

Potchara Dockchaam

65199160192

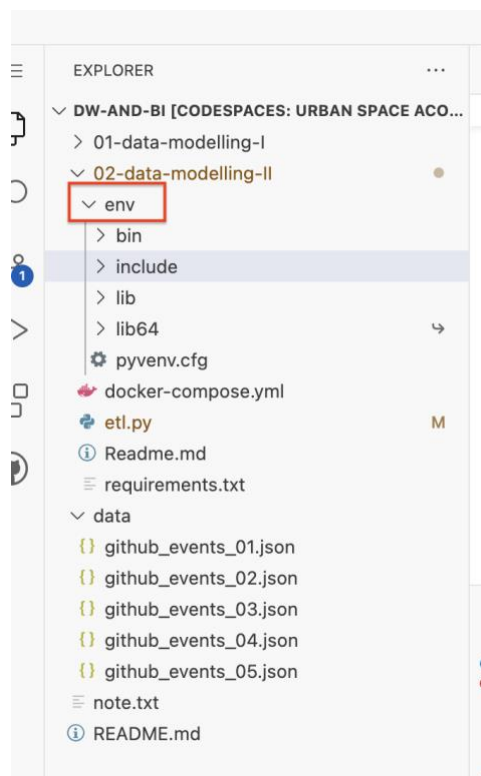
Documentation

1. เปลี่ยน Directory ให้อยู่ใน Folder 02-data-modelling-II

```
$ cd 02-data-modelling-II
```

2. สร้าง Environment Python เพื่อรองรับการติดตั้ง Library ให้เหมาะสมกับงาน

```
$ python -m venv env
```



3. Active Env. ที่สร้างไว้ให้สามารถใช้งานได้

```
$ source env/bin/activate
```

```
o (env) @Fooklnwza007 → /workspaces/dw-and-bi/02-data-modelling-II (main) $
```

4. Install Library ที่เกี่ยวข้องกับการใช้งาน

```
$ pip install -r requirements.txt
```



5. เริ่มการใช้งาน Cassandra ด้วย Docker

```
$ docker compose up
```

6. หากต้องการสร้าง table, เพิ่มข้อมูล, Query ข้อมูล สามารถทำได้ใน etl.py

```
$ python ety.py
```

```
EXPLORER
DW-AND-BI [CODESPACES: URBAN SPACE ACO...
  > 01-data-modelling-I
  > 02-data-modelling-II
    env
    bin
    include
    lib
    lib64
    pyvenv.cfg
    docker-compose.yml
    etl.py
    README.md
    requirements.txt
  data
    github_events_01.json
    github_events_02.json
    github_events_03.json
    github_events_04.json
    github_events_05.json
    note.txt
    README.md
  OUTLINE

02-data-modelling-II > etl.py
4 from typing import List
5
6 from cassandra.cluster import Cluster
7
8
9 table_drop_events = "DROP TABLE events"
10 table_drop_actor = "DROP TABLE actor"
11 table_drop_repo = "DROP TABLE repo"
12
13 table_create_events = """
14 CREATE TABLE IF NOT EXISTS events
15 (
16     event_id varchar,
17     type varchar,
18     public boolean,
19     created_at timestamp,
20     actor_id varchar,
21     actor_login varchar,
22     actor_display_login varchar,
23     actor_url varchar,
24     repo_id varchar,
25     repo_name varchar,
26     repo_url varchar,
27     PRIMARY KEY (
28         (event_id,
29         actor_id,
30         repo_id
31     )
32 )
33 """
```

```
EXPLORER
DW-AND-BI [CODESPACES: URBAN SPACE ACO...
  > 01-data-modelling-I
  > 02-data-modelling-II
    env
    bin
    include
    lib
    lib64
    pyvenv.cfg
    docker-compose.yml
    etl.py
    README.md
    requirements.txt
  data
    github_events_01.json
    github_events_02.json
    github_events_03.json
    github_events_04.json
    github_events_05.json
    note.txt
    README.md
  OUTLINE

02-data-modelling-II > etl.py
115 # Insert data into tables here
116 insert_statement = f"""
117 INSERT INTO events (
118     event_id,
119     type,
120     public,
121     created_at,
122     actor_id,
123     actor_login,
124     actor_display_login,
125     actor_url,
126     repo_id,
127     repo_name,
128     repo_url
129 ) VALUES ({'each["id"]', 'each["type"]', 'each["public"]', 'each["created_at"]',
130 'each["actor"]["id"]', 'each["actor"]["login"]', 'each["actor"]["display_login"]', 'each["actor"]["u
131 'each["repo"]["id"]', 'each["repo"]["name"]', 'each["repo"]["url"]}')
132
133 session.execute(insert_statement)
134
135
136 insert_statement = f"""
137 INSERT INTO actor (
138     actor_id,
139     actor_login
140 ) VALUES ({'each["actor"]["id"]', 'each["actor"]["login"]}')
141
142 session.execute(insert_statement)
143
144
```

```
EXPLORER
DW-AND-BI [CODESPACES: URBAN SPACE ACO...
  > 01-data-modelling-I
  > 02-data-modelling-II
    env
    bin
    include
    lib
    lib64
    pyvenv.cfg
    docker-compose.yml
    etl.py
    README.md
    requirements.txt
  data
    github_events_01.json
    github_events_02.json
    github_events_03.json
    github_events_04.json
    github_events_05.json
    note.txt
    README.md
  OUTLINE

02-data-modelling-II > etl.py
182
183 drop_tables(session)
184 create_tables(session)
185 process(session, filepath="../data")
186 #insert_sample_data(session)
187
188 # Select data in Cassandra and print them to stdout
189 query = """
190 SELECT event_id, actor_id, created_at FROM events
191 WHERE created_at >= '2022-08-17T15:51:04Z'
192 AND created_at <= '2022-08-17T15:51:05Z'
193 AND type = 'IssuesEvent'
194 ALLOW FILTERING;
195
196
197 try:
198     rows = session.execute(query)
199 except Exception as e:
200     print(e)
201
202 for row in rows:
203     print(row)
204
205
206 if __name__ == "__main__":
207     main()
```

23487929580 IssuesEvent 21071133 2022-08-17T15:51:05Z

```
Row(event_id='23487929428', actor_id='21071133', created_at=datetime.datetime(2022, 8, 17, 15, 51, 5))
Row(event_id='23487929580', actor_id='21972349', created_at=datetime.datetime(2022, 8, 17, 15, 51, 5))
```