GLS University
Assignment
Big Data Analytics Using Hadoop

Q.1	Develop a MapReduce job that will split each line of the input text file in to tokens and then count the occurrences of each unique token. The input text file should be available on the HDFS. The output will be in the form of: Hadoop 5 Hotspot 2
Q.2	Develop a MapReduce job that will analyze the minimum temperature for each year. The input text file should be available on HDFS. The output will be in the form of: 2014 -1 2015 20
Q.3	Develop a MapReduce job that will split each line of the input text file in to tokens and then find the average count of all the tokens in the input file. The input text file should be available on the HDFS. The output will be in the form of: Hadoop 5 Hotspot 2 AverageCount = 3.5
Q.4	Develop a MapReduce job that will take text file input and will produce the output: (1) Count of the total token having length greater than or equal to four characters. Total count for token = 5
Q.5	Develop a MapReduce job that will count the total number of females voters in the record file Voters.txt. The input text file should be available on the HDFS. The file must have fields : ID,NAME,GENDER,AGE. The output will be in the form of: No. of female voters are : 5
Q.6	Develop a MapReduce job that will count the number of reviews given by each unique user for the Musical Instruments. The file should be available on HDFS. Its availabe on Moodle (Musical_instruments_reviews.csv). The output should be in the form of A00625243BI8W1SSZNLMD 3 A10044ECXDUVKS 2

	A102MU6ZC9H1N6 3
	A109JTUZXO61UY 3
	A109ME7C09HM2M 4
	A10APIDAZISWQF 1
	A10E3QH2FQUBLF 3
	A10FM4ILBIMJJ7 5
	A10H2F00ZOT8S2 4
	A10HYGDU2NITYQ 3
	A10KH8EN77ZKWH 3
	A10N243R7A5ZW3 1
Q. 7	Perform the following on the movies.csv dataset (uploaded on moodle)
7.1	Write a mapreduce job to display all the details of the comedy movies.
7.2	Write a mapreduce job to find the count of the Documentary movies released in the year 1995.
7.3	Write a mapreduce job that will count the total number of missing records where 'genres' are missing. (Note: Remove genre from few records before processing.)
7.4	Write a mapreduce job to display only titles of the movie having "Gold" anywhere in the title.
7.5	Write a mapreduce that will display the count of the movies which belongs to both Drama and Romantic genre.