

**CS 3006 Parallel and Distribute Computing**  
**Assignment #3**  
**Fall 2022**

**Due Date: Wednesday 7<sup>th</sup> December, 23:59**

**CLO#4**

**Q1: [HDFS]** write a survey on some HDFS client application, explaining the whole HDFS architecture as per client application

**Q2: [HDFS]** What is Google File System (GFS), differentiate between GFS and HDFS.

**Q3: [MapReduce]** Write and implement MAP and Reduce Functions for the following problems. Also do write input and output of each of map() and reduce().

- a) **Maintaining a paragraph-oriented inverted index of terms in a text file**
- b) **Calculating each word's frequency in a corpus of text**
- c) **Calculating n-Itemset Frequency count from a set of transactions data files**