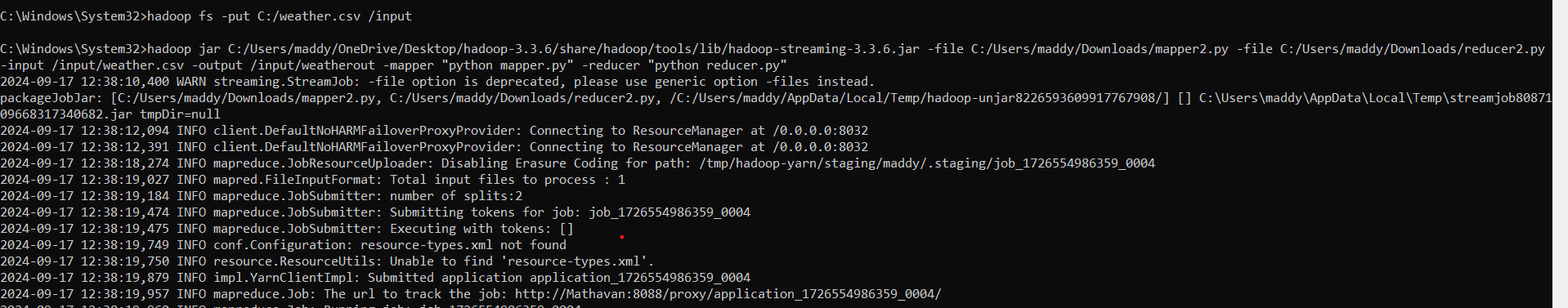
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **EX.NO.3 210701154**  **IMPLEMENT A MAPREDUCE PROGRAM TO PROCESS A WEATHER DATASET**      **AIM:**  To implement a MapReduce python program to process a weather dataset in Hadoop.      **PROCEDURE:**   1. Open command prompt as administrator and start the Hadoop by using the command:   start-all.cmd   1. Create a new directory in the Hadoop file systems using the command:   hadoop fs -mkdir /weather   1. Upload the input text file into the weather directory using the command:   hadoop fs -put  C:/Users/gjega/OneDrive/Documents/hadoop\_weather/WeatherPrediction/sample\_weather.txt  /weather   1. Create the mapper and reducer files. 2. To execute the files with Hadoop streaming run the following command:   hadoop jar C:/hadoop-3.3.6/share/hadoop/tools/lib/hadoop-streaming-3.3.6.jar ^ -file C:/Users/maddy/Documents/ hadoop\_weather /WeatherPrediction/mapper.py ^ -file   |  |  | | --- | --- | | /weather/sample\_weather.txt ^ -output /weather/output ^ -mapper "python mapper.py" ^ - | | | reducer "python reducer.py" |  |   C:/Users/maddy/Documents/ hadoop\_weather /WeatherPrediciton/reducer.py ^ -input      **MAPPER.PY:**  import sys for line in sys.stdin:  # Strip whitespace and skip empty lines line = line.strip() if not line: continue    fields = line.split(',') if len(fields) < 2:  continue # Skip lines that don't have enough fields | |
| **EX.NO.3**  date = fields[0]  year = date[:4] # Extract the year (first 4 characters of date) temperature = fields[1]    # Print the year and temperature print(f"{year}\t{temperature}")    **REDUCER.PY:**    import sys    current\_year = None current\_sum = 0.0  current\_count = 0    for line in sys.stdin:  line = line.strip()  year, temperature = line.split('\t')    # Skip non-numeric temperatures try:  temperature = float(temperature) except ValueError:  continue    if current\_year == year:  current\_sum += temperature current\_count += 1 else:  if current\_year:  # Output the average temperature for the previous year print(f"{current\_year}\t{current\_sum / current\_count:.2f}")    current\_year = year current\_sum = temperature current\_count = 1    # Output the average temperature for the last year if current\_year == year: print(f"{current\_year}\t{current\_sum / current\_count:.2f}") | **210701154** |

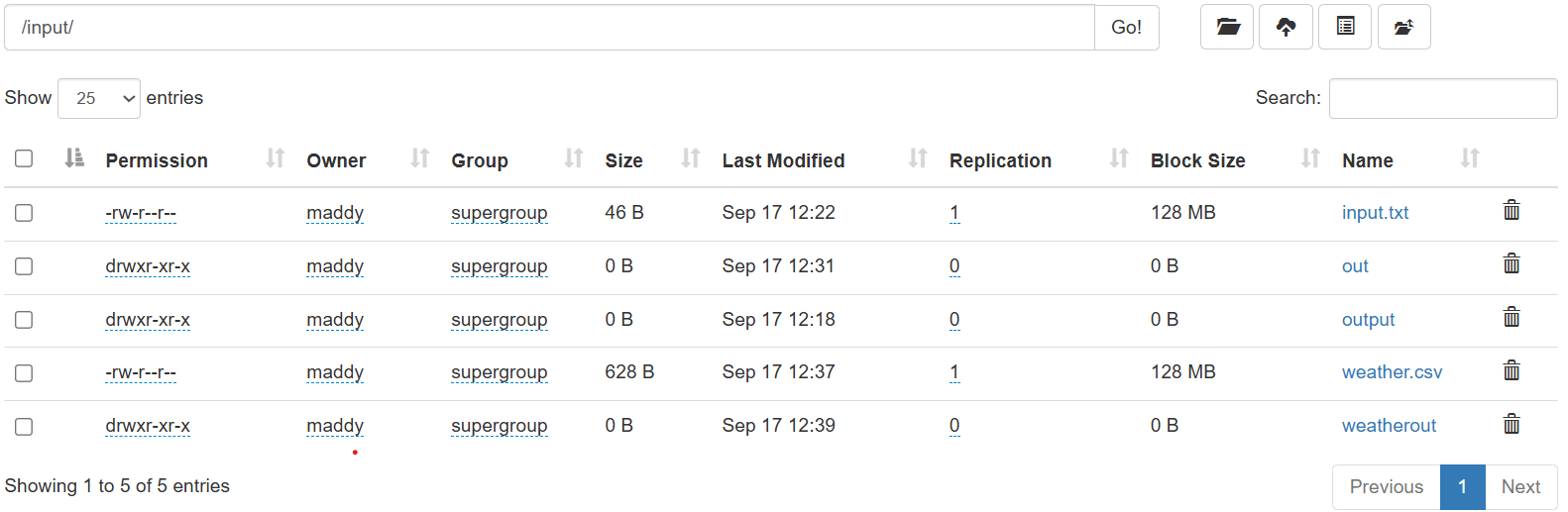
**EX.NO.3**

**210701**

**154**

**OUTPUT:**





**EX.NO.3**

**210701**

**154**

**RESULT:**

Thus

the

implementation

of

the

MapReduce

python

program

to

process

a

weather

dataset

in

Hadoop is executed successfully.