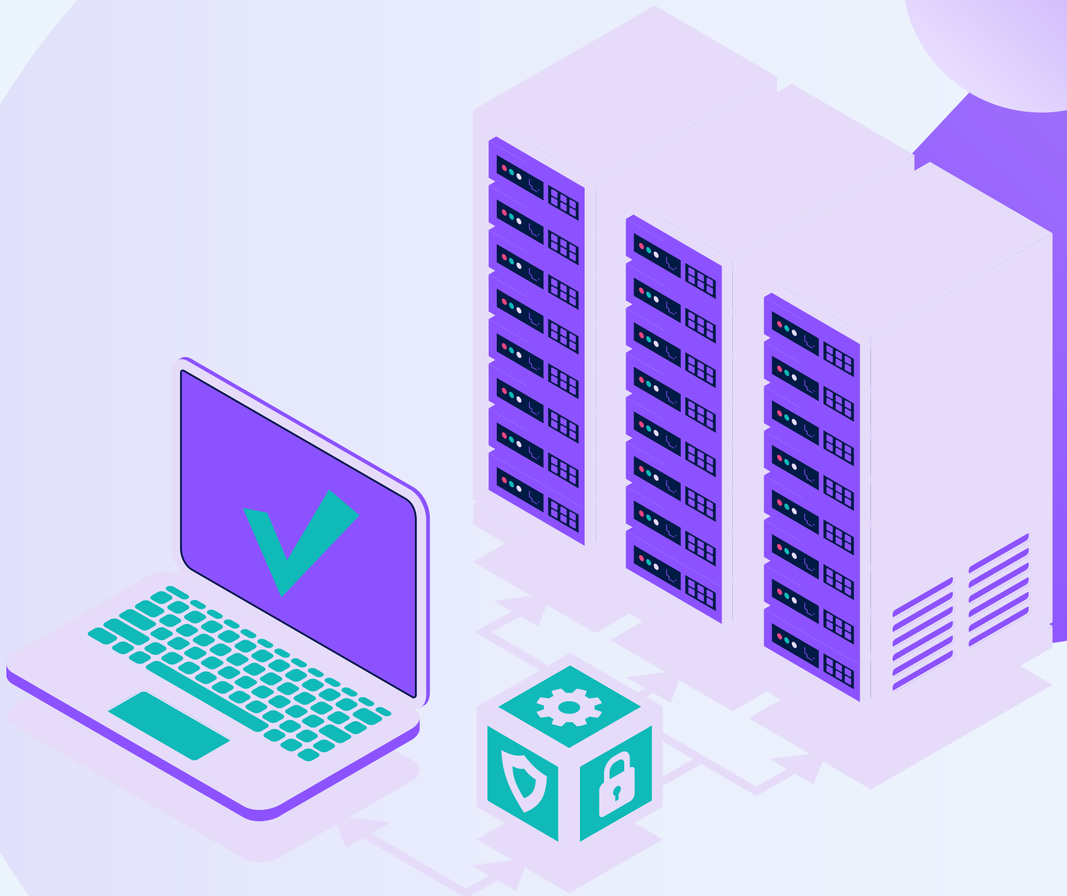
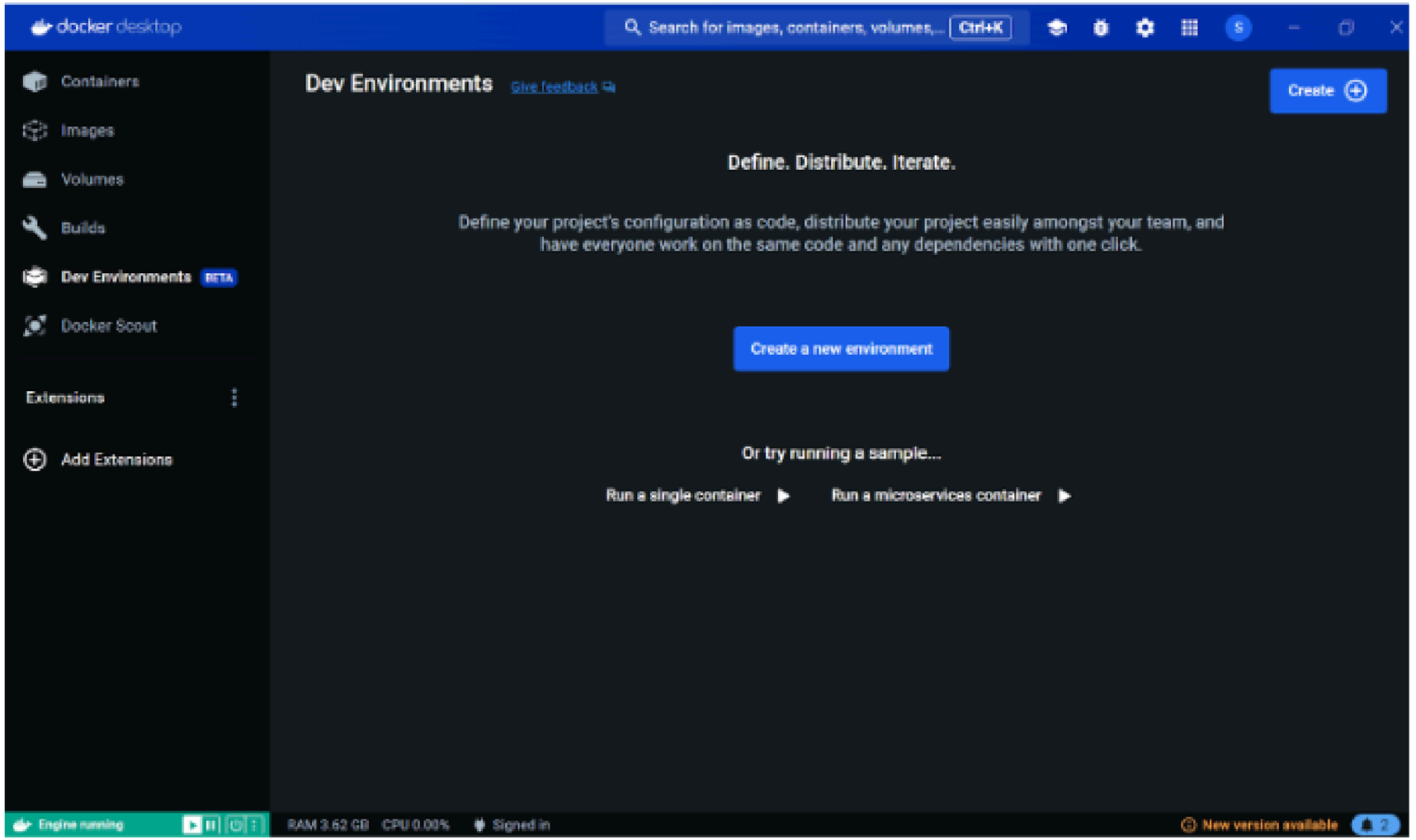
Project 4

Kelompok 8 - Infinity



1. Open and start your Docker



2. Create the docker-compose.yml and create MySQL and Postgres Container declaratively

```
(env) PS C:\Users\hallo\OneDrive\Documents\Kuliah\Stupen\project4> git clone https://github.com/ayyoubmaul/docker_etl.git
Cloning into 'docker_etl'...
remote: Enumerating objects: 170, done.
remote: Counting objects: 100% (170/170), done.
remote: Compressing objects: 100% (75/75), done.
remote: Total 170 (delta 89), reused 139 (delta 75), pack-reused 0
Receiving objects: 100% (170/170), 414.63 KiB | 387.00 KiB/s, done.
Resolving deltas: 100% (89/89), done.
```

docker-compose.yml

docker_etl > examples > etl > docker-compose.yml

```yaml
version: "3.7"

services:
  db-mysql:
    image: mysql:8
    container_name: db-mysql
    environment:
      MYSQL_ROOT_PASSWORD: mysql
      MYSQL_USER: mysql
      MYSQL_PASSWORD: mysql
      MYSQL_DATABASE: operational
    volumes:
      # - ./init.sql:/docker-entrypoint-initdb.d/init.sql
      - ./data:/docker-entrypoint-initdb.d/init_data
    ports:
      - "3305:3306" # local:docker

  db-psql:
    image: postgres:11
    container_name: db-postgres
    environment:
      POSTGRES_USER: postgres
      POSTGRES_PASSWORD: postgres
    ports:
      - "5432:5432"

```

3. Run this command to make MySQL and Postgres Image run as a Container detached mode: docker-compose up -d run in background non-detached mode: docker-compose up run in foreground

```
(env) PS C:\Users\hallo\OneDrive\Documents\Kuliah\Stupen\project4> cd docker_etl
(env) PS C:\Users\hallo\OneDrive\Documents\Kuliah\Stupen\project4\docker_etl> cd examples
(env) PS C:\Users\hallo\OneDrive\Documents\Kuliah\Stupen\project4\docker_etl\examples> cd etl
(env) PS C:\Users\hallo\OneDrive\Documents\Kuliah\Stupen\project4\docker_etl\examples\etl> docker-compose up
error during connect: this error may indicate that the docker daemon is not running: Get "http://%2F%2F.%2Fpipe%2Fdocker_engine/v1.24/con
tainers/json?all=1&filters=%7B%22label%22%3A%7B%22com.docker.compose.config-hash%22%3Atrue%2C%22com.docker.compose.project%3Detl%22%3Atru
e%7D%7D": open //./pipe/docker_engine: The system cannot find the file specified.
(env) PS C:\Users\hallo\OneDrive\Documents\Kuliah\Stupen\project4\docker_etl\examples\etl> docker-compose up
[+] Running 25/12
 ✓ db-psql 13 layers [⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿]      0B/0B      Pulled                                                    184.6s
 ✓ db-mysql 10 layers [⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿⣿]      0B/0B      Pulled                                                    128.2s
```

```
[+] Running 3/3
 ✓ Network etl_default     Created
 ✓ Container db-mysql      Created
 ✓ Container db-postgres   Created
Attaching to db-mysql, db-postgres
db-postgres  | The files belonging to this database system will be owned by user "postgres".
db-postgres  | This user must also own the server process.
db-postgres  |
db-postgres  | The database cluster will be initialized with locale "en_US.utf8".
db-postgres  | The default database encoding has accordingly been set to "UTF8".
db-postgres  | The default text search configuration will be set to "english".
db-postgres  |
db-postgres  | Data page checksums are disabled.
db-postgres  |
db-postgres  | fixing permissions on existing directory /var/lib/postgresql/data ... ok
db-postgres  | creating subdirectories ... ok
db-postgres  | selecting default max_connections ... 100
db-postgres  | selecting default shared buffers ... 128MB
```

4. Login to MySQL database: mysql --local-infile=1 -uroot -pmysql operational < /docker-entrypoint-initdb.d/init.sql

```
sh-4.4# mysql --local-infile=1 -uroot -pmysql operational
mysql: [Warning] Using a password on the command line interface can be insecure.
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 14
Server version: 8.3.0 MySQL Community Server - GPL

Copyright (c) 2000, 2024, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql>
```

Membuat file init.sql

```
mysql> CREATE DATABASE IF NOT EXISTS operational;
```

```
mysql> SET GLOBAL local_infile=1;
```

```
mysql> USE operational;
```

```
mysql> CREATE TABLE IF NOT EXISTS youtube (
    -> `rank` INT(5),
    -> youtuber VARCHAR(50),
    -> subscribers BIGINT,
    -> video_views VARCHAR(50),
    -> category VARCHAR(50),
    -> title VARCHAR(50),
    -> uploads INT(10),
    -> country VARCHAR(50),
    -> abbreviation VARCHAR(50),
    -> channel_type VARCHAR(50),
    -> video_views_rank INT(10),
    -> country_rank INT(11),
    -> channel_type_rank VARCHAR(50),
    -> video_views_for_the_last_30_days INT(10),
    -> lowest_monthly_earnings INT(10),
    -> highest_monthly_earnings INT(10),
    -> lowest_yearly_earnings INT(10),
    -> highest_yearly_earnings INT(10),
    -> subscribers_for_last_30_days INT(10),
    -> created_year INT(10),
    -> created_month VARCHAR(50),
    -> created_date INT(5),
    -> gross_tertiary_education_enrollment_percent FLOAT,
    -> population BIGINT,
    -> unemployment_rate FLOAT,
    -> urban_population BIGINT,
    -> latitude FLOAT,
    -> longitude FLOAT
    -> );
Query OK, 0 rows affected, 13 warnings (0.01 sec)
```

```
2024-04-29 12:48:25 2024-04-29 05:48:25+00:00 [Note] [Entrypoint]: Creating database operational
2024-04-29 12:48:25 2024-04-29 05:48:25+00:00 [Note] [Entrypoint]: Creating user mysql
2024-04-29 12:48:25 2024-04-29 05:48:25+00:00 [Note] [Entrypoint]: Giving user mysql access to schema operational
```

5. If error Loading local data is disabled; this must be enabled on both the client and server sides Run this query: SET GLOBAL local_infile=1; inside mysql database then re-run LOAD DATA LOCAL INFILE query.

```
mysql> LOAD DATA LOCAL INFILE '/docker-entrypoint-initdb.d/init_data/glo
bal_youtube_stat.csv'
    -> INTO TABLE youtube
    -> FIELDS TERMINATED BY ','
    -> ENCLOSED BY '"'
    -> LINES TERMINATED BY '\n'
    -> IGNORE 1 ROWS;
Query OK, 995 rows affected, 1265 warnings (0.05 sec)
Records: 995  Deleted: 0  Skipped: 0  Warnings: 1265
```

6. Go to script directory and setup python virtual environment using python3 -m venv env

```
PS C:\Users\hallo\OneDrive\Documents\Kuliah\Stupen\project4\docker_etl\examples> cd etl
PS C:\Users\hallo\OneDrive\Documents\Kuliah\Stupen\project4\docker_etl\examples\etl> cd scripts
PS C:\Users\hallo\OneDrive\Documents\Kuliah\Stupen\project4\docker_etl\examples\etl\scripts> pytho
n -m venv env
```

7. Activate env using source env/bin/activate in Windows we can use env/Scripts/activate

```
PS C:\Users\hallo\OneDrive\Documents\Kuliah\Stupen\project4\docker_etl\examples\etl\scripts> .\env
\Scripts\activate
```

8. Install python requirements.txt using python3 -m pip install -r requirements.txt or pip install -r requirements.txt

```
(env) PS C:\Users\hallo\OneDrive\Documents\Kuliah\Stupen\project4\docker_etl\examples\etl\scripts>
 pip install -r requirements.txt
Collecting pandas (from -r requirements.txt (line 1))
  Downloading pandas-2.2.2-cp312-cp312-win_amd64.whl.metadata (19 kB)
Collecting mysql-connector-python (from -r requirements.txt (line 2))
  Downloading mysql_connector_python-8.3.0-cp312-cp312-win_amd64.whl.metadata (2.0 kB)
Collecting Flask-SQLAlchemy (from -r requirements.txt (line 3))
  Downloading flask_sqlalchemy-3.1.1-py3-none-any.whl.metadata (3.4 kB)
Collecting psycopg2-binary (from -r requirements.txt (line 4))
  Using cached psycopg2_binary-2.9.9-cp312-cp312-win_amd64.whl.metadata (4.6 kB)
Collecting numpy>=1.26.0 (from pandas->-r requirements.txt (line 1))
  Using cached numpy-1.26.4-cp312-cp312-win_amd64.whl.metadata (61 kB)
Collecting python-dateutil>=2.8.2 (from pandas->-r requirements.txt (line 1))
  Using cached python_dateutil-2.9.0.post0-py2.py3-none-any.whl.metadata (8.4 kB)
Collecting pytz>=2020.1 (from pandas->-r requirements.txt (line 1))
```

```
# psql -Upostgres
psql (11.16 (Debian 11.16-1.pgdg90+1))
Type "help" for help.

postgres=# \l
                                  List of databases
   Name    |  Owner   | Encoding |  Collate   |   Ctype    |   Access privileges
-----------+----------+----------+------------+------------+-----------------------
 postgres  | postgres | UTF8     | en_US.utf8 | en_US.utf8 |
 template0 | postgres | UTF8     | en_US.utf8 | en_US.utf8 | =c/postgres          +
           |          |          |            |            | postgres=CTc/postgres
 template1 | postgres | UTF8     | en_US.utf8 | en_US.utf8 | =c/postgres          +
           |          |          |            |            | postgres=CTc/postgres
(3 rows)

postgres=# create database data_warehouse; query
CREATE DATABASE
postgres-# \c data_warehouse
You are now connected to database "data_warehouse" as user "postgres".
```

9. Create database in postgresql:
    - docker exec -it <postgres_container_id> bash
    - login to postgres db using root user: psql -Upostgres
    - list databases run: \l
    - create database named data_warehouse using: create database data_warehouse; query
    - Choose database using \c data_warehouse
    - List tables inside database: \d

```
(env) PS C:\Users\hallo\OneDrive\Documents\Kuliah\Stupen\project4\docker_etl\examples\etl\scripts>
 python3 /Users/hallo/OneDrive/Documents/Kuliah/Stupen/project4/docker_etl/examples/etl/scripts/et
l.py
C:\Users\hallo\OneDrive\Documents\Kuliah\Stupen\project4\docker_etl\examples\etl\scripts\etl.py:14
: UserWarning: pandas only supports SQLAlchemy connectable (engine/connection) or database string
URI or sqlite3 DBAPI2 connection. Other DBAPI2 objects are not tested. Please consider using SQLAl
chemy.
  result_dataFrame = pd.read_sql(query, mydb)
      rank                    youtuber  subscribers  ... urban_population latitude longitude
0        1                     T-Series    245000000  ...        471031528  20.5937  78.96290
1        2                YouTube Movies    170000000  ...        270663028  37.0902 -95.71290
2        3                     MrBeast    166000000  ...        270663028  37.0902 -95.71290
3        4  Cocomelon - Nursery Rhymes    162000000  ...        270663028  37.0902 -95.71290
4        5                   SET India    159000000  ...        471031528  20.5937  78.96290
...    ...                         ...          ...  ...              ...      ...       ...
1985   991                Natan por A♦     12300000  ...        183241641 -14.2350 -51.92530
1986   992    Free Fire India Official     12300000  ...        471031528  20.5937  78.96290
1987   993                       Panda     12300000  ...         55908316  55.3781  -3.43597
1988   994                  RobTopGames     12300000  ...          9021165  60.1282  18.64350
1989   995                Make Joke Of      12300000  ...        471031528  20.5937  78.96290
```

9. Create database in postgresql:
   - docker exec -it <postgres_container_id> bash
   - login to postgres db using root user: psql -Upostgres
   - list databases run: \l
   - create database named data_warehouse using: create database data_warehouse;
   - query
   - Choose database using \c data_warehouse
   - List tables inside database: \d