The Course Project

The course project includes 4 parts.

Part 1:

The first part is to develop a Mapper and Reducer application to calculate average visibility distance (meters) for each USAF weather station ID from NCDC records (note: 999999 indicates missing value, and [01459] indicate good quality value).

Command to execute python code with Hadoop streaming:

[student57@msba-hadoop-name Project]$ hadoop jar hadoop-streaming-2.7.3.jar -file /home/student57/Project/mapper1.py -mapper /home/student57/Project/mapper1.py -file /home/student57/Project/reducer1.py -reducer /home/student57/Project/reducer1.py -input /home/57student57/Project/Data/ -output /home/57student57/Project/Part1_Output/

MapReduce job execution:

```
[student57@msba-hadoop-name Project]$ hadoop jar hadoop-streaming-2.7.3.jar -file /home/student57/Project/mapper1.py    -mapper /home/student57/Project/mapper1.py -file /home/student57/Project/reducer1.py
   -reducer /home/student57/Project/reducer1.py -input /home/57student57/Project/Data/ -output /home/57student57/Project/Part1_Output/
22/11/30 10:47:20 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [/home/student57/Project/mapper1.py, /home/student57/Project/reducer1.py, /tmp/hadoop-unjar4770182122590039O1/] [] /tmp/streamjob8638484016762179173.jar tmpDir=null
22/11/30 10:47:21 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
22/11/30 10:47:21 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
22/11/30 10:47:22 INFO mapred.FileInputFormat: Total input files to process : 50
22/11/30 10:47:22 INFO mapreduce.JobSubmitter: number of splits:50
22/11/30 10:47:22 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
22/11/30 10:47:22 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1669743171306_0237
22/11/30 10:47:22 INFO impl.YarnClientImpl: Submitted application application_1669743171306_0237
22/11/30 10:47:22 INFO mapreduce.Job: The url to track the job: http://msba-hadoop-name:8088/proxy/application_1669743171306_0237/
22/11/30 10:47:22 INFO mapreduce.Job: Running job: job_1669743171306_0237
22/11/30 10:47:28 INFO mapreduce.Job: Job job_1669743171306_0237 running in uber mode : false
22/11/30 10:47:28 INFO mapreduce.Job:  map 0% reduce 0%


22/11/30 10:48:32 INFO mapreduce.Job:  map 94% reduce 28%
22/11/30 10:48:34 INFO mapreduce.Job:  map 98% reduce 28%
22/11/30 10:48:35 INFO mapreduce.Job:  map 100% reduce 33%
22/11/30 10:48:36 INFO mapreduce.Job:  map 100% reduce 100%
22/11/30 10:48:36 INFO mapreduce.Job: Job job_1669743171306_0237 completed successfully
22/11/30 10:48:36 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=254841
                FILE: Number of bytes written=10981332
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=5134180
                HDFS: Number of bytes written=173
                HDFS: Number of read operations=153
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=50
                Launched reduce tasks=1
                Data-local map tasks=50
                Total time spent by all maps in occupied slots (ms)=277980
                Total time spent by all reduces in occupied slots (ms)=48627
                Total time spent by all map tasks (ms)=277980
                Total time spent by all reduce tasks (ms)=48627
                Total vcore-milliseconds taken by all map tasks=277980
                Total vcore-milliseconds taken by all reduce tasks=48627
                Total megabyte-milliseconds taken by all map tasks=284651520
                Total megabyte-milliseconds taken by all reduce tasks=49794048
        Map-Reduce Framework
                Map input records=36404
                Map output records=50
                Map output bytes=254635
                Map output materialized bytes=255135
                Input split bytes=6400
                Combine input records=0
                Combine output records=0
                Reduce input groups=40
                Reduce shuffle bytes=255135
                Reduce input records=50
                Reduce output records=15
                Spilled Records=100
                Shuffled Maps =50
                Failed Shuffles=0
                Merged Map outputs=50
                GC time elapsed (ms)=9915
                CPU time spent (ms)=30940
                Physical memory (bytes) snapshot=17856946176
                Virtual memory (bytes) snapshot=154103181312
                Total committed heap usage (bytes)=13430161408
        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=5127780
        File Output Format Counters
                Bytes Written=173
22/11/30 10:48:36 INFO streaming.StreamJob: Output directory: /home/57student57/Project/Part1_Output/
[student57@msba-hadoop-name Project]$
```

Output:

```
[student57@msba-hadoop-name Project]$ hdfs dfs -ls /home/57student57/Project/
Found 2 items
drwxr-xr-x   - student57 supergroup          0 2022-11-30 00:02 /home/57student57/Project/Data
drwxr-xr-x   - student57 supergroup          0 2022-11-30 10:48 /home/57student57/Project/Part1_Output
[student57@msba-hadoop-name Project]$ hdfs dfs -ls /home/57student57/Project/Part1_Output/
Found 2 items
-rw-r--r--   5 student57 supergroup          0 2022-11-30 10:48 /home/57student57/Project/Part1_Output/_SUCCESS
-rw-r--r--   5 student57 supergroup        173 2022-11-30 10:48 /home/57student57/Project/Part1_Output/part-00000
[student57@msba-hadoop-name Project]$ hdfs dfs -cat /home/57student57/Project/Part1_Output/part-00000
014270  17137
012620  26542
038040  14158
030910  11362
034970  5803
023610  37068
029110  0
029350  0
028970  0
033020  12318
028360  0
014030  33686
032620  8316
011060  24848
029700  0
[student57@msba-hadoop-name Project]$
```

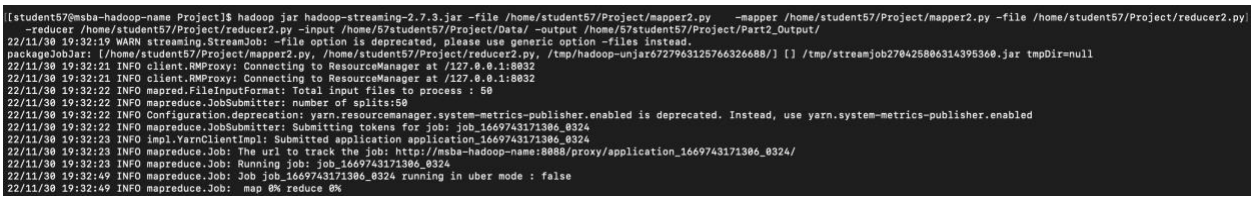
Part 2:

The second part is to develop a Mapper and Reducer application to retrieve USAF weather station ID and sky ceiling height (meters) from NCDC records (note: 99999 indicates missing value, and [01459] indicate good quality value) and then write the USAF weather station ID and sky ceiling height data into a text file.

Command to execute python code with Hadoop streaming:

[student57@msba-hadoop-name Project]$ hadoop jar hadoop-streaming-2.7.3.jar -file /home/student57/Project/mapper2.py -mapper /home/student57/Project/mapper2.py -file /home/student57/Project/reducer2.py -reducer /home/student57/Project/reducer2.py -input /home/57student57/Project/Data/ -output /home/57student57/Project/Part2_Output/

MapReduce job execution:



```
22/11/30 19:34:07 INFO mapreduce.Job:  map 98% reduce 29%
22/11/30 19:34:08 INFO mapreduce.Job:  map 98% reduce 33%
22/11/30 19:34:09 INFO mapreduce.Job:  map 100% reduce 33%
22/11/30 19:34:10 INFO mapreduce.Job:  map 100% reduce 100%
22/11/30 19:34:10 INFO mapreduce.Job: Job job_1669743171306_0324 completed successfully
22/11/30 19:34:10 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=54582
                FILE: Number of bytes written=10580763
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=5134180
                HDFS: Number of bytes written=1287
                HDFS: Number of read operations=153
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=50
                Launched reduce tasks=1
                Data-local map tasks=50
                Total time spent by all maps in occupied slots (ms)=346422
                Total time spent by all reduces in occupied slots (ms)=51016
                Total time spent by all map tasks (ms)=346422
                Total time spent by all reduce tasks (ms)=51016
                Total vcore-milliseconds taken by all map tasks=346422
                Total vcore-milliseconds taken by all reduce tasks=51016
                Total megabyte-milliseconds taken by all map tasks=354736128
                Total megabyte-milliseconds taken by all reduce tasks=52240384
        Map-Reduce Framework
                Map input records=36404
                Map output records=3411
                Map output bytes=47754
                Map output materialized bytes=54876
                Input split bytes=6400
                Combine input records=0
                Combine output records=0
                Reduce input groups=99
                Reduce shuffle bytes=54876
                Reduce input records=3411
                Reduce output records=99
                Spilled Records=6822
                Shuffled Maps =50
                Failed Shuffles=0
                Merged Map outputs=50
                GC time elapsed (ms)=12609
                CPU time spent (ms)=34640
                Physical memory (bytes) snapshot=17942900736
                Virtual memory (bytes) snapshot=154233430016
                Total committed heap usage (bytes)=14994636800
        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=5127780
        File Output Format Counters
                Bytes Written=1287
22/11/30 19:34:10 INFO streaming.StreamJob: Output directory: /home/57student57/Project/Part2_Output/
[student57@msba-hadoop-name Project]$
```

Output:

```
[[student57@msba-hadoop-name Project]$ hdfs dfs -ls /home/57student57/Project/
Found 3 items
drwxr-xr-x   - student57 supergroup          0 2022-11-30 00:02 /home/57student57/Project/Data
drwxr-xr-x   - student57 supergroup          0 2022-11-30 19:23 /home/57student57/Project/Part1_Output
drwxr-xr-x   - student57 supergroup          0 2022-11-30 19:34 /home/57student57/Project/Part2_Output
[[student57@msba-hadoop-name Project]$ hdfs dfs -ls /home/57student57/Project/Part2_Output/
Found 2 items
-rw-r--r--   5 student57 supergroup          0 2022-11-30 19:34 /home/57student57/Project/Part2_Output/_SUCCESS
-rw-r--r--   5 student57 supergroup       1287 2022-11-30 19:34 /home/57student57/Project/Part2_Output/part-00000
[[student57@msba-hadoop-name Project]$ hdfs dfs -cat /home/57student57/Project/Part2_Output/part-00000
014270  00780
014270  01500
014270  03000
014270  07500
014270  00150
014270  00450
014270  00240
014270  00015
014270  01230
014270  22000
012620  01500
012620  03000
012620  07500
012620  00150
012620  00060
012620  00450
012620  00240
012620  00015
012620  01230
012620  22000
038040  00780
038040  09000
038040  00150
038040  00060
038040  00450
038040  03600
038040  00240
038040  01230
038040  22000
030910  00780
030910  01500
030910  03000
030910  07500
030910  00150
030910  02400
030910  00450
030910  00240
030910  00015
030910  01230
030910  22000
034970  00780
034970  01500
034970  03000
034970  07500
034970  00150
034970  00060
034970  00450
034970  00240
034970  00015
034970  01230
034970  22000
023610  00780
```

```
023610   02400
023610   00150
023610   00060
023610   00450
023610   00240
023610   00015
023610   01230
023610   22000
033020   00780
033020   01500
033020   03000
033020   07500
033020   00150
033020   02100
033020   00060
033020   00450
033020   00240
033020   00015
033020   01230
033020   22000
014030   00780
014030   01500
014030   03000
014030   07500
014030   00150
014030   02100
014030   00450
014030   00240
014030   00015
014030   01230
014030   22000
032620   03000
032620   07500
032620   00450
032620   00240
032620   01230
032620   22000
011060   00780
011060   01500
011060   03000
011060   07500
011060   00060
011060   00450
011060   00240
011060   00015
011060   01230
011060   22000
[student57@msba-hadoop-name Project]$
```

Converting the map-reduce output to a text file:

```
[[student57@msba-hadoop-name Project]$ mkdir Part2_Output
```

```
[student57@msba-hadoop-name Project]$ ls
Data  hadoop-streaming-2.7.3.jar  mapper1.py  mapper2.py  Part1_Output  Part2_Output  reducer1.py  reducer2.py
[student57@msba-hadoop-name Project]$ cd Part2_Output
[student57@msba-hadoop-name Part2_Output]$ hdfs dfs -copyToLocal /home/57student57/Project/Part2_Output/* /home/student57/Project/Part2_Output/
[student57@msba-hadoop-name Part2_Output]$ ls
part-00000  _SUCCESS
[student57@msba-hadoop-name Part2_Output]$ cat part-00000 > station_sky_ceiling_height.txt
[student57@msba-hadoop-name Part2_Output]$ ls
part-00000  station_sky_ceiling_height.txt  _SUCCESS
[student57@msba-hadoop-name Part2_Output]$ cat station_sky_ceiling_height.txt | head -20
014270  00780
014270  01500
014270  03000
014270  07500
014270  00150
014270  00450
014270  00240
014270  00015
014270  01230
014270  22000
012620  01500
012620  03000
012620  07500
012620  00150
012620  00060
012620  00450
012620  00240
012620  00015
012620  01230
012620  22000
[student57@msba-hadoop-name Part2_Output]$
```

(Note: There are duplicate combinations of station_id and sky_ceiling_height are present when we consider records from all data files, only unique combinations of are selected from all the files. As keeping duplicates will give wrong average sky_ceiling_height)

Part 3:

The third part is to load the text file into Pig and get the average sky ceiling height for each USAF weather station ID.

Loading data into Pig:

grunt> station_height = LOAD 'Part2_Output/station_sky_ceiling_height.txt' AS(station:chararray, height:int);

grunt> DUMP station_height;

```
[[student57@msba-hadoop-name Project]$ pig -x local
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.9.0/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hbase-1.4.9/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
22/11/30 22:14:58 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
22/11/30 22:14:58 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2022-11-30 22:14:58,625 [main] INFO  org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2022-11-30 22:14:58,625 [main] INFO  org.apache.pig.Main - Logging error messages to: /home/student57/Project/pig_1669875298624.log
2022-11-30 22:14:58,644 [main] INFO  org.apache.pig.impl.util.Utils - Default bootup file /home/student57/.pigbootup not found
2022-11-30 22:14:58,780 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-11-30 22:14:58,782 [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
2022-11-30 22:14:58,963 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-30 22:14:58,979 [main] INFO  org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-d2fc4415-dbcd-46e2-a17b-426dcc8ca113
2022-11-30 22:14:58,979 [main] WARN  org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
[grunt> station_height = LOAD 'Part2_Output/station_sky_ceiling_height.txt' AS(station:chararray, height:int);
2022-11-30 22:17:40,591 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
grunt> DUMP station_height;
```

```
2022-11-30 22:21:01,971 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-11-30 22:21:01,973 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-11-30 22:21:01,973 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-11-30 22:21:01,979 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-11-30 22:21:01,981 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-30 22:21:01,981 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-11-30 22:21:01,993 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-11-30 22:21:01,993 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(014270,780)
(014270,1500)
(014270,3000)
(014270,7500)
(014270,150)
(014270,450)
(014270,240)
(014270,15)
(014270,1230)
(014270,22000)
(012620,1500)
(012620,3000)
```

Grouping records by Station:

grunt> st_grouped_height = GROUP station_height BY station;

grunt> DUMP st_grouped_height;

```
[grunt> st_grouped_height = GROUP station_height BY station;
[grunt> DUMP st_grouped_height;
```

```
2022-11-30 22:32:25,040 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-11-30 22:32:25,041 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-11-30 22:32:25,042 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-11-30 22:32:25,046 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-11-30 22:32:25,046 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-30 22:32:25,047 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-11-30 22:32:25,056 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-11-30 22:32:25,056 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(011060,{(011060,1230),(011060,15),(011060,240),(011060,450),(011060,60),(011060,22000),(011060,7500),(011060,3000),(011060,1500),(011060,780)})
(012620,{(012620,450),(012620,1500),(012620,3000),(012620,7500),(012620,150),(012620,60),(012620,240),(012620,15),(012620,1230),(012620,22000)})
(014030,{(014030,450),(014030,2100),(014030,150),(014030,7500),(014030,3000),(014030,1500),(014030,780),(014030,22000),(014030,1230),(014030,15),(014030,240)})
(014270,{(014270,780),(014270,22000),(014270,1230),(014270,15),(014270,240),(014270,450),(014270,150),(014270,7500),(014270,3000),(014270,1500)})
(023610,{(023610,1230),(023610,240),(023610,450),(023610,60),(023610,150),(023610,2400),(023610,780),(023610,15),(023610,22000)})
(030910,{(030910,15),(030910,240),(030910,450),(030910,2400),(030910,150),(030910,7500),(030910,3000),(030910,780),(030910,1500),(030910,22000),(030910,1230)})
(032620,{(032620,22000),(032620,3000),(032620,7500),(032620,450),(032620,240),(032620,1230)})
(033020,{(033020,22000),(033020,1230),(033020,15),(033020,780),(033020,1500),(033020,3000),(033020,7500),(033020,450),(033020,150),(033020,2100),(033020,60),(033020,240)})
(034970,{(034970,15),(034970,240),(034970,450),(034970,60),(034970,150),(034970,7500),(034970,3000),(034970,1500),(034970,780),(034970,1230),(034970,22000)})
(038040,{(038040,9000),(038040,150),(038040,60),(038040,3600),(038040,240),(038040,1230),(038040,22000),(038040,780),(038040,450)})
grunt>
```

Calculating average height for each station:

grunt> st_avg_height = FOREACH st_grouped_height GENERATE group, AVG(station_height.height);

grunt> DUMP st_avg_height;

```
grunt> st_avg_height = FOREACH st_grouped_height GENERATE group, AVG(station_height.height);
grunt> DUMP st_avg_height;
```

```
2022-11-30 22:39:49,795 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-11-30 22:39:49,796 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-11-30 22:39:49,797 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-11-30 22:39:49,799 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-11-30 22:39:49,800 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-30 22:39:49,800 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-11-30 22:39:49,810 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-11-30 22:39:49,810 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(011060,3677.5)
(012620,3614.5)
(014030,3542.2727272727275)
(014270,3686.5)
(023610,3036.1111111111113)
(030910,3569.5454545454545)
(032620,5736.666666666667)
(033020,3252.083333333335)
(034970,3356.818181818182)
(038040,4167.777777777777)
grunt>
```

Part 4:

The fourth part is to load the text file into Hive and get the highest and lowest sky ceiling height for each USAF weather station ID.

Creating table and loading data in HIVE:

CREATE TABLE Station_Height_57 (station STRING, height INT) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t';

LOAD DATA LOCAL INPATH 'Project/Part2_Output/station_sky_ceiling_height.txt' OVERWRITE INTO TABLE Station_Height_57;

```
[student57@msba-hadoop-name ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive-2.3.2/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.9.0/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/usr/local/hive-2.3.2/lib/hive-common-2.3.2.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> CREATE TABLE Station_Height_57 (station STRING, height INT) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t';
OK
Time taken: 10.35 seconds
hive> LOAD DATA LOCAL INPATH 'Project/Part2_Output/station_sky_ceiling_height.txt' OVERWRITE INTO TABLE Station_Height_57;
Loading data to table default.station_height_57
OK
Time taken: 0.863 seconds
hive>
```

Querying HIVE table to get the highest, lowest sky ceiling height for each weather station:

SELECT station, MAX(height), MIN(height) FROM Station_Height_57 GROUP BY station;

```
hive> SELECT station, MAX(height), MIN(height) FROM Station_Height_57 GROUP BY station;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = student57_20221130230255_aca95e12-3245-493f-b765-0ec260c9996c
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1669743171306_0352, Tracking URL = http://msba-hadoop-name:8088/proxy/application_1669743171306_0352/
Kill Command = /usr/local/hadoop/bin/hadoop job  -kill job_1669743171306_0352
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-11-30 23:03:02,456 Stage-1 map = 0%,  reduce = 0%
2022-11-30 23:03:08,888 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.93 sec
2022-11-30 23:03:13,019 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.07 sec
MapReduce Total cumulative CPU time: 4 seconds 70 msec
Ended Job = job_1669743171306_0352
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.07 sec   HDFS Read: 10189 HDFS Write: 368 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 70 msec
OK
011060  22000   15
012620  22000   15
014030  22000   15
014270  22000   15
023610  22000   15
030910  22000   15
032620  22000   240
033020  22000   15
034970  22000   15
038040  22000   60
Time taken: 19.484 seconds, Fetched: 10 row(s)
hive>
```