitted by:** Abdullah Al Raqibul Islam  **UNCC ID:** 801151189 |

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

|  |
| --- |
| There are numerous advantages of using MapReduce in wide ranges of applications. Among them here is the few notably,   * Scalability: Surely this should be the top reason of using MapReduce in any application. High scalability comes through the ability of storing and distributing the data and processing efforts among lots of servers. * Fastness: MapReduce application can utilize distributed file system which accelerate the processing speed of a server, compare with storing the whole data in a single file system. * Cost-effectiveness: MapReduce is a very much cost-effective solution for the applications that need to store and process data of exponential growths. * Wide variety of usefulness: Any application that has huge chunk of unstructured data, can utilize the MapReduce functionality. |

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

|  |
| --- |
| MapReduce executes in two steps,   * Map step: In this step, unstructured data is been processed and saved to the Hadoop File System (HDFS). The main goal is to process the data and create small chunks of relative data. * Reduce step: Reduce phase take the input from the output generated in the map stage. Mostly apply the business logic in this step to retrieve more meaningful information.   In the video 2, a MapReduce example demonstrated where in the mapper phase, raw data has been broken into chunks and distributed among the mapper nodes. All the mappers run parallelly and produced indexed data as the output. |

1. What happens in the Reducer phase?

|  |
| --- |
| In the video 2, the Reducer phase receive the indexed data (by store) as input, from the mappers. Each store (i.e. NYC, MIAMI, LA, etc.) is been assigned to a single reducer. Reducers go to the mappers and receive the data of the corresponding stores. Once the reducer receives the data, it combines each of the small chunk of data (by store) and do the aggregation for them. |

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

|  |
| --- |
| As I mentioned earlier, any application that has huge chunk of unstructured data, can utilize the MapReduce functionality. For example,   * Search Engine can use MapReduce to index the search strings * Social Media can use MapReduce to find the key persons in the network |