

Implement a MapReduce program to process a weather dataset

Aim:

To Implement a MapReduce program to process a weather dataset

Procedure:

1. Open the terminal and start Hadoop using start-all.sh command
2. Open the browser and go to the URL localhost:9870.
3. In the terminal using the command `hadoop fs -mkdir /user` create a directory called user.
4. Upload the sample_weather.txt file to hdfs using the command `hadoop fs -put sample_weather.txt /user`.

Then perform the mapreduce operation using the command

```
hadoop jar /path/to/hadoop-streaming.jar \  
-files /path/to/mapper.py, /path/to/reducer.py \ -input /path/to/input \  
-output /path/to/output \  
-mapper mapper.py \  
-reducer reducer.py
```

5. Check the output using the command `hadoop fs -cat /user/output/part-00000`.

Output:

```
jesper — -zsh — 78x15
[jesper@j ~ % start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as jesper in 10 seconds
.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [0.0.0.0]
Starting datanodes
Starting secondary namenodes [j.local]
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further deta
ils.
Starting resourcemanager
Starting nodemanagers
jesper@j ~ %
```

localhost

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse Directory

/ Go!

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	jesper	supergroup	0 B	Aug 19 10:46	0	0 B	ex-2
drwxr-xr-x	jesper	supergroup	0 B	Aug 23 08:50	0	0 B	ex-3
drwxr-xr-x	jesper	supergroup	0 B	Aug 23 09:03	0	0 B	ex-3_test
drwxr-xr-x	jesper	supergroup	0 B	Aug 16 17:15	0	0 B	samp
drwxr-xr-x	jesper	supergroup	0 B	Aug 23 09:44	0	0 B	three
drwxr-xr-x	jesper	supergroup	0 B	Aug 23 10:03	0	0 B	tmp
drwxr-xr-x	jesper	supergroup	0 B	Aug 29 10:21	0	0 B	user

Showing 1 to 7 of 7 entries

Previous 1 Next

Hadoop, 2024.

```
jesper — -zsh — 78x15
[jesper@j ~ % hadoop fs -mkdir /ex-1
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further deta
ils.
jesper@j ~ %
```

```
jesper — -zsh — 92x15
jesper@j ~ % hadoop fs -put /Users/jesper/Documents/sem-7/DA/lab/ex-1/input.txt /ex-1
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
jesper@j ~ %
```

```
jesper — -zsh — 126x22
jesper@j ~ % hadoop jar /Users/jesper/hadoop-3.4.0/share/hadoop/tools/lib/hadoop-streaming-3.4.0.jar\
-files /Users/jesper/Documents/sem-7/DA/lab/ex-1/mapper.py,/Users/jesper/Documents/sem-7/DA/lab/ex-1/reducer.py\
-input /ex-1/input.txt\
-output /ex-1/output\
-mapper mapper.py\
-reducer reducer.py
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
packageJobJar: [/var/folders/l9/jpyrrxs5hd30rmbdjy22q4w0000gn/T/hadoop-unjar73079381717837866/] [] /var/folders/l9/jpyrrxs5h
d30rmbdjy22q4w0000gn/T/streamjob4036673018412328904.jar tmpDir=null
jesper@j ~ %
```

```
jesper — -zsh — 99x14
jesper@j ~ % hadoop fs -cat /ex-2/output/part-00000
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
690190_200602_section1 53.87166666666666 25.899999999999995 7.774999999999998
690190_200602_section2 54.761250000000001 25.900000000000006 7.774999999999999
690190_200602_section3 53.250416666666667 25.899999999999995 7.774999999999996
690190_200602_section4 52.44708333333333 25.900000000000006 7.774999999999999
jesper@j ~ %
```

Result:

Thus the Installation, Configuration and run Hadoop and HDFS is successfully executed.