

The Course Project

Submitted By: Vanshika Gupta

The course project includes 3 parts. The first part is to develop a Mapper and Reducer application to retrieve Year and Temperature from original NCDC records (i.e., the dataset we are using for this class) and then write the Year and Temperature data into a text file. The second part is to load the text file into Pig and get the highest and lowest temperatures for each year. The third part is to load the text file into Hive and get the average temperature for each year. You need to turn in 1) the three java files (mapper, reducer and main), 2) the commands from converting them into a Jar file to running the Jar file in Hadoop, 3) the text file including Year and Temperature data created by you, 4) the screenshot of the text file being created, 5) the screenshot of the final Pig output showing the year and the highest and lowest temperatures, and 6) the screenshot of the final Hive output showing the year and average temperature.

The original dataset for this project is available on Blackboard.

HADOOP

Connection to the server:

```
(base) Vanshikas-MacBook-Air:~ vanshikagupta$ ssh student10@msba-hadoop-name.csueastbay.edu
student10@msba-hadoop-name.csueastbay.edu's password:
Last login: Sun Dec  5 13:27:26 2021 from 205.154.255.167
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory
[student10@msba-hadoop-name ~]$ cd project
[student10@msba-hadoop-name project]$ ls -ltr
total 22984
-rw-r--r--. 1 student10 hadoop 130868 Sep 24 18:30 commons-cli-2.0.jar
-rw-r--r--. 1 student10 hadoop 1512169 Sep 24 18:30 hadoop-mapreduce-client-core-2.6.1.jar
-rw-r--r--. 1 student10 hadoop 3318727 Sep 24 18:30 hadoop-common-2.6.1.jar
-rw-r--r--. 1 student10 hadoop 3409548 Oct 29 19:15 pig-0.11.0.jar
-rw-r--r--. 1 student10 hadoop 15137704 Nov 11 16:23 hive-exec-0.13.0.jar
-rw-r--r--. 1 student10 hadoop 717 Dec  5 15:18 YearAndTemperatureReducer.java
-rw-r--r--. 1 student10 hadoop 1123 Dec  5 15:18 YearAndTemperatureMapper.java
-rw-r--r--. 1 student10 hadoop 1167 Dec  5 15:18 YearAndTemperature.java
drwxr-xr-x. 2 student10 hadoop 4096 Dec  5 15:20 ProjectData
[student10@msba-hadoop-name project]$
```

Compiling the Mapper, Reducer and application java file:

```
javac -cp src/:hadoop-common-2.6.1.jar:hadoop-mapreduce-client-core-2.6.1.jar:commons-cli-2.0.jar -d . YearAndTemperature.java YearAndTemperatureMapper.java
YearAndTemperatureReducer.java
```

```
[student10@msba-hadoop-name project]$ javac -cp src/:hadoop-common-2.6.1.jar:hadoop-mapreduce-client-core-2.6.1.jar:commons-cli-2.0.jar -d . YearAndTemperature.java YearAndTemperatureMapper.java YearAndTemperatureReducer.java
Note: YearAndTemperature.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
[student10@msba-hadoop-name project]$ ls -ltr
total 22996
-rw-r--r--. 1 student10 hadoop 130868 Sep 24 18:30 commons-cli-2.0.jar
-rw-r--r--. 1 student10 hadoop 1512169 Sep 24 18:30 hadoop-mapreduce-client-core-2.6.1.jar
-rw-r--r--. 1 student10 hadoop 3318727 Sep 24 18:30 hadoop-common-2.6.1.jar
-rw-r--r--. 1 student10 hadoop 3409548 Oct 29 19:15 pig-0.11.0.jar
-rw-r--r--. 1 student10 hadoop 15137704 Nov 11 16:23 hive-exec-0.13.0.jar
-rw-r--r--. 1 student10 hadoop 717 Dec  5 15:18 YearAndTemperatureReducer.java
-rw-r--r--. 1 student10 hadoop 1123 Dec  5 15:18 YearAndTemperatureMapper.java
-rw-r--r--. 1 student10 hadoop 1167 Dec  5 15:18 YearAndTemperature.java
drwxr-xr-x. 2 student10 hadoop 4096 Dec  5 15:20 ProjectData
-rw-r--r--. 1 student10 hadoop 1438 Dec  5 15:27 YearAndTemperature.class
-rw-r--r--. 1 student10 hadoop 1884 Dec  5 15:27 YearAndTemperatureMapper.class
-rw-r--r--. 1 student10 hadoop 1617 Dec  5 15:27 YearAndTemperatureReducer.class
[student10@msba-hadoop-name project]$
```

Creating the JAR file:

`jar -cvf year-temperature.jar ./YearAndTemperature*.class`

```
[student10@msba-hadoop-name project]$ jar -cvf year-temperature.jar ./YearAndTemperature*.class
added manifest
adding: YearAndTemperature.class(in = 1438) (out= 804)(deflated 44%)
adding: YearAndTemperatureMapper.class(in = 1884) (out= 808)(deflated 57%)
adding: YearAndTemperatureReducer.class(in = 1617) (out= 670)(deflated 58%)
[student10@msba-hadoop-name project]$ ls -tlr
total 23000
-rw-r--r--. 1 student10 hadoop 130868 Sep 24 18:30 commons-cli-2.0.jar
-rw-r--r--. 1 student10 hadoop 1512169 Sep 24 18:30 hadoop-mapreduce-client-core-2.6.1.jar
-rw-r--r--. 1 student10 hadoop 3318727 Sep 24 18:30 hadoop-common-2.6.1.jar
-rw-r--r--. 1 student10 hadoop 3409548 Oct 29 19:15 pig-0.11.0.jar
-rw-r--r--. 1 student10 hadoop 15137704 Nov 11 16:23 hive-exec-0.13.0.jar
-rw-r--r--. 1 student10 hadoop 717 Dec 5 15:18 YearAndTemperatureReducer.java
-rw-r--r--. 1 student10 hadoop 1123 Dec 5 15:18 YearAndTemperatureMapper.java
-rw-r--r--. 1 student10 hadoop 1167 Dec 5 15:18 YearAndTemperature.java
drwxr-xr-x. 2 student10 hadoop 4096 Dec 5 15:20 ProjectData
-rw-r--r--. 1 student10 hadoop 1438 Dec 5 15:27 YearAndTemperature.class
-rw-r--r--. 1 student10 hadoop 1884 Dec 5 15:27 YearAndTemperatureMapper.class
-rw-r--r--. 1 student10 hadoop 1617 Dec 5 15:27 YearAndTemperatureReducer.class
-rw-r--r--. 1 student10 hadoop 3066 Dec 5 15:29 year-temperature.jar
[student10@msba-hadoop-name project]$
```

Creating the HDFS directory:

`hdfs dfs -mkdir /home/student10/project_input/`

Copying the project data into HDFS:

`hdfs dfs -copyFromLocal /home/student10/project/ProjectData /home/student10/project_input/`

```
[student10@msba-hadoop-name project]$ hdfs dfs -mkdir /home/student10/project_input/
[student10@msba-hadoop-name project]$ hdfs dfs -copyFromLocal /home/student10/project/ProjectData /home/student10/project_input/
[student10@msba-hadoop-name project]$
```

Running the JAR file:

hadoop jar year-temperature.jar YearAndTemperature
/home/student10/project_input/ProjectData /home/student10/project_output

```
[student10@msba-hadoop-name project]$ hadoop jar year-temperature.jar YearAndTemperature /home/student10/project_input/ProjectData /home/student10/project_output
21/12/05 15:35:08 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
21/12/05 15:35:08 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
21/12/05 15:35:09 INFO input.FileInputFormat: Total input files to process : 50
21/12/05 15:35:09 INFO mapreduce.JobSubmitter: number of splits:50
21/12/05 15:35:09 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
21/12/05 15:35:09 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1628901148502_1425
21/12/05 15:35:09 INFO impl.YarnClientImpl: Submitted application application_1628901148502_1425
21/12/05 15:35:09 INFO mapreduce.Job: The url to track the job: http://msba-hadoop-name:8088/proxy/application_1628901148502_1425/
21/12/05 15:35:09 INFO mapreduce.Job: Running job: job_1628901148502_1425
21/12/05 15:35:15 INFO mapreduce.Job: Job job_1628901148502_1425 running in uber mode : false
21/12/05 15:35:15 INFO mapreduce.Job: map 0% reduce 0%
21/12/05 15:35:24 INFO mapreduce.Job: map 4% reduce 0%
21/12/05 15:35:25 INFO mapreduce.Job: map 12% reduce 0%
21/12/05 15:35:32 INFO mapreduce.Job: map 18% reduce 0%
21/12/05 15:35:33 INFO mapreduce.Job: map 24% reduce 0%
21/12/05 15:35:40 INFO mapreduce.Job: map 32% reduce 0%
21/12/05 15:35:41 INFO mapreduce.Job: map 36% reduce 0%
21/12/05 15:35:48 INFO mapreduce.Job: map 42% reduce 0%
21/12/05 15:35:49 INFO mapreduce.Job: map 46% reduce 0%
21/12/05 15:35:55 INFO mapreduce.Job: map 56% reduce 0%
21/12/05 15:35:59 INFO mapreduce.Job: map 56% reduce 19%
21/12/05 15:36:02 INFO mapreduce.Job: map 66% reduce 19%
21/12/05 15:36:05 INFO mapreduce.Job: map 66% reduce 22%
21/12/05 15:36:09 INFO mapreduce.Job: map 76% reduce 22%
21/12/05 15:36:11 INFO mapreduce.Job: map 76% reduce 25%
21/12/05 15:36:17 INFO mapreduce.Job: map 84% reduce 25%
21/12/05 15:36:18 INFO mapreduce.Job: map 86% reduce 29%
21/12/05 15:36:25 INFO mapreduce.Job: map 96% reduce 29%
21/12/05 15:36:28 INFO mapreduce.Job: map 100% reduce 100%
21/12/05 15:36:29 INFO mapreduce.Job: Job job_1628901148502_1425 completed successfully
21/12/05 15:36:29 INFO mapreduce.Job: Counters: 49
    File System Counters
      FILE: Number of bytes read=398888
      FILE: Number of bytes written=11067976
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=423063
```

•
•

```

HDFS: Number of bytes written=314049
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)=316509
  Total time spent by all reduces in occupied slots (ms)=46578
  Total time spent by all map tasks (ms)=316509
  Total time spent by all reduce tasks (ms)=46578
  Total vcore-milliseconds taken by all map tasks=316509
  Total vcore-milliseconds taken by all reduce tasks=46578
  Total megabyte-milliseconds taken by all map tasks=324105216
  Total megabyte-milliseconds taken by all reduce tasks=47695872
Map-Reduce Framework
  Map input records=36404
  Map output records=36262
  Map output bytes=326358
  Map output materialized bytes=399182
  Input split bytes=7750
  Combine input records=0
  Combine output records=0
  Reduce input groups=10
  Reduce shuffle bytes=399182
  Reduce input records=36262
  Reduce output records=36262
  Spilled Records=72524
  Shuffled Maps =50
  Failed Shuffles=0
  Merged Map outputs=50
  GC time elapsed (ms)=12696
  CPU time spent (ms)=32780
  Physical memory (bytes) snapshot=17071669248
  Virtual memory (bytes) snapshot=154005131264
  Total committed heap usage (bytes)=15754330112
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=415313
File Output Format Counters
  Bytes Written=314049
[student10@msba-hadoop-name project]$

```

Verifying the output directory:

`hdfs dfs -ls /home/student10/project_output`

```

[student10@msba-hadoop-name project]$ hdfs dfs -ls /home/student10/project_output
Found 2 items
-rw-r--r--  5 student10 supergroup          0 2021-12-05 15:36 /home/student10/project_output/_SUCCESS
-rw-r--r--  5 student10 supergroup    314049 2021-12-05 15:36 /home/student10/project_output/part-r-00000
[student10@msba-hadoop-name project]$

```

Displaying the last 20 records of the output file:

```
hdfs dfs -text /home/student10/project_output/part-r-00000 | tail -20
```

```
[student10@msba-hadoop-name project]$ hdfs dfs -text /home/student10/project_output/part-r-00000 | tail -20
1930 50
1930 72
1930 61
1930 28
1930 28
1930 50
1930 39
1930 39
1930 39
1930 28
1930 50
1930 50
1930 72
1930 39
1930 61
1930 50
1930 39
1930 50
1930 22
1930 50
[student10@msba-hadoop-name project]$
```

Copying the output file data to local file:

```
hdfs dfs -copyToLocal /home/student10/project_output/part-r-00000
/home/student10/project/YearTemperatureData.txt
```

```
[student10@msba-hadoop-name project]$ hdfs dfs -copyToLocal /home/student10/project_output/part-r-00000
/home/student10/project/YearTemperatureData.txt
[student10@msba-hadoop-name project]$ ls -ltr
total 23308
-rw-r--r--. 1 student10 hadoop 130868 Sep 24 18:30 commons-cli-2.0.jar
-rw-r--r--. 1 student10 hadoop 1512169 Sep 24 18:30 hadoop-mapreduce-client-core-2.6.1.jar
-rw-r--r--. 1 student10 hadoop 3318727 Sep 24 18:30 hadoop-common-2.6.1.jar
-rw-r--r--. 1 student10 hadoop 3409548 Oct 29 19:15 pig-0.11.0.jar
-rw-r--r--. 1 student10 hadoop 15137704 Nov 11 16:23 hive-exec-0.13.0.jar
-rw-r--r--. 1 student10 hadoop 717 Dec 5 15:18 YearAndTemperatureReducer.java
-rw-r--r--. 1 student10 hadoop 1123 Dec 5 15:18 YearAndTemperatureMapper.java
-rw-r--r--. 1 student10 hadoop 1167 Dec 5 15:18 YearAndTemperature.java
drwxr-xr-x. 2 student10 hadoop 4096 Dec 5 15:20 ProjectData
-rw-r--r--. 1 student10 hadoop 1438 Dec 5 15:27 YearAndTemperature.class
-rw-r--r--. 1 student10 hadoop 1884 Dec 5 15:27 YearAndTemperatureMapper.class
-rw-r--r--. 1 student10 hadoop 1617 Dec 5 15:27 YearAndTemperatureReducer.class
-rw-r--r--. 1 student10 hadoop 3066 Dec 5 15:29 year-temperature.jar
-rw-r--r--. 1 student10 hadoop 314049 Dec 5 15:57 YearTemperatureData.txt
[student10@msba-hadoop-name project]$
```

PIG

pig -x local

```
[student10@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]
21/12/05 16:25:08 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
21/12/05 16:25:08 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2021-12-05 16:25:08,250 [main] INFO org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2021-12-05 16:25:08,250 [main] INFO org.apache.pig.Main - Logging error messages to: /home/student10/project/pig_1638750308248.log
2021-12-05 16:25:08,268 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/student10/.pigbootup not found
2021-12-05 16:25:08,376 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2021-12-05 16:25:08,378 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
2021-12-05 16:25:08,552 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-12-05 16:25:08,568 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-9146a279-7a57-4106-be26-b931de97da66
2021-12-05 16:25:08,568 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
grunt>
```

Loading the output text file with year and temperature data:

```
records = LOAD 'YearTemperatureData.txt'
AS (year:int, temperature:int);
```

```
grunt> records = LOAD 'YearTemperatureData.txt'
>> AS (year:int, temperature:int);
2021-12-05 16:30:19,501 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
grunt>
```

Populating the loaded records:

DUMP records;

```
student10@msba-hadoop-name:~/project
(1930,50)
(1930,11)
(1930,11)
(1930,28)
(1930,61)
(1930,28)
(1930,39)
(1930,61)
(1930,11)
(1930,11)
(1930,-22)
(1930,-28)
(1930,-22)
(1930,-39)
(1930,-50)
(1930,-28)
(1930,-28)
(1930,-22)
(1930,-11)
(1930,22)
(1930,22)
(1930,28)
(1930,39)
(1930,50)
(1930,22)
(1930,22)
(1930,50)
(1930,72)
(1930,61)
(1930,28)
(1930,28)
(1930,50)
(1930,39)
(1930,39)
(1930,39)
(1930,28)
(1930,50)
(1930,50)
(1930,72)
(1930,39)
(1930,39)
(1930,61)
(1930,50)
(1930,39)
(1930,50)
(1930,22)
(1930,50)
grunt>
```

Grouping the records year wise:

grouped_records = GROUP records BY year;

```
student10@msba-hadoop-name:~/project
(1930,50)
(1930,39)
(1930,50)
(1930,22)
(1930,50)
grunt> grouped_records =GROUP records BY year;
grunt>
```



```
DUMP grouped_records;
```

[illegible]

Getting the highest temperatures for each year:

```
highest_temperature = FOREACH grouped_records GENERATE group,
MAX(records.temperature);
```

```
student10@msba-hadoop-name:~/project
,11),(1930,11),(1930,-22),(1930,-28),(1930,-22),(1930,-39),(1930,-50),(1930,-28),(1930,-28),(1930,-
22),(1930,-11),(1930,22),(1930,22),(1930,28),(1930,39),(1930,50),(1930,22),(1930,22),(1930,50),(193
0,72),(1930,61),(1930,28),(1930,28),(1930,50),(1930,39),(1930,39),(1930,39),(1930,28),(1930,50),(19
30,50),(1930,72),(1930,39),(1930,61),(1930,50),(1930,39),(1930,50),(1930,22),(1930,50))}
grunt> highest_temperature = FOREACH grouped_records GENERATE group,
>> MAX(records.temperature);
grunt>
```

Populating the highest temperatures for each year:

DUMP highest_temperature;

```
student10@msba-hadoop-name:~/project
2021-12-05 16:36:59,550 [main] INFO org.apache.pig.backend
.hadoop.executionengine.util.MapRedUtil - Total input paths
to process : 1
(1921,283)
(1922,278)
(1923,294)
(1924,294)
(1925,317)
(1926,261)
(1927,489)
(1928,178)
(1929,178)
(1930,228)
grunt>
```

Getting the lowest temperatures for each year:

lowest_temperature = FOREACH grouped_records GENERATE group,
MIN(records.temperature);

```
student10@msba-hadoop-name:~/project
(1929,178)
(1930,228)
grunt> lowest_temperature = FOREACH grouped_records GENERATE group,
>> MIN(records.temperature);
grunt>
```

Populating the lowest temperatures for each year:

DUMP lowest_temperature;

```
student10@msba-hadoop-name:~/project
2021-12-05 16:39:48,236 [main] INFO org.apache.pig.backend.hado
op.executionengine.util.MapRedUtil - Total input paths to proces
s : 1
(1921,-417)
(1922,-400)
(1923,-394)
(1924,-456)
(1925,-378)
(1926,-411)
(1927,-322)
(1928,-178)
(1929,-122)
(1930,-139)
grunt>
```

HIVE

schematool -dbType derby -initSchema

hive

```
student10@msba-hadoop-name:~/project
drwxr-xr-x. 2 student10 hadoop 4096 Dec  5 15:20 ProjectData
-rw-r--r--. 1 student10 hadoop 1438 Dec  5 15:27 YearAndTemperature.class
-rw-r--r--. 1 student10 hadoop 1884 Dec  5 15:27 YearAndTemperatureMapper.class
-rw-r--r--. 1 student10 hadoop 1617 Dec  5 15:27 YearAndTemperatureReducer.class
-rw-r--r--. 1 student10 hadoop 3066 Dec  5 15:29 year-temperature.jar
-rw-r--r--. 1 student10 hadoop 314049 Dec  5 15:57 YearTemperatureData.txt
[student10@msba-hadoop-name project]$ schematool -dbType derby -initSchema
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/Static
LoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.9.0/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/o
rg/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]
Metastore connection URL:      jdbc:derby;;databaseName=metastore_db;create=true
Metastore Connection Driver :  org.apache.derby.jdbc.EmbeddedDriver
Metastore connection User:     APP
Starting metastore schema initialization to 2.3.0
Initialization script hive-schema-2.3.0.derby.sql
Initialization script completed
schemaTool completed
[student10@msba-hadoop-name project]$ 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/Static
LoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.9.0/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/o
rg/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]

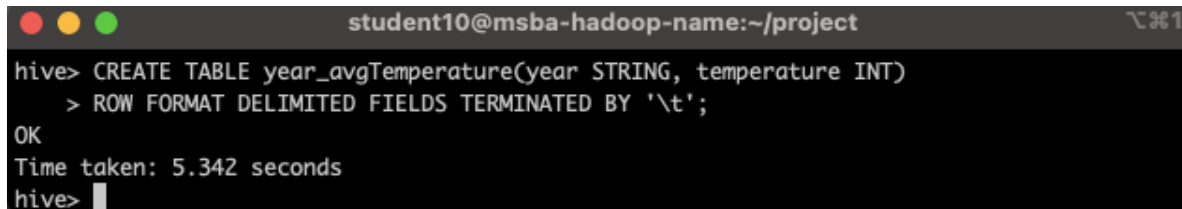
Logging initialized using configuration in jar:file:/usr/local/hive-2.3.2/lib/hive-common-2.3.2.jar!/hive-log
4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a differen
t execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive>
```

set hive.metastore.warehouse.dir;

```
hive> set hive.metastore.warehouse.dir;
hive.metastore.warehouse.dir=/user/hive/warehouse
hive>
```

Creating a table:

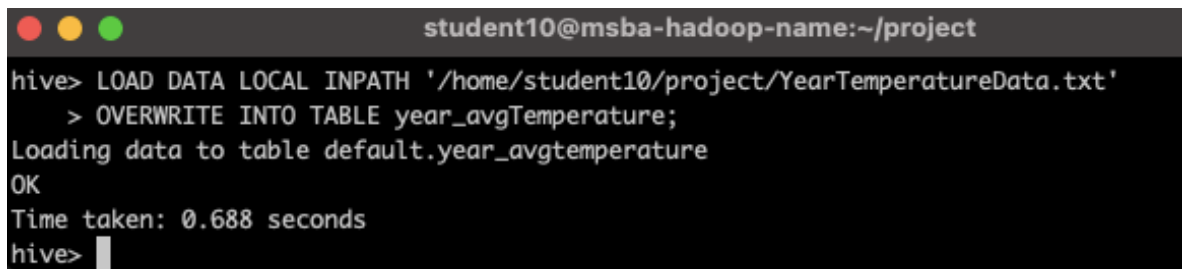
```
CREATE TABLE year_avgTemperature(year STRING, temperature INT)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t';
```

A terminal window with a dark background and light gray text. The title bar shows 'student10@msba-hadoop-name:~/project' and a window icon with three colored circles (red, yellow, green). The terminal content shows the Hive command to create a table, followed by 'OK' and the execution time.

```
student10@msba-hadoop-name:~/project
hive> CREATE TABLE year_avgTemperature(year STRING, temperature INT)
> ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t';
OK
Time taken: 5.342 seconds
hive> █
```

Loading the output text file with year and temperature data:

```
LOAD DATA LOCAL INPATH '/home/student10/project/YearTemperatureData.txt'
OVERWRITE INTO TABLE year_avgTemperature;
```

A terminal window with a dark background and light gray text. The title bar shows 'student10@msba-hadoop-name:~/project' and a window icon with three colored circles (red, yellow, green). The terminal content shows the Hive command to load data, followed by 'Loading data to table default.year_avgtemperature', 'OK', and the execution time.

```
student10@msba-hadoop-name:~/project
hive> LOAD DATA LOCAL INPATH '/home/student10/project/YearTemperatureData.txt'
> OVERWRITE INTO TABLE year_avgTemperature;
Loading data to table default.year_avgtemperature
OK
Time taken: 0.688 seconds
hive> █
```

Populating the average temperatures for each year:

```
SELECT year, AVG(temperature)
FROM year_avgTemperature
GROUP BY year;
```

```
student10@msba-hadoop-name:~/project
hive> SELECT year, AVG(temperature)
> FROM year_avgTemperature
> GROUP BY year;
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 = student10_20211205165845_fe941b23-1b53-4e3f-abc4-d36999e6962a
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_1628901148502_1437, Tracking URL = http://msba-hadoop-name:8088/proxy/application_1628901148502_1437/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1628901148502_1437
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-12-05 16:58:53,603 Stage-1 map = 0%, reduce = 0%
2021-12-05 16:58:58,781 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.83 sec
2021-12-05 16:59:04,954 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.02 sec
MapReduce Total cumulative CPU time: 5 seconds 20 msec
Ended Job = job_1628901148502_1437
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.02 sec HDFS Read: 322845 HDFS Write: 441 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 20 msec
OK
1921 32.696049743964885
1922 22.96797482413921
1923 23.003109566489847
1924 34.101427776747634
1925 33.29253567508233
1926 18.740430394210627
1927 79.7364185110664
1928 36.446247464503045
1929 45.38611449451888
1930 62.13735899137359
Time taken: 20.725 seconds, Fetched: 10 row(s)
hive>
```