FAKULTI SAINS KOMPUTER & TEKNOLOGI MAKLUMAT

Faculty of Computer Science & Information

Climate Analysis using Weather Data Repository Datasets

Kod & Nama Kursus ***Course Code & Name***	WQD7009 BIG DATA APPLICATIONS AND ANALYTICS
Nama ***Name***	MUHAMAD RAFIQ IQBAL BIN SAMSUDIN (17206926)
Pensyarah ***Lecturer***	DR. RIYAZ AHAMED ARIYALURAN HABEEB MOHAMED
Tarikh Hantar ***Submission Date***	1 DECEMBER 2024

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Dataset Overview

Dataset Name : Global Weather Repository

Description : This dataset provides overall weather data for major city in all country in the world .

Key features Selected :

Collumn	Description
No	Row Number (ROW-KEY)
country	The country where the data was recorded.
location_name	The city name of the country the data was recorded.
updated_time	The time and date of the data when it was recorded
Time	The time of the data recorded.
Date	The date of the data recorded.
temperature_celsius	The recorded temperature in degree celcius.
humidity	The humidity percentage of the area.
air_quality_PM2.5	Pm 2.5 air quality index
air_quality_PM10	Pm 10 air quality index
air_quality_Carbon_Monoxide	Air quality measurement: Carbon Monoxide
air_quality_Nitrogen_dioxide	Air quality measurement: Nitrogen Dioxide
air_quality_Sulphur_dioxide	Air quality measurement: Sulphur Dioxide

CHAPTER 2: METHODOLOGY

2.1 Hbase Table Design

This section describes the Hbase schema and DDL Languange use to create the table.

Collumn Family	Columns
location_info	country , location_name
timestamps	updated_time ,Time , Date ,
weather_data	humidity, temperature_celcius
air_quality	air_quality_PM2.5, air_quality_PM10, air_quality_Carbon_Monoxide, air_quality_Nitrogen_dioxide, air_quality_Sulphur_dioxide.
air_speed	wind_mph , wind_degree

2.2 Importing Data Into Hbase

1.**Start** the Hbase .

```
[cloudera@quickstart Desktop]$ hbase shell
2024-11-29 00:18:31,042 INFO  [main] Configuration.deprecation: hadoop.native.li
b is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.10.0, rUnknown, Fri Jan 20 12:13:18 PST 2017
```

2.**Create** Hbase Table Name Weather_Data and propose collumn family.

```
hbase(main):005:0> create 'Weather_Data','location_info','timestamps','weather_d
ata','air_quality','air_speed'
0 row(s) in 1.2930 seconds
```

3. Check back the **list** of table created.

```
=> Hbase::Table - Weather_Data
hbase(main):005:0> list
TABLE
Weather_Data
weather_data
2 row(s) in 0.0060 seconds
```

4. **Exit** Hbase to go to the Hdfs to upload the csv

```
=> ["Weather_Data", "weather_data"]
hbase(main):006:0> exit
[cloudera@quickstart Desktop]$
```

5. Input the Csv file into Hfds Path (**Mkdir, put,ls**)

```
[cloudera@quickstart Desktop]$ hdfs dfs -mkdir /Dataset
[cloudera@quickstart Desktop]$ hdfs dfs -put /home/cloudera/Desktop/WeatherDataReduce.c
sv /Dataset
[cloudera@quickstart Desktop]$ hdfs dfs -ls /
Found 16 items
drwxr-xr-x   - cloudera supergroup          0 2024-11-28 20:42 /Bdaa
drwxr-xr-x   - cloudera supergroup          0 2024-11-28 23:06 /Csvfile
drwxr-xr-x   - cloudera supergroup          0 2024-11-29 00:46 /Dataset
drwxr-xr-x   - cloudera supergroup          0 2024-11-10 20:00 /Sparkdata
drwxrwxrwx   - hdfs     supergroup          0 2017-04-05 04:27 /benchmarks
drwxr-xr-x   - hbase    supergroup          0 2024-11-29 00:25 /hbase
drwxr-xr-x   - cloudera supergroup          0 2024-10-18 07:27 /inputfolder1
drwxr-xr-x   - cloudera supergroup          0 2024-10-21 00:21 /inputfolder2
drwxr-xr-x   - cloudera supergroup          0 2024-11-10 20:09 /my_spark_output1
drwxr-xr-x   - cloudera supergroup          0 2024-10-18 07:30 /out1
drwxr-xr-x   - cloudera supergroup          0 2024-10-21 00:23 /out2
drwxr-xr-x   - solr     solr                0 2017-04-05 04:29 /solr
-rw-r--r--   1 cloudera supergroup    4418139 2024-11-10 19:10 /sparkdata
drwxrwxrwt   - hdfs     supergroup          0 2024-10-17 01:46 /tmp
drwxr-xr-x   - hdfs     supergroup          0 2017-04-05 04:29 /user
drwxr-xr-x   - hdfs     supergroup          0 2017-04-05 04:29 /var
```

6. Check the csv file in the hdfs path by using **cat command**

```
[cloudera@quickstart Desktop]$ hdfs dfs -cat /Dataset/WeatherDataReduce.csv
```

Output :

cloudera@quickstart:~/Desktop

File Edit View Search Terminal Help

```
37775,Oman,Muscat,27/11/2024 12:45,12:45,27/11/2024,29.2,58,25.53,44.77,384.8,40.7,37.5
55
37776,Pakistan,Islamabad,27/11/2024 13:30,13:30,27/11/2024,22.2,16,69.005,71.225,2325.4
5,97.495,16.28
37777,Palau,Airai,27/11/2024 17:45,17:45,27/11/2024,28.7,78,4.625,7.215,185,0.555,0.37
37778,Bolivia,National,27/11/2024 4:45,04:45,27/11/2024,22.3,98,13.875,14.06,423.65,0.5
55,0.185
37779,Panama,Panama City,27/11/2024 3:30,03:30,27/11/2024,25.1,94,5.365,6.29,1011.95,8.
325,2.96
37780,Papua New Guinea,Port Moresby,27/11/2024 18:45,18:45,27/11/2024,27.7,84,6.475,7.4
,273.8,1.48,0.37
37781,Paraguay,Aurora,27/11/2024 6:00,06:00,27/11/2024,28.4,59,7.384,7.384,468.223,0.67
1,0.168
37782,Peru,Lima,27/11/2024 3:45,03:45,27/11/2024,17.3,84,29.415,40.515,566.1,19.98,5.36
5
37783,Philippines,Manila,27/11/2024 16:45,16:45,27/11/2024,32.1,56,34.41,37,1402.3,27.5
65,35.52
37784,Poland,Warsaw,27/11/2024 9:45,09:45,27/11/2024,6.2,93,58.09,63.64,556.85,63.455,3
2.19
37785,Portugal,Lisbon,27/11/2024 8:45,08:45,27/11/2024,9.4,100,29.785,39.035,316.35,38.
11,3.33
37786,Qatar,Doha,27/11/2024 11:45,11:45,27/11/2024,32.4,49,53.465,74.74,825.1,57.535,18
.13
37787,Romania,Bucharest,27/11/2024 10:45,10:45,27/11/2024,4.1,81,150.59,161.505,1269.1,
95.09,29.6
37788,Russia,Moscow,27/11/2024 11:30,11:30,27/11/2024,-3.7,93,141.895,183.52,1546.6,120
.62,227.55
```

7. Upload the csv file from the HDFS path to the Hbase Table that we have create.

```
[cloudera@quickstart Desktop]$ hbase org.apache.hadoop.hbase.mapreduce.ImportTsv
 -Dimporttsv.separator=',' -Dimporttsv.columns=HBASE_ROW_KEY,location_info:count
ry,location_info:location_name,timestamps:updated_time,timestamps:Time,timestamp
s:Date,weather_data:temperature_celsius,weather_data:humidity,air_quality:air_qu
ality_PM2.5,air_quality:air_quality_PM10,air_quality:air_quality_Carbon_Monoxide
,air_quality:air_quality_Nitrogen_dioxide,air_quality:air_quality_Sulphur_dioxid
e Weather_Data /Dataset/WeatherDataReduce.csv
```

Hbase table (Weather_Data) and Hdfs Path (/Dataset/WeatherDataReduce.csv)

8. **Describe** the table to check the collumn families on the table

```
hbase(main):003:0> describe 'Weather_Data'
Table Weather_Data is ENABLED
```

Output:

```
COLUMN FAMILIES DESCRIPTION
{NAME => 'air_quality', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REP
LICATION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '
0', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_ME
MORY => 'false', BLOCKCACHE => 'true'}
{NAME => 'air_speed', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLI
CATION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0'
, TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMO
RY => 'false', BLOCKCACHE => 'true'}
{NAME => 'location_info', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', R
EPLICATION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS =>
 '0', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_
MEMORY => 'false', BLOCKCACHE => 'true'}
{NAME => 'timestamps', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPL
ICATION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0
', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEM
ORY => 'false', BLOCKCACHE => 'true'}
{NAME => 'weather_data', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', RE
PLICATION_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS =>
'0', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_M
EMORY => 'false', BLOCKCACHE => 'true'}
5 row(s) in 0.0870 seconds
```

As we can see here, the table have 5 collumn family that we have created .

9. Check the 1st row and last row of the data using **get** command.

```
hbase(main):005:0> get 'Weather_Data','1'
COLUMN                                                  CELL
0 row(s) in 0.0050 seconds

hbase(main):006:0> get 'Weather_Data','00001'
COLUMN                                                  CELL
 air_quality:air_quality_Carbon_Monoxide                timestamp=1732915791330, value=277
 air_quality:air_quality_Nitrogen_dioxide               timestamp=1732915791330, value=1.1
 air_quality:air_quality_PM10                           timestamp=1732915791330, value=26.6
 air_quality:air_quality_PM2.5                          timestamp=1732915791330, value=8.4
 air_quality:air_quality_Sulphur_dioxide                timestamp=1732915791330, value=0.2
 air_speed:wind_degree                                  timestamp=1732915791330, value=338
 air_speed:wind_mph                                     timestamp=1732915791330, value=8.3
 location_info:country                                  timestamp=1732915791330, value=Afghanistan
 location_info:location_name                            timestamp=1732915791330, value=Kabul
 timestamps:Date                                        timestamp=1732915791330, value=16/5/2024
 timestamps:Time                                        timestamp=1732915791330, value=13:15
 timestamps:updated_time                                timestamp=1732915791330, value=16/5/2024 13:15
 weather_data:humidity                                  timestamp=1732915791330, value=24
 weather_data:temperature_celsius                       timestamp=1732915791330, value=26.6
14 row(s) in 0.0040 seconds
```

```
hbase(main):007:0> get 'Weather_Data','38034'
COLUMN                                                  CELL
 air_quality:air_quality_Carbon_Monoxide                timestamp=1732915791330, value=458.8
 air_quality:air_quality_Nitrogen_dioxide               timestamp=1732915791330, value=0.74
 air_quality:air_quality_PM10                           timestamp=1732915791330, value=12.95
 air_quality:air_quality_PM2.5                          timestamp=1732915791330, value=12.95
 air_quality:air_quality_Sulphur_dioxide                timestamp=1732915791330, value=0.925
 air_speed:wind_degree                                  timestamp=1732915791330, value=324
 air_speed:wind_mph                                     timestamp=1732915791330, value=10.1
 location_info:country                                  timestamp=1732915791330, value=Zimbabwe
 location_info:location_name                            timestamp=1732915791330, value=Harare
 timestamps:Date                                        timestamp=1732915791330, value=28/11/2024
 timestamps:Time                                        timestamp=1732915791330, value=12:45
 timestamps:updated_time                                timestamp=1732915791330, value=28/11/2024 12:45
 weather_data:humidity                                  timestamp=1732915791330, value=50
 weather_data:temperature_celsius                       timestamp=1732915791330, value=26.2
14 row(s) in 0.0030 seconds

hbase(main):008:0>
```

2.3 Deleting collumn

Since there are air_speed collumn that are most probably not use in our analysis, we can drop the collumn.

1.**Disable** the table : Before making any structural changes to an Hbase table, we disable it first

```
=> ["Weather_Data", "weather_Data", "weather_data"]
hbase(main):008:0> disable 'weather_Data'
0 row(s) in 2.2890 seconds
```

2. **Alter** the table to drop the Collumn family of air_speed.

```
hbase(main):005:0* alter 'Weather_Data', {NAME => 'air_speed',METHOD => 'delete'}
Updating all regions with the new schema...
1/1 regions updated.
Done.
0 row(s) in 1.9290 seconds
```

3. **Enable** the table .

```
hbase(main):011:0> enable 'Weather_Data'
0 row(s) in 1.2590 seconds
```

4. Check first row of the Hbase table using **get** command.

```
hbase(main):011:0> get 'Weather_Data','00001'
COLUMN                                          CELL
 air_quality:air_quality_Carbon_Monoxide        timestamp=1732915791330, value=277
 air_quality:air_quality_Nitrogen_dioxide       timestamp=1732915791330, value=1.1
 air_quality:air_quality_PM10                   timestamp=1732915791330, value=26.6
 air_quality:air_quality_PM2.5                  timestamp=1732915791330, value=8.4
 air_quality:air_quality_Sulphur_dioxide        timestamp=1732915791330, value=0.2
 location_info:country                          timestamp=1732915791330, value=Afghanistan
 location_info:location_name                    timestamp=1732915791330, value=Kabul
 timestamps:Date                                timestamp=1732915791330, value=16/5/2024
 timestamps:Time                                timestamp=1732915791330, value=13:15
 timestamps:updated_time                        timestamp=1732915791330, value=16/5/2024 13:15
 weather_data:humidity                          timestamp=1732915791330, value=24
 weather_data:temperature_celsius               timestamp=1732915791330, value=26.6
12 row(s) in 0.0070 seconds

hbase(main):012:0>
```

Since the air_speed collumn family has been drop, there are no more wind_mph and wind_degree. The overall row reduce from 14 to 12.

CHAPTER 3: ANALYSIS AND RESULT DISCUSSION

3.1 GLOBAL TEMPERATURE CHANGE

3.1.1 Top 2 Global Highest Temperature recorded

Query & Output :

```
hbase(main):036:0> scan 'Weather_Data', {COLUMNS => ['location_info:country', 'weather_data:temperature_celsius', 'timestamps:update
d_time'], FILTER => "SingleColumnValueFilter('weather_data', 'temperature_celsius', >, 'binary:47.9')"}
ROW                                  COLUMN+CELL
 06897                               column=location_info:country, timestamp=1732952792686, value=Kuwait
 06897                               column=timestamps:updated_time, timestamp=1732952792686, value=19/6/2024 16:45
 06897                               column=weather_data:temperature_celsius, timestamp=1732952792686, value=49.2
 07271                               column=location_info:country, timestamp=1732952792686, value=Iraq
 07271                               column=timestamps:updated_time, timestamp=1732952792686, value=21/6/2024 16:30
 07271                               column=weather_data:temperature_celsius, timestamp=1732952792686, value=48.4
 07461                               column=location_info:country, timestamp=1732952792686, value=Iraq
 07461                               column=timestamps:updated_time, timestamp=1732952792686, value=22/6/2024 16:45
 07461                               column=weather_data:temperature_celsius, timestamp=1732952792686, value=49.1
 08426                               column=location_info:country, timestamp=1732952792686, value=Kuwait
 08426                               column=timestamps:updated_time, timestamp=1732952792686, value=27/6/2024 16:15
 08426                               column=weather_data:temperature_celsius, timestamp=1732952792686, value=48.9
 09401                               column=location_info:country, timestamp=1732952792686, value=Kuwait
 09401                               column=timestamps:updated_time, timestamp=1732952792686, value=2/7/2024 16:15
 09401                               column=weather_data:temperature_celsius, timestamp=1732952792686, value=48.1
 09976                               column=location_info:country, timestamp=1732952792686, value=Iraq
 09976                               column=timestamps:updated_time, timestamp=1732952792686, value=5/7/2024 15:45
 09976                               column=weather_data:temperature_celsius, timestamp=1732952792686, value=48.3
 11146                               column=location_info:country, timestamp=1732952792686, value=Iraq
 11146                               column=timestamps:updated_time, timestamp=1732952792686, value=11/7/2024 15:45
 11146                               column=weather_data:temperature_celsius, timestamp=1732952792686, value=49.1
 11339                               column=location_info:country, timestamp=1732952792686, value=Iraq
 11339                               column=timestamps:updated_time, timestamp=1732952792686, value=12/7/2024 15:45
 11339                               column=weather_data:temperature_celsius, timestamp=1732952792686, value=48.8
 No                                  column=location_info:country, timestamp=1732952792686, value=country
 No                                  column=timestamps:updated_time, timestamp=1732952792686, value=updated_time
 No                                  column=weather_data:temperature_celsius, timestamp=1732952792686, value=temperature_celsius
9 row(s) in 0.0730 seconds
```

Discussion: From the query, the table has been specifies to only showing country, temperature, and the timestamps. Also, by using the filter of 'SingleColumnValueFilter', the output generated will be only showing for temperature that is more than 47.9 . From this output, we can see that the country that have highest temperature recorded was in Kuwait with 49.2 degrees celcius while the second highest temperature recorded was in Iraq with 49.1 degrees. This extreme temperature indicated that some region are experiencing unpredictable heat waves, which a clear sign of global warming. Next to get deeper understanding, we retrieve the row of both country using **get** command.

```
hbase(main):003:0> get 'Weather_Data','06897'
COLUMN                               CELL
 air_quality:air_quality_Carbo       timestamp=1732952792686, value=00208.6
 n_Monoxide
 air_quality:air_quality_Nitro       timestamp=1732952792686, value=003.7
 gen_dioxide
 air_quality:air_quality_PM10        timestamp=1732952792686, value=187.0
 air_quality:air_quality_PM2.5       timestamp=1732952792686, value=033.6
 air_quality:air_quality_Sulph       timestamp=1732952792686, value=000.8
 ur_dioxide
 location_info:country               timestamp=1732952792686, value=Kuwait
 location_info:location_name         timestamp=1732952792686, value=Kuwait City
 timestamps:Date                     timestamp=1732952792686, value=19/6/2024
 timestamps:Time                     timestamp=1732952792686, value=16:45
 timestamps:updated_time             timestamp=1732952792686, value=19/6/2024 16:45
 weather_data:humidity               timestamp=1732952792686, value=004
 weather_data:temperature_cels       timestamp=1732952792686, value=49.2
 ius
12 row(s) in 0.0810 seconds
```

```
hbase(main):004:0> get 'Weather_Data','07461'
COLUMN                          CELL
 air_quality:air_quality_Carbo  timestamp=1732952792686, value=00208.6
 n_Monoxide
 air_quality:air_quality_Nitro  timestamp=1732952792686, value=003.8
 gen_dioxide
 air_quality:air_quality_PM10   timestamp=1732952792686, value=047.9
 air_quality:air_quality_PM2.5  timestamp=1732952792686, value=016.1
 air_quality:air_quality_Sulph  timestamp=1732952792686, value=002.3
 ur_dioxide
 location_info:country          timestamp=1732952792686, value=Iraq
 location_info:location_name    timestamp=1732952792686, value=Baghdad
 timestamps:Date                timestamp=1732952792686, value=22/6/2024
 timestamps:Time                timestamp=1732952792686, value=16:45
 timestamps:updated_time        timestamp=1732952792686, value=22/6/2024 16:45
 weather_data:humidity          timestamp=1732952792686, value=008
 weather_data:temperature_cels  timestamp=1732952792686, value=49.1
 ius
12 row(s) in 0.0250 seconds
```

Looking deeper onto this 2 rows using get command, we can see that the effected city for the heat waves is Kuwait City and Baghdad,Iraq .

3.1.2 Trend of Malaysia Temperature Changes

1.**Scan** : To obtain weather data table with column of country, temperature ,and the date and time. Also use filter to only show Malaysia country.

```
hbase(main):002:0> scan 'Weather_Data', {COLUMNS => ['location_info:country', 'w
eather_data:temperature_celsius','timestamps:updated_time'], FILTER => "SingleCo
lumnValueFilter('location_info', 'country', =, 'substring:Malaysia')"}
```

Output :

Since the output have 190 Rows , the screenshot shows below only show the highest temperature recorded for each month from May 2024 to November 2024.

May 2024

```
03024                    column=location_info:country, timestamp=1732952792686, value=Malaysia
03024                    column=timestamps:updated_time, timestamp=1732952792686, value=30/5/2024 21:45
03024                    column=weather_data:temperature_celsius, timestamp=1732952792686, value=31.0
```

June 2024

```
07871                    column=location_info:country, timestamp=1732952792686, value=Malaysia
07871                    column=timestamps:updated_time, timestamp=1732952792686, value=24/6/2024 21:45
07871                    column=weather_data:temperature_celsius, timestamp=1732952792686, value=30.4
```

July 2024

```
12483                    column=location_info:country, timestamp=1732952792686, value=Malaysia
12483                    column=timestamps:updated_time, timestamp=1732952792686, value=18/7/2024 20:45
12483                    column=weather_data:temperature_celsius, timestamp=1732952792686, value=32.2
```

August 2024

```
20257                    column=location_info:country, timestamp=1732952792686, value=Malaysia
20257                    column=timestamps:updated_time, timestamp=1732952792686, value=29/8/2024 20:00
20257                    column=weather_data:temperature_celsius, timestamp=1732952792686, value=32.1
```

September 2024

```
25131                    column=location_info:country, timestamp=1732952792686, value=Malaysia
25131                    column=timestamps:updated_time, timestamp=1732952792686, value=23/9/2024 17:45
25131                    column=weather_data:temperature_celsius, timestamp=1732952792686, value=33.1
```

October 2024

```
26886                    column=location_info:country, timestamp=1732952792686, value=Malaysia
26886                    column=timestamps:updated_time, timestamp=1732952792686, value=2/10/2024 17:30
26886                    column=weather_data:temperature_celsius, timestamp=1732952792686, value=34.3
```

November 2024

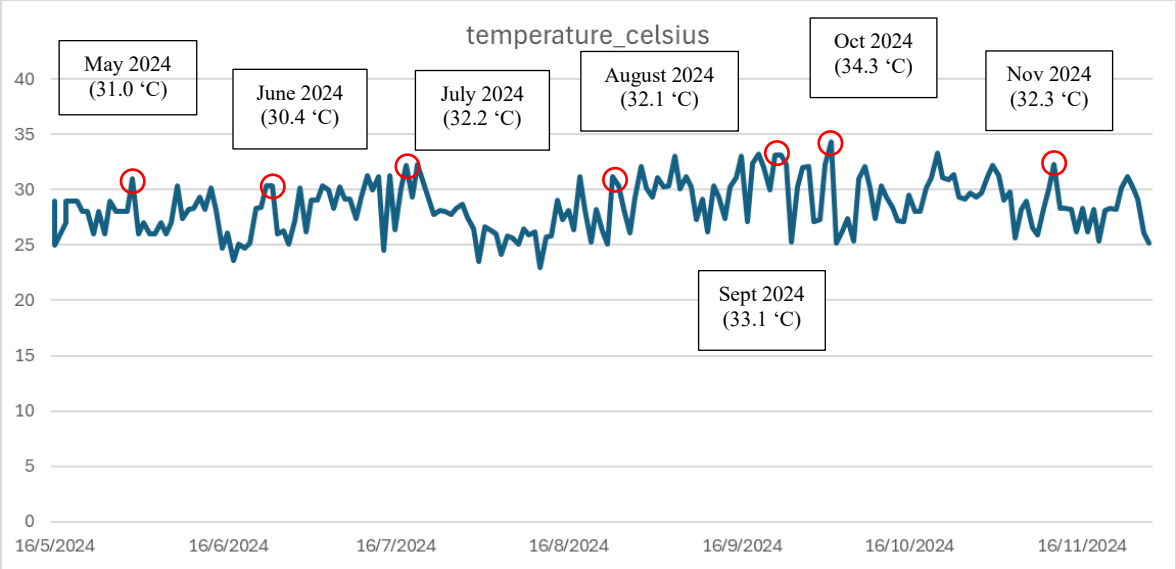
```
34628                    column=location_info:country, timestamp=1732952792686, value=Malaysia
34628                    column=timestamps:updated_time, timestamp=1732952792686, value=11/11/2024 17:00
34628                    column=weather_data:temperature_celsius, timestamp=1732952792686, value=32.3
```

Discussion : The highest temperature data from each month from May to November 2024 in Malaysia reflects the potential impact of climate change. There is a positive increment in warming particularly from July to October. With the peak temperature achieved 34.3 C in October, this indicates the result of global warming to Malaysia. While moderate temperature in May and June are alligned with the wind mooson cooling effect. These patterns indicate the importance of adressing climate change, as an increase in temperature could lead to many consequences in every aspect.

2. **Echo** : Export the output for easier trend visualization

```
echo "scan 'Weather_Data', {COLUMNS => ['location_info:country',
'weather_data:temperature_celsius','timestamps:updated_time'], FILTER => "SingleColumnValueFilter('location_info',
'country', =, 'substring:Malaysia')"} " | hbase shell >  malaysia_temps.txt
```

Output :



Discussion :

This graph shows the trend for Malaysia temperatures changes showing there is clear trend of increasing and sustained heat with the peak temperature of 34.3°C in October. The slow and steady heat rise indicates the intensity of summer conditions. With october temperature reaching the highest to 34. 3°C, it surpasses the typical seasonal expectations. Despite the transition to November, the temperature still elevates at 32.3 °C indicating summer season has prolonged. This pattern of rising and sustaining peak temperature showcase to use that this is the impact of climat change posing significant risk to public health, ecosystem, and energy system. This highlights the urgency of climates adaptation efforts are needed.

3.2 AIR QUALITY INDEX

3.1.1 Discovering Countries with high Air Quality Index (Pm2.5) and its Causes.

1.**Scan** the table of weather data and its air quality column together with time and location.Filter use to set a minimum of 500 micrograms per cubic meter.

```
hbase(main):011:0> scan 'Weather_Data', {COLUMNS => ['location_info:country', 'air_quality:air_quality_PM2.5', 'timestamps:updated_t
ime'], FILTER => "SingleColumnValueFilter('air_quality', 'air_quality_PM2.5', >, 'binary:500.0')"}
ROW                                  COLUMN+CELL
 00231                               column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=714.1
 00231                               column=location_info:country, timestamp=1732952792686, value=Chile
 00231                               column=timestamps:updated_time, timestamp=1732952792686, value=16/5/2024 10:00
 00469                               column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=798.5
 00469                               column=location_info:country, timestamp=1732952792686, value=Indonesia
 00469                               column=timestamps:updated_time, timestamp=1732952792686, value=17/5/2024 23:00
 00618                               column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=554.7
 00618                               column=location_info:country, timestamp=1732952792686, value=Chile
 00618                               column=timestamps:updated_time, timestamp=1732952792686, value=18/5/2024 10:30
 01007                               column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=502.1
 01007                               column=location_info:country, timestamp=1732952792686, value=Chile
 01007                               column=timestamps:updated_time, timestamp=1732952792686, value=20/5/2024 10:30
 01245                               column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=874.9
 01245                               column=location_info:country, timestamp=1732952792686, value=Indonesia
 01245                               column=timestamps:updated_time, timestamp=1732952792686, value=21/5/2024 21:45
 01397                               column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=579.3
 01397                               column=location_info:country, timestamp=1732952792686, value=Chile
 01397                               column=timestamps:updated_time, timestamp=1732952792686, value=22/5/2024 10:15
 01787                               column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=557.9
 01787                               column=location_info:country, timestamp=1732952792686, value=Chile
 01787                               column=timestamps:updated_time, timestamp=1732952792686, value=24/5/2024 10:30
 01982                               column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=608.0
 01982                               column=location_info:country, timestamp=1732952792686, value=Chile
 01982                               column=timestamps:updated_time, timestamp=1732952792686, value=25/5/2024 10:00
 02567                               column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=981.8
 02567                               column=location_info:country, timestamp=1732952792686, value=Chile
 02567                               column=timestamps:updated_time, timestamp=1732952792686, value=28/5/2024 10:30
 02762                               column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=767.1
 02762                               column=location_info:country, timestamp=1732952792686, value=Chile
 02762                               column=timestamps:updated_time, timestamp=1732952792686, value=29/5/2024 10:00
 02805                               column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=659.9
 02805                               column=location_info:country, timestamp=1732952792686, value=Indonesia
 02805                               column=timestamps:updated_time, timestamp=1732952792686, value=29/5/2024 21:00
 03194                               column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=650.2
 03194                               column=location_info:country, timestamp=1732952792686, value=Indonesia
 03194                               column=timestamps:updated_time, timestamp=1732952792686, value=31/5/2024 21:15
 03346                               column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=843.7
 03346                               column=location_info:country, timestamp=1732952792686, value=Chile
 03346                               column=timestamps:updated_time, timestamp=1732952792686, value=1/6/2024 10:15
 03541                               column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=646.8
 03541                               column=location_info:country, timestamp=1732952792686, value=Chile
```

Discussion :

From the output, it is observed that Chile and Indonesia having the highest Pm2.5 air quality recorded exceeding 500 µg/m³, the factors to this can be study from the overall data quality of each data record. 2 Query is use to obtain the overall column information for both of this row to have deeper understanding of the data and its posible causes.

2. **Get** = Obtain the overall row of the highest value of PM2.5 for Indonesia(Row= 01245) and Chile (Row=02567) to understand better on overall air quality data.

For Indonesia :

```
hbase(main):012:0> get 'Weather_Data','01245'
COLUMN                                CELL
 air_quality:air_quality_Carbon_Mo    timestamp=1732952792686, value=37597.7
 noxide
 air_quality:air_quality_Nitrogen_    timestamp=1732952792686, value=252.3
 dioxide
 air_quality:air_quality_PM10         timestamp=1732952792686, value=1072.8
 air_quality:air_quality_PM2.5        timestamp=1732952792686, value=874.9
 air_quality:air_quality_Sulphur_d    timestamp=1732952792686, value=150.7
 ioxide
 location_info:country                timestamp=1732952792686, value=Indonesia
 location_info:location_name          timestamp=1732952792686, value=Jakarta
 timestamps:Date                      timestamp=1732952792686, value=21/5/2024
 timestamps:Time                      timestamp=1732952792686, value=21:45
 timestamps:updated_time              timestamp=1732952792686, value=21/5/2024 21:45
 weather_data:humidity                timestamp=1732952792686, value=079
 weather_data:temperature_celsius     timestamp=1732952792686, value=29.0
12 row(s) in 0.0160 seconds
```

For Chile :

```
hbase(main):013:0> get 'Weather_Data','02567'
COLUMN                                CELL
 air_quality:air_quality_Carbon_Mo    timestamp=1732952792686, value=10467.5
 noxide
 air_quality:air_quality_Nitrogen_    timestamp=1732952792686, value=175.5
 dioxide
 air_quality:air_quality_PM10         timestamp=1732952792686, value=1167.0
 air_quality:air_quality_PM2.5        timestamp=1732952792686, value=981.8
 air_quality:air_quality_Sulphur_d    timestamp=1732952792686, value=070.6
 ioxide
 location_info:country                timestamp=1732952792686, value=Chile
 location_info:location_name          timestamp=1732952792686, value=Santiago
 timestamps:Date                      timestamp=1732952792686, value=28/5/2024
 timestamps:Time                      timestamp=1732952792686, value=10:30
 timestamps:updated_time              timestamp=1732952792686, value=28/5/2024 10:30
 weather_data:humidity                timestamp=1732952792686, value=087
 weather_data:temperature_celsius     timestamp=1732952792686, value=07.0
12 row(s) in 0.0030 seconds
```

Discussion :

Based on the output above for both Indonesia and chile, both air quality are influenced by the Carbon Monoxide which is 37597.7 µg/m³ for Indonesia and 10467.5 µg/m³ for Chile. In Indonesia , this are probably driven by slash-and-burn agriculture, urban emissions, and forest fires, worsened by prolonged dry seasons linked to global warming. While for Chile,

this is also influenced by the geographic location of Santiago,Chile which in position between two mountains ranges, the Andes and Cordilera de la Costa. Air pocket is created where stale air accumulates and takes longer to disperse. This is also influnced by their industrial activities, wildfires, and the air pocket which trap pollutants near the area. These concerning pollution levels highlight the urgent need for focused mitigation measures, such as more stringent emission regulations, sustainable land management techniques, and more robust adaption plans to safeguard ecosystems and public health against the effects of climate change.

3.1.2 Malaysia Air Quality Trends

1**. Scan**: To display all Pm2.5 air quality values that is in Malaysia together with the date and time and the coutry column.

```
hbase(main):017:0> scan 'Weather_Data', {COLUMNS => ['location_info:country', 'air_quality:air_quality_PM2.5','timestamps:updated_t
ime'], FILTER => "SingleColumnValueFilter('location_info', 'country', =, 'substring:Malaysia')"}
ROW                             COLUMN+CELL
 00104                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=040.5
 00104                          column=location_info:country, timestamp=1732952792686, value=Malaysia
 00104                          column=timestamps:updated_time, timestamp=1732952792686, value=16/5/2024 16:45
 00299                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=131.5
 00299                          column=location_info:country, timestamp=1732952792686, value=Malaysia
 00299                          column=timestamps:updated_time, timestamp=1732952792686, value=16/5/2024 22:00
 00494                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=094.3
 00494                          column=location_info:country, timestamp=1732952792686, value=Malaysia
 00494                          column=timestamps:updated_time, timestamp=1732952792686, value=18/5/2024 0:00
 00686                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=054.2
 00686                          column=location_info:country, timestamp=1732952792686, value=Malaysia
 00686                          column=timestamps:updated_time, timestamp=1732952792686, value=18/5/2024 22:30
 00880                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=070.8
 00880                          column=location_info:country, timestamp=1732952792686, value=Malaysia
 00880                          column=timestamps:updated_time, timestamp=1732952792686, value=19/5/2024 22:00
 01075                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=136.0
 01075                          column=location_info:country, timestamp=1732952792686, value=Malaysia
 01075                          column=timestamps:updated_time, timestamp=1732952792686, value=20/5/2024 22:45
 01270                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=129.9
 01270                          column=location_info:country, timestamp=1732952792686, value=Malaysia
 01270                          column=timestamps:updated_time, timestamp=1732952792686, value=21/5/2024 22:45
 01465                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=104.0
 01465                          column=location_info:country, timestamp=1732952792686, value=Malaysia
 01465                          column=timestamps:updated_time, timestamp=1732952792686, value=22/5/2024 22:15
 01660                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=111.7
 01660                          column=location_info:country, timestamp=1732952792686, value=Malaysia
 01660                          column=timestamps:updated_time, timestamp=1732952792686, value=23/5/2024 22:00
 01855                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=051.9
 01855                          column=location_info:country, timestamp=1732952792686, value=Malaysia
 01855                          column=timestamps:updated_time, timestamp=1732952792686, value=24/5/2024 22:30
 02050                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=034.4
 02050                          column=location_info:country, timestamp=1732952792686, value=Malaysia
 02050                          column=timestamps:updated_time, timestamp=1732952792686, value=25/5/2024 22:00
 02245                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=026.0
 02245                          column=location_info:country, timestamp=1732952792686, value=Malaysia
 02245                          column=timestamps:updated_time, timestamp=1732952792686, value=26/5/2024 22:15
 02440                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=075.6
 02440                          column=location_info:country, timestamp=1732952792686, value=Malaysia
 02440                          column=timestamps:updated_time, timestamp=1732952792686, value=27/5/2024 22:30
 02635                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=124.4
 02635                          column=location_info:country, timestamp=1732952792686, value=Malaysia
```

*Overall Row is 196 .

2. **Scan**: Since the data is too long , we filter the value to data that is above 55.5 µg/m3 (Unhealthy Level) for analysing .

```
hbase(main):026:0> scan 'Weather_Data', {COLUMNS => ['location_info:country', 'air_quality:air_quality_PM2.5', 'timestamps:updated_
time'], FILTER => "(SingleColumnValueFilter('location_info', 'country', =, 'substring:Malaysia') AND SingleColumnValueFilter('air_q
uality', 'air_quality_PM2.5', >, 'binary:055.5'))"}
ROW                              COLUMN+CELL
 00299                           column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=131.5
 00299                           column=location_info:country, timestamp=1732952792686, value=Malaysia
 00299                           column=timestamps:updated_time, timestamp=1732952792686, value=16/5/2024 22:00
 00494                           column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=094.3
 00494                           column=location_info:country, timestamp=1732952792686, value=Malaysia
 00494                           column=timestamps:updated_time, timestamp=1732952792686, value=18/5/2024 0:00
 00880                           column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=070.8
 00880                           column=location_info:country, timestamp=1732952792686, value=Malaysia
 00880                           column=timestamps:updated_time, timestamp=1732952792686, value=19/5/2024 22:00
 01075                           column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=136.0
 01075                           column=location_info:country, timestamp=1732952792686, value=Malaysia
 01075                           column=timestamps:updated_time, timestamp=1732952792686, value=20/5/2024 22:45
 01270                           column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=129.9
 01270                           column=location_info:country, timestamp=1732952792686, value=Malaysia
 01270                           column=timestamps:updated_time, timestamp=1732952792686, value=21/5/2024 22:45
 01465                           column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=104.0
 01465                           column=location_info:country, timestamp=1732952792686, value=Malaysia
 01465                           column=timestamps:updated_time, timestamp=1732952792686, value=22/5/2024 22:15
 01660                           column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=111.7
 01660                           column=location_info:country, timestamp=1732952792686, value=Malaysia
 01660                           column=timestamps:updated_time, timestamp=1732952792686, value=23/5/2024 22:00
 02440                           column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=075.6
 02440                           column=location_info:country, timestamp=1732952792686, value=Malaysia
 02440                           column=timestamps:updated_time, timestamp=1732952792686, value=27/5/2024 22:30
 02635                           column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=124.4
 02635                           column=location_info:country, timestamp=1732952792686, value=Malaysia
 02635                           column=timestamps:updated_time, timestamp=1732952792686, value=28/5/2024 22:30
 02830                           column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=097.2
 02830                           column=location_info:country, timestamp=1732952792686, value=Malaysia
 02830                           column=timestamps:updated_time, timestamp=1732952792686, value=29/5/2024 22:00
 03609                           column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=085.2
 03609                           column=location_info:country, timestamp=1732952792686, value=Malaysia
```

The filtered data for Malaysia shows PM2.5 values consistently exceeding the unhealthy threshold of 55.5 µg/m³, with readings reaching as high as 136.0 µg/m³ on May 19, 2024, and 129.9 µg/m³ on May 20, 2024. These elevated levels suggest persistent and severe air pollution, likely caused by urban emissions, industrial activities, and regional transboundary haze from forest fires. Such air quality levels pose significant health risks, including respiratory and cardiovascular issues, while also reflecting broader environmental challenges exacerbated by climate change. To better understand its cause, we obtain the highest value recorded first to analyse from that.

3. **Scan**: Obtain the highest Pm2.5 value in Malaysia by using filter to only show value that is more than 200 .

```
hbase(main):006:0> scan 'Weather_Data', {COLUMNS => ['location_info:country', 'air_quality:air_quality_PM2.5', 'timestamps:updated_
time'], FILTER => "(SingleColumnValueFilter('location_info', 'country', =, 'substring:Malaysia') AND SingleColumnValueFilter('air_q
uality', 'air_quality_PM2.5', >, 'binary:200'))"}
ROW                           COLUMN+CELL
 03999                        column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=241.0
 03999                        column=location_info:country, timestamp=1732952792686, value=Malaysia
 03999                        column=timestamps:updated_time, timestamp=1732952792686, value=4/6/2024 22:15
 04194                        column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=247.6
 04194                        column=location_info:country, timestamp=1732952792686, value=Malaysia
 04194                        column=timestamps:updated_time, timestamp=1732952792686, value=5/6/2024 22:15
 05557                        column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=376.0
 05557                        column=location_info:country, timestamp=1732952792686, value=Malaysia
 05557                        column=timestamps:updated_time, timestamp=1732952792686, value=12/6/2024 22:15
 06720                        column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=207.5
 06720                        column=location_info:country, timestamp=1732952792686, value=Malaysia
 06720                        column=timestamps:updated_time, timestamp=1732952792686, value=18/6/2024 22:00
 08634                        column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=387.8
 08634                        column=location_info:country, timestamp=1732952792686, value=Malaysia
 08634                        column=timestamps:updated_time, timestamp=1732952792686, value=28/6/2024 21:15
 09219                        column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=239.4
 09219                        column=location_info:country, timestamp=1732952792686, value=Malaysia
 09219                        column=timestamps:updated_time, timestamp=1732952792686, value=1/7/2024 21:15
 09414                        column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=291.4
 09414                        column=location_info:country, timestamp=1732952792686, value=Malaysia
 09414                        column=timestamps:updated_time, timestamp=1732952792686, value=2/7/2024 21:15
 10974                        column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=203.0
 10974                        column=location_info:country, timestamp=1732952792686, value=Malaysia
 10974                        column=timestamps:updated_time, timestamp=1732952792686, value=10/7/2024 21:00
 18307                        column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=242.2
 18307                        column=location_info:country, timestamp=1732952792686, value=Malaysia
 18307                        column=timestamps:updated_time, timestamp=1732952792686, value=19/8/2024 20:30
9 row(s) in 0.1200 seconds
```

As we can observed here, the query reveals there are significantly some data recorded above 200 µg/m³, on multiple dates between June and August 2024. The highest recorded value was 387.8 µg/m³ recorded on 28 June 2024. Query below shows the overall air quality for this particular row to better understand it cause.

4. **Get** : Detail air quality index on the highest pm2.5 recorded row (08634)

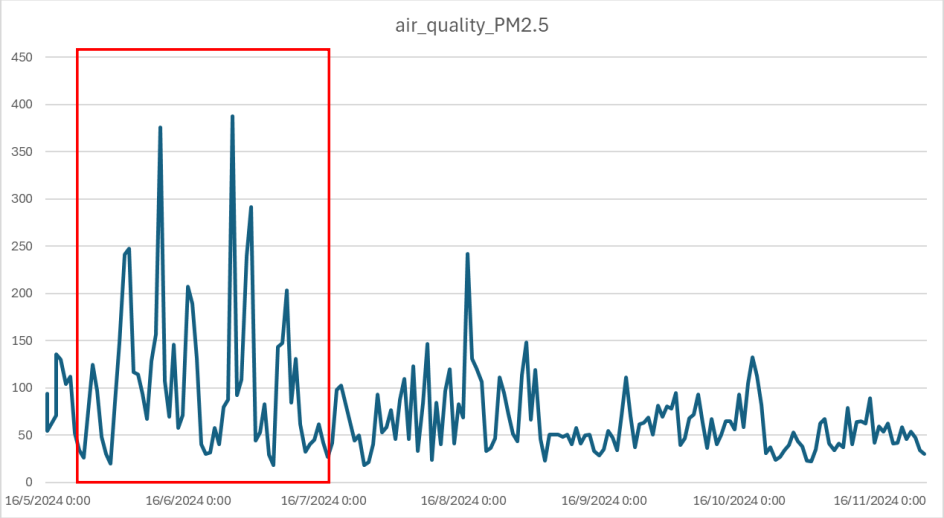
```
hbase(main):002:0> get 'Weather_Data' ,'08634'
COLUMN                                CELL
 air_quality:air_quality_Carbon_M     timestamp=1732952792686, value=10040.3
 onoxide
 air_quality:air_quality_Nitrogen     timestamp=1732952792686, value=224.8
 _dioxide
 air_quality:air_quality_PM10         timestamp=1732952792686, value=419.4
 air_quality:air_quality_PM2.5        timestamp=1732952792686, value=387.8
 air_quality:air_quality_Sulphur_     timestamp=1732952792686, value=064.9
 dioxide
 location_info:country                timestamp=1732952792686, value=Malaysia
 location_info:location_name          timestamp=1732952792686, value=Kuala Lumpur
 timestamps:Date                      timestamp=1732952792686, value=28/6/2024
 timestamps:Time                      timestamp=1732952792686, value=21:15
 timestamps:updated_time              timestamp=1732952792686, value=28/6/2024 21:15
 weather_data:humidity                timestamp=1732952792686, value=089
 weather_data:temperature_celsius     timestamp=1732952792686, value=27.2
12 row(s) in 0.2440 seconds
```

As we can see here, the highest value recorded is Carbon Monoxide, which could indicate the activity of Fuel-burning appliances, Motor vehicles, including power plants and wildfire.

5. **Echo** : Export output file , this is the overall pattern of the air quality in Malaysia from May- November 2024

Export the output using echo command

```
[cloudera@quickstart Desktop]$ echo "scan 'Weather_Data', {COLUMNS => ['location_info:country', 'air_quality:air_quality_PM2.5','ti
mestamps:updated_time'], FILTER => \"SingleColumnValueFilter('location_info', 'country', =, 'substring:Malaysia')\"}" | hbase shell
 > malaysia_airquality.txt
2024-11-30 16:38:12,545 INFO  [main] Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.availab
le
```



Discussion: The air quality trend in Malaysia from May to November 2024 reveals a distinct pattern of fluctuating PM2.5 levels, with several severe spikes indicating unhealthy to hazardous air quality during specific periods. The highest peaks were observed in **June and July**, coinciding with the dry season, which is often associated with **transboundary haze from forest fires in the region (Indonesia slash and burn activity) . To study this, we obtain the pm2.5 value in Indonesia in the query below to confirm the statement**. After July, the air quality gradually improved, showing fewer extreme spikes, likely due to seasonal changes and rainfall reducing airborne particulates. However, occasional moderate increases were still observed in August and October, indicating persistent localized sources of pollution such as industrial emissions or agricultural burning. Overall, while the trend shows some improvement toward the latter months, it underscores the critical need for proactive measures to mitigate pollution during high-risk periods.

6. **Scan** : To confirm the statement of slash and burn activity in Indonesia, we obtained the overall air quality value of Neighbouring country which is Indonesia.

The Query :

```
hbase(main):005:0> scan 'Weather_Data', {COLUMNS => ['location_info:country', 'air_quality:air_quality_PM2.5','timest
amps:updated_time'], FILTER => "SingleColumnValueFilter('location_info', 'country', =, 'substring:Indonesia')"}
```

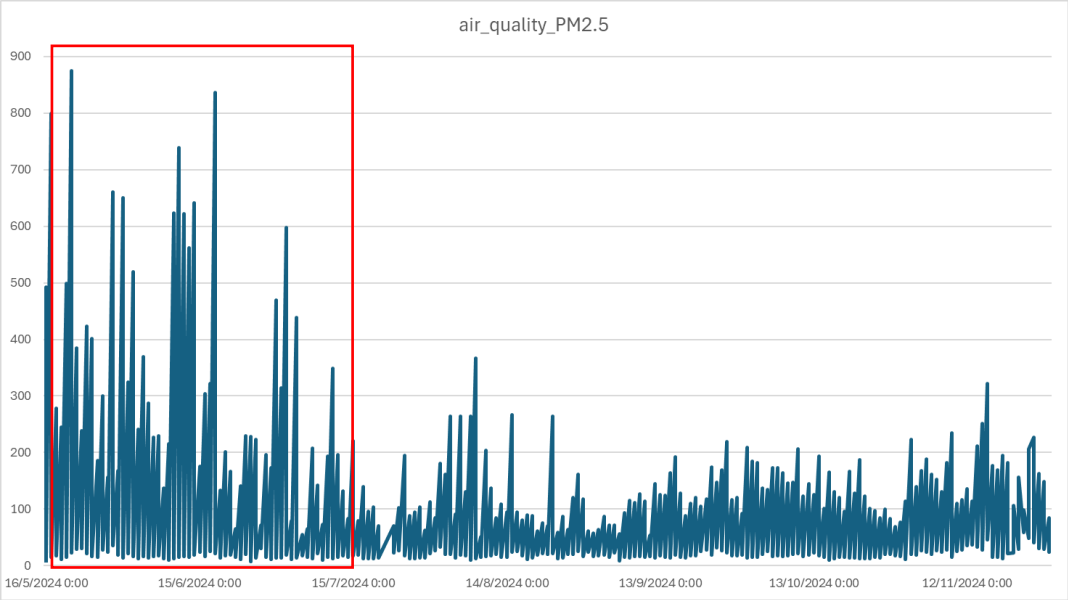
Output Snippet :

```
 37138                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=206.8
 37138                          column=location_info:country, timestamp=1732952792686, value=Indonesia
 37138                          column=timestamps:updated_time, timestamp=1732952792686, value=24/11/2024 15:30
 37152                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=048.8
 37152                          column=location_info:country, timestamp=1732952792686, value=Indonesia
 37152                          column=timestamps:updated_time, timestamp=1732952792686, value=24/11/2024 15:15
 37333                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=227.4
 37333                          column=location_info:country, timestamp=1732952792686, value=Indonesia
 37333                          column=timestamps:updated_time, timestamp=1732952792686, value=25/11/2024 15:30
 37347                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=040.5
 37347                          column=location_info:country, timestamp=1732952792686, value=Indonesia
 37347                          column=timestamps:updated_time, timestamp=1732952792686, value=25/11/2024 15:45
 37528                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=162.4
 37528                          column=location_info:country, timestamp=1732952792686, value=Indonesia
 37528                          column=timestamps:updated_time, timestamp=1732952792686, value=26/11/2024 16:00
 37542                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=030.5
 37542                          column=location_info:country, timestamp=1732952792686, value=Indonesia
 37542                          column=timestamps:updated_time, timestamp=1732952792686, value=26/11/2024 16:00
 37723                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=148.6
 37723                          column=location_info:country, timestamp=1732952792686, value=Indonesia
 37723                          column=timestamps:updated_time, timestamp=1732952792686, value=27/11/2024 15:45
 37737                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=029.8
 37737                          column=location_info:country, timestamp=1732952792686, value=Indonesia
 37737                          column=timestamps:updated_time, timestamp=1732952792686, value=27/11/2024 15:45
 37918                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=084.2
 37918                          column=location_info:country, timestamp=1732952792686, value=Indonesia
 37918                          column=timestamps:updated_time, timestamp=1732952792686, value=28/11/2024 17:45
 37932                          column=air_quality:air_quality_PM2.5, timestamp=1732952792686, value=024.4
 37932                          column=location_info:country, timestamp=1732952792686, value=Indonesia
 37932                          column=timestamps:updated_time, timestamp=1732952792686, value=28/11/2024 18:00
393 row(s) in 0.6330 seconds
```

7. **Echo** :Export the output file .

```
[cloudera@quickstart Desktop]$ echo "scan 'Weather_Data', {COLUMNS => ['location_info:country', 'air_quality:air_qual
ity_PM2.5', 'timestamps:updated_time'], FILTER => \"SingleColumnValueFilter('location_info', 'country', =, 'substring
:Indonesia')\"}" | hbase shell > Indonesia_airquality.txt
2024-11-30 18:05:22,020 INFO  [main] Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.nati
ve.lib.available
```

Output:



Discussion :

This graph shows that the Pm2.5 levels in Indonesia, revealing its significantly high value when compared to Malaysia in the same period. During highlighted period, Indonesia air quality of pm2.5 frequently exceed 500 µg/m³, with peaks achieving 800µg/m³. These extreme levels indicate their severe air pollution are likely driven by its slash and burn agricultural practices during this dry period.

When putting both graph side by side, a clear connection emerge. This correlation between Malaysia and Indonesia air quality suggest that the transboundary haze originating from Indonesia slash and burn practices significantly impacts Malaysia air Quality.

In conclusion, there is a highlights on the need for collaborative efforts to adress slash and burn practices between these countries, as it not only effect local environment in Indonesia but also pose significant threat of environmental for their neighbouring countries like Malaysia