# Big Data Analysis Experiment 5

## Name – Kush Parihar

## Roll No: 2203023

**Class: ISA-1-A**

# Execute Map Reduce program for the weather forecasting data and word count

## Title:

Analyzing Weather Forecasting Data and Word Count using MapReduce

## Problem Statement:

Discuss the challenge of handling large-scale weather forecasting data and the necessity of analyzing this data to derive meaningful insights for better decision- making in various sectors like agriculture, disaster management, etc.

## Aim:

To efficiently process and analyze weather forecasting data along with performing a word count operation to understand the frequency of certain terms, using the MapReduce programming model.

## Theory:

Explain the concept of Big Data, the MapReduce programming model, and its significance in handling large datasets. You may want to include how MapReduce works, its advantages, and how it's suited for analyzing weather forecasting data.

## Word Count Analysis Algorithm:

### Map Phase:

**Input Reading**: Read a portion of the text data.

**Processing**: Split each line into words.

**Key-Value Emitting**: Emit a key-value pair for each word where the key is the word and the value is 1.

import sys def mapper():

for line in sys.stdin:

words = line.strip().split() for word in words:

print(f'{word}\t1')

if name == " main ": mapper()

### Shuffle and Sort Phase:

**Grouping**: Group all the key-value pairs by key (word).

### Reduce Phase:

**Aggregation**: Sum the counts for each word.

**Output Emitting**: Emit the word and its count to the output file.

import sys def reducer():

current\_word = None current\_count = 0

for line in sys.stdin:

word, count = line.strip().split('\t') count = int(count)

if current\_word == word: current\_count += count

else:

if current\_word: print(f'{current\_word}\t{current\_count}')

current\_word = word current\_count = count

if current\_word == word: print(f'{current\_word}\t{current\_count}')

if name == " main ": reducer()

### Running the Program:

Run the MapReduce job using Hadoop Streaming with the command:

hadoop jar $HADOOP\_HOME/share/hadoop/tools/lib/hadoop-streaming-\* \

-files mapper.py,reducer.py \

-mapper mapper.py \

-reducer reducer.py \

-input /user/hadoop/input-data \

-output /user/hadoop/output-results

## Applications:

Discuss various fields where analyzing weather data could be beneficial like agriculture, disaster management, urban planning, etc. Also, mention how word count analysis can be used in data mining, text analytics, etc.

## Conclusion:

Summarize the findings from executing the MapReduce program on weather forecasting data and word count. Discuss the efficiency and scalability of MapReduce and how it can be employed for further analysis.