

Welcome to Capstone Project with Dynamic DAG



Team D - Data Engineer
batch 01

Table of Contents



Team Introduction

About Project

Dynamic DAG

Report

Live Demo

OUR TEAM

"The Future"



Muhammad Rifa



Karno

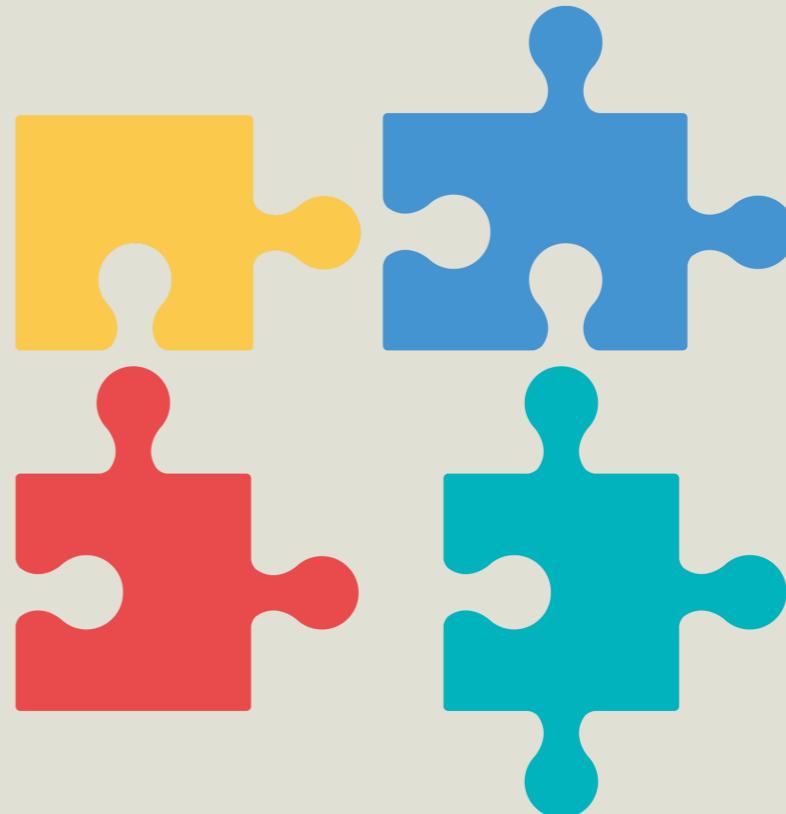


Yoga Martafian

About the Project

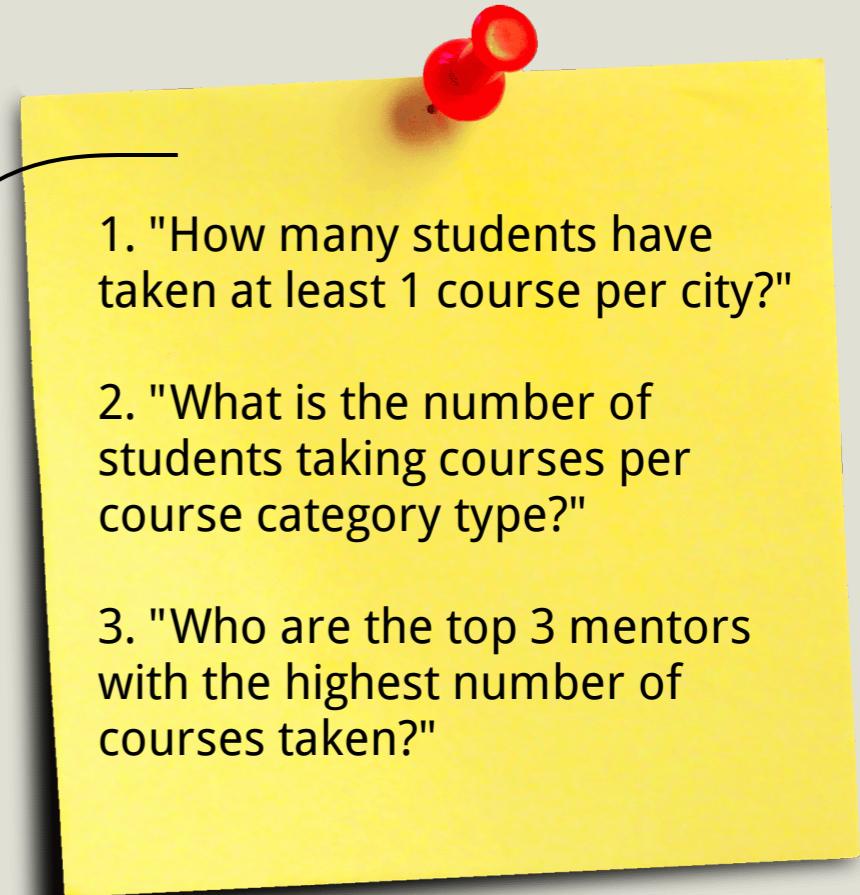
An edu-tech platform called "pinterskuy" provides online courses facilitated by professional mentors, and anyone can enroll in these courses. As the business gains momentum, the management level aims to conduct monitoring and evaluation of their online courses.

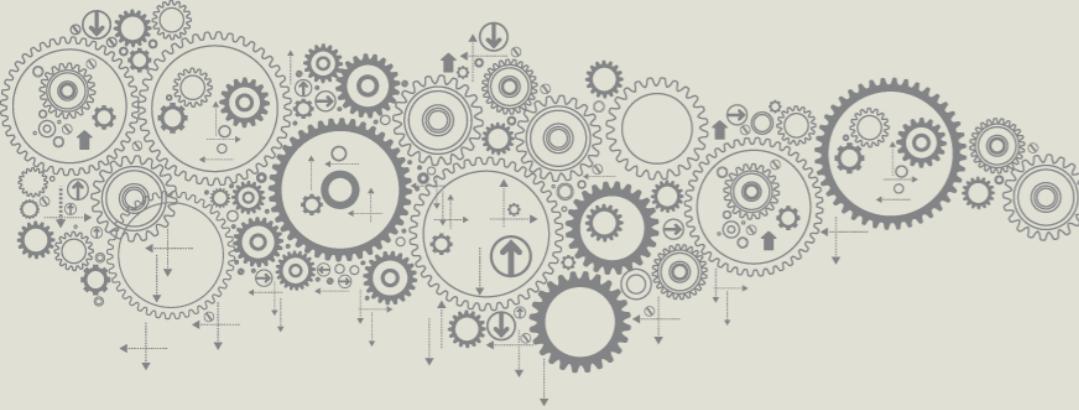
Therefore, the information that has been stored in different sources to date is intended to be consolidated into a ***single source of truth*** for subsequent analysis.



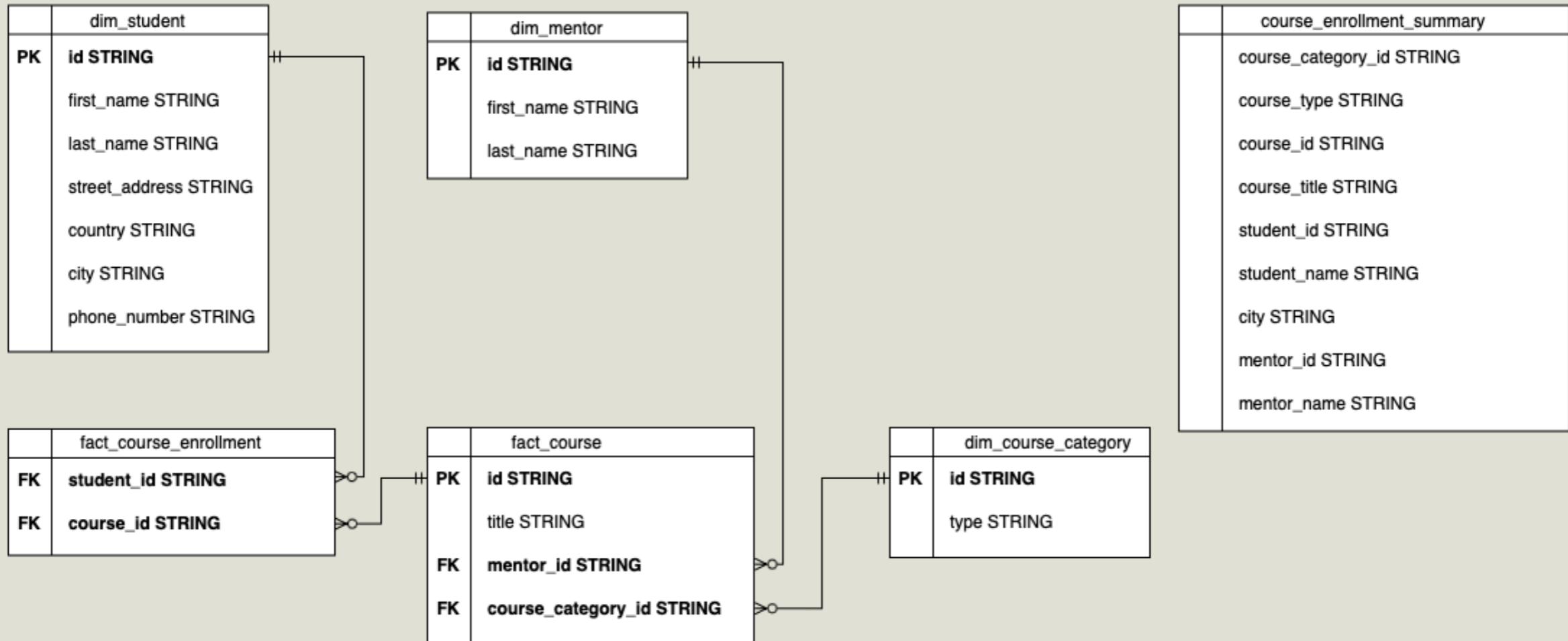
Our Goals

- Successfully ingested data from multiple sources into BigQuery.
- The tables created in BigQuery align with the specified requirements and visualization needs.
- Orchestrated ELT/ETL pipeline using Apache Airflow.
- Successfully created visualizations that address the questions posed.
- Created project documentation





ERD



Tools and Framework

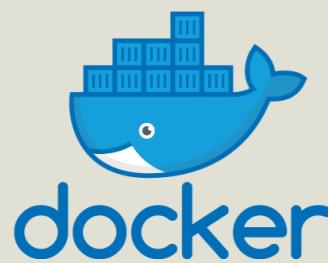
Deployment &
Orchestration



Apache
Airflow



Google Cloud Shell



docker

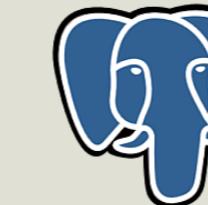


pythonTM

Data Warehousing



Google
Big Query



Postgre**SQL**

Data Storage and
Management

Data Visualization and
Reporting



Looker Studio



draw.io



GitHub



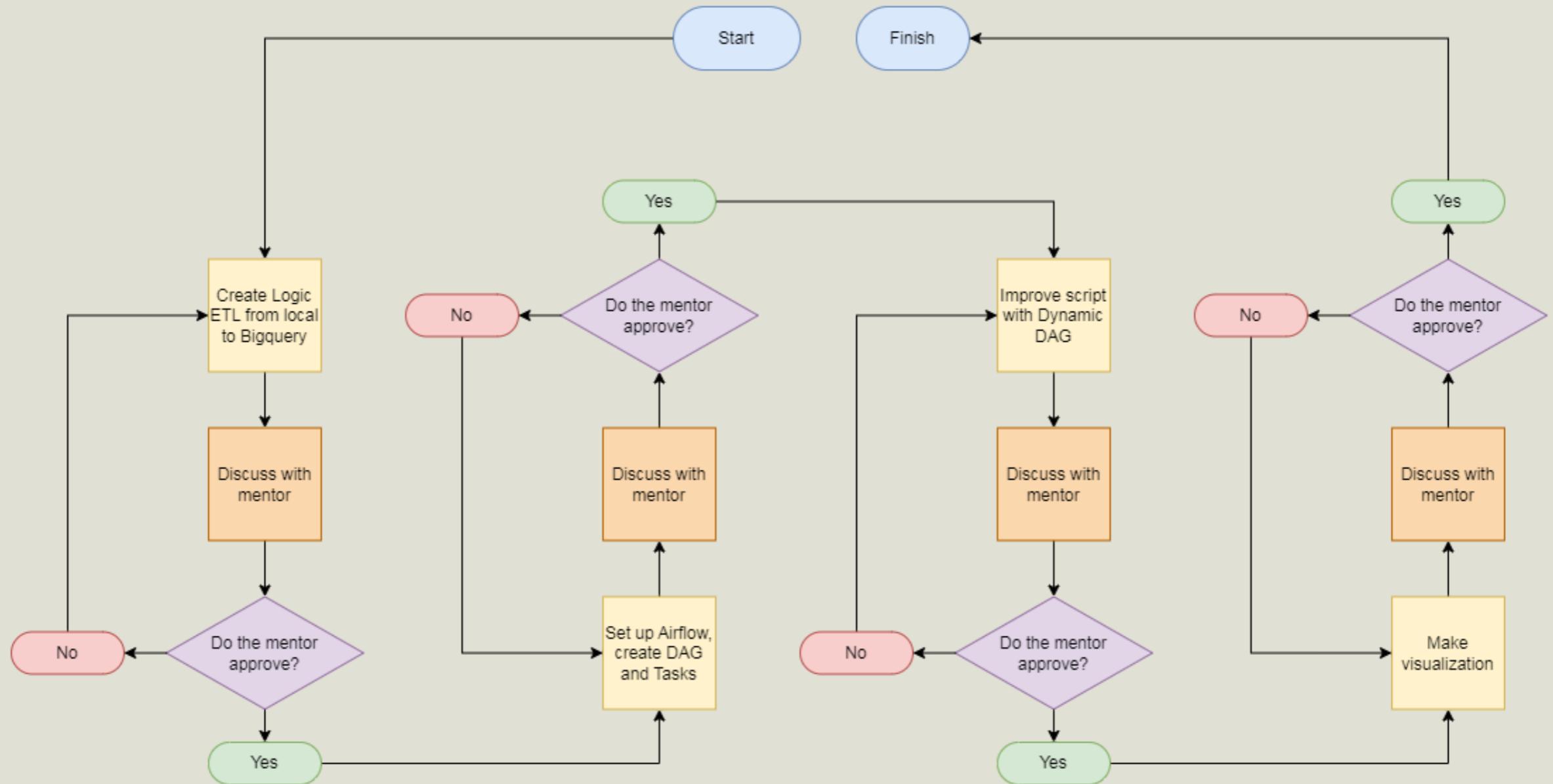
Trello



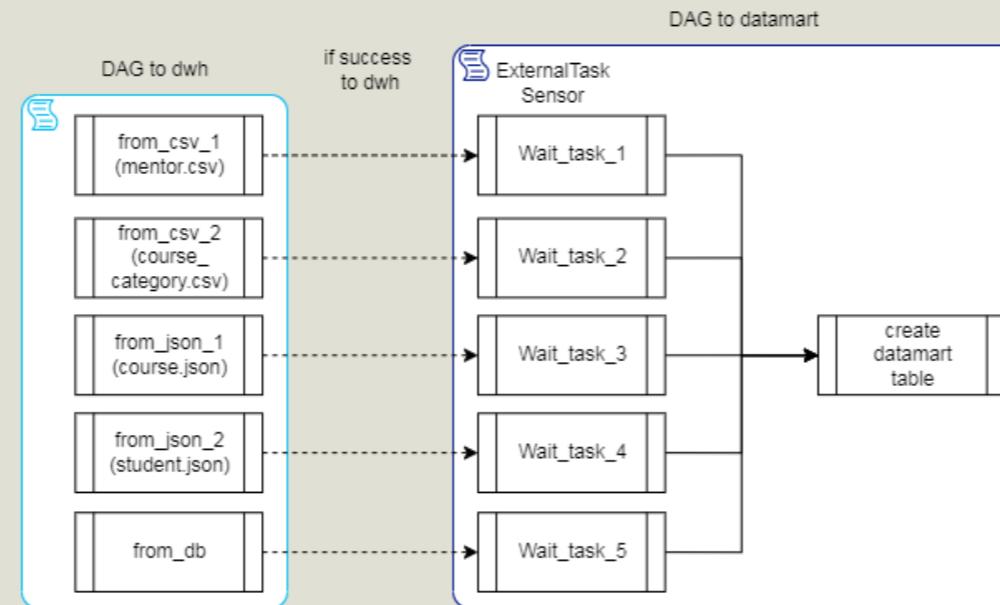
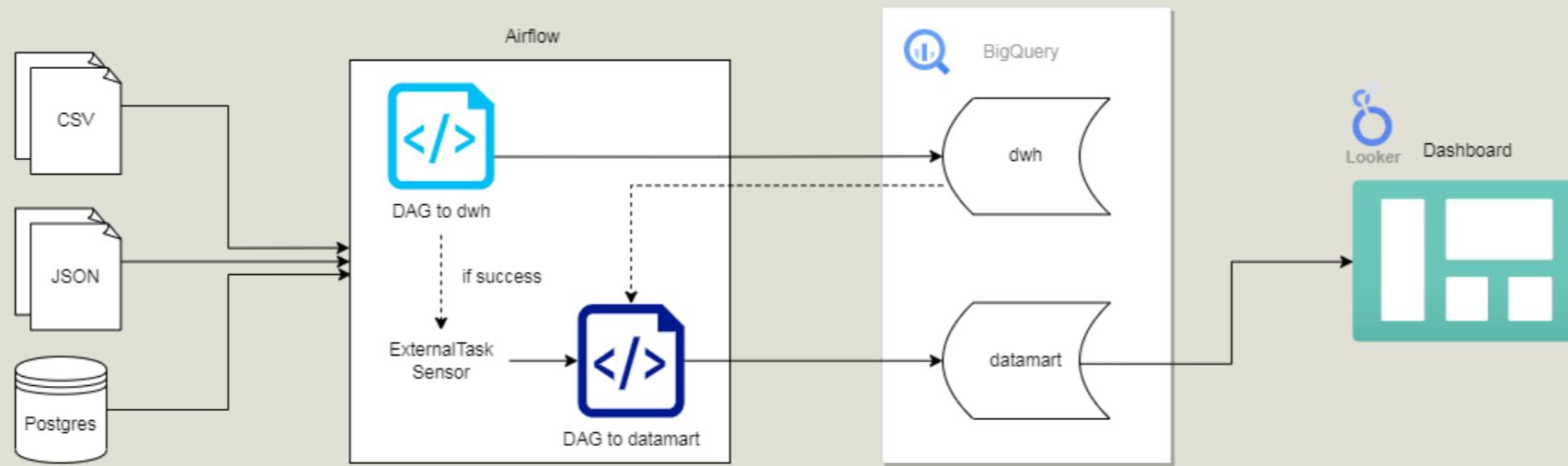
GitHub Actions

Continuous Integration

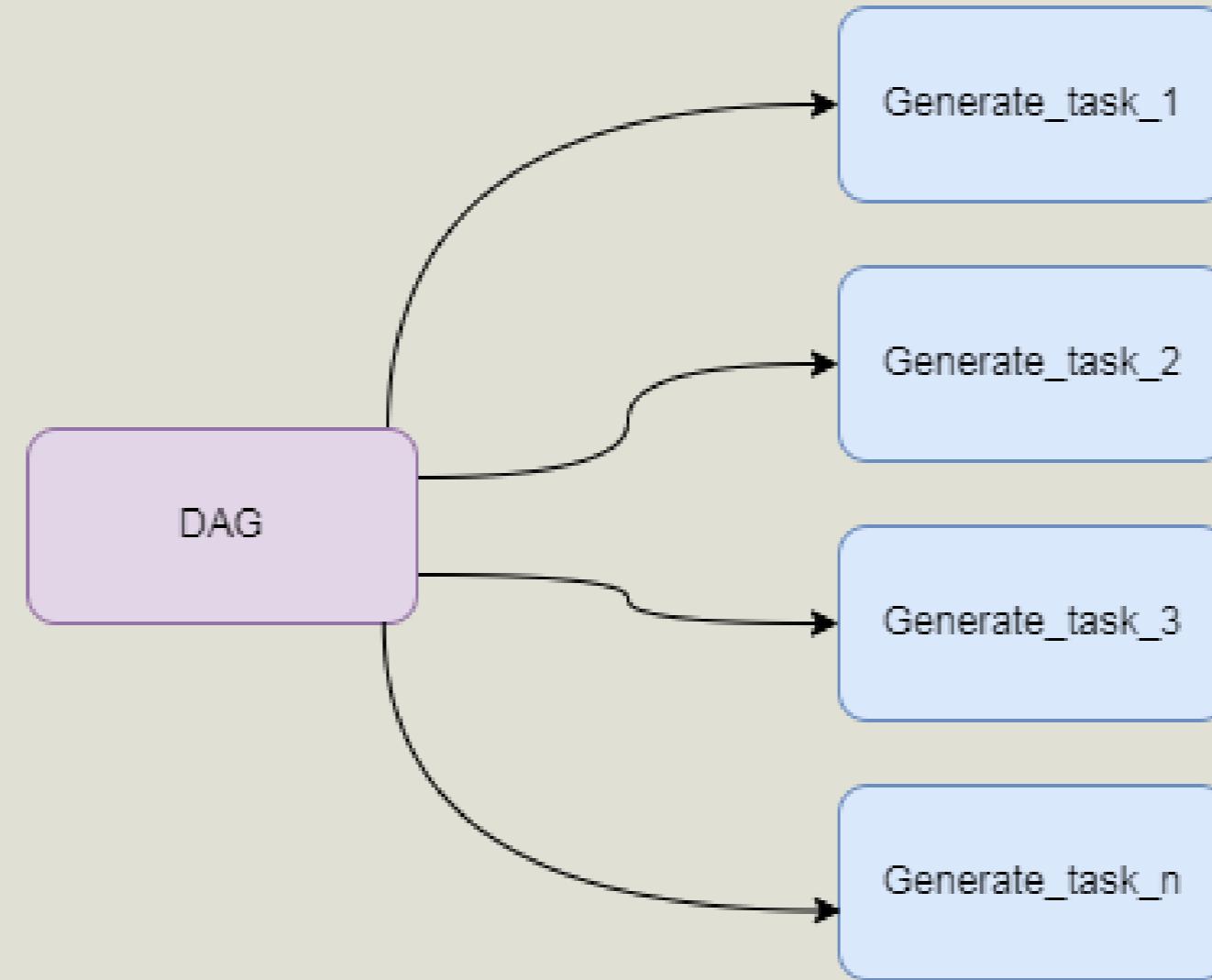
Workflow Project



How does it work?

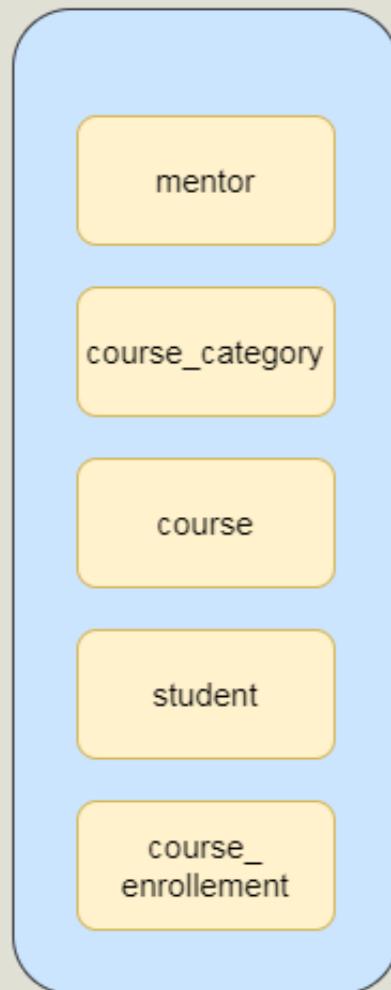


What is Dynamic DAG?

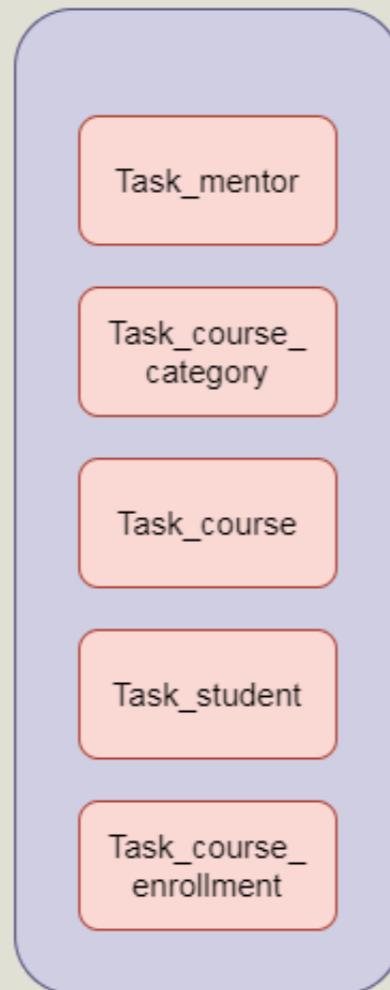


How do we do that?

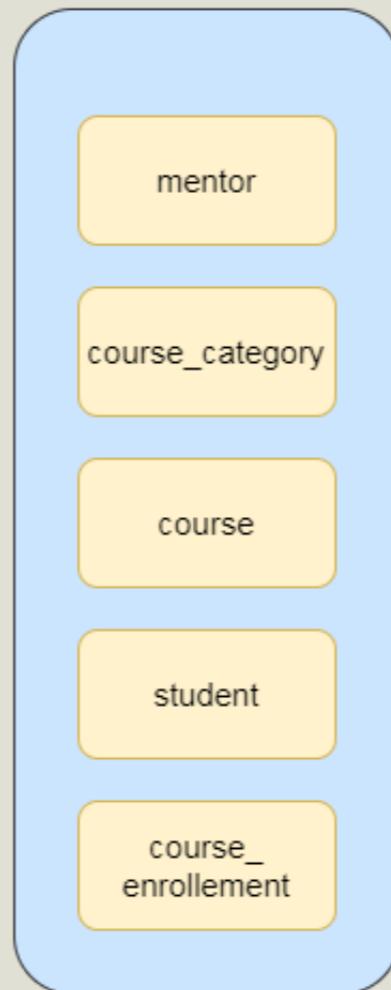
Tables



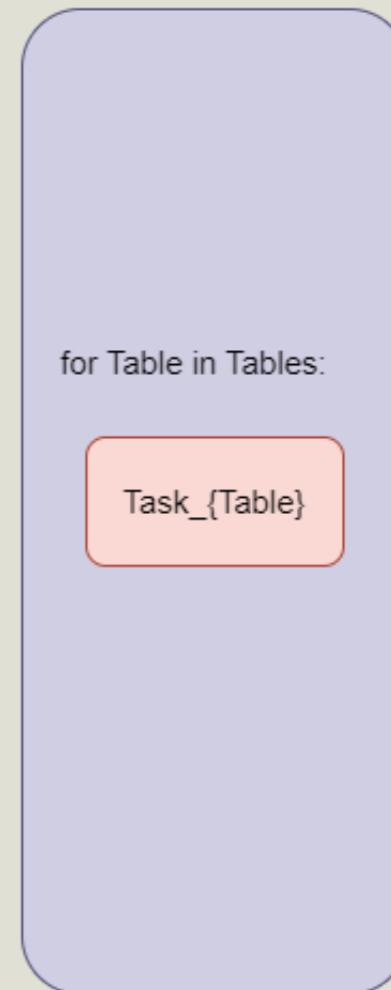
DAG



Tables



Dynamic DAG

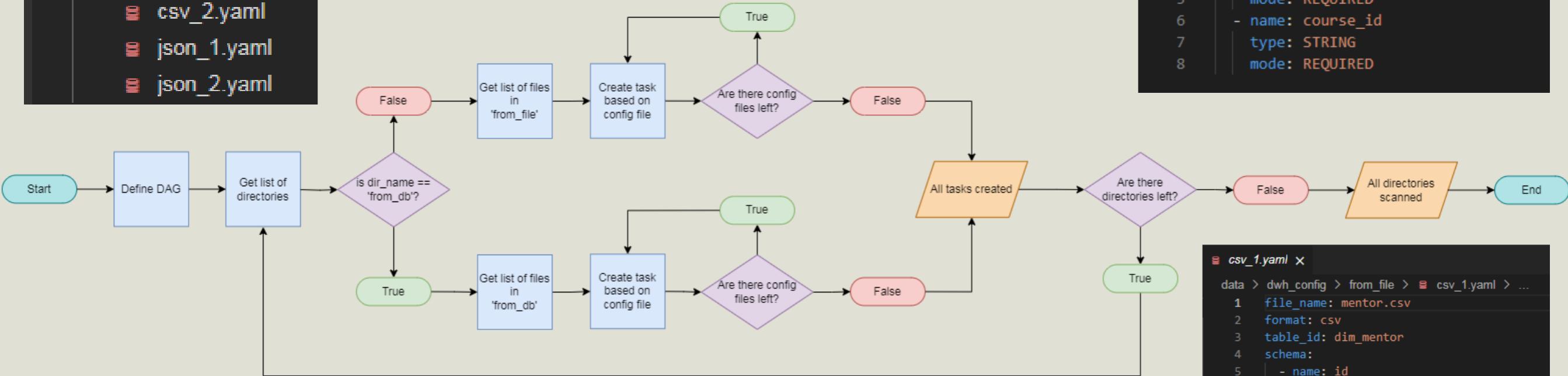


What's inside the code?

```

dwh_config
  from_db
    db.yaml
  from_file
    csv_1.yaml
    csv_2.yaml
    json_1.yaml
    json_2.yaml
  
```

* DAG to dwh



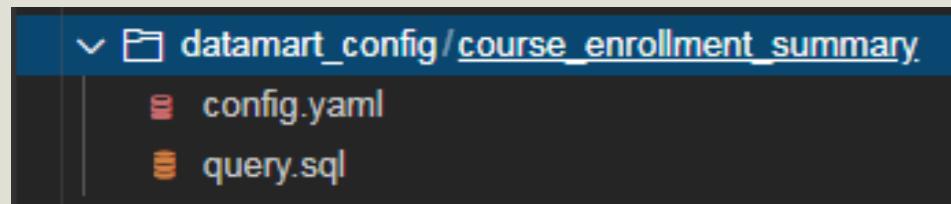
```

db.yaml
  table_id: fact_course_enrollment
  schema:
    - name: student_id
      type: STRING
      mode: REQUIRED
    - name: course_id
      type: STRING
      mode: REQUIRED
  
```

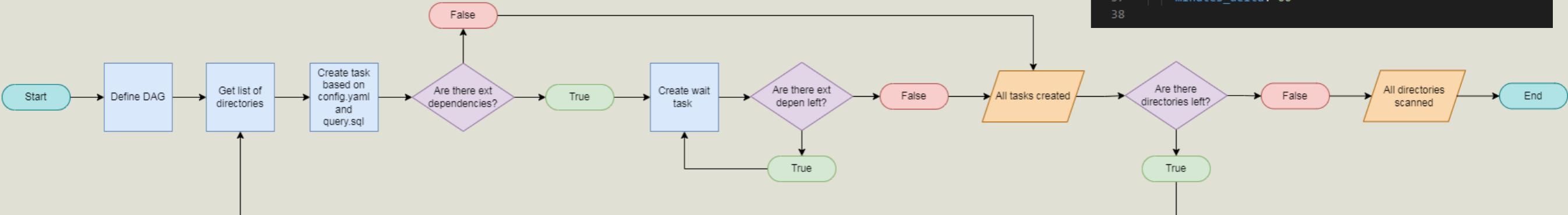
```

csv_1.yaml
  file_name: mentor.csv
  format: csv
  table_id: dim_mentor
  schema:
    - name: id
      type: STRING
      mode: REQUIRED
    - name: first_name
      type: STRING
    - name: last_name
      type: STRING
  
```

What's inside the code?



* DAG to datamart



```

config.yaml x
data > datamart_config > course_enrollment_summary > config.yaml > ...
22 external_task_dependencies:
23   - dag_id: dag_etl_to_dwh
24     task_id: from_db
25     minutes_delta: 60
26   - dag_id: dag_etl_to_dwh
27     task_id: from_csv_1
28     minutes_delta: 60
29   - dag_id: dag_etl_to_dwh
30     task_id: from_csv_2
31     minutes_delta: 60
32   - dag_id: dag_etl_to_dwh
33     task_id: from_json_1
34     minutes_delta: 60
35   - dag_id: dag_etl_to_dwh
36     task_id: from_json_2
37     minutes_delta: 60
38
  
```

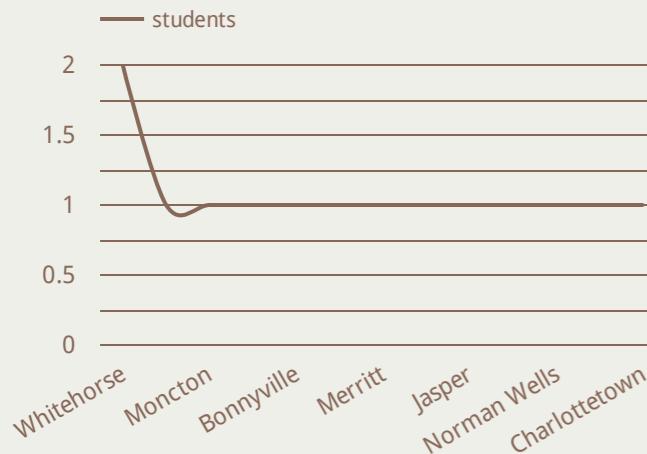
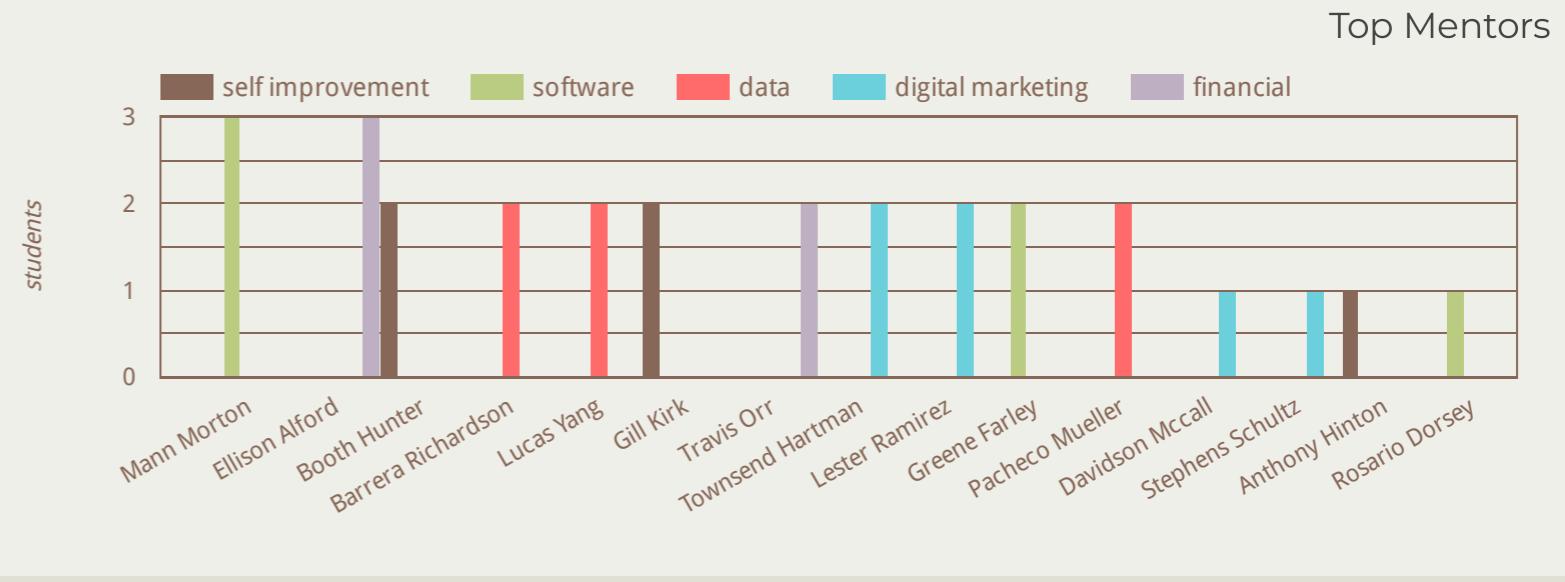
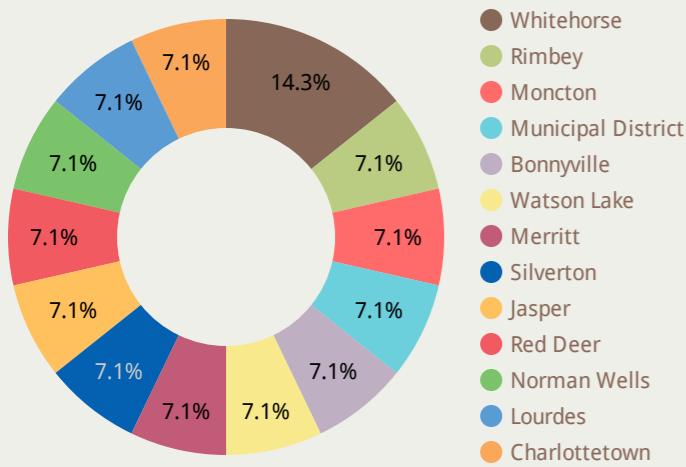
Config File and Query Example

```
config.yaml x  
data > datamart_config > course_enrollment_summary > config.yaml > ...  
1 dataset: datamart  
2 table_id: course_enrollment_summary  
3 schema:  
4   - name: course_category_id  
5     type: STRING  
6   - name: course_type  
7     type: STRING  
8   - name: course_id  
9     type: STRING  
10  - name: course_title  
11    type: STRING  
12  - name: student_id  
13    type: STRING  
14  - name: student_name  
15    type: STRING  
16  - name: city  
17    type: STRING  
18  - name: mentor_id  
19    type: STRING  
20  - name: mentor_name  
21    type: STRING
```

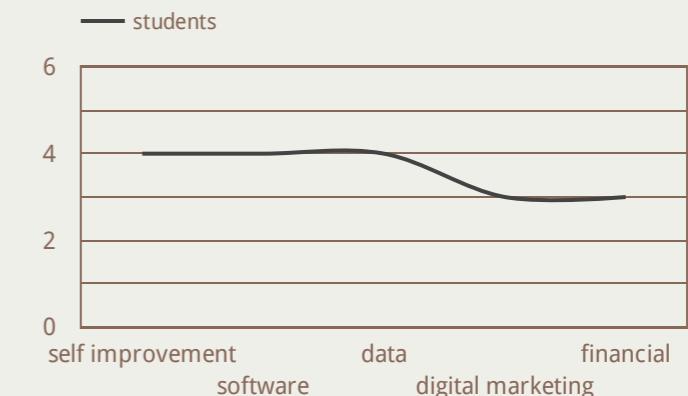
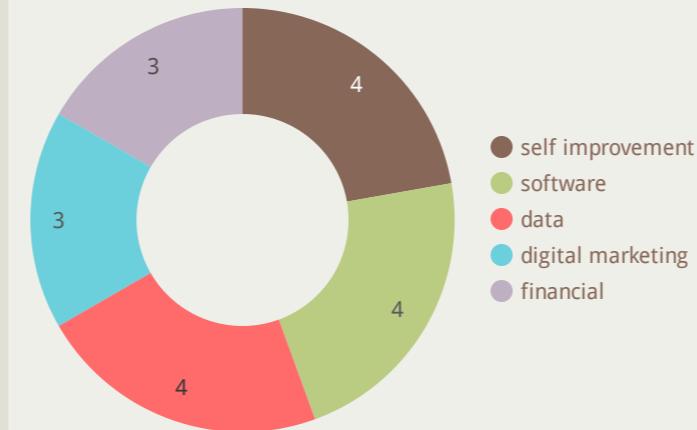
```
query.sql x  
data > datamart_config > course_enrollment_summary > query.sql  
1 select  
2   course_category_id  
3   , type as course_type  
4   , course_id  
5   , title as course_title  
6   , student_id  
7   , concat(s.first_name, ' ', s.last_name) as student_name  
8   , city  
9   , mentor_id  
10  , concat(m.first_name, ' ', m.last_name) as mentor_name  
11  from `dwh.dim_student` as s  
12  join `dwh.fact_course_enrollment` as ce  
13  on ce.student_id = s.id  
14  join `dwh.fact_course` as c  
15  on c.id = ce.course_id  
16  join `dwh.dim_course_category` as cc  
17  on cc.id = c.course_category_id  
18  join `dwh.dim_mentor` m  
19  on c.mentor_id = m.id;
```

Pinter Skuy Online Course Report

Student Per City



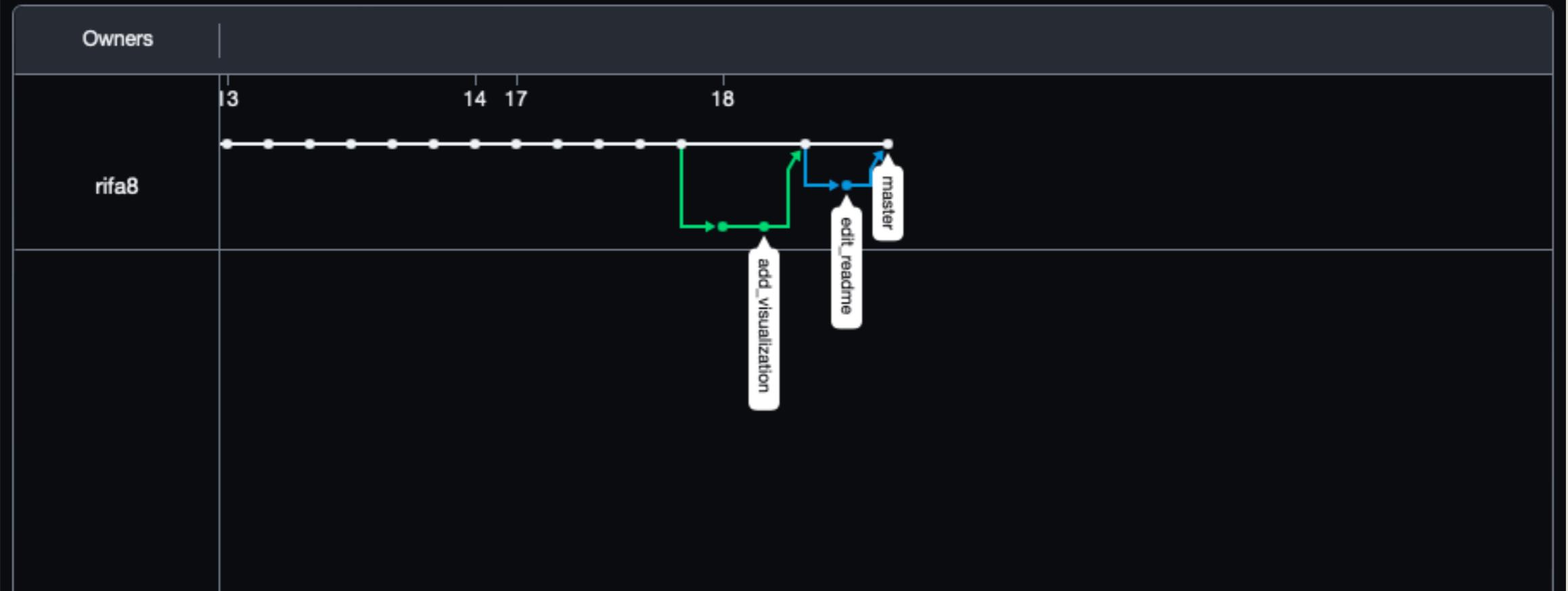
Students Per Course Category



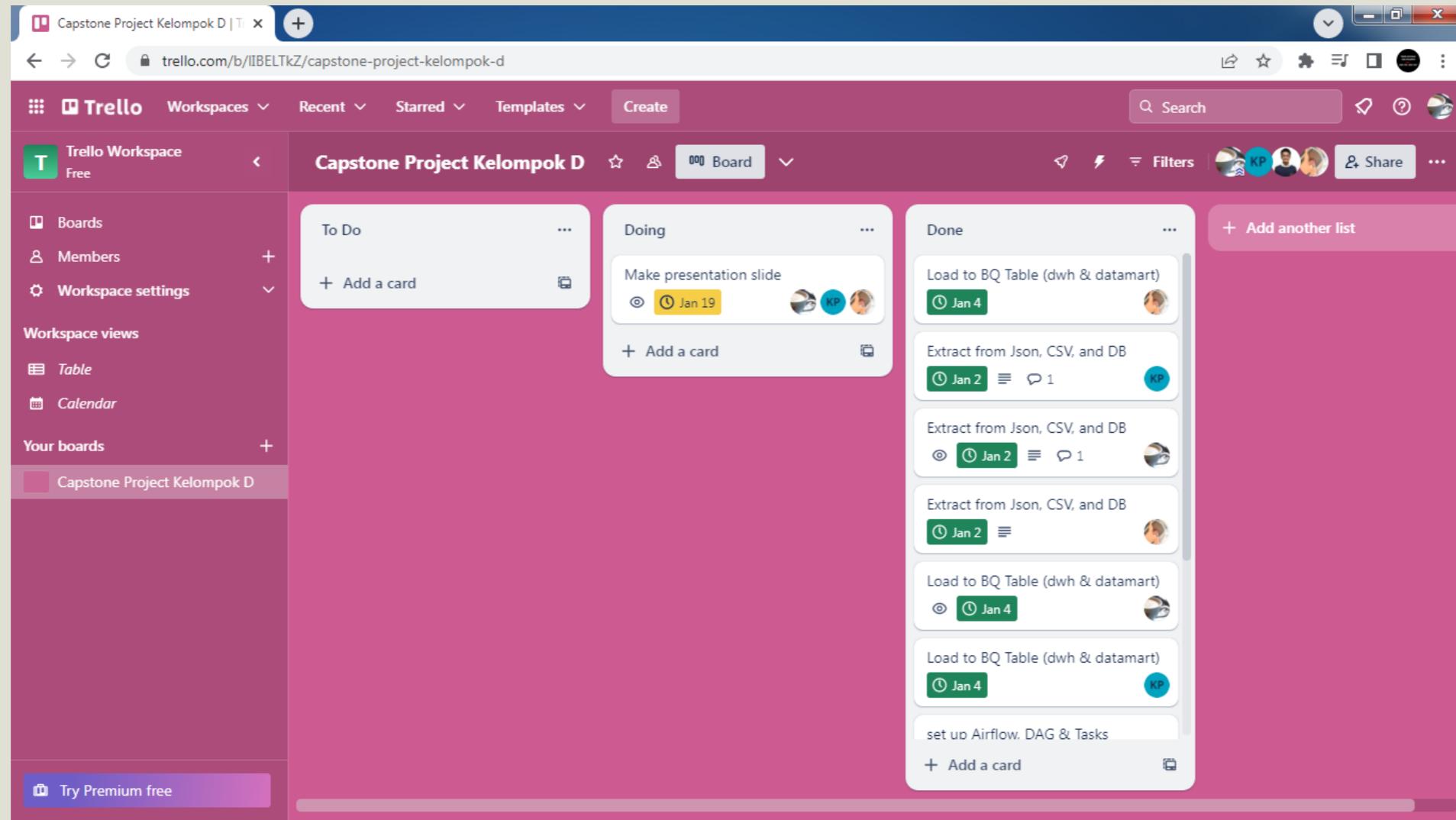
GIT WORKFLOW

Network graph

Timeline of the most recent commits to this repository and its network ordered by most recently pushed to.



Project Management and Teamwork



Project Demonstration

JUST DO IT!!!





Thank You

Team D - "The Future"
Data Engineer 01
Alterra Academy