Database Programming

DigitalSkola
Uncover The World Of Digital Skills With Us

Table of Content

What will We Learn Today?

1. File reading and writing in Python
2. Database Programming in Python
3. - Connect database
4. - Do a SQL command
5. SQL command become Dataframe
6. Data wrangling using Pandas

DigitalSkola
Uncover The World Of Digital Skills With Us

Reading file in Python

There are several files that can be read in the Python.

You have learned:

- *CSV*
- *Excel*
- *Text*

Text File in Python

Terdapat beberapa mode dalam mengeksekusi file text di dalam Python

Character	Meaning
`'r'`	open for reading (default)
`'w'`	open for writing, truncating the file first
`'x'`	open for exclusive creation, failing if the file already exists
`'a'`	open for writing, appending to the end of file if it exists
`'b'`	binary mode
`'t'`	text mode (default)
`'+'`	open for updating (reading and writing)

source: *Python Docs*

Membuat file text di Python

```
# Writing text in python
with open('untitled.txt', 'w') as writer:
  writer.write('Welcome back to my channel')
```

```
['Welcome back to my channel']
```

Execute text file

```
path = '/content/saham.txt'
```

Dalam membuka file text

```
# reading all text in file text
with open(path, 'r') as text_file:
  print(text_file.readlines())
```

```
['Saham (stock) merupakan salah satu instrumen pasar keuangan yang paling popular.
```

Menambahkan text pada existing file

```
# appending text into the existing file
with open(path, 'a') as text_file:
  text_file.write("\nDigital Skola Batch 11")
```

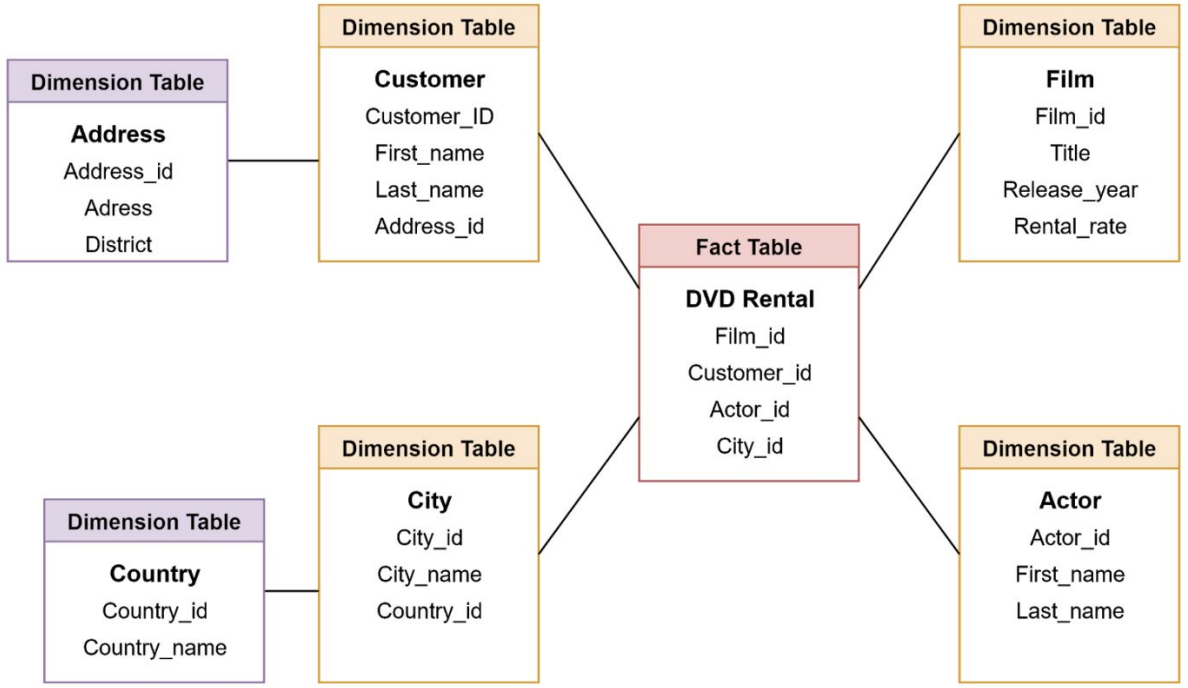
```
# show the result
with open(path, 'r') as text_file:
  for line in text_file:
    print(line, end="")
```

```
Saham (stock) merupakan salah satu instrumen pasar keuangan yang paling popular.

Saham dapat didefinisikan sebagai tanda penyertaan modal seseorang atau pihak (ba
Digital Skola Batch 11
```

DigitalSkola
Uncover The World Of Digital Skills With Us

Database Programming



source: *Introduction of Data Modeling in Your Data Career*

Put Postgre to Python

Psycopg library, adaptor database postgreSQL untuk digunakan di dalam pemrograman Python

```
!pip install psycopg2
```

What needs to be prepared?

Siapkan dulu alat-alat dan bahan untuk keperluan ~~masak~~ database programming

Cloud server yg terkoneksi dengan PostgreSQL:

host: digitalskoladb.c04me33o8tni.ap-southeast-1.rds.amazonaws.com

database: sandbox

Your access:

User Admin: **********

Password: **********

DigitalSkola
Uncover The World Of Digital Skills With Us

Connect your database

Connect first

```
conn = psycopg2.connect(
          host = "digitalskoladb.c04me33o8tni.ap-southeast-1.rds.amazonaws.com",
          database = "sandbox",
          user = "**********",
          password = "**********")
```

Store the database table menggunakan cursor (Temporary Memory)

```
cur = conn.cursor()
```

DigitalSkola
Uncover The World Of Digital Skills With Us

Result of query

Execute query and read the result in pandas dataframe

```
sql = """SELECT * FROM batch_11.cb_stations"""
data = pd.read_sql_query(sql, conn)
data.head()
```

	station_id	name	short_name	latitude	longitude	region_id	rental_methods	capacity	eightd_has_key_dispenser	r
0	128	MacDougal St & Prince St	5687.04	40.727104	-74.002970	71.0	CREDITCARD,KEY	0	False	
1	224	Spruce St & Nassau St	5137.10	40.711464	-74.005520	71.0	CREDITCARD,KEY	0	False	
2	229	Great Jones St	5636.11	40.727436	-73.993790	71.0	CREDITCARD,KEY	0	False	
3	410	Suffolk St & Stanton St	5445.02	40.720665	-73.985176	71.0	CREDITCARD,KEY	0	False	
4	434	9 Ave & W 18 St	6190.08	40.743176	-74.003660	71.0	CREDITCARD,KEY	0	False	

Make it more challenging

Use Common Table Expression (CTE) (data1)

```
sql = """WITH TABS AS
              (SELECT *
               FROM batch_11.cb_trips
              )

         SELECT *
         from TABS"""
data = pd.read_sql_query(sql, conn)
data.head()
```

CTE more challenging (data2)

```
sql = """WITH TABS AS
              (SELECT b.tripduration, a.name as station_name, b.starttime
               FROM batch_11.cb_stations as a
               JOIN batch_11.cb_trips as b
               on a.station_id = b.end_station_id)

         SELECT station_name, sum(tripduration) as total_trip_duration
         from TABS
         where station_name like '%Clermont%'
         group by 1
         having sum(tripduration) < 3000000
         """
data2 = pd.read_sql_query(sql, conn)
data2.head(7)
```

Applied Data Wrangling

Sekarang saatnya memanggil the almighty pandas

untuk melakukan pengolahan data

Pengolahan data menggunakan pandas

Exercise 1

Lakukan query dan ambil table **cb_trips,** filter gender = 0 dan tripduration = 634 seconds. Buat output ke dalam pandas dan lakukan statation_start apa saja yang ada dan berapa banyak kemudian urutkan dari yang terbesar sampai terkecil

Exercise 2

Lakukan query dan ambil table **cb_trips,** filter end_station_name yang mengandung tulisan broadway (pastikan set lower di field-nya) dan ambil tripduration di bawah dari atau sama dengan 500 seconds. Buat output untuk mencari tau proporsi dari usertype dan tampilkan dalam visualisasi

Homework

CB_Stations X CB_trips

lakukan:

- Menggabungkan table cb_station dan cb_trips berdasarkan start_station_id
- Ambil hanya kolom name, region_id, rental_methods, tripduration
- Filter tripduration di atas 1 hari (satuan tripduration adalah second)
- Lakukan pivot table guna melihat station mana aja dan tunjukan rata-rata tripduration

DigitalSkola

Uncover The World Of Digital Skills With Us

Thank YOU