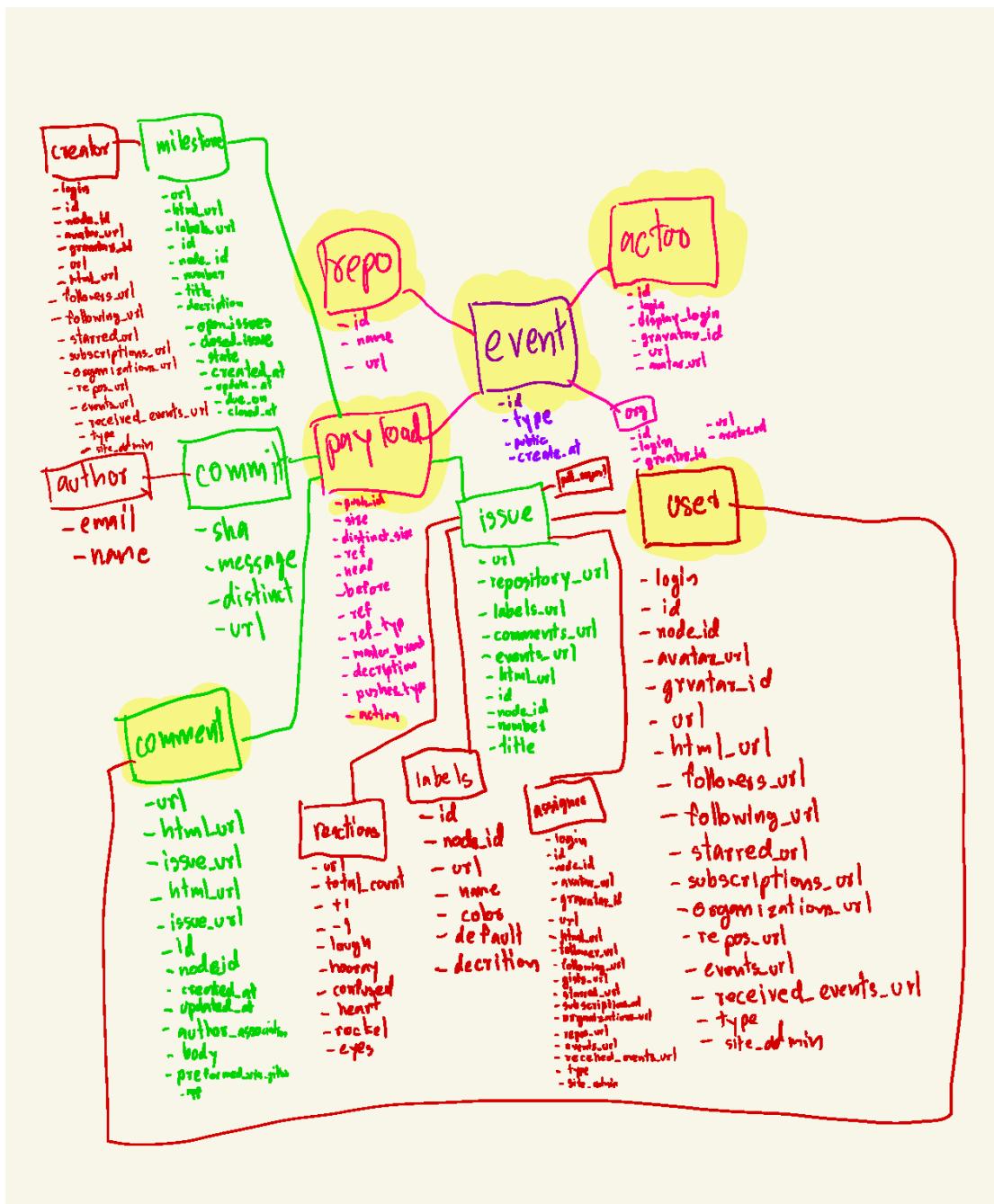


## การสร้างแบบจำลองข้อมูลด้วย Postgres (SQL)

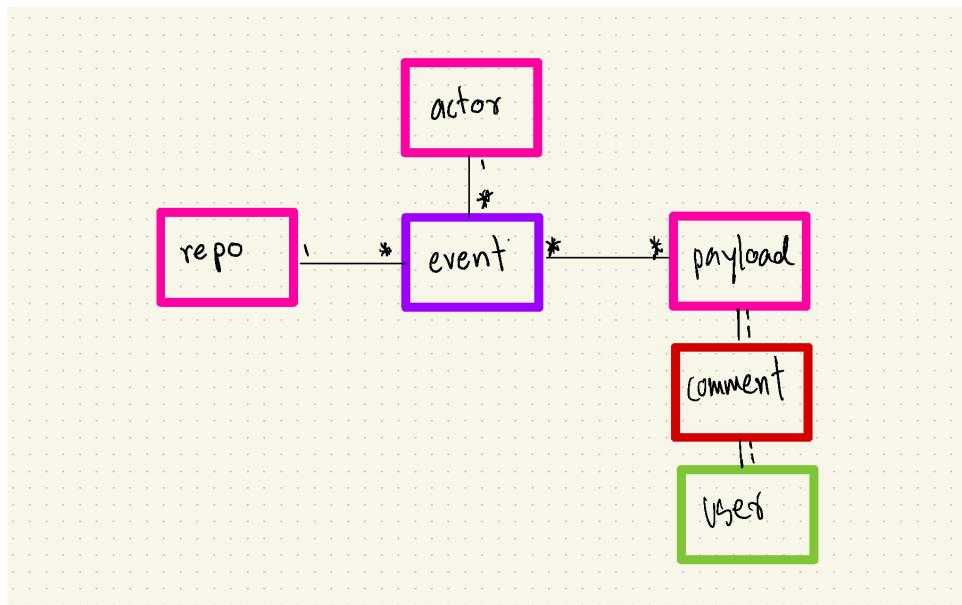
## วิธีการดำเนินการ

## 1. วิเคราะห์โครงสร้างข้อมูลของ JSON ไฟล์ ได้ดังรูป



จากภาพ ข้อมูลใน Json เป็นข้อมูล เหตุการณ์ หรือ Event ของ Github โดยมีโครงสร้างที่ยืดหยุ่น โดยแต่ละ record จะมีข้อมูลที่ไม่เท่ากัน และไม่เหมือนกัน

2. ออกแบบตารางที่ต้องการสร้าง (เนื่องจากมี Column เป็นจำนวนมากจึงเลือกเพียงตัวอย่างเท่านั้น)  
โดยเลือก ออกแบบตารางในลักษณะ Snowflake เนื่องจากมีตาราง Dimension มากกว่า 1 Level  
ได้ตารางดังรูป



จากรูป ได้ทำการเลือกตารางดังนี้

**Fact Table :** event

**Dimension Table :**

Level 1 repo, actor, payload

Level 2 comment

Level 3 user

โดย ตาราง event คือเป็นตารางหลักที่เก็บ measurements หรือ metrics ของ business process ไว้ จึงเลือกสร้างตารางดังกล่าว ซึ่งจะมีการเชื่อมกับตารางต่างๆ

ตาราง repo คือ ตารางที่เก็บข้อมูลในส่วนของพื้นที่ workspace ในการที่ user จะสร้างขึ้น เพื่อเก็บ code หรือไฟล์ต่างๆ

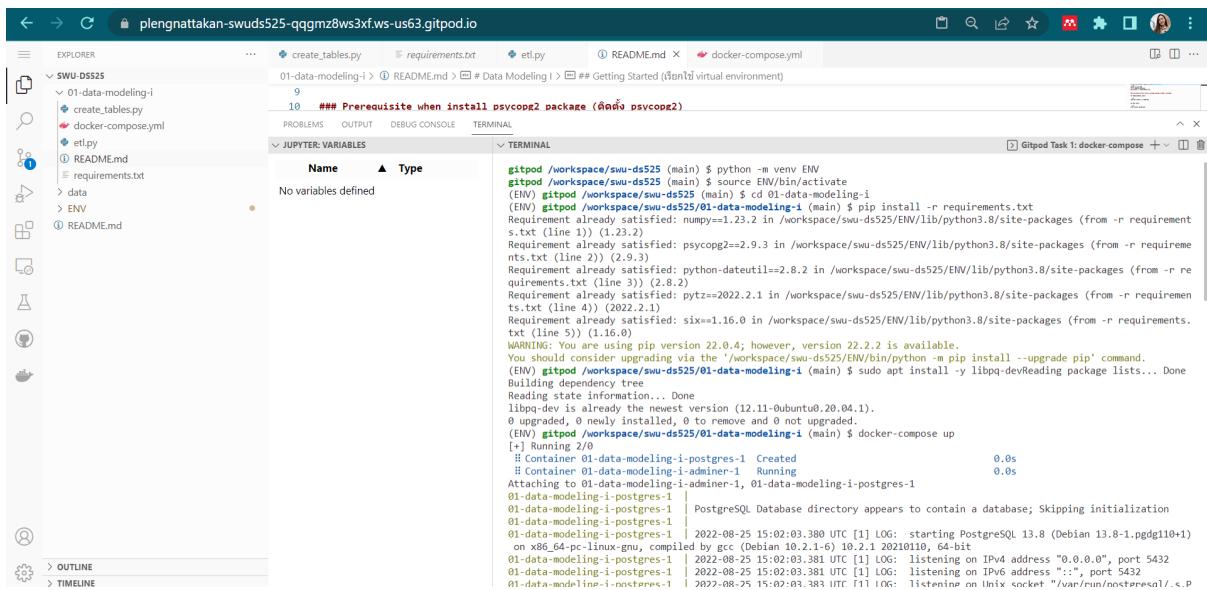
ตาราง actor คือ ตารางที่เก็บข้อมูลผู้กระทำ event ได้

ตาราง payload คือ ตารางที่เก็บการ push หรือ การกระทำใดๆ กับข้อมูล เพื่อเก็บเป็น version

ตาราง comment คือ ตารางที่เก็บข้อมูลการแสดงความคิดเห็น ที่เชื่อมโยงกับตาราง payload

ตาราง user คือ ตารางที่เก็บข้อมูลของ user ที่ทำการแสดง ความคิดเห็นที่เชื่อมกับตาราง comment

3. ทำการเตรียม Gitpod สำหรับการเขียนโปรแกรม โดยทำการ Sync Github account และ Gitpod เข้าด้วยกัน ติดตั้ง Postgresql เชื่อมกับ Gitpod เพื่อสามารถสร้างตาราง SQL บน Gitpod ได้ และทำการเตรียม environment



The screenshot shows the Gitpod interface with the following details:

- EXPLORER:** Shows the workspace structure:
  - SWU-DS525
  - 01-data-modeling-i
    - create\_tables.py
    - requirements.txt
    - etl.py
    - README.md
    - docker-compose.yml
  - REQUIREMENTS
  - data
  - ENV
  - README.md
- TERMINAL:** Shows the command-line output of running docker-compose up:

```
gitpod /workspace/swu-ds525 (main) $ python -m venv ENV
gitpod /workspace/swu-ds525 (main) $ cd 01-data-modeling-i
(ENV) gitpod /workspace/swu-ds525/01-data-modeling-i (main) $ pip install -r requirements.txt
Requirement already satisfied: numpy==1.23.2 in /workspace/swu-ds525/ENV/lib/python3.8/site-packages (from -r requirements.txt (line 1)) (1.23.2)
Requirement already satisfied: psycopg2==2.9.3 in /workspace/swu-ds525/ENV/lib/python3.8/site-packages (from -r requirements.txt (line 2)) (2.9.3)
Requirement already satisfied: python-dateutil==2.8.2 in /workspace/swu-ds525/ENV/lib/python3.8/site-packages (from -r requirements.txt (line 3)) (2.8.2)
Requirement already satisfied: pytz==2022.2.1 in /workspace/swu-ds525/ENV/lib/python3.8/site-packages (from -r requirements.txt (line 4)) (2022.2.1)
Requirement already satisfied: six==1.16.0 in /workspace/swu-ds525/ENV/lib/python3.8/site-packages (from -r requirements.txt (line 5)) (1.16.0)
WARNING: You are using pip version 22.0.4; however, version 22.2.2 is available.
You should consider upgrading via the 'pip install --upgrade pip' command.
(ENV) gitpod /workspace/swu-ds525/01-data-modeling-i (main) $ sudo apt install -y libpq-dev
Reading package lists... Done
Building dependency tree
Reading state information... Done
libpq-dev is already the newest version (12.11-0ubuntu0.20.04.1).
0 upgraded, 0 newly installed, 0 to remove and 0 not upgraded.
(ENV) gitpod /workspace/swu-ds525/01-data-modeling-i (main) $ docker-compose up
[*] Running 2/0
  ┌─────────────────────────────────────────────────────────────────┐
  ┌ Container 01-data-modeling-i-postgres-1 Created           0.0s
  ┌ Container 01-data-modeling-i-admeiner-1 Running           0.0s
  └─────────────────────────────────────────────────────────┘
Attaching to 01-data-modeling-i-admeiner-1, 01-data-modeling-i-postgres-1
01-data-modeling-i-postgres-1 | PostgreSQL Database directory appears to contain a database; Skipping initialization
01-data-modeling-i-postgres-1 | PostgreSQL is ready to accept connections
01-data-modeling-i-postgres-1 | 2022-08-25 15:02:03.388 UTC [1] LOG:  starting PostgreSQL 13.8 (Debian 13.8-1.pgdg110+1)
on port "5432" pid=11646 compiled by gcc (Debian 10.2.1-6) 10.2.1-82210110, 64-bit
01-data-modeling-i-postgres-1 | 2022-08-25 15:02:03.381 UTC [1] LOG:  listening on IPv4 address "0.0.0.0", port 5432
01-data-modeling-i-postgres-1 | 2022-08-25 15:02:03.381 UTC [1] LOG:  listening on IPv6 address "::", port 5432
01-data-modeling-i-postgres-1 | 2022-08-25 15:02:03.383 UTC [1] LOG:  listening on Unix socket "/var/run/postgresql/.s.PGSQL.5432"
```

4. แก้ไขไฟล์ docker-compose.yml เพื่อทำการแก้ไขค่า Config ที่ เชื่อมกับ Postgresql (DB,username,password)

```

version: "3.9"
services:
  postgres:
    image: postgres:13
    environment:
      POSTGRES_USER: postgres
      POSTGRES_PASSWORD: postgres
      POSTGRES_DB: postgres
    volumes:
      - postgres-data-volume:/var/lib/postgresql/data
    ports:
      - 5432:5432
    adminer:
      image: adminer:4.8.1
      restart: always
    depends_on:
      - postgres
    ports:
      - 8080:8080
volumes:
  postgres-data-volume:

```

## 5. แก้ไขไฟล์ create\_tables.py เพื่อทำการสร้างตารางไปยัง Postgresql โดยสร้างตามที่ออกแบบไว้ในข้อ 2

```

#Drop table ให้สัมภ์แล้ว Table เท่านั้นในเครื่องไม่
5 table_drop_repo = "DROP TABLE IF EXISTS Repo"
6 table_drop_actor = "DROP TABLE IF EXISTS Actor"
7 table_drop_user = "DROP TABLE IF EXISTS UserT"
8 table_drop_comment = "DROP TABLE IF EXISTS Comment"
9 table_drop_payload = "DROP TABLE IF EXISTS Payload"
10 table_drop_event = "DROP TABLE IF EXISTS Event"

#สร้างคำสั่ง sql สร้างตาราง
15 table_create_repo = """
CREATE TABLE IF NOT EXISTS Repo (
    repo_id BIGINT NOT NULL,
    name VARCHAR(100) NOT NULL,
    url VARCHAR(200) NOT NULL,
    PRIMARY KEY (repo_id)
)
"""

table_create_actor = """
CREATE TABLE IF NOT EXISTS Actor (
    actor_id BIGINT NOT NULL,
    login VARCHAR(100) NOT NULL,
    display_login VARCHAR(100) NOT NULL,
    gravatar_id VARCHAR(50),
    url VARCHAR(200) NOT NULL,
    avtar_url VARCHAR(200) NOT NULL,
)
"""

```

## 6. แก้ไขไฟล์ etl.py เพื่อทำการดึงข้อมูลจากลิงค์ json เข้าไปยัง ตารางที่สร้างไว้

```

1 #import library
2 import glob
3 import json
4 import os
5 from typing import List
6
7 import psycopg2
8
9
10
11 #สร้างคำแนะนำ sql สำหรับงาน
12 table_insert_repo = """
13     INSERT INTO Repo (repo_id,name,url) VALUES %s
14     ON CONFLICT (repo_id) DO NOTHING
15 """
16
17 table_insert_actor = """
18     INSERT INTO Actor (
19         actor_id,login,display_login,gravatar_id,avatar_url
20     ) VALUES %s
21     ON CONFLICT (actor_id) DO NOTHING
22 """
23
24 table_insert_user = """
25     INSERT INTO UserT (
26         user_id,login
27     ) VALUES %s
28     ON CONFLICT (user_id) DO NOTHING
29 """
30
31 table_insert_comment = """
32     INSERT INTO Comment VALUES %s
33     ON CONFLICT (comment_id) DO NOTHING

```

## 7. ทดสอบโปรแกรม

ภาษา: ภาษาไทย

PostgreSQL > postgres > postgres > public > เลือก: event

ออกจากระบบ

เลือก: event

เลือกชื่อ模版 แสดงโครงสร้าง เปรียบเทียบตารางแล้ว รายการใหม่

เลือก ค้นหา เรียงลำดับ จำนวน 50 จำนวนของรายการ ดำเนินการ

SELECT \* FROM "event" LIMIT 50 (0.001 วินาที) แก้ไข

Modify	event_id	type	public	create_at	repo_id	actor_id	push_id
<input type="checkbox"/>	23487929637	IssueCommentEvent	true	2022-08-17 15:51:05	75340147	1696078	NULL
<input type="checkbox"/>	23487929676	PushEvent	true	2022-08-17 15:51:05	525526680	66924041	10754961780
<input type="checkbox"/>	23487929674	PushEvent	true	2022-08-17 15:51:05	512700417	74971347	10754961752
<input type="checkbox"/>	23487929661	PushEvent	true	2022-08-17 15:51:05	525851446	91143053	10754961791
<input type="checkbox"/>	23487929682	PushEvent	true	2022-08-17 15:51:05	505029866	60825784	10754961801
<input type="checkbox"/>	23487929673	PushEvent	true	2022-08-17 15:51:05	525845306	102357035	10754961758
<input type="checkbox"/>	23487929588	PushEvent	true	2022-08-17 15:51:05	274724694	46447321	10754961749
<input type="checkbox"/>	23487929636	CreateEvent	true	2022-08-17 15:51:05	523708252	41898282	NULL
<input type="checkbox"/>	23487929580	IssuesEvent	true	2022-08-17 15:51:05	524701461	21972349	NULL
<input type="checkbox"/>	23487929591	PushEvent	true	2022-08-17 15:51:05	164541747	17869732	10754961756

หน้า 1 2 3 รวมผล 150 แก้ไข บันทึก Selected (0) ล้าง (150)

Project : 01-data-modeling-i  
Nattakan Towpunwong (64199130043)