SURUTHI S

225229141

CREATING INPUT DIRECTORIES IN HDFS

```
[cloudera@quickstart ~]$ hdfs dfs -mkdir py_mr
[cloudera@quickstart ~]$ hdfs dfs -mkdir py_mr/input
[cloudera@quickstart ~]$  hdfs dfs -copyFromLocal /home/cloudera/Desktop/cloudera_f
[cloudera@quickstart ~]$ hdfs dfs -ls py_mr/input/
Found 1 items
-rw-r--r--   1 cloudera cloudera      12053 2023-08-09 10:59 py_mr/input/sample_wea
[cloudera@quickstart ~]$
```

2. HADOOP COMMAND FOR RUNNING MAP AND REDUCE FILE

hadoop jar /usr/lib/hadoop-0.20-mapreduce/contrib/streaming/hadoop-streaming-2.6.0-mr1-cdh5.13.0.jar \

   -files /home/cloudera/Desktop/cloudera_fs/python/max_temperature_map.py,/home/cloudera/Desktop/cloudera_fs/python/max_temperature_reduce.py \

   -mapper "python max_temperature_map.py" \

   -reducer "python max_temperature_reduce.py" \

   -input /weather_py/in00/sample_weather.txt \

   -output /weather_py/out00

3. GIVING PERMISSIONS

chmod +x /home/cloudera/Desktop/cloudera_fs/python/max_temperature_map.py

chmod +x /home/cloudera/Desktop/cloudera_fs/python/max_temperature_reduce.py

PROCESS:

```
[cloudera@quickstart ~]$ hadoop jar /usr/lib/hadoop-0.20-mapreduce/contrib/streamin
g/hadoop-streaming-2.6.0-mr1-cdh5.13.0.jar \
>      -files /home/cloudera/Desktop/cloudera_fs/python/max_temperature_map.py,/home
/cloudera/Desktop/cloudera_fs/python/max_temperature_reduce.py \
>      -mapper "python max_temperature_map.py" \
>      -reducer "python max_temperature_reduce.py" \
>      -input py_mr/input \
>      -output py_mr/output
packageJobJar: [] [/usr/lib/hadoop-mapreduce/hadoop-streaming-2.6.0-cdh5.13.0.jar]
/tmp/streamjob8366516004575266989.jar tmpDir=null
23/08/09 11:03:33 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:80
32
23/08/09 11:03:33 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:80
32
23/08/09 11:03:34 INFO mapred.FileInputFormat: Total input paths to process : 1
23/08/09 11:03:35 INFO mapreduce.JobSubmitter: number of splits:2
23/08/09 11:03:36 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_16915
62778358_0007
23/08/09 11:03:36 INFO impl.YarnClientImpl: Submitted application application_16915
62778358_0007
23/08/09 11:03:36 INFO mapreduce.Job: The url to track the job: http://quickstart.c
loudera:8088/proxy/application_1691562778358_0007/
```

```
23/08/09 11:03:34 INFO mapred.FileInputFormat: Total input paths to process
23/08/09 11:03:35 INFO mapreduce.JobSubmitter: number of splits:2
23/08/09 11:03:36 INFO mapreduce.JobSubmitter: Submitting tokens for job: j
62778358_0007
23/08/09 11:03:36 INFO impl.YarnClientImpl: Submitted application applicati
62778358_0007
23/08/09 11:03:36 INFO mapreduce.Job: The url to track the job: http://quic
loudera:8088/proxy/application_1691562778358_0007/
23/08/09 11:03:36 INFO mapreduce.Job: Running job: job_1691562778358_0007
23/08/09 11:03:46 INFO mapreduce.Job: Job job_1691562778358_0007 running ir
de : false
23/08/09 11:03:46 INFO mapreduce.Job:  map 0% reduce 0%
23/08/09 11:04:05 INFO mapreduce.Job:  map 50% reduce 0%
23/08/09 11:04:07 INFO mapreduce.Job:  map 100% reduce 0%
```

OUTPUT:

```
[cloudera@quickstart ~]$ hdfs dfs -ls py_mr/output/
Found 2 items
-rw-r--r--   1 cloudera cloudera          0 2023-08-09 11:04 py_mr/output/_SUCCESS
-rw-r--r--   1 cloudera cloudera         10 2023-08-09 11:04 py_mr/output/part-0000
0
[cloudera@quickstart ~]$ █
```

```
0
[cloudera@quickstart ~]$ hdfs dfs -cat py_mr/output/part-00000
2006    22.0
[cloudera@quickstart ~]$ S█
```