Testing Jupyter Notebook

import library

In [440]:
```
import pandas as pd
import numpy as np
import pymysql as sql
from datetime import datetime, timedelta
```

connect to DB

In [515]:
```
db = sql.connect(host='localhost',user='root',password='',database='ict')
db2 = sql.connect(host='localhost',user='root',password='',database='ictn', autocommit=True)
cur = db.cursor()
cur2 = db2.cursor()
```

ETL Read DB and reshape

In [442]:
```
q = 'select * from daily'
cur.execute(q)
dt = cur.fetchall()
df = pd.DataFrame(dt)
df.rename(columns={0:'DATE',1:'SITEID',2:'NOP',3:'CP',4:'KABUPATEN',5:'CLASS',6:'OUTAGES',7:'AVAIL',8
df
```

Out[442]:

	DATE	SITEID	NOP	CP	KABUPATEN	CLASS	OUTAGES	AVAIL	TARGET	ACHIEVE	MAIN_P
0	2023-01-01	COG130	DENPASAR	BALI TIMUR	BADUNG	Bronze	0	100.000	95.0	1	E
1	2023-01-01	COG102	DENPASAR	BALI TIMUR	BADUNG	Bronze	0	100.000	95.0	1	E
2	2023-01-01	COG114	MATARAM	SUMBAWA	SUMBAWA BARAT	Silver	8	99.990	97.0	1	E
3	2023-01-01	COG123	KUPANG	WAINGAPU	SUMBA TIMUR	Bronze	86399	0.001	95.0	0	E
4	2023-01-01	BLI065	DENPASAR	BALI TIMUR	BANGLI	Bronze	0	100.000	95.0	1	HA PROBL
...	...	...	...	...	...	...	...	...	...	...	
524327	2023-04-30	WKB170	KUPANG	WAINGAPU	SUMBA BARAT DAYA	Gold	0	100.000	98.4	1	
524328	2023-04-30	WKB172	KUPANG	WAINGAPU	SUMBA BARAT DAYA	Silver	0	100.000	97.0	1	
524329	2023-04-30	WKB186	KUPANG	WAINGAPU	SUMBA BARAT	Silver	0	100.000	97.0	1	
524330	2023-04-30	WKB187	KUPANG	WAINGAPU	SUMBA BARAT DAYA	Gold	0	100.000	98.4	1	
524331	2023-04-30	WKB189	KUPANG	WAINGAPU	SUMBA BARAT	Silver	0	100.000	97.0	1	

524332 rows × 13 columns

CREATE DIM TABLE SITEID

In [443]:
```
df2 = df.loc[:,'SITEID']
df2 = df2.drop_duplicates()
df2 = df2.sort_values()
df2 = df2.reset_index(drop=True)
df2 = pd.DataFrame(df2)
a = []
b = 1
for row in df2['SITEID']:
  a.append(b)
  b+=1
df2['ID'] = a
df2
```

Out[443]:

	SITEID	ID
0	APR001	1
1	APR002	2
2	APR004	3
3	APR005	4
4	APR006	5
...	...	...
4386	WKB170	4387
4387	WKB172	4388
4388	WKB186	4389
4389	WKB187	4390
4390	WKB189	4391

4391 rows × 2 columns

CREATE DIM TABLE IN DB and CSV

In [444]:
```
q1 = 'DROP TABLE IF EXISTS SITE'
cur2.execute(q1)
q2 = 'create table SITE (SITEID varchar(6), ID_SITE int AUTO_INCREMENT PRIMARY KEY );'
cur2.execute(q2)
df2.to_csv('D:/Test Data/SITE.csv',index=False)

q5 = 'insert into SITE (SITEID) select distinct(siteid) from ict.daily order by Siteid'
cur2.execute(q5)
```

Out[444]: 4391

CREATE DIM TABLE CLASS

In [445]:
```
df3 = df.loc[:,'CLASS']
df3 = df3.drop_duplicates()
df3 = df3.reset_index(drop=True)
df3 = pd.DataFrame(df3)
a = []
b=1
for row in df3['CLASS']:
    a.append(b)
    b+=1
df3['ID'] = a
df3
```

Out[445]:

	CLASS	ID
0	Bronze	1
1	Silver	2
2	Gold	3
3	Platinum	4
4	Diamond	5

INSERT DIM TABLE CLASS TO DB AND EXPORT TO CSV

In [446]:
```
q1 = 'DROP TABLE IF EXISTS CLASS'
cur2.execute(q1)
q2 = 'create table CLASS (CLASS varchar(20), ID_CLASS int AUTO_INCREMENT PRIMARY KEY );'
cur2.execute(q2)
df3.to_csv('D:/Test Data/CLASS.csv',index=False)

q5 = 'insert into class(class) select distinct(class) from ict.daily order by Siteid'
cur2.execute(q5)
```

Out[446]: 5

CREATE DIM TABLE NOP

In [447]:
```
df4 = df.loc[:,'NOP']
df4 = df4.drop_duplicates()
df4 = df4.reset_index(drop=True)
df4 = pd.DataFrame(df4)

a= []
b=1
for row in df4['NOP']:
    a.append(b)
    b+=1
df4['ID'] = a

df4
```

Out[447]:

	NOP	ID
0	DENPASAR	1
1	MATARAM	2
2	KUPANG	3
3	FLORES	4

INSERT TO DB AND EXPORT CSV

In [448]:
```
q1 = 'DROP TABLE IF EXISTS NOP'
cur2.execute(q1)
q2 = 'create table NOP (NOP varchar(100), ID_NOP int AUTO_INCREMENT PRIMARY KEY );'
cur2.execute(q2)
df4.to_csv('D:/Test Data/NOP.csv',index=False)

q5 = 'insert into NOP(NOP) select distinct(NOP) from ict.daily'
cur2.execute(q5)
```

Out[448]: 4

CREATE DIM TABLE CP

In [449]:
```
df5 = df.loc[:,'CP']
df5 = df5.drop_duplicates()
df5 = df5.reset_index(drop=True)
df5 = df5.sort_values()
df5 = pd.DataFrame(df5)

a= []
b=1
for row in df5['CP']:
    a.append(b)
    b+=1
df5['ID'] = a

df5
```

Out[449]:

	CP	ID
7	BALI BARAT	1
0	BALI TIMUR	2
3	BIMA	3
4	KUPANG	4
8	MATARAM	5
5	MAUMERE	6
6	RUTENG	7
1	SUMBAWA	8
2	WAINGAPU	9

INSERT TO DB AND CSV

In [450]:
```
q1 = 'DROP TABLE IF EXISTS CP'
cur2.execute(q1)
q2 = 'create table CP (CP varchar(100), ID_CP int AUTO_INCREMENT PRIMARY KEY );'
cur2.execute(q2)
df5.to_csv('D:/Test Data/CP.csv',index=False)

q5 = 'insert into CP(CP) select distinct(CP) from ict.daily'
cur2.execute(q5)
```

Out[450]: 9

CREATE DIM TABLE MAIN PROBLEM

In [451]:
```
df6 = df.loc[:,'MAIN_PROBLEM']
df6 = df6.drop_duplicates()
df6 = df6.sort_values(ascending=False)
df6 = df6.reset_index(drop=True)
df6 = pd.DataFrame(df6)

a= []
b=1
for row in df6['MAIN_PROBLEM']:
    a.append(b)
    b+=1
df6['ID'] = a

df6
```

Out[451]:

	MAIN_PROBLEM	ID
0	TRANSMISSION PROBLEM*(TR)	1
1	POWER PROBLEM*(PW)	2
2	OTHERS*(OT)	3
3	HARDWARE PROBLEM*(HW)	4
4	EXCLUDE	5
5		6

INSERT TO DB AND EXPORT CSV

In [460]:
```
q1 = 'DROP TABLE IF EXISTS MAIN_PROBLEM'
cur2.execute(q1)
q2 = 'create table MAIN_PROBLEM (MAIN_PROBLEM varchar(100), ID_MP int AUTO_INCREMENT PRIMARY KEY );'
cur2.execute(q2)
df6.to_csv('D:/Test Data/MAIN_PROBLEM.csv',index=False)

q5 = 'insert into MAIN_PROBLEM(MAIN_PROBLEM) select distinct(CATEGORY_PROBLEM) from ict.daily order by
cur2.execute(q5)
```

Out[460]: 6

CREATE DIM TABLE KABUPATEN

In [453]:
```
df7 = df.loc[:,'KABUPATEN']
df7 = df7.drop_duplicates()
df7 = df7.sort_values()
df7 = df7.reset_index(drop=True)
df7 = pd.DataFrame(df7)

a= []
b=1
for row in df7['KABUPATEN']:
    a.append(b)
    b+=1
df7['ID'] = a

df7
```

Out[453]:

	KABUPATEN	ID
0	ALOR	1
1	BADUNG	2
2	BANGLI	3
3	BELU	4
4	BIMA	5
5	BULELENG	6
6	DOMPU	7
7	ENDE	8
8	FLORES TIMUR	9
9	GIANYAR	10
10	JEMBRANA	11
11	KARANGASEM	12
12	KLUNGKUNG	13
13	KOTA BIMA	14
14	KOTA DENPASAR	15
15	KOTA KUPANG	16
16	KOTA MATARAM	17
17	KUPANG	18
18	LEMBATA	19
19	LOMBOK BARAT	20
20	LOMBOK TENGAH	21
21	LOMBOK TIMUR	22
22	LOMBOK UTARA	23
23	MALAKA	24
24	MANGGARAI	25
25	MANGGARAI BARAT	26
26	MANGGARAI TIMUR	27
27	NAGEKEO	28
28	NGADA	29
29	ROTE NDAO	30
30	SABU RAIJUA	31
31	SIKKA	32
32	SUMBA BARAT	33
33	SUMBA BARAT DAYA	34
34	SUMBA TENGAH	35
35	SUMBA TIMUR	36
36	SUMBAWA	37
37	SUMBAWA BARAT	38
38	TABANAN	39
39	TIMOR TENGAH SELATAN	40
40	TIMOR TENGAH UTARA	41

INSERT TO DB AND CSV

In [454]:
```
q1 = 'DROP TABLE IF EXISTS KABUPATEN'
cur2.execute(q1)
q2 = 'create table KABUPATEN (KABUPATEN varchar(100), ID_KAB int AUTO_INCREMENT PRIMARY KEY );'
cur2.execute(q2)
df7.to_csv('D:/Test Data/KABUPATEN.csv',index=False)

q5 = 'insert into KABUPATEN(KABUPATEN) select distinct(KAB) from ict.daily'
cur2.execute(q5)
```

Out[454]: 41

CREATE DIM TABLE DATE

In [477]:
```
df8 = df.loc[:,'DATE']
df8 = df8.drop_duplicates()
df8 = pd.to_datetime(df8)
df8 = df8.sort_values()
df8 = df8.reset_index(drop=True)
a=[]
b=[]
c=[]
d=1
e=[]
for val in df8:
    vals = val + timedelta(days=3)
    a.append(vals.weekofyear)
    b.append(val.year)
    c.append(d)
    d+=1
    e.append(str(val.month_name())[:3])
df8 = pd.DataFrame(df8)
df8['WEEK']=a
df8['YEAR']=b
df8['MONTH']=e
df8['ID']=c
df8
```

Out[477]:

	DATE	WEEK	YEAR	MONTH	ID
0	2023-01-01	1	2023	Jan	1
1	2023-01-02	1	2023	Jan	2
2	2023-01-03	1	2023	Jan	3
3	2023-01-04	1	2023	Jan	4
4	2023-01-05	1	2023	Jan	5
...	...	...	...	...	...
115	2023-04-26	17	2023	Apr	116
116	2023-04-27	17	2023	Apr	117
117	2023-04-28	18	2023	Apr	118
118	2023-04-29	18	2023	Apr	119
119	2023-04-30	18	2023	Apr	120

120 rows × 5 columns

INSERT TO DB AND CSV

In [480]:
```
q1 = 'DROP TABLE IF EXISTS DATE'
cur2.execute(q1)
q2 = 'create table DATE (DATE date, WEEK INT, MONTH VARCHAR(50), YEAR int, ID_DATE int AUTO_INCREMENT
cur2.execute(q2)
df8.to_csv('D:/Test Data/DATE.csv',index=False)

q5 = 'insert into DATE(DATE, week, year, month) select distinct(DATE), case when mid(yearweek(date_ad
cur2.execute(q5)
```

Out[480]: 120

CREATE FACT TABLE

In [ ]:
```
que = 'DROP TABLE IF EXISTS fact'
cur2.execute(que)
que2 = 'CREATE TABLE fact (IDDATE INT, IDSITE INT, IDCLASS INT, IDNOP INT, IDCP INT, IDKAB INT, OUTAGE
```

MIGRATE DATA FROM ONE BIG DATA TABLE TO FACT TABLE

In [ ]:
```
query = 'INSERT INTO FACT (IDDATE, IDSITE, IDCLASS, IDNOP, IDCP, IDKAB, OUTAGE, AVAILABILITY, TARGET,
cur2.execute(query)
```

RESULT

BEFORE

In [490]:
```
q = 'Select * from ict.daily where month(date) = 4'
cur2.execute(q)
dt = cur2.fetchall()
tbb = pd.DataFrame(dt)
print(tbb)
```

```
                0       1         2           3                4       5
0       2023-04-01  APR130  DENPASAR  BALI TIMUR        KARANGASEM    Gold  \
1       2023-04-01  APR011  DENPASAR  BALI TIMUR        KARANGASEM  Bronze
2       2023-04-01  APR150  DENPASAR  BALI TIMUR        KARANGASEM  Bronze
3       2023-04-01  BMA123   MATARAM        BIMA         KOTA BIMA    Gold
4       2023-04-01  BIM075   MATARAM        BIMA              BIMA    Gold
...            ...     ...       ...         ...               ...     ...
131095  2023-04-30  WKB170    KUPANG    WAINGAPU  SUMBA BARAT DAYA    Gold
131096  2023-04-30  WKB172    KUPANG    WAINGAPU  SUMBA BARAT DAYA  Silver
131097  2023-04-30  WKB186    KUPANG    WAINGAPU       SUMBA BARAT  Silver
131098  2023-04-30  WKB187    KUPANG    WAINGAPU  SUMBA BARAT DAYA    Gold
131099  2023-04-30  WKB189    KUPANG    WAINGAPU       SUMBA BARAT  Silver

          6       7     8   9                         10              11
0       3843   95.55  98.4   0  TRANSMISSION PROBLEM*(TR)  RADIO IP TELKOM  \
1          0  100.00  95.0   1         POWER PROBLEM*(PW)              MBP
2       8790   89.82  95.0   0         POWER PROBLEM*(PW)              MBP
3        192   99.77  98.4   1         POWER PROBLEM*(PW)          BATTERY
4          0  100.00  98.4   1  TRANSMISSION PROBLEM*(TR)  RADIO IP TELKOM
...      ...     ...   ...  ..                        ...              ...
131095     0  100.00  98.4   1
131096     0  100.00  97.0   1
131097     0  100.00  97.0   1
131098     0  100.00  98.4   1
131099     0  100.00  97.0   1

                            12
0                RADIO IP TELKOM
1                TIM TELAT JALAN
2                TIM TELAT JALAN
3       PLN OFF, BATTERY DEGRADED
4                RADIO IP TELKOM
...                          ...
131095
131096
131097
131098
131099

[131100 rows x 13 columns]
```

AFTER

In [521]:
```
q = 'select fact.* from fact join (select * from date)b on fact.iddate = b.id_date where month(date) =
cur2.execute(q)
dt = cur2.fetchall()
tba = pd.DataFrame(dt)
print(tba)
```

```
         0     1     2   3   4   5   6      7     8   9   10
0        91     1   5   1   1  15   0  100.0  95.0   1   6
1        91     2   5   1   1  15   0  100.0  95.0   1   6
2        91     3   5   1   1  15   0  100.0  95.0   1   6
3        91     4   4   1   1  15   0  100.0  97.0   1   6
4        91     5   3   1   1  15   0  100.0  98.4   1   6
...     ...   ...  ..  ..  ..  ..  ..    ...   ...  ..  ..
131095  120  4387   3   3   3  25   0  100.0  98.4   1   6
131096  120  4388   4   3   3  25   0  100.0  97.0   1   6
131097  120  4389   4   3   3  19   0  100.0  97.0   1   6
131098  120  4390   3   3   3  25   0  100.0  98.4   1   6
131099  120  4391   4   3   3  19   0  100.0  97.0   1   6

[131100 rows x 11 columns]
```

In [514]:
```
cur2.close()
db2.close()
```

In [518]:
```
qiu = 'select ref.date, siteid, ref.class, ref.nop, ref.cp, ref.kabupaten, outage, availability, targe
cur2.execute(qiu)
dt = cur2.fetchall()
tba2 = pd.DataFrame(dt)
print(tba2)
```

```
                0       1       2         3           4            5   6
0       2023-04-01  APR001  Bronze  DENPASAR  BALI TIMUR   KARANGASEM   0  \
1       2023-04-02  APR001  Bronze  DENPASAR  BALI TIMUR   KARANGASEM   0
2       2023-04-03  APR001  Bronze  DENPASAR  BALI TIMUR   KARANGASEM   0
3       2023-04-04  APR001  Bronze  DENPASAR  BALI TIMUR   KARANGASEM   0
4       2023-04-05  APR001  Bronze  DENPASAR  BALI TIMUR   KARANGASEM   0
...            ...     ...     ...       ...         ...          ...  ..
131095  2023-04-26  WKB189  Silver    KUPANG    WAINGAPU  SUMBA BARAT   0
131096  2023-04-27  WKB189  Silver    KUPANG    WAINGAPU  SUMBA BARAT   0
131097  2023-04-28  WKB189  Silver    KUPANG    WAINGAPU  SUMBA BARAT   0
131098  2023-04-29  WKB189  Silver    KUPANG    WAINGAPU  SUMBA BARAT   0
131099  2023-04-30  WKB189  Silver    KUPANG    WAINGAPU  SUMBA BARAT   0

           7     8   9  10
0       100.0  95.0   1
1       100.0  95.0   1
2       100.0  95.0   1
3       100.0  95.0   1
4       100.0  95.0   1
...       ...   ...  .. ..
131095  100.0  97.0   1
131096  100.0  97.0   1
131097  100.0  97.0   1
131098  100.0  97.0   1
131099  100.0  97.0   1

[131100 rows x 11 columns]
```

CONCLUSION MORE COMPLICATED IN SQL SYNTAX BUT MORE FAST IN PROCCESSING