

Topic : 03-building-a-data-warehouse

Name : Potchara Dockchaam

ID : 65199160192

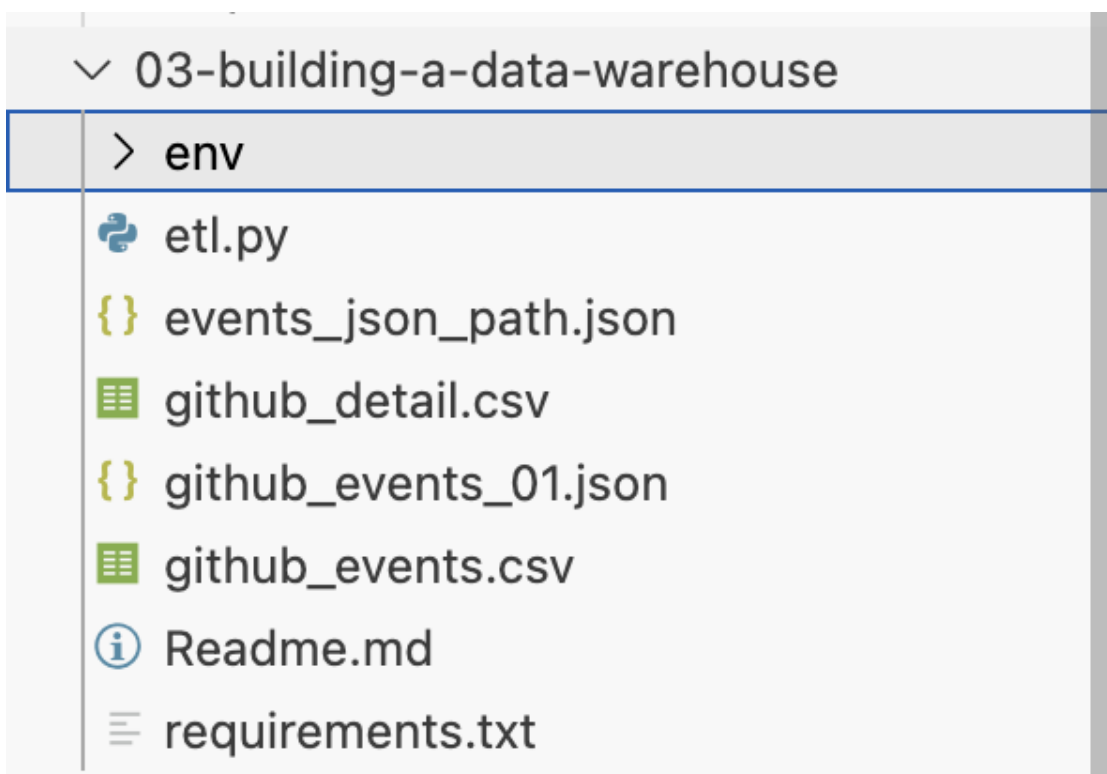
Documentation

1. เปลี่ยน Directory ให้อยู่ใน Folder 03-building-a-data-warehouse

```
$ cd 03-building-a-data-warehouse
```

2. สร้าง Environment Python เพื่อรองรับการติดตั้ง Library ให้เหมาะสมกับงาน

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



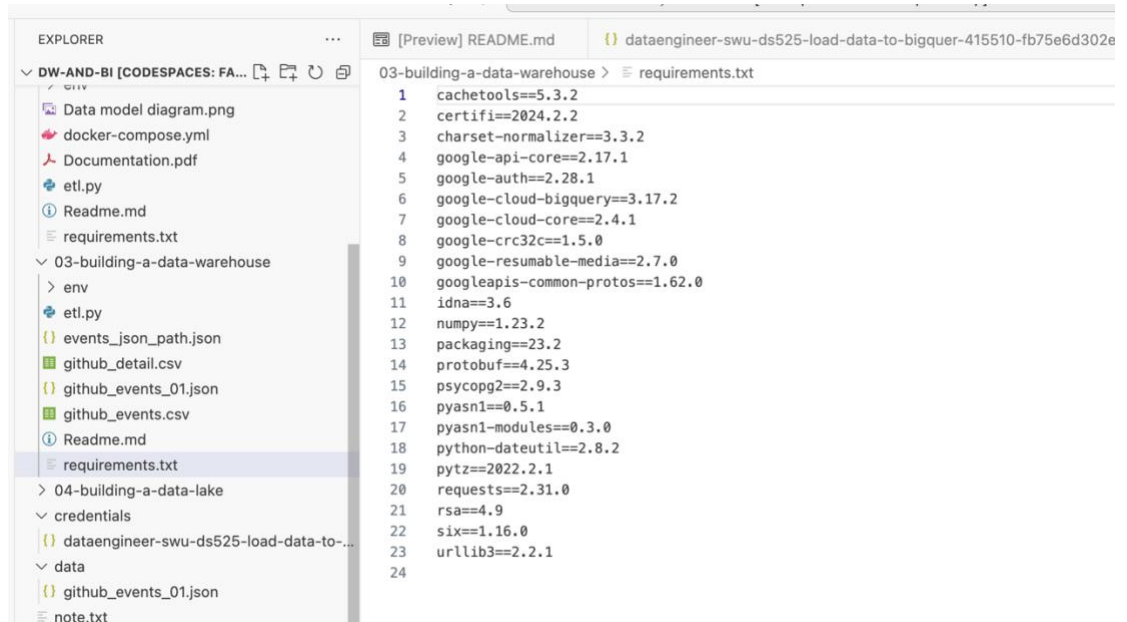
3. Active Env. ที่สร้างไว้ให้สามารถใช้งานได้

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


```
● @Fooklnwza007 → /workspaces/dw-and-bi/03-building-a-data-warehouse (main) $ source env/bin/activate
○ (env) @Fooklnwza007 → /workspaces/dw-and-bi/03-building-a-data-warehouse (main) $
```


4. Install Library ที่เกี่ยวข้องกับการใช้งาน

\$ pip install -r requirements.txt



5. สร้าง Key ใน Google query เพื่อให้สามารถสร้าง Script python เพื่อ Google bigquery ได้

 dataengineer ▼ iam

[←](#) Create service account

1

Service account details

Service account name

swu-ds525-load-data-to-bigquery

Display name for this service account

Service account ID *

swu-ds525-load-data-to-big-653

X ↺

Email address: swu-ds525-load-data-to-big-653@dataengineer-415510.iam.gserviceaccount.com

Service account description

Describe what this service account will do

CREATE AND CONTINUE

2

Grant this service account access to project (optional)

3

Grant users access to this service account (optional)

2

Grant this service account access to project (optional)

Grant this service account access to dataengineer so that it has permission to complete specific actions on the resources in your project. [Learn more](#)

Role

BigQuery Admin

Administer all BigQuery resources and data

IAM condition (optional) ?

+ ADD IAM CONDITION



+ ADD ANOTHER ROLE

CONTINUE

3

Grant users access to this service account (optional)

DONE

CANCEL

← Create service account

✓ Service account details

✓ Grant this service account access to project (optional)

3 Grant users access to this service account (optional)

Grant access to users or groups that need to perform actions as this service account. [Learn more](#)

Service account users role



Grant users the permissions to deploy jobs and VMs with this service account

Service account admins role



Grant users the permission to administer this service account

DONE

CANCEL

6. จากนั้นนำไฟล์ JSON แล้วเก็บไว้ในคอมฯของเรา

<

>

data model3

☰

↕

View

☰

☰

☰

☰

Group

📄

📄

Share

🏷️

🏷️

Edit Tags

⋮








⋮

Action

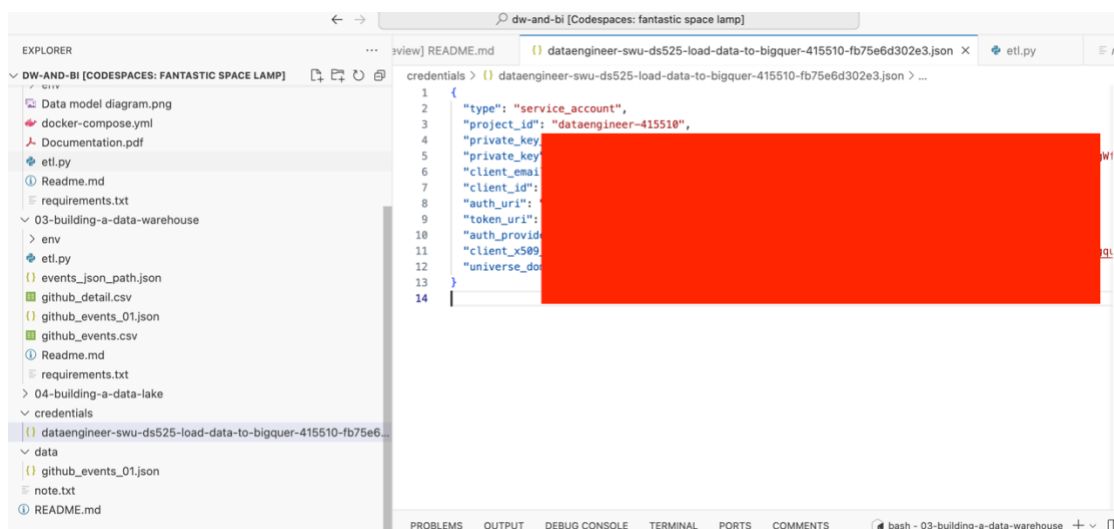
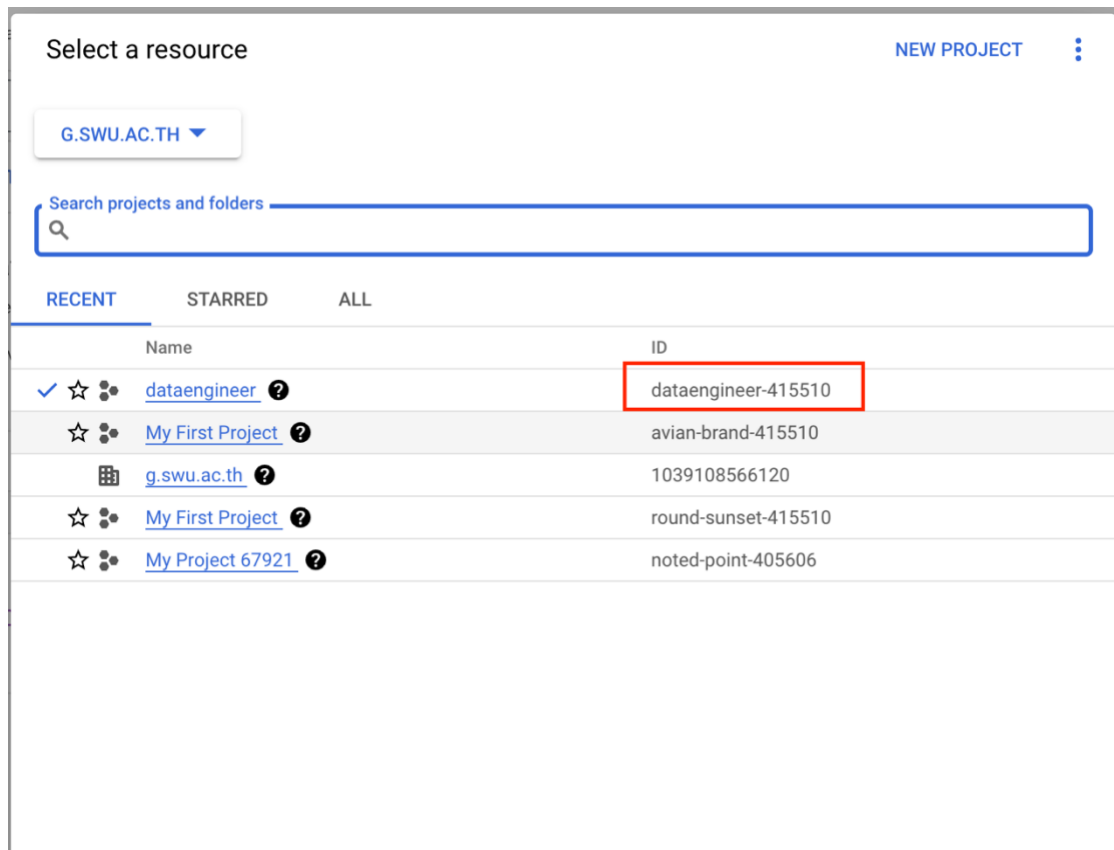
🔍

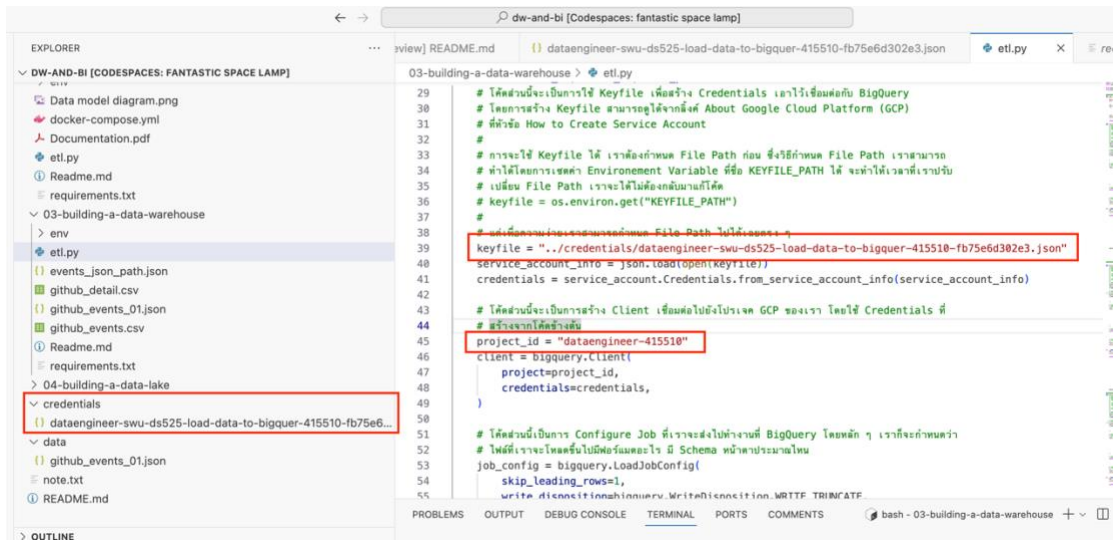
Search

Back/Forward

Name	Date Modified	Size	Kind
 dataengineer-415510-fb75e6d302e3.json	Today 18:38	2 KB	JSON
 etl_bigquery.py	Today 17:35	4 KB	Python script
 etl.py	15 Feb BE 2567 22:54	2 KB	Python script
 events_json_path.json	15 Feb BE 2567 22:54	110 bytes	JSON
 github_detail.csv	Today 19:32	70 KB	CSV Document
 github_events_01.json	15 Feb BE 2567 22:54	67 KB	JSON
 requirements.txt	Today 18:22	445 bytes	Plain Text

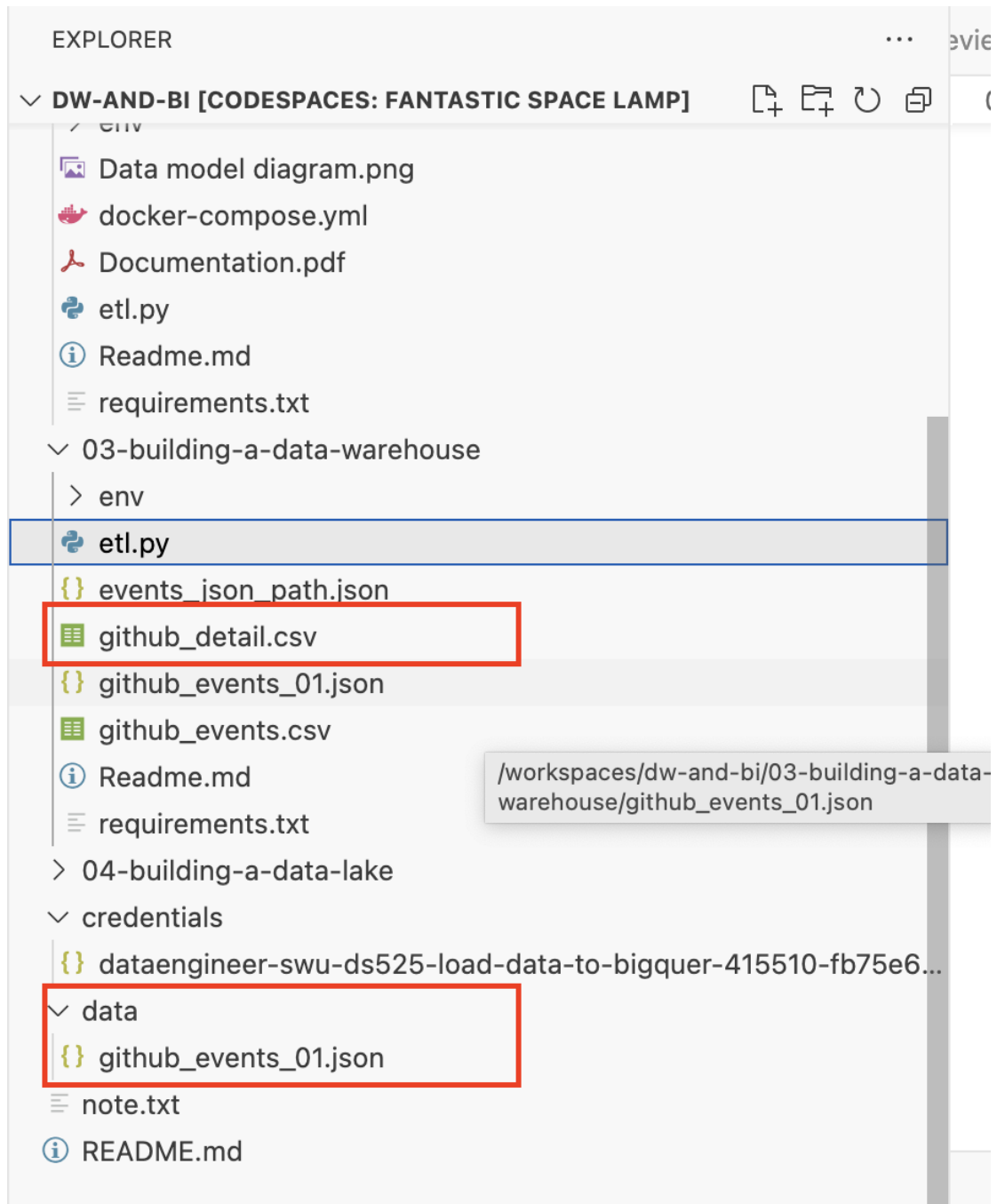
7. สร้าง Folder จัดเก็บไฟล์ JSON ใน code space หากอยู่ folder เดียวกับไฟล์ etl.py ไม่จำเป็นต้องมี ../credentials/ สามารถใช้ชื่อไฟล์ได้เลย project_id มาจาก project name บน Bigquery หรือสามารถดูได้จากไฟล์ Key ใน JSON ได้เช่นกัน





8. การสร้าง tables และเพิ่มข้อมูลสามารถทำได้ในไฟล์ etl.py ข้อมูลที่เข้ามาจะขึ้นอยู่กับโครงสร้าง และ ข้อมูลที่กำหนดในไฟล์ etl.py ในงานนี้จะให้ file_path = github_detail.csv และ ใช้ข้อมูล github_events_01.json

```
$ python etl.py
```



Google Cloud

dataengineer

bigquery

Search

Explorer

Type to search

Viewing resources.

SHOW STARRED ONLY

dataengineer-415510

External connections

github

actor

events

repo

SUMMARY

actor

dataengineer-415510.github

Last modified

Feb 26, 2024, 11:10:17 PM UTC+7

Data

asia-southeast1

actor

QUERY

SHARE

COPY

SNAPSHOT

DELETE

EXPORT

REFRESH

SCHEMA

DETAILS

PREVIEW

LINEAGE

DATA PROFILE

DATA QUALITY

Filter

Enter property name or value

Field name

Type

Mode

Key

Collation

Default Value

Policy Tags

Description

id

STRING

NULLABLE

-

-

-

-

-

name

STRING

NULLABLE

-

-

-

-

-

display_login

STRING

NULLABLE

-

-

-

-

-

url

STRING

NULLABLE

-

-

-

-

-

EDIT SCHEMA

VIEW ROW ACCESS POLICIES

Free trial status: \$10,640.40 credit and 91 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

Google Cloud

dataengineer

bigquery

Search

Your BigQuery projects will have new capabilities after February 14, 2024. Services and roles will be enabled automatically to help with these changes. [Learn more](#)

DISMISS

ACTIVATE

Explorer

Type to search

Viewing resources.

SHOW STARRED ONLY

dataengineer-415510

External connections

github

actor

events

repo

SUMMARY

events

dataengineer-415510.github

Last modified

Feb 26, 2024, 11:10:13 PM UTC+7

Data

asia-southeast1

events

QUERY

SHARE

COPY

SNAPSHOT

DELETE

EXPORT

REFRESH

SCHEMA

DETAILS

PREVIEW

LINEAGE

DATA PROFILE

DATA QUALITY

Filter

Enter property name or value

Field name

Type

Mode

Key

Collation

Default Value

Policy Tags

Description

id

STRING

NULLABLE

-

-

-

-

-

type

STRING

NULLABLE

-

-

-

-

-

actor_id

STRING

NULLABLE

-

-

-

-

-

created_at

TIMESTAMP

NULLABLE

-

-

-

-

-

EDIT SCHEMA

VIEW ROW ACCESS POLICIES

Google Cloud

dataengineer

bigquery

Search

Your BigQuery projects will have new capabilities after February 14, 2024. Services and roles will be enabled automatically to help with these changes. [Learn more](#)

DISMISS

Explorer

Type to search

Viewing resources.

SHOW STARRED ONLY

dataengineer-415510

External connections

github

actor

events

repo

SUMMARY

repo

dataengineer-415510.github

Last modified

Feb 26, 2024, 11:10:21 PM UTC+7

Data

asia-southeast1

repo

QUERY

SHARE

COPY

SNAPSHOT

DELETE

EXPORT

REFRESH

SCHEMA

DETAILS

PREVIEW

LINEAGE

DATA PROFILE

DATA QUALITY

Filter

Enter property name or value

Field name

Type

Mode

Key

Collation

Default Value

Policy Tags

Description

id

STRING

NULLABLE

-

-

-

-

-

name

STRING

NULLABLE

-

-

-

-

-

url

STRING

NULLABLE

-

-

-

-

-

EDIT SCHEMA

VIEW ROW ACCESS POLICIES