Youtube Watch History Analyzer

Yehor Kotsiuba
ČVUT–FIT
kotsiyeh@fit.cvut.cz

December 25, 2023

1 About Project

Django project that allows you to upload your YouTube watch history in json format. After file being processed, you can visualize your watch history. Insights about your interests and how they have changed over the years.

2 Project Architecture

2.1 Technologies

Django * PostgreSQL, MongoDB * RabbitMQ, Celery * Docker, Docker Compose * Plotly * Bootstrap

2.2 Project Logic

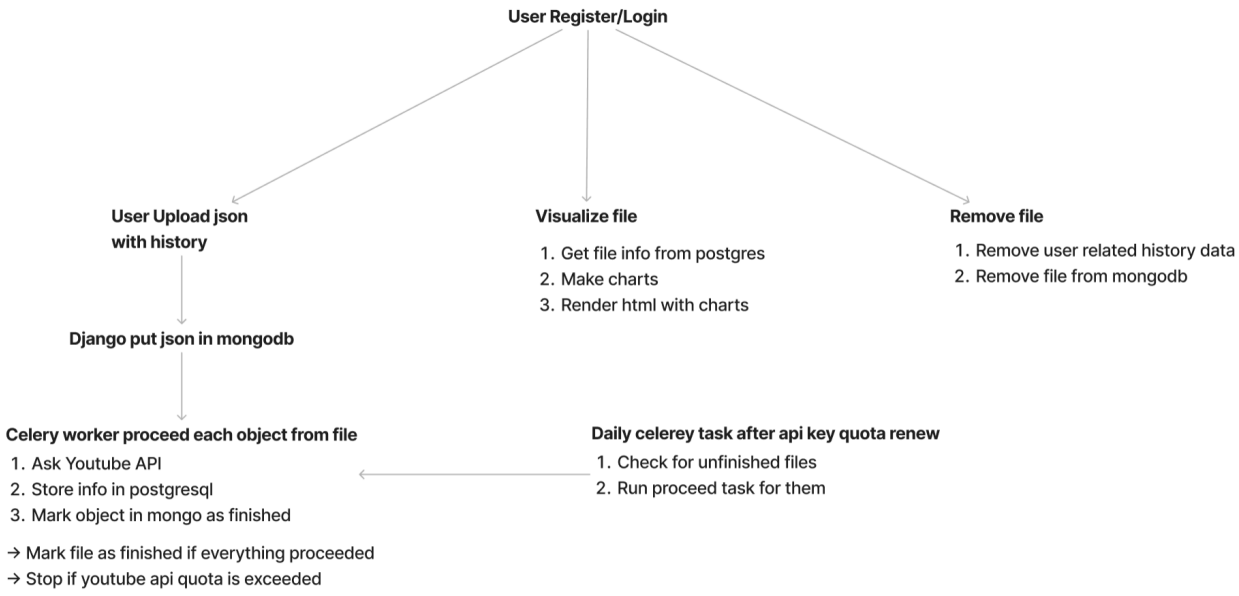

Image 1: Project Logic

Databases

PostgreSQL

PostgreSQL is used as the main database and is responsible for storing processed data.

The database is designed in a way that information about a video is stored independently of the user. This approach allows us to avoid gathering the same data each time.

Model	Description
WatchRecord model	'User watched some video at some time'
UserProfile (Or File) model	'Uploaded file in MongoDB'. Relates to website user.
Video model	'YouTube video'
Channel model	'YouTube channel'
Category model	'YouTube video category'. Predefined list of categories.

Table 1: Database Models

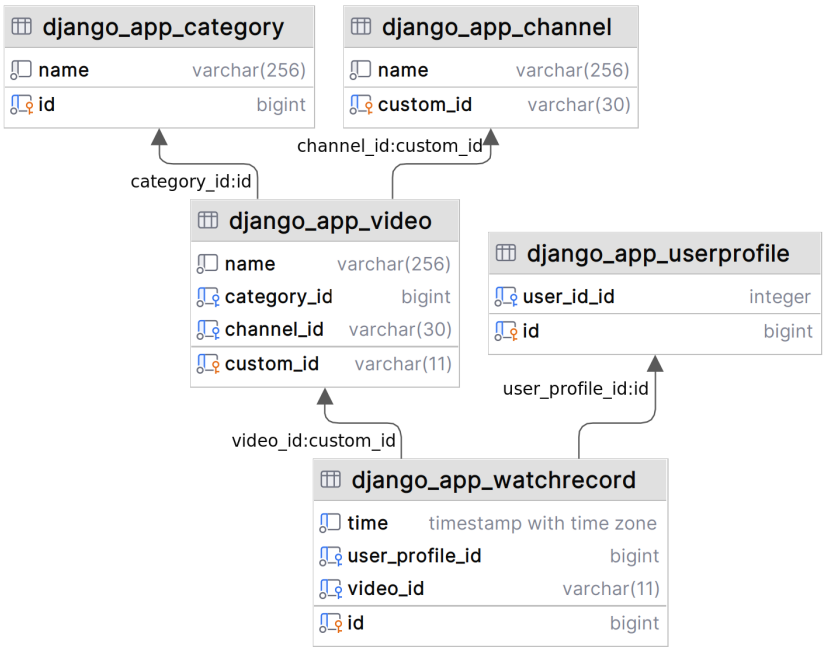

Image 2: Postgres DB diagram.
Image doesn't contain django generated default tables.

MongoDB

Is used as history object storage from Json.

- **File - 'Uploaded file'.**
  - **Status:** True, indicating that each video's status is True (Proceeded).
- **Video - 'Watch Record json object'.**
  - **Host:** Relates to File.
  - **Status:** True if the video is proceeded.

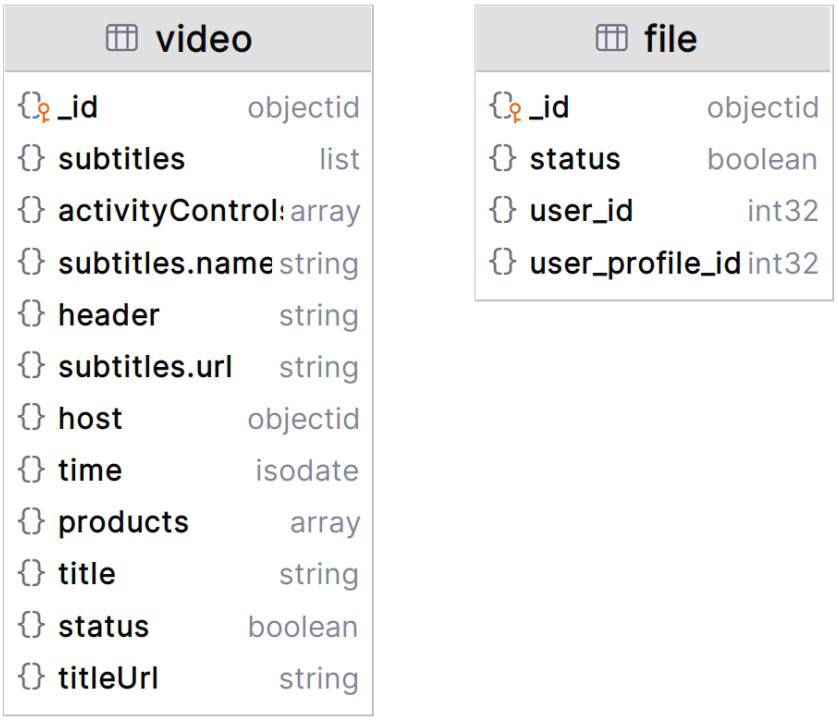

Image 3: MongoDB diagram

3 Docker

For this project was used docker compose to run web, postgres, mongodb, rabbitmq and celery containers. Dockerfiles and entrypoint scripts was used for better container management and automation. For better configurability was used environment variables. Helpful guide for docker compose [1].

4 Visualization

Different charts are used to visualize our processed data. Using plotly on backend we generate charts and send them to frontend. Useful article about watch history analysis [2], [3].

5 Project Future

What to improve:

- Better UI.
- Use another YouTube API to get more info for analysis.
- Add more charts.

Project can be scaled to platform that allows users to analyze their life on the internet.

6 Lessons Learned

This project had a big impact on my knowledge of Django and Docker.
App and databases architecture design practice.
Learned how to use workers (RabbitMQ and Celery).
Run server on virtual machine using OpenStack.

References

[1] The Dev Timeline. Add mongodb and postgresql in django using docker. online. [cit. 2023–12–24] https://dev.to/thedevtimeline/add-mongodb-and-postgresql-in-django-using-docker-55j6.

[2] Saúl Buentello. Explore your activity on youtube with r: How to analyze and visualize your personal data history. online. [cit. 2023–12–24] https://towardsdatascience.com/explore-your-activity-on-youtube-with-r-how-to-analyze-and-visualize-your-personal-data-history-b171aca632bc.

[3] neerajb22. My youtube history analysis. online. [cit. 2023–12–24] https://jovian.com/neerajb22/my-youtube-history-analysis.