Expt-3

# AIM:

Map Reduce program to process a weather dataset.

To implement MapReduce program to process a weather dataset.

# PROCEDURE:

## Create Weather Dataset:

nano weather\_data.txt

## Example content:

20220101 30.5

20220102 29.8

## Mapper Program (mapper.py):

#!/usr/bin/env python3 import sys

for line in sys.stdin: line = line.strip()

month = line[4:6] # Extracting month temp = line[7:11] # Extracting temperature print(f'{month}\t{temp}')

## Reducer Program (reducer.py):

#!/usr/bin/env python3 import sys

current\_month = None current\_max\_temp = -float('inf')

for line in sys.stdin: line = line.strip()

month, temp = line.split('\t')

try:

temp = float(temp) except ValueError:

continue

if current\_month == month:

current\_max\_temp = max(current\_max\_temp, temp) else:

if current\_month: print(f'{current\_month}\t{current\_max\_temp}')

current\_month = month current\_max\_temp = temp

if current\_month == month: print(f'{current\_month}\t{current\_max\_temp}')

## Run the Program:

hdfs dfs -mkdir /weatherdata

hdfs dfs -copyFromLocal weather\_data.txt /weatherdata

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

-input /weatherdata/weather\_data.txt \

-output /weatherdata/output \

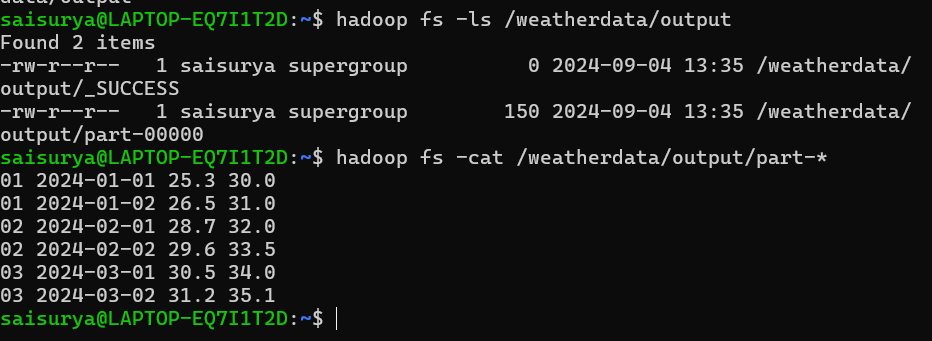
-mapper mapper.py \

-reducer reducer.py

## Check Output:

hdfs dfs -cat /weatherdata/output/part-00000

# OUTPUT:



**RESULT:**

Thus, the program for weather dataset using Map Reduce has been executed successfully.