

Corteva Assessment - Data Engineer

By: Ronak Sankaranarayanan

Part 1:

Database Modeling:

Yield Data model:

Column Name	Data Type	Constraints	Description
Y_YEAR	INT	Primary Key	Year of the production
YIELD	INT	YIELD >= 0	Yield produced in the year in USA

Weather Data model:

Column Name	Data Type	Constraints	Description
W_DATE	DATE	Primary Key	-
MAX_TEMP	INT	MAX_TEMP >= -2731	Maximum temperature recorded on that day.
MIN_TEMP	INT	MIN_TEMP >= -2731	Minimum temperature recorded on that day.
PRECIPITATION	INT	PRECIPITATION >= 0	Precipitation recorded on that day
STATION_ID	VARCHAR(11)	Primary Key	Station ID

Log Data model:

Column Name	Data Type	Constraints	Description
START_TIME	DATETIME(4)	Milliseconds to the decimal 4.	Time before Ingestion starts
END_TIME	DATETIME(4)	Milliseconds to the decimal 4.	Time after Ingestion ends
ROWS_AFFECTED	INT		Number of rows ingested.

The Tables in the Database corteva_database have been created according to the model mentioned above.

Data Ingestion:

Yield Data:

Procedure:

- 1) All the files under the directory the target have been stored in a Temporary Table #Temp_files
- 2) All the files other than .txt are removed
- 3) A while loop is initiated with the condition #Temp_files is not empty
 - a) The File on top of the table is taken
 - b) BULK INSERT is used to read the data from RAW TXT file and into another Temporary File #Temp_yield
 - c) Data is loaded from #Temp_yield into YIELD with removal conditions of Null data.
 - d) Log has been recorded with start time, end time, and rows affected and Inserted into Log Table
 - e) #Temp_yield is truncated
 - f) The current file is removed from the #Temp_files

Data Loading Yield...LAPTOP\ronak (70) X

```

BEGIN CATCH
    PRINT 'Failed processing : ' + @FILENAME
END CATCH

DELETE FROM #TEMP_FILES WHERE FileName = @FILENAME
TRUNCATE TABLE #TEMP_YIELD

END

SELECT * FROM YIELD;

SELECT * FROM LOG_YIELD;

```

table CORTEVA_DATABASE.dbo.LOG_YIELD

110 %

Results Messages

	Y_YEAR	YIELD
1	1985	225447
2	1986	208944
3	1987	181143
4	1988	125194
5	1989	191320
6	1990	201534
7	1991	189868
8	1992	240719
9	1993	160086

	START_TIME	END_TIME	ROWS_AFFECTED
1	2022-12-18 19:24:49.4933	2022-12-18 19:24:49.4933	30

Query executed suc... RONAK_LAPTOP\MSSQLSERVER01 ... RONAK_LAPTOP\ronak (70) CORTEVA_DATABASE 00:00:00 31 rows

Weather Data Ingestion:

Procedure:

- 1) All the files under the directory the target have been stored in a Temporary Table #Temp_files
- 2) All the files other than .txt are removed
- 3) A while loop is initiated with the condition #Temp_files is not empty
 - a) The File on top of the table is taken
 - b) BULK INSERT is used to read the data from RAW TXT file and into another Temporary File #Temp_weather
 - c) Data is loaded from #Temp_weather into WEATHER with removal conditions of Null data.
 - d) Log has been recorded with start time, end time, and rows affected and Inserted into Log Table
 - e) #Temp_weather is truncated
 - f) The current file is removed from the #Temp_files

The screenshot shows a SQL Server Enterprise Manager window titled "Data Loading Weat...LAPTOP\ronak (56)". The script editor contains the following T-SQL code:

```

END TRY
BEGIN CATCH
    PRINT 'Failed processing : ' + @FILENAME
END CATCH
--Delete the top files and the next file will be on Top
DELETE FROM #TEMP_FILES WHERE FileName = @FILENAME

-- Clear the temporary table
TRUNCATE TABLE #TEMP_WEATHER
END
-- View Number of Rows currently in Weather

```

The Results pane shows the execution results of the script. The first result is a scalar value representing the number of rows present in the #TEMP_WEATHER table.

Num of Rows Present
1669107

The second result is a table showing the execution details of the script, including the start and end times and the number of rows affected by each statement.

	START_TIME	END_TIME	ROWS_AFFECTED
1	2022-12-18 19:27:19.5958	2022-12-18 19:27:19.6584	10421
2	2022-12-18 19:27:19.6604	2022-12-18 19:27:19.7112	10184
3	2022-12-18 19:27:19.7112	2022-12-18 19:27:19.7586	10588
4	2022-12-18 19:27:19.7586	2022-12-18 19:27:19.8109	9116
5	2022-12-18 19:27:19.8109	2022-12-18 19:27:19.8705	10411
6	2022-12-18 19:27:19.8705	2022-12-18 19:27:19.9208	10863
7	2022-12-18 19:27:19.9208	2022-12-18 19:27:19.9680	10204
8	2022-12-18 19:27:19.9680	2022-12-18 19:27:20.0155	9180
9	2022-12-18 19:27:20.0155	2022-12-18 19:27:20.0826	10856

NOTE:

A comment has been left on the places to change the directory path in the SQL file. Please check the comment and change the directory in those statements for successful execution.

All the SQL files are also stored inside SQL Commands in .txt form for quick review

Data Aggregation:

Average of Min and Max temperature:

A Select statement has been used to extract the data with aggregation functions AVG and SUM to find the Average of MIN_TEMP and MAX_TEMP and the TOTAL of PRECIPITATION per annum in each station.

Data Aggregation for MAX_TEMP

	YEAR	STATION_ID	AVG_MAX_TEMP
1	1985	USC00124837	141
2	1985	USC00254985	154
3	1994	USC00259090	165
4	1994	USC00112193	180
5	1995	USC00250435	169
6	1995	USC00258133	166
7	2001	USC00254985	174
8	2001	USC00112140	177

Data Aggregation for MIN_TEMP

	YEAR	STATION_ID	AVG_MIN_TEMP
1	1989	USC00116610	58
2	1989	USC00255310	19
3	1990	USC00116579	73
4	1990	USC00131635	56
5	1990	USC00137161	40
6	1990	USC00258915	37
7	1992	USC00115326	28
8	1992	USC00115515	61

Data Aggregation for Precipitation

	YEAR	STATION_ID	TOTAL_PRECIPITATION_TEMP
1	1985	USC00135796	9743
2	1993	USC00258480	9861
3	1993	USC00332791	8539
4	1994	USC00115326	9246
5	1994	USC00116446	9812
6	2000	USC00126580	14562
7	2000	USC00129080	13144
8	2000	USC00255470	4264

Data Aggregation on all tables

	YEAR	STATION_ID	AVG_MIN_TEMP	AVG_MAX_TEMP	TOTAL_PRECIPITATION_TEMP
1	1991	USC00112193	63	17	8045
2	2001	USC00112140	66	17	10034
3	2002	USC00257515	43	18	3629
4	2007	USC00119241	63	18	6144
5	1991	USC00252840	44	17	5307
6	1995	USC00124837	54	14	9420
7	1995	USC00113335	70	17	9415
8	2002	USC00258480	40	16	7401

REST API:

Django is used to create the REST API. The SQL Server is connected with the help of pyodbc with the database cortevea_database. A cursor is used to execute the SQL Statements at each function to return the JSON object.

/api/weather/ -

A simple select statement is used to extract all the data from the table WEATHER in the form of a list. The data is then looped to get a key-value pair for each data and added to a list.

The final list is then parsed into a JSON object and returned to the request using JsonResponse.

Key-Value pair structure :

```
{  
W_DATE : MMDDYYYY,  
MAX_TEMP : INT  
MIN_TEMP : INT  
PRECIPITATION : INT  
STATION_ID : String  
}
```

Final JSON Data :

```
{  
Weather_data : [data]  
}
```

```
127.0.0.1:8000/api/weather/
{"Weather data": "[{"W_DATE": "\u01011985", "MAX_TEMP": -2.2, "MIN_TEMP": -12.8, "PRECIPITATION": 94, "STATION_ID": "\u0000110072"}, {"W_DATE": "\u01011985", "MAX_TEMP": 18.3, "MIN_TEMP": -1.1, "PRECIPITATION": 94, "STATION_ID": "\u0000110187"}, {"W_DATE": "\u01011985", "MAX_TEMP": 1.1, "MIN_TEMP": -12.2, "PRECIPITATION": 348, "STATION_ID": "\u0000110338"}, {"W_DATE": "\u01011985", "MAX_TEMP": 4.4, "MIN_TEMP": -5.0, "PRECIPITATION": 295, "STATION_ID": "\u0000111280"}, {"W_DATE": "\u01011985", "MAX_TEMP": -1.1, "MIN_TEMP": -2.2, "PRECIPITATION": 203, "STATION_ID": "\u0000111436"}, {"W_DATE": "\u01011985", "MAX_TEMP": 6.1, "MIN_TEMP": -2.2, "PRECIPITATION": 484, "STATION_ID": "\u0000112140"}, {"W_DATE": "\u01011985", "MAX_TEMP": 4.4, "MIN_TEMP": -4.4, "PRECIPITATION": 328, "STATION_ID": "\u0000112193"}, {"W_DATE": "\u01011985", "MAX_TEMP": 0.0, "MIN_TEMP": -7.8, "PRECIPITATION": 178, "STATION_ID": "\u0000112348"}, {"W_DATE": "\u01011985", "MAX_TEMP": 12.8, "MIN_TEMP": -2.8, "PRECIPITATION": 18, "STATION_ID": "\u0000112483"}, {"W_DATE": "\u01011985", "MAX_TEMP": 0.0, "MIN_TEMP": -6.7, "PRECIPITATION": 345, "STATION_ID": "\u0000113335"}]"]}
```

/api/yield/

A simple select statement is used to extract all the data from the table YIELD in the form of a list. The data is then looped to get a key-value pair for each data and added to a list.

The final list is then parsed into a JSON object and returned to the request using JsonResponse.

Key-Value pair structure :

```
{
Y_YEAR : INT,
YIELD : INT
}
```

Final JSON Data :

```
{
Yield_data : [data]
}
```

```
127.0.0.1:8000/api/yield.json
{"Yield_Data": "[{"Y_YEAR": 1985, "YIELD": 225447}, {"Y_YEAR": 1986, "YIELD": 208944}, {"Y_YEAR": 1987, "YIELD": 181143}, {"Y_YEAR": 1988, "YIELD": 125194}, {"Y_YEAR": 1989, "YIELD": 191320}, {"Y_YEAR": 1990, "YIELD": 201534}, {"Y_YEAR": 1991, "YIELD": 189868}, {"Y_YEAR": 1992, "YIELD": 240719}, {"Y_YEAR": 1993, "YIELD": 160986}, {"Y_YEAR": 1994, "YIELD": 255295}, {"Y_YEAR": 1995, "YIELD": 187970}, {"Y_YEAR": 1996, "YIELD": 234518}, {"Y_YEAR": 1997, "YIELD": 233864}, {"Y_YEAR": 1998, "YIELD": 247882}, {"Y_YEAR": 1999, "YIELD": 239549}, {"Y_YEAR": 2000, "YIELD": 251854}, {"Y_YEAR": 2001, "YIELD": 241377}, {"Y_YEAR": 2002, "YIELD": 227767}, {"Y_YEAR": 2003, "YIELD": 256229}, {"Y_YEAR": 2004, "YIELD": 299876}, {"Y_YEAR": 2005, "YIELD": 282263}, {"Y_YEAR": 2006, "YIELD": 267503}, {"Y_YEAR": 2007, "YIELD": 331177}, {"Y_YEAR": 2008, "YIELD": 305911}, {"Y_YEAR": 2009, "YIELD": 331921}, {"Y_YEAR": 2010, "YIELD": 315618}, {"Y_YEAR": 2011, "YIELD": 312789}, {"Y_YEAR": 2012, "YIELD": 273192}, {"Y_YEAR": 2013, "YIELD": 351272}, {"Y_YEAR": 2014, "YIELD": 361091}]"}
```

/api/weather/year/station_id/

A Select statement with WHERE clause is used to extract the data from WEATHER table and filter them based on the year and station_id given by the user through the GET Request.

The data is then looped to get a key-value pair for each data and added to a list.

The final list is then parsed into a JSON object and returned to the request using JsonResponse.

Key-Value pair structure :

Final JSON Data :

Weather Data for year 1985 and Station id USC00110072

Weather Data for the year 1987 and station_id USC00110072

[illegible]

Weather Data for the year 1987 and station id USC00112193

```
127.0.0.1:8000/api/weather/1987/USC00112193
127.0.0.1:8000/api/weather/1987/USC00112193

{"Weather_stats": [{"W_DATE": "01011987", "MAX_TEMP": 0.0, "MIN_TEMP": -4.4, "PRECIPITATION": 13, "STATION_ID": "USC00112193"}, {"W_DATE": "01021987", "MAX_TEMP": -0.6, "MIN_TEMP": -4.4, "PRECIPITATION": 18, "STATION_ID": "USC00112193"}, {"W_DATE": "01031987", "MAX_TEMP": 3.3, "MIN_TEMP": -6.7, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01041987", "MAX_TEMP": 6.1, "MIN_TEMP": -1.1, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01051987", "MAX_TEMP": 6.7, "MIN_TEMP": -3.9, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01061987", "MAX_TEMP": 12.8, "MIN_TEMP": 1.1, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01071987", "MAX_TEMP": 4.4, "MIN_TEMP": 1.1, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01081987", "MAX_TEMP": 1.7, "MIN_TEMP": -3.9, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01091987", "MAX_TEMP": -0.6, "MIN_TEMP": -6.7, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01101987", "MAX_TEMP": 0.0, "MIN_TEMP": -3.9, "PRECIPITATION": 185, "STATION_ID": "USC00112193"}, {"W_DATE": "01111987", "MAX_TEMP": 0.0, "MIN_TEMP": -5.6, "PRECIPITATION": 13, "STATION_ID": "USC00112193"}, {"W_DATE": "01121987", "MAX_TEMP": 5.0, "MIN_TEMP": -7.2, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01131987", "MAX_TEMP": 8.5, "MIN_TEMP": -7.8, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01141987", "MAX_TEMP": 6.1, "MIN_TEMP": 2.2, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01151987", "MAX_TEMP": 2.2, "MIN_TEMP": -0.6, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01161987", "MAX_TEMP": 1.7, "MIN_TEMP": -5.7, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01171987", "MAX_TEMP": 1.1, "MIN_TEMP": -2.0, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01181987", "MAX_TEMP": -1.1, "MIN_TEMP": -5.6, "PRECIPITATION": 102, "STATION_ID": "USC00112193"}, {"W_DATE": "01191987", "MAX_TEMP": -0.6, "MIN_TEMP": -10.6, "PRECIPITATION": 241, "STATION_ID": "USC00112193"}, {"W_DATE": "01201987", "MAX_TEMP": -3.3, "MIN_TEMP": -16.7, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01211987", "MAX_TEMP": 0.6, "MIN_TEMP": -15.6, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01221987", "MAX_TEMP": 0.0, "MIN_TEMP": -11.1, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01231987", "MAX_TEMP": -9.4, "MIN_TEMP": -23.9, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01241987", "MAX_TEMP": -7.8, "MIN_TEMP": -26.7, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01251987", "MAX_TEMP": -9.4, "MIN_TEMP": -28.6, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01261987", "MAX_TEMP": -6.7, "MIN_TEMP": -21.7, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01271987", "MAX_TEMP": 0.0, "MIN_TEMP": -18.3, "PRECIPITATION": 5, "STATION_ID": "USC00112193"}, {"W_DATE": "01281987", "MAX_TEMP": 2.8, "MIN_TEMP": -7.8, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01291987", "MAX_TEMP": 5.6, "MIN_TEMP": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01301987", "MAX_TEMP": 4.4, "MIN_TEMP": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "01311987", "MAX_TEMP": 5.6, "MIN_TEMP": 0.6, "PRECIPITATION": 5, "STATION_ID": "USC00112193"}, {"W_DATE": "02011987", "MAX_TEMP": 10.0, "MIN_TEMP": -0.6, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02021987", "MAX_TEMP": 9.4, "MIN_TEMP": -1.7, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02031987", "MAX_TEMP": 6.7, "MIN_TEMP": -4.4, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02041987", "MAX_TEMP": 4.4, "MIN_TEMP": -2.8, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02051987", "MAX_TEMP": 0.6, "MIN_TEMP": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02061987", "MAX_TEMP": 10.0, "MIN_TEMP": 0.6, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02071987", "MAX_TEMP": 7.2, "MIN_TEMP": -3.9, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02081987", "MAX_TEMP": 3.3, "MIN_TEMP": -11.1, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02091987", "MAX_TEMP": 0.0, "MIN_TEMP": 0.0, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02101987", "MAX_TEMP": 11.1, "MIN_TEMP": 0.0, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02111987", "MAX_TEMP": 16.7, "MIN_TEMP": -2.8, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02121987", "MAX_TEMP": 8.3, "MIN_TEMP": -0.6, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02131987", "MAX_TEMP": 8.9, "MIN_TEMP": -6.7, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02141987", "MAX_TEMP": 10.6, "MIN_TEMP": 1.7, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02151987", "MAX_TEMP": 1.7, "MIN_TEMP": -7.2, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02161987", "MAX_TEMP": 0.6, "MIN_TEMP": -1.7, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02171987", "MAX_TEMP": 6.7, "MIN_TEMP": -7.2, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02181987", "MAX_TEMP": 6.7, "MIN_TEMP": -7.2, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02191987", "MAX_TEMP": 8.3, "MIN_TEMP": -6.7, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}, {"W_DATE": "02201987", "MAX_TEMP": 8.9, "MIN_TEMP": -7.8, "PRECIPITATION": 0, "STATION_ID": "USC00112193"}]
```