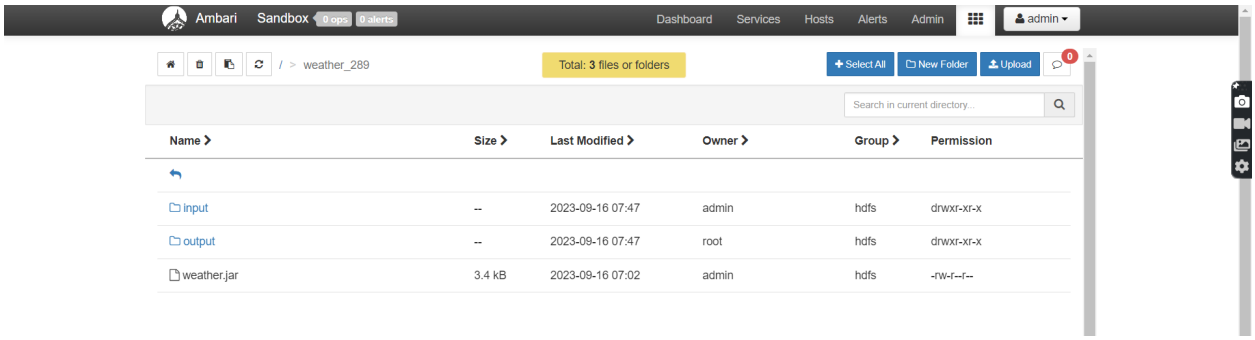
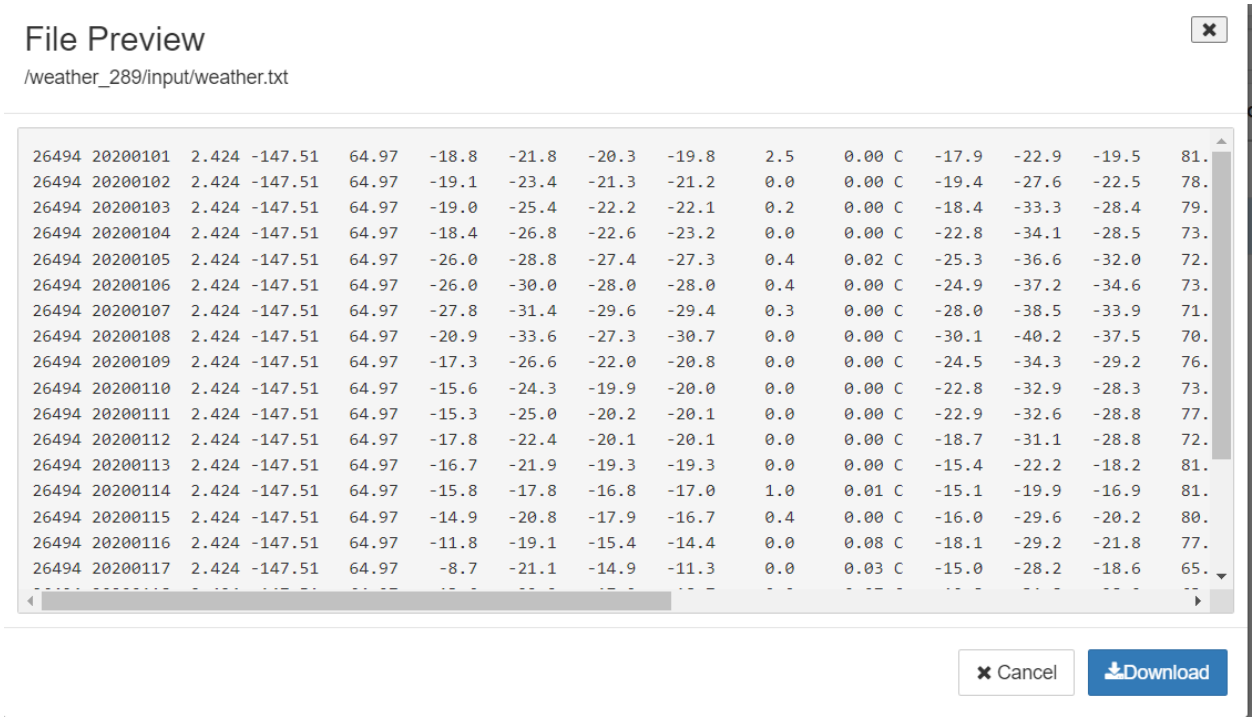
Weather data analysis for analyzing hot and cold days using MapReduce.

Folder:



Input data File Preview:



Sandbox Execution:

```
[root@sandbox ~]# hadoop fs -get /weather_289
[root@sandbox ~]# hadoop jar weather.jar /weather_289/input /weather_289/output
23/09/16 02:22:55 INFO impl.TimelineClientImpl: Timeline service address: http://sandbox.hortonworks.com:8188/ws/v1/timeline/
23/09/16 02:22:55 INFO client.RMProxy: Connecting to ResourceManager at sandbox.hortonworks.com/172.17.0.2:8050
23/09/16 02:22:55 INFO client.AHSProxy: Connecting to Application History server at sandbox.hortonworks.com/172.17.0.2:10200
23/09/16 02:22:56 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with T
oolRunner to remedy this.
23/09/16 02:22:56 INFO input.FileInputFormat: Total input paths to process : 1
23/09/16 02:22:56 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library
23/09/16 02:22:56 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo library [hadoop-lzo rev 7a4b57bedce694048432dd5bf5b90a6c8ccdba80]
23/09/16 02:22:57 INFO mapreduce.JobSubmitter: number of splits:1
23/09/16 02:22:58 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1694516065863_0004
23/09/16 02:22:58 INFO impl.YarnClientImpl: Submitted application application_1694516065863_0004
23/09/16 02:22:58 INFO mapreduce.Job: The url to track the job: http://sandbox.hortonworks.com:8088/proxy/application_1694516065863_0004/
23/09/16 02:22:58 INFO mapreduce.Job: Running job: job_1694516065863_0004
23/09/16 02:23:06 INFO mapreduce.Job: Job job_1694516065863_0004 running in uber mode : false
23/09/16 02:23:06 INFO mapreduce.Job:  map 0% reduce 0%
23/09/16 02:23:14 INFO mapreduce.Job:  map 100% reduce 0%
23/09/16 02:23:20 INFO mapreduce.Job:  map 100% reduce 100%
23/09/16 02:23:21 INFO mapreduce.Job: Job job_1694516065863_0004 completed successfully
23/09/16 02:23:21 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=13533
                FILE: Number of bytes written=316261
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=79916
                HDFS: Number of bytes written=12797
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=6457
                Total time spent by all reduces in occupied slots (ms)=2837
                Total time spent by all map tasks (ms)=6457
```

```
                Total megabyte-milliseconds taken by all reduce tasks=709250
        Map-Reduce Framework
                Map input records=366
                Map output records=365
                Map output bytes=12797
                Map output materialized bytes=13533
                Input split bytes=130
                Combine input records=0
                Combine output records=0
                Reduce input groups=365
                Reduce shuffle bytes=13533
                Reduce input records=365
                Reduce output records=365
                Spilled Records=730
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=146
                CPU time spent (ms)=1020
                Physical memory (bytes) snapshot=296239104
                Virtual memory (bytes) snapshot=3893121024
                Total committed heap usage (bytes)=135266304
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=79786
        File Output Format Counters
                Bytes Written=12797
[root@sandbox ~]# hadoop fs -cat /weather_289/output/part-r-00000
The Day is Cold Day :20200101   -21.8
The Day is Cold Day :20200102   -23.4
```

Output:

```
[root@sandbox ~]# hadoop fs -cat /weather_289/output/part-r-00000
The Day is Cold Day :20200101   -21.8
The Day is Cold Day :20200102   -23.4
The Day is Cold Day :20200103   -25.4
The Day is Cold Day :20200104   -26.8
The Day is Cold Day :20200105   -28.8
The Day is Cold Day :20200106   -30.0
The Day is Cold Day :20200107   -31.4
The Day is Cold Day :20200108   -33.6
The Day is Cold Day :20200109   -26.6
The Day is Cold Day :20200110   -24.3
The Day is Cold Day :20200111   -25.0
The Day is Cold Day :20200112   -22.4
The Day is Cold Day :20200113   -21.9
The Day is Cold Day :20200114   -17.8
The Day is Cold Day :20200115   -20.8
The Day is Cold Day :20200116   -19.1
The Day is Cold Day :20200117   -21.1
The Day is Cold Day :20200118   -22.2
The Day is Cold Day :20200119   -22.8
The Day is Cold Day :20200120   -24.0
The Day is Cold Day :20200121   -26.3
The Day is Cold Day :20200122   -23.3
The Day is Cold Day :20200123   -25.1
The Day is Cold Day :20200124   -24.4
The Day is Cold Day :20200125   -25.7
The Day is Cold Day :20200126   -29.2
The Day is Cold Day :20200127   -31.7
The Day is Cold Day :20200128   -30.5
The Day is Cold Day :20200129   -28.8
The Day is Cold Day :20200130   -28.0
The Day is Cold Day :20200131   -29.8
The Day is Cold Day :20200201   -31.8
The Day is Cold Day :20200202   -30.7
The Day is Cold Day :20200203   -13.3
The Day is Cold Day :20200204   -12.9
The Day is Cold Day :20200205   -12.8
The Day is Cold Day :20200206   -9.3
```

```
The Day is Cold Day :20201223   -9.5
The Day is Cold Day :20201224   -8.4
The Day is Cold Day :20201225   -6.5
The Day is Cold Day :20201226   -14.1
The Day is Cold Day :20201227   -13.8
The Day is Cold Day :20201228   -2.5
The Day is Cold Day :20201229   -8.2
The Day is Cold Day :20201230   -13.8
The Day is Cold Day :20201231   -15.5
The Day is Hot Day :20201104    9999.0
The Day is Hot Day :20201105    9999.0
The Day is Hot Day :20201106    9999.0
The Day is Hot Day :20201107    9999.0
The Day is Hot Day :20201108    9999.0
The Day is Hot Day :20201109    9999.0
The Day is Hot Day :20201110    9999.0
The Day is Hot Day :20201111    9999.0
[root@sandbox ~]#
```

Output file preview:

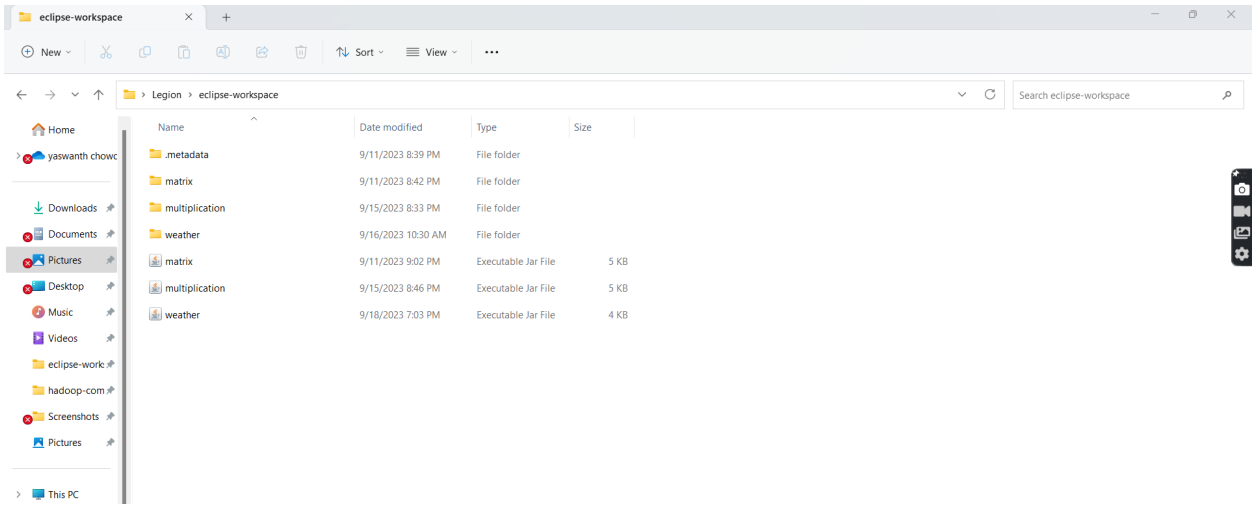
File Preview

/weather_289/output/part-r-00000

```
The Day is Cold Day :20200101   -21.8
The Day is Cold Day :20200102   -23.4
The Day is Cold Day :20200103   -25.4
The Day is Cold Day :20200104   -26.8
The Day is Cold Day :20200105   -28.8
The Day is Cold Day :20200106   -30.0
The Day is Cold Day :20200107   -31.4
The Day is Cold Day :20200108   -33.6
The Day is Cold Day :20200109   -26.6
The Day is Cold Day :20200110   -24.3
The Day is Cold Day :20200111   -25.0
The Day is Cold Day :20200112   -22.4
The Day is Cold Day :20200113   -21.9
The Day is Cold Day :20200114   -17.8
The Day is Cold Day :20200115   -20.8
The Day is Cold Day :20200116   -19.1
The Day is Cold Day :20200117   -21.1
The Day is Cold Day :20200118   -22.2
```

Cancel Download

Jar File path location:



AP20110010289,
T. YASWANTH CHOWDARY.