**Guvi\_CAPSTON\_Project\_YOUTUBE-DATA-HARVESTING-AND-WAREHOUSING**

*YouTube Data Harvesting and Warehousing* is a project that aims to provide users with the ability to access and analyze data from numerous YouTube channels. SQL, MongoDB, and Streamlit are used in the project to develop a user-friendly application that allows users to retrieve, save, and query YouTube channel and video data.

**TOOLS AND LIBRARIES USED:**

This project requires the following components:

1. **STREAMLIT:** The Streamlit library was used to create a user-friendly UI that enables users to interact with the program and carry out data retrieval and analysis operations.
2. **PYTHON:** Python is a powerful programming language renowned for being easy to learn and understand. It is the primary language employed in this project for the development of the complete application, including data retrieval, processing, analysis, and visualization.
3. **GOOGLE API CLIENT:** The googleapiclient library in Python facilitates communication with different Google APIs. Its primary purpose in this project is to interact with YouTube's Data API v3, allowing the retrieval of essential information like channel details, video specifics, and comments. By utilizing googleapiclient, developers can easily access and manipulate YouTube's extensive data resources through code.
4. **MONGODB:** MongoDB is built on a scale-out architecture that has become popular with developers of all kinds for developing scalable applications with evolving data schemas. As a document database, MongoDB makes it easy for developers to store structured or unstructured data. It uses a JSON-like format to store documents.
5. **MYSQL:** MySQL is an open-source relational database management system (RDBMS) known for its reliability and extensive features. It provides a platform for storing and managing structured data, offering support for various data types and advanced SQL capabilities.

**YOUTUBE DATA SCRAPPING AND ITS ETHICAL PERSPECTIVE:**

When engaging in the scraping of YouTube content, it is crucial to approach it ethically and responsibly. Respecting YouTube's terms and conditions, obtaining appropriate authorization, and adhering to data protection regulations are fundamental considerations. The collected data must be handled responsibly, ensuring privacy, confidentiality, and preventing any form of misuse or misrepresentation. Furthermore, it is important to take into account the potential impact on the platform and its community, striving for a fair and sustainable scraping process. By following these ethical guidelines, we can uphold integrity while extracting valuable insights from YouTube data.

**REQUIRED LIBRARIES:**

1. googleapiclient.discovery
2. streamlit
3. pymysql
4. pymongo
5. pandas

**FEATURES:**

The YouTube Data Harvesting and Warehousing application includes the following functions:

* Retrieval of channel and video data from YouTube using the YouTube API.
* Storage of data in a MongoDB database as a data lake.
* Migration of data from the data lake to a MySQL database for efficient querying and analysis.
* Search and retrieval of data from the MySQL database using different search options.