

# YouTube Data Analysis App with SQL, MongoDB, and Streamlit

## Overview

This project aims to develop a comprehensive Streamlit application for harvesting, warehousing, and analyzing YouTube data using the YouTube Data API, MongoDB, and SQL databases. The application allows users to input YouTube channel IDs, retrieve and store data in MongoDB, and perform advanced analysis through SQL queries within the Streamlit interface.

## Problem Statement

The primary objectives are to:

1. Create a Streamlit app enabling users to analyze data from multiple YouTube channels.
2. Utilize the YouTube Data API to retrieve channel and video information.
3. Store the harvested data in a MongoDB data lake for its flexibility with unstructured data.
4. Facilitate the migration of selected channel data from MongoDB to a SQL database for in-depth analysis.
5. Enable users to search, retrieve, and analyze data from the SQL database, incorporating features like joining tables for a holistic view.

## Technology Stack

1. Python (for Streamlit application development)
2. MySQL (for SQL data warehousing)
3. MongoDB (for storing YouTube data in a data lake)
4. Google API Client Library (for accessing YouTube Data API)

## Approach

- 1. Streamlit Application Setup:-** Develop a Streamlit application with an intuitive interface allowing users to input YouTube channel IDs, view details, and select channels for further analysis.
- 2. YouTube Data Retrieval:** - Utilize the YouTube Data API V3 with the Google API client library to retrieve channel and video data based on user input.
- 3. MongoDB Data Storage:** - Design a MongoDB data lake to store harvested data efficiently. Implement a method to store data in three different collections: channels, videos, and comments.
- 4. SQL Data Warehousing:** - Transfer selected data from MongoDB to a SQL database (e.g., MySQL or PostgreSQL) to facilitate more complex analysis.
- 5. SQL Data Analysis:** - Utilize SQL queries to join tables and retrieve specific channel data. Establish proper foreign key relationships for efficient querying.
- 6. Streamlit Data Visualization:-** Leverage Streamlit's capabilities to visualize data through charts and graphs, providing users with a user-friendly environment for data analysis.

By integrating these technologies and following this structured approach, the application aims to provide users with a seamless experience for exploring and analyzing YouTube channel data.