

The Course Project

The project includes 3 parts. The first part is to develop a python application to retrieve year and temperature from original NCDC records (i.e. the dataset we are using for this project) 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 the lowest temperatures for each year. The third part is to load the text file into hive and get the average temperature for each year.

1) Python files (mapper and reducer):

- Mapper: **Project_Map.py**
- Reducer: **Project_Reduce.py**

2) the commands for executing the Python application in Hadoop:

Copying CourseProjectData into Hadoop file:

```
hdfs dfs -copyFromLocal /home/student76/CourseProjectData  
/home/76student76/CourseProjectData
```

Running the Python files to get temperature & year as an output file:

```
/home/76student76/CourseProjectData/* \output /home/76student76/Project_output  
\mapper /home/student76/Project_Map.py \reducer /home/student76/Project_Reduce.py  
\file /home/student76/Project_Map.py \file /home/student76/Project_Reduce.py
```

Listing contents of the output directory in HDFS:

```
hdfs dfs -ls /home/76student76/Project_output/
```

Converting output to a text file with a new name:

```
hdfs dfs -copyToLocal /home/76student76/Project_output/part-00000  
/home/student76/Project_output_
```

```
[student76@msba-hadoop-name ~]$ hadoop jar /home/student76/hadoop-streaming-2.9.0.jar -input /home/76student76/CourseProjectData/* -output /home/76student76/
Project_output_ts1799 -mapper /home/student76/Project_Map_ts1799.py -reducer /home/student76/Project_Reduce_ts1799.py -file /home/student76/Project_Map_ts179
9.py -file /home/student76/Project_Reduce_ts1799.py
24/05/05 19:06:55 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [/home/student76/Project_Map_ts1799.py, /home/student76/Project_Reduce_ts1799.py, /tmp/hadoop-unjar3537842120054451390/] [] /tmp/streamjob4818
405284772310499.jar tmpDir=null
24/05/05 19:06:55 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
24/05/05 19:06:56 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
24/05/05 19:06:56 INFO mapred.FileInputFormat: Total input files to process : 160
24/05/05 19:06:56 INFO mapreduce.JobSubmitter: number of splits:160
24/05/05 19:06:56 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publi
sher.enabled
24/05/05 19:06:56 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1714512003524_1324
24/05/05 19:06:57 INFO impl.YarnClientImpl: Submitted application application_1714512003524_1324
24/05/05 19:06:57 INFO mapreduce.Job: The url to track the job: http://msba-hadoop-name:8088/proxy/application_1714512003524_1324/
24/05/05 19:06:57 INFO mapreduce.Job: Running job: job_1714512003524_1324
24/05/05 19:07:03 INFO mapreduce.Job: Job job_1714512003524_1324 running in uber mode : false
24/05/05 19:07:03 INFO mapreduce.Job: map 0% reduce 0%
24/05/05 19:07:12 INFO mapreduce.Job: map 4% reduce 0%
24/05/05 19:07:21 INFO mapreduce.Job: map 8% reduce 0%
24/05/05 19:07:28 INFO mapreduce.Job: map 9% reduce 0%
24/05/05 19:07:29 INFO mapreduce.Job: map 11% reduce 0%
24/05/05 19:07:37 INFO mapreduce.Job: map 13% reduce 0%
24/05/05 19:07:38 INFO mapreduce.Job: map 15% reduce 0%
24/05/05 19:07:45 INFO mapreduce.Job: map 16% reduce 0%
24/05/05 19:07:47 INFO mapreduce.Job: map 19% reduce 0%
24/05/05 19:07:55 INFO mapreduce.Job: map 21% reduce 0%
24/05/05 19:07:58 INFO mapreduce.Job: map 22% reduce 0%
24/05/05 19:08:03 INFO mapreduce.Job: map 23% reduce 0%
24/05/05 19:08:04 INFO mapreduce.Job: map 24% reduce 0%
24/05/05 19:08:05 INFO mapreduce.Job: map 25% reduce 0%
24/05/05 19:08:06 INFO mapreduce.Job: map 26% reduce 0%
24/05/05 19:08:11 INFO mapreduce.Job: map 27% reduce 0%
24/05/05 19:08:12 INFO mapreduce.Job: map 29% reduce 9%
24/05/05 19:08:18 INFO mapreduce.Job: map 31% reduce 10%
24/05/05 19:08:19 INFO mapreduce.Job: map 32% reduce 10%
24/05/05 19:08:24 INFO mapreduce.Job: map 32% reduce 11%
24/05/05 19:08:25 INFO mapreduce.Job: map 34% reduce 11%
24/05/05 19:08:26 INFO mapreduce.Job: map 35% reduce 11%
24/05/05 19:08:28 INFO mapreduce.Job: map 36% reduce 11%
24/05/05 19:08:30 INFO mapreduce.Job: map 36% reduce 12%
24/05/05 19:08:32 INFO mapreduce.Job: map 37% reduce 12%
24/05/05 19:08:33 INFO mapreduce.Job: map 38% reduce 12%
24/05/05 19:08:34 INFO mapreduce.Job: map 39% reduce 12%
24/05/05 19:08:36 INFO mapreduce.Job: map 39% reduce 13%
24/05/05 19:08:39 INFO mapreduce.Job: map 40% reduce 13%
24/05/05 19:08:40 INFO mapreduce.Job: map 42% reduce 13%
24/05/05 19:08:42 INFO mapreduce.Job: map 42% reduce 14%

Total time spent by all reduces in occupied slots (ms)=172433
Total time spent by all map tasks (ms)=937137
Total time spent by all reduce tasks (ms)=172433
Total vcore-milliseconds taken by all map tasks=937137
Total vcore-milliseconds taken by all reduce tasks=172433
Total megabyte-milliseconds taken by all map tasks=959628288
Total megabyte-milliseconds taken by all reduce tasks=176571392

Map-Reduce Framework
Map input records=239794
Map output records=215258
Map output bytes=2367838
Map output materialized bytes=2799314
Input split bytes=23200
Combine input records=0
Combine output records=0
Reduce input groups=1
Reduce shuffle bytes=2799314
Reduce input records=215258
Reduce output records=215258
Spilled Records=430516
Shuffled Maps =160
Failed Shuffles=0
Merged Map outputs=160
GC time elapsed (ms)=31157
CPU time spent (ms)=120320
Physical memory (bytes) snapshot=56328265728
Virtual memory (bytes) snapshot=486618705920
Total committed heap usage (bytes)=47868542976

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=4370358
File Output Format Counters
Bytes Written=2367838
24/05/05 19:10:47 INFO streaming.StreamJob: Output directory: /home/76student76/Project_output_ts1799
[student76@msba-hadoop-name ~]$ hdfs dfs -ls /home/76student76/Project_output_ts1799/_
Found 2 items
-rw-r--r-- 5 student76 supergroup 0 2024-05-05 19:10 /home/76student76/Project_output_ts1799/_SUCCESS
-rw-r--r-- 5 student76 supergroup 2367838 2024-05-05 19:10 /home/76student76/Project_output_ts1799/part-00000
[student76@msba-hadoop-name ~]$ hdfs dfs -copyToLocal /home/76student76/Project_output_ts1799/part-00000 /home/student76/Project_output_ts1799.txt
[student76@msba-hadoop-name ~]$ 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]
```

3) the text file including Year and Temperature data created by you:

- File Name: **Project_output**

4) the screenshot of the text file being created:

```
Total time spent by all reduces in occupied slots (ms)=172433
Total time spent by all map tasks (ms)=937137
Total time spent by all reduce tasks (ms)=172433
Total vcore-milliseconds taken by all map tasks=937137
Total vcore-milliseconds taken by all reduce tasks=172433
Total megabyte-milliseconds taken by all map tasks=959628288
Total megabyte-milliseconds taken by all reduce tasks=176571392
Map-Reduce Framework
  Map input records=239794
  Map output records=215258
  Map output bytes=2367838
  Map output materialized bytes=2799314
  Input split bytes=23200
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=2799314
  Reduce input records=215258
  Reduce output records=215258
  Spilled Records=430516
  Shuffled Maps =160
  Failed Shuffles=0
  Merged Map outputs=160
  GC time elapsed (ms)=31157
  CPU time spent (ms)=120320
  Physical memory (bytes) snapshot=56328265728
  Virtual memory (bytes) snapshot=486618705920
  Total committed heap usage (bytes)=47868542976
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=4370358
File Output Format Counters
  Bytes Written=2367838
24/05/05 19:10:47 INFO streaming.StreamJob: Output directory: /home/76student76/Project_output_ts1799
[student76@msba-hadoop-name ~]$ hdfs dfs -ls /home/76student76/Project_output_ts1799/
Found 2 items
-rw-r--r-- 5 student76 supergroup          0 2024-05-05 19:10 /home/76student76/Project_output_ts1799/ SUCCESS
-rw-r--r-- 5 student76 supergroup    2367838 2024-05-05 19:10 /home/76student76/Project_output_ts1799/part-00000
[student76@msba-hadoop-name ~]$ hdfs dfs -copyToLocal /home/76student76/Project_output_ts1799/part-00000 /home/student76/Project_output_ts1799.txt
[student76@msba-hadoop-name ~]$ 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]
```

```
1940      -0022
1940      +0000
1940      +0028
1940      +0050
1940      +0039
1940      +0022
1940      +0022
1940      +0000
1940      +0011
1940      +0011
1940      +0011
1940      -0028
1940      -0072
1940      -0022
1940      -0039
1940      -0078
1940      -0022
1940      -0028
1940      -0028
1940      -0078
1940      -0078
1940      -0072
1940      -0028
1940      +0011
1940      +0000
1940      -0011
1940      -0022
1940      -0028
1940      -0039
1940      -0072
1940      -0028
1940      -0039
1940      -0011
1940      -0011
1940      +0000
1940      -0011
1940      +0022
1940      +0022
1940      +0028
```

Ln 1, Col 1	23,67,838 characters
-------------	----------------------

- 5) the screenshot of the final Pig output showing the year and the highest and lowest temperatures:

1. **Open Pig in local mode:**

```
pig -x local
```

2. **Load the data from the text file:**

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

3. **Group the records by year:**

```
grouped_records = GROUP records BY year;
```

4. **Find the maximum temperature for each year:**

```
maxtemp = FOREACH grouped_records GENERATE group AS year,  
MAX(records.temperature) AS max_temperature;
```

5. **Display the result of maximum temperatures:**

```
DUMP maxtemp;
```

6. **Find the minimum temperature for each year:**

```
mintemp = FOREACH grouped_records GENERATE group AS year,  
MIN(records.temperature) AS min_temperature;
```

7. **Display the result of minimum temperatures:**

```
DUMP mintemp;
```

```
[student76@muba-hadoop-name ~]$ 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]
24/05/05 19:13:43 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
24/05/05 19:13:43 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2024-05-05 19:13:43,362 [main] INFO org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2024-05-05 19:13:43,362 [main] INFO org.apache.pig.Main - Logging error messages to: /home/student76/pig/1714961623361.log
2024-05-05 19:13:43,381 [main] INFO org.apache.pig.impl.util.Util - Default bootstrap file /home/student76/pig/bootstrap not found
2024-05-05 19:13:43,517 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-05-05 19:13:43,519 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
2024-05-05 19:13:43,798 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2024-05-05 19:13:43,817 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-61c9b0f5-3aad-426d-8216-ae51bb492b0f
2024-05-05 19:13:43,818 [main] WARN org.apache.pig.PigServer - AT3 is disabled since yarn.timeline-service.enabled set to false
grunt> records = LOAD 'Project_output.txt' AS (year:chararray, temperature:int);
2024-05-05 19:14:01,477 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
grunt> grouped_records = GROUP records BY year;
grunt> maxtemp = FOREACH grouped_records GENERATE group, MAX(records.temperature);
grunt> DUMP maxtemp;
2024-05-05 19:15:07,256 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2024-05-05 19:15:07,273 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY
2024-05-05 19:15:07,293 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2024-05-05 19:15:07,333 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - (RULES_ENABLED={AddForEach, ColumnMapKeyPrune, ConstantC
alculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, Predi
catePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter})
2024-05-05 19:15:07,409 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen) of size 699400192 to monitor. collectionUsa
geThreshold = 489580128, usageThreshold = 489580128
2024-05-05 19:15:07,523 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic?
false
2024-05-05 19:15:07,540 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOptimizerUtil - Choosing to move algebraic foreach to combine
r
2024-05-05 19:15:07,571 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2024-05-05 19:15:07,571 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2024-05-05 19:15:07,594 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2024-05-05 19:15:07,608 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - session.id is deprecated. Instead, use dfs.metrics.session-id
2024-05-05 19:15:07,609 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Initializing JVM Metrics with processName=JobTracker, sessionId=
2024-05-05 19:15:07,635 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2024-05-05 19:15:07,641 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use
mapreduce.reduce.markreset.buffer.percent
2024-05-05 19:15:07,641 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.p
```

Highest Temperature: 361

```
HadoopVersion  PigVersion  UserId StartedAt  FinishedAt  Features
2.9.0  0.17.0  student76  2024-05-05 19:15:07  2024-05-05 19:15:10  GROUP_BY

Success!

Job Stats (time in seconds):
JobId  Maps  Reduces MaxMapTime  MinMapTime  AvgMapTime  MedianMapTime  MaxReduceTime  MinReduceTime  AvgReduceTime  MedianReducetime  A
lias  Feature Outputs
job_local1498912576_0001  1  1  n/a  n/a  n/a  n/a  n/a  n/a  grouped_records,maxtemp,records  GROUP_BY,COMB
INER  file:/tmp/temp306273774/tmp-1433279772,

Input(s):
Successfully read 215258 records from: "file:///home/student76/Project_output_tsl799.txt"

Output(s):
Successfully stored 1 records in: "file:/tmp/temp306273774/tmp-1433279772"

Counters:
Total records written : 1
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1498912576_0001

2024-05-05 19:15:10,063 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2024-05-05 19:15:10,064 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2024-05-05 19:15:10,064 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2024-05-05 19:15:10,070 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2024-05-05 19:15:10,073 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2024-05-05 19:15:10,073 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2024-05-05 19:15:10,085 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2024-05-05 19:15:10,085 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1940,361)
```

Lowest temperature: -311

```
HadoopVersion  PigVersion  UserId StartedAt  FinishedAt  Features
2.9.0  0.17.0  student76  2024-05-05 19:15:57  2024-05-05 19:15:58  GROUP_BY

Success!

Job Stats (time in seconds):
JobId  Maps  Reduces MaxMapTime  MinMapTime  AvgMapTime  MedianMapTime  MaxReduceTime  MinReduceTime  AvgReduceTime  MedianReducetime  A
lias  Feature Outputs
job_local1633695278_0002  1  1  n/a  n/a  n/a  n/a  n/a  n/a  grouped_records,mintemp,records  GROUP_BY,COMB
INER  file:/tmp/temp306273774/tmp2091934347,

Input(s):
Successfully read 215258 records from: "file:///home/student76/Project_output_tsl799.txt"

Output(s):
Successfully stored 1 records in: "file:/tmp/temp306273774/tmp2091934347"

Counters:
Total records written : 1
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1633695278_0002

2024-05-05 19:15:58,969 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2024-05-05 19:15:58,970 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2024-05-05 19:15:58,971 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2024-05-05 19:15:58,975 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2024-05-05 19:15:58,976 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2024-05-05 19:15:58,977 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2024-05-05 19:15:58,987 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2024-05-05 19:15:58,988 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1940,-311)
```

6) the screenshot of the final Hive output showing the year and average temperature:

1. **Open Hive:**
hive

2. **Drop the table if it exists to ensure a clean start:**
DROP TABLE IF EXISTS Project_table;
3. **Create a new table named Project_table to store the data:**
CREATE TABLE Project_table (year STRING, temperature INT) ROW FORMAT
DELIMITED FIELDS TERMINATED BY '\t';
4. **Load the output text data into the newly created table:**
LOAD DATA LOCAL INPATH 'Project_output.txt' OVERWRITE INTO TABLE
Project_table;
5. **Run the final query to calculate the average temperature for each year:**
SELECT year, AVG(temperature) FROM Project_table GROUP BY year;

Average Temperature: 74.9186371

```
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 H
ve 1.X releases.
hive> DROP TABLE IF EXISTS Project_table_ts1799;
OK
Time taken: 11.378 seconds
hive> CREATE TABLE Project_table_ts1799 (year STRING, temperature INT) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t';
OK
Time taken: 2.212 seconds
hive> LOAD DATA LOCAL INPATH 'Project_output_ts1799.txt' OVERWRITE INTO TABLE Project_table_ts1799;
Loading data to table default.project_table_ts1799
OK
Time taken: 0.951 seconds
hive> SELECT year, AVG(temperature)
> FROM Project_table_ts1799
> 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 = student76_20240505192702_6f392c47-1617-43ec-a219-b12a98e46045
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.reducers=<number>
Starting Job = job_1714512003524_1337, Tracking URL = http://msba-hadoop-name:8088/proxy/application_1714512003524_1337/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1714512003524_1337
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2024-05-05 19:27:12,353 Stage-1 map = 0%, reduce = 0%
2024-05-05 19:27:18,643 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.86 sec
2024-05-05 19:27:25,904 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.31 sec
MapReduce Total cumulative CPU time: 6 seconds 310 msec
Ended Job = job_1714512003524_1337
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.31 sec HDFS Read: 2376648 HDFS Write: 122 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 310 msec
OK
1940 74.91863717027938
Time taken: 24.57 seconds, Fetched: 1 row(s)
```