# Exp.No: 3

# Map Reduce program to process a weather dataset

## AIM:

To implement MapReduce program to process a weather dataset.

#### PROCEDURE:

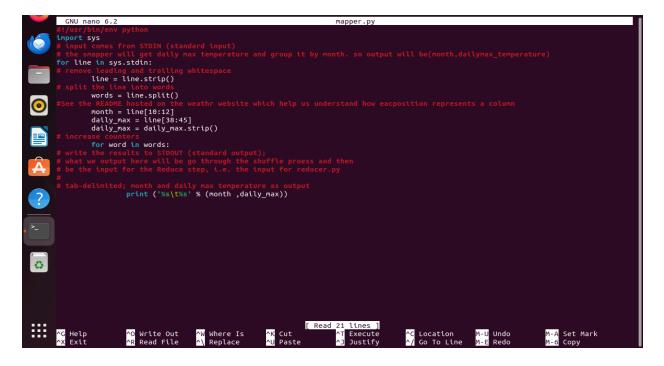
**Step 1: Create Data File:** Create a file named "word\_count\_data.txt" and populate it with text data that you wish to analyse. Login with your hadoop user.

## **Output:**

	GNU nano 6.2 dataset.txt													ı	
	26494 20240101	2.424 -147.51	64.97	-8.6	-15.7	-12.2	-13.5	0.0	0.00 C	-13.2	-22.0	-19.5	-9999.0	-9999.0	-9 <mark>&gt;</mark>
	26494 20240102	2.424 -147.51	64.97	-6.9	-10.1	-8.5	-8.3	0.0	0.01 C	-10.9	-16.6	-12.8	-9999.0	-9999.0	-9 <mark>&gt;</mark>
	26494 20240103	2.424 -147.51	64.97	10.0	-15.7	-12.9	-13.5	0.0	0.00 C	-14.4	-20.1	-17.4	-9999.0	-9999.0	- 9 <mark>&gt;</mark>
	26494 20240104	2.424 -147.51		13.6	-16.4	-15.0	-15.6	0.2	0.03 C	-14.7	-19.5		-9999.0		
	26494 20240105	2.424 -147.51		13.5	-20.5	-17.0	-17.3	0.0	0.00 C	-16.7	-25.2		-9999.0		
	26494 20240106	2.424 -147.51		10.4	-21.7	-16.0	-17.7	0.0	0.00 C	-16.1	-25.7		-9999.0		
	26494 20240107	2.424 -147.51	64.97	-2.9	-10.7	-6.8	-5.8	0.0	0.01 C	-6.2	-17.8		-9999.0		
	26494 20240108	2.424 -147.51	64.97	-4.8	-12.5	-8.6	-8.0	5.5	0.00 C	-6.2	-12.0		-9999.0		
	26494 20240109	2.424 -147.51		12.5	-20.0	-16.2	-15.8	0.2	0.00 C	-12.0	-27.6		-9999.0		
	26494 20240110	2.424 -147.51		15.2	-20.9	-18.1	-17.6	0.0	0.00 C	-16.6	-31.2		-9999.0		
	26494 20240111	2.424 -147.51		10.2	-17.1	-13.7	-13.6	0.0	0.04 C	-15.7	-25.2		-9999.0		
	26494 20240112	2.424 -147.51		14.5	-19.3	-16.9	-17.2	0.0	0.00 C	-18.4	-23.9		-9999.0		
	26494 20240113	2.424 -147.51		15.9	-20.9	-18.4	-18.5	0.0	0.01 C	-16.5	-28.2		-9999.0		
	26494 20240114	2.424 -147.51		14.9	-18.8	-16.8	-17.2	3.1	0.02 C	-14.3	-19.5		-9999.0		
	26494 20240115	2.424 -147.51	64.97	-7.6	-15.2	-11.4	-11.8	2.6	0.02 C	-7.8	-18.4		-9999.0		
	26494 20240116	2.424 -147.51	64.97	-3.8	-9.8	-6.8	-7.4	1.8	0.00 C	-4.9	-16.3		-9999.0		
	26494 20240117	2.424 -147.51	64.97	-8.7	-16.2	-12.5	-12.5	0.0	0.13 C	-14.0	-24.4		-9999.0		
<del></del>	26494 20240118	2.424 -147.51		12.2	-17.3	-14.7	-14.7	0.0	0.18 C	-11.8	-25.9		-9999.0		
	26494 20240119	2.424 -147.51		11.6	-17.4	-14.5	-15.3	0.0	0.01 C	-11.7	-24.7		-9999.0		
	26494 20240120	2.424 -147.51		14.0	-21.1	-17.5	-18.6	0.0	0.03 C	-18.4	-26.8		-9999.0		
2	26494 20240121	2.424 -147.51		19.1	-27.0	-23.1	-22.2	0.0	0.05 C	-22.8	-34.6		-9999.0		
	26494 20240122	2.424 -147.51		23.5	-29.0	-26.3	-26.4	0.0	0.05 C	-30.6	-35.8		-9999.0		
	26494 20240123	2.424 -147.51		23.5	-31.7	-27.6	-26.9	0.0	0.22 C	-28.1	-36.6		-9999.0		
	26494 20240124	2.424 -147.51		23.9	-33.6	-28.7	-28.1	0.0	0.23 C	-24.4	-37.4		-9999.0		
-<	26494 20240125	2.424 -147.51		25.9	-31.0	-28.5	-28.7	0.0	0.07 C	-30.8	-37.6		-9999.0		
	26494 20240126	2.424 -147.51		28.0	-36.9	-32.4	-33.2	0.0	0.03 C	-34.7	-42.2		-9999.0		
	26494 20240127	2.424 -147.51		36.2	-40.4	-38.3	-38.5	0.0	0.26 C	-38.7	-43.7		-9999.0		
	26494 20240128	2.424 -147.51		31.7	-38.3	-35.0	-34.9	1.4	0.12 C	-28.5	-43.3		-9999.0		
	26494 20240129	2.424 -147.51		29.8	-33.7	-31.7	-31.9	0.7	0.00 C	-26.9	-32.5		-9999.0		
46	26494 20240130	2.424 -147.51		27.6	-32.6	-30.1	-29.5	0.0	0.08 C	-25.4	-37.5		-9999.0		
	26494 20240131	2.424 -147.51		27.7	-32.6	-30.1	-30.2	0.0	0.15 C	-29.2	-40.1		-9999.0		
	26494 20240201	2.424 -147.51		31.4	-36.6	-34.0	-34.1	0.0	0.24 C	-36.4	-41.8		-9999.0		
	26494 20240202	2.424 -147.51		33.8	-38.8	-36.3	-36.1	0.0	0.33 C	-36.9	-45.1		-9999.0		
<i>y</i>	26494 20240203	2.424 -147.51						-9999.0	-9999.00 U					-9999.0	
	26494 20240204	2.424 -147.51							-9999.00 U				-9999.0		
	26494 20240205	2.424 -147.51	64.97	16.1	-20.2	-18.1	-17.9	0.0	0.65 C	-18.2	-25.1	-22.2	-9999.0	-9999.0	-9>
:::	^G Help ^X Exit	^O Write Out ^R Read File	^W Where		^K Cut ^U Pas		d 239 li ^T Exe ^J Jus		^C Locat ^/ Go To		M-U Undo M-E Redo		I-A Set M I-6 Copy	ark	

**Step 2: Mapper Logic - mapper.py:** Create a file named "mapper.py" to implement the logic for the mapper. The mapper will read input data from STDIN, split lines into words, and output each word with its count.

nano mapper.py



**Step 3: Reducer Logic - reducer.py:** Create a file named "reducer.py" to implement the logic for the reducer. The reducer will aggregate the occurrences of each word and generate the final output.

# nano reducer.py

```
GNU nano 6.2

# //ws/pen/ferr python

for reys

freducer of import itengetter

tener sys

freducer of import itengetter

freducer logic will get all the daily max temperature for a month and find max temperature of the nonth

## will ensure that key are sorted(month)

## will ensure that key are sorted(month)

## prive the imput we got from spper.py

## parse the imput we got from apper.py

## month, daily_max = iloas(daily_max)

## weept ValueTor:

## daily_max = float(daily_max)

## weept ValueTor:

## daily_max = float(daily_max)

## weept ValueTor:

## daily_max = float(daily_max)

## import discard that line

## continue

## this IF-switch only works because Hadoop shuffle process sorts map output

## furrent_month == month:

## furrent_month == month:

## furrent_month == month:

## current_max = daily_max

## current_max = daily_max

## current_month == month:

## current_month
```

**Step 4: Prepare Hadoop Environment:** Start the Hadoop daemons and create a directory in HDFS to store your data.

#### start-all.sh

**Step 6: Make Python Files Executable:** Give executable permissions to your mapper.py and reducer.py files.

### chmod 777 mapper.py reducer.py

**Step 7: Run the program using Hadoop Streaming:** Download the latest hadoop-streaming jar file and place it in a location you can easily access.

Then run the program using Hadoop Streaming.

hadoop fs -mkdir -p /weatherdata hadoop fs -copyFromLocal /home/sx/Downloads/dataset.txt /weatherdata

hdfs dfs -ls /weatherdata hadoop jar

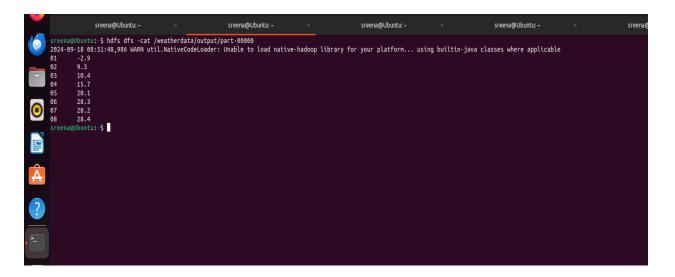
/home/sx/hadoop-3.2.3/share/hadoop/tools/lib/hadoop-streaming-3.2.3.jar  $\$  -input /weatherdata/dataset.txt  $\$  -output /weatherdata/output  $\$  -file

"/home/sx/Downloads/mapper.py" \ -mapper "python3 mapper.py" \ -file

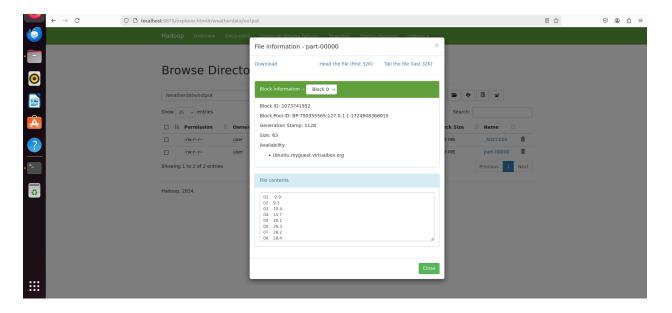
hdfs dfs -text /weatherdata/output/\* > /home/vboxuser/Downloads/outputfile.txt

**Step 8: Check Output:** Check the output of the program in the specified HDFS output directory.

<sup>&</sup>quot;/home/sx/Downloads/reducer.py" \ -reducer "python3 reducer.py"



**Step 9:** The result in the browser is as follows:



## **RESULT:**

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