



# **BIG DATA ANALYSIS**

**Hadoop MapReduce for Climate  
Data Analytics**

NAME - SAHIL  
ROLL NO. -107121086  
ELECTRICAL AND ELECTRONICS ENGINEERING

# Task 2

- (b) How many Map and Reduce tasks did running Word Count on Gberg-100M.txt produce? Run it again on Gberg-200M.txt and Gberg-500M.txt and write your observations. Additionally, run the following command on the cluster: `$ hdfs getconf -confKey dfs.blocksize`

Text file	Map Tasks	Reduce Tasks	Input file Size	Block size
Gberg-100M.txt	1	1	104.3 Mb	134.21 Mb
Gberg-200M.txt	3	1	209.7 Mb	134.21 Mb
Gberg-500M.txt	4	1	524.3 Mb	134.21 Mb

After running given command for each Cluster :

1. Gberg-100.txt

```
hadoop@SahilHP:~/hadoop-3.2.3/bin$ hdfs getconf -confKey dfs.blocksize
2023-11-26 19:24:50,822 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
134217728
```

2. Gberg-200.txt

```
hadoop@SahilHP:~/hadoop-3.2.3/bin$ hdfs getconf -confKey dfs.blocksize
2023-11-26 19:36:38,404 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
134217728
```

3. Gberg-500.txt

```
hadoop@SahilHP:~/hadoop-3.2.3/bin$ hdfs getconf -confKey dfs.blocksize
2023-11-26 20:43:00,671 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
134217728
```

- (c) What is the link between the input size, the number of Map tasks, and the size of a block on HDFS?

In above table in part (b), we can clearly see that as input size increases map tasks increases but block size on HDFS remain same for all three input files.