The Maxis Task

Task to do:

1. Create data pipeline to ingest data from public website(airport DB: link), store data in raw form in Google Storage, and ingest the data into BigQuery for analysis purpose. (link : https://raw.githubusercontent.com/jpatokal/openflights/master/data/airports.dat (https://raw.githubusercontent.com/jpatokal/openflights/master/data/airports.dat))
2. Share snapshot of the codes, and provide screenshot of BigQuery showing how many airports there are in Malaysia.
3. Material/Codes to be shared for reference. Publish in Git/Google Cloud Source Repo.

packages to install before the program

In [ ]:

```
!pip install google-cloud-storage
!pip install google-cloud-bigquery
!pip install pyarrow
```

The Airport Data Pipeline

Download raw data file from URL

In [20]:

```
from google.cloud import storage
import pandas as pd
import requests
import io

#Download file from url
url="https://raw.githubusercontent.com/jpatokal/openflights/master/data/airports.dat"
file = requests.get(url)
open('C:/Users/weionn.chong/Documents/Onn/Work/R&D/maxis_task/airport.dat', 'wb').write
(myfile.content)
print("File download complete")
```

```
File download complete
```

Upload raw data file to storage

In [22]:

```
#store raw data to google cloud storage
client = storage.Client.from_service_account_json('onn-research-dev-storage.json')   #i
nitializing google storage with credentials
bucket = client.get_bucket('maxis_task')
blob = bucket.blob('airport.dat')
blob.upload_from_filename('airport.dat')
print("Upload to google storage complete")
```

```
Upload to google storage complete
```

Data Transformation

In [63]:

```
#peek the data file from url
data = pd.read_csv('airport.dat', header = None)
data.head()
```

Out[63]:

	0	1	2	3	4	5	6	7	8	9	10	
0	1	Goroka Airport	Goroka	Papua New Guinea	GKA	AYGA	-6.081690	145.391998	5282	10	U	Pacifi
1	2	Madang Airport	Madang	Papua New Guinea	MAG	AYMD	-5.207080	145.789001	20	10	U	Pacifi
2	3	Mount Hagen Kagamuga Airport	Mount Hagen	Papua New Guinea	HGU	AYMH	-5.826790	144.296005	5388	10	U	Pacifi
3	4	Nadzab Airport	Nadzab	Papua New Guinea	LAE	AYNZ	-6.569803	146.725977	239	10	U	Pacifi
4	5	Port Moresby Jacksons International Airport	Port Moresby	Papua New Guinea	POM	AYPY	-9.443380	147.220001	146	10	U	Pacifi

In [53]:

```
print(len(data))
```

```
7698
```

In [47]:

```
#Transformation
#Renaming the column names
data.rename(columns = {1: 'airport_name',
                       2: 'city',
                       3: 'country',
                       4: 'airport_code',
                       5: 'iso_code',
                       6: 'latitude',
                       7: 'longitude',
                       8: 'elevation',
                       9: 'timezone',
                       10: 'unknown',
                       11: 'continent/region',
                       12: 'type',
                       13: 'source'}, inplace=True)
```

In [59]:

```
#reset index and droping the initial index column
data = data.reset_index().drop(columns = 0)
```

Upload dataframe to BigQuery table

In [62]:

```
from google.cloud import bigquery
#Initialize the project id and staging table destination
table_destination_firm = 'onn-research-dev.maxis_task.airport_db'

#initializing the client with google credentials
client = bigquery.Client.from_service_account_json('onn-research-dev-bigquery.json')

#Setting the schema for the bigquery table
job_config_airport = bigquery.LoadJobConfig(scheme=[
                        bigquery.SchemaField("no", "STRING"),
                        bigquery.SchemaField("airport_name", "STRING"),
                        bigquery.SchemaField("city", "STRING"),
                        bigquery.SchemaField("country", "STRING"),
                        bigquery.SchemaField("airport_code", "STRING"),
                        bigquery.SchemaField("iso_code", "STRING"),
                        bigquery.SchemaField("latitude", "STRING"),
                        bigquery.SchemaField("longitude", "STRING"),
                        bigquery.SchemaField("elevation", "STRING"),
                        bigquery.SchemaField("timezone", "STRING"),
                        bigquery.SchemaField("Unknown", "STRING"),
                        bigquery.SchemaField("continent/region", "STRING"),
                        bigquery.SchemaField("type", "STRING"),
                        bigquery.SchemaField("source", "STRING")],
                        write_disposition="WRITE_TRUNCATE"
                    )
#passing dataframe to bigquery table
job_airport = client.load_table_from_dataframe(data, table_destination_firm, job_config
=job_config_airport)

job_airport.result()
print('Upload to bigquery table is completed')
```

Out[62]:

```
<google.cloud.bigquery.job.LoadJob at 0x2a9b6789400>
```

View Result (Extra)

Show all airport in Malaysia and the total number

In [67]:

```
#run sql job to query all the airports in malaysia
sql = """
    SELECT distinct *
    FROM `onn-research-dev.maxis_task.airport_db`
    WHERE country = @country
"""
query_config = bigquery.QueryJobConfig(
                    query_parameters=[
                        bigquery.ScalarQueryParameter('country','STRING','Malaysia')
                    ])

result_my = client.query(sql, job_config=query_config).to_dataframe()

display(result_my) #diplay airport details in malaysia
print(f"Total number of airport in Malaysia is {len(result_my)}")  #number of airport i
n malaysia
```

	airport_name	city	country	airport_code	iso_code	latitude	longitude	elevati
0	Bintulu Airport	Bintulu	Malaysia	BTU	WBGB	3.12385	113.019997	
1	Kuching International Airport	Kuching	Malaysia	KCH	WBGG	1.48470	110.347000	
2	Limbang Airport	Limbang	Malaysia	LMN	WBGJ	4.80830	115.010002	
3	Marudi Airport	Marudi	Malaysia	MUR	WBGM	4.17898	114.329002	1
4	Miri Airport	Miri	Malaysia	MYY	WBGR	4.32201	113.987000	
5	Sibu Airport	Sibu	Malaysia	SBW	WBGS	2.26160	111.985001	1
6	Lahad Datu Airport	Lahad Datu	Malaysia	LDU	WBKD	5.03225	118.323997	
7	Kota Kinabalu International Airport	Kota Kinabalu	Malaysia	BKI	WBKK	5.93721	116.051003	
8	Labuan Airport	Labuan	Malaysia	LBU	WBKL	5.30068	115.250000	1
9	Tawau Airport	Tawau	Malaysia	TWU	WBKW	4.32016	118.127998	
10	Kluang Airport	Kluang	Malaysia	\N	WMAP	2.04139	103.306999	1
11	Sultan Abdul Halim Airport	Alor Setar	Malaysia	AOR	WMKA	6.18967	100.398003	
12	Butterworth Airport	Butterworth	Malaysia	BWH	WMKB	5.46592	100.390999	
13	Sultan Ismail Petra Airport	Kota Bahru	Malaysia	KBR	WMKC	6.16685	102.292999	
14	Kuantan Airport	Kuantan	Malaysia	KUA	WMKD	3.77539	103.209000	
15	Kerteh Airport	Kerteh	Malaysia	KTE	WMKE	4.53722	103.427002	
16	Simpang Airport	Simpang	Malaysia	\N	WMKF	3.11225	101.703003	
17	Sultan Azlan Shah Airport	Ipoh	Malaysia	IPH	WMKI	4.56797	101.092003	1
18	Senai International Airport	Johor Bahru	Malaysia	JHB	WMKJ	1.64131	103.669998	1
19	Kuala Lumpur International Airport	Kuala Lumpur	Malaysia	KUL	WMKK	2.74558	101.709999	
20	Langkawi International Airport	Langkawi	Malaysia	LGK	WMKL	6.32973	99.728699	
21	Malacca Airport	Malacca	Malaysia	MKZ	WMKM	2.26336	102.251999	
22	Sultan Mahmud Airport	Kuala Terengganu	Malaysia	TGG	WMKN	5.38264	103.102997	
23	Penang International Airport	Penang	Malaysia	PEN	WMKP	5.29714	100.277000	
24	Pulau Tioman Airport	Tioman	Malaysia	TOD	WMBT	2.81818	104.160004	

	airport_name	city	country	airport_code	iso_code	latitude	longitude	elevati
25	Sultan Abdul Aziz Shah International Airport	Kuala Lumpur	Malaysia	SZB	WMSA	3.13058	101.549004	
26	LTS Pulau Redang Airport	Redang	Malaysia	RDN	WMPR	5.76528	103.007004	
27	Mulu Airport	Mulu	Malaysia	MZV	WBMU	4.04833	114.805000	
28	Sandakan Airport	Sandakan	Malaysia	SDK	WBKS	5.90090	118.058998	
29	Belaga Airport	Belaga	Malaysia	BLG	WBGC	2.65000	113.766998	2
30	Long Lellang Airport	Long Datih	Malaysia	LGL	WBGF	3.42100	115.153999	14
31	Long Seridan Airport	Long Seridan	Malaysia	ODN	WBGI	3.96700	115.050003	6
32	Mukah Airport	Mukah	Malaysia	MKM	WBGK	2.90639	112.080002	
33	Bakalalan Airport	Bakalalan	Malaysia	BKM	WBGQ	3.97400	115.617996	29
34	Lawas Airport	Lawas	Malaysia	LWY	WBGW	4.84917	115.407997	
35	Bario Airport	Bario	Malaysia	BBN	WBGZ	3.73389	115.478996	33
36	Tomanggong Airport	Tomanggong	Malaysia	TMG	WBKM	5.40257	118.657630	
37	Kudat Airport	Kudat	Malaysia	KUD	WBKT	6.92250	116.835999	
38	Pulau Pangkor Airport	Pangkor Island	Malaysia	PKG	WMPA	4.24472	100.553001	
39	Long Akah Airport	Long Akah	Malaysia	LKH	WBGL	3.30000	114.782997	2

```
Total number of airport in Malaysia is 40
```

In [ ]: