**YOUTUBE DATA HARVESTING AND WAREHOUSING**

**Synopsis**

1. Introduction
2. Features
3. Project Overview
4. Key Components
5. Technologies Used
6. Installation
7. Usage
8. Conclusion

**Introduction**

YouTube Data Harvesting is a Python-based project designed to extract and store comprehensive information from YouTube channels. The project focuses on retrieving data such as channel details, playlists, videos, and comments. This tool is useful for content analysis, audience engagement, and understanding the dynamics of a YouTube channel.

**Features**

**1. Channel Data Extraction:**

* Retrieve key details about YouTube channels.

**2. Playlist and Video Information:**

* Collect data on playlists, videos, and associated metadata.

**3. Comments Harvesting:**

* Capture comments from videos for sentiment analysis and engagement insights.

**Project Overview**

The YouTube Data Harvesting project is designed to empower users with the ability to extract and analyze comprehensive information from YouTube channels. This project primarily focuses on fetching data related to channels, playlists, videos, and comments, providing valuable insights into content, audience engagement, and overall channel dynamics.

**Key Components**

**1. Channel Data Extraction**

* **Objective**: Retrieve essential details about YouTube channels.
* **Functionality:** Extract and store key information such as channel name, description, subscriber count, and more.

**2. Playlist Information**

* **Objective:** Collect data on playlists and associated metadata.
* **Functionality**: Fetch details on playlists, individual videos, view counts, likes, dislikes, and publication dates.

**3. Video Information**

* **Objective:** Collect data on videos and associated metadata.
* **Functionality:** Extract and store key information such as Video Id, Channel Id, Channel Name, Video Title, and etc.

**4. Comments Harvesting**

* **Objective:** Capture comments from videos for sentiment analysis and engagement insights.
* **Functionality:** Extract comments along with user details, timestamps, and replies.

**Technologies Used**

The YouTube Data Harvesting project leverages a set of cutting-edge technologies to efficiently extract, process, and analyze data from YouTube. Here are the key technologies used in the project:

**1. Python:**

* **Description:** The primary programming language for scripting and automation.
* **Purpose:** Used for writing the main data extraction script, handling API requests, and managing data processing.

**2. YouTube Data API:**

* **Description:** API provided by Google for interacting with YouTube data.
* **Purpose:** Enables fetching information about channels, playlists, videos, and comments.

**3. MySQL:**

* **Description:** Open-source relational database management system.
* **Purpose:** Stores structured data related to channels, playlists, and videos.

**4. MongoDB:**

* **Description**: NoSQL document database.
* **Purpose**: Stores unstructured or semi-structured data, particularly details about channel descriptions.

**5. Pandas:**

* **Description:** Data manipulation library for Python.
* **Purpose**: Facilitates data cleaning, transformation, and analysis.

**6. Streamlit:**

* **Description:** Python library for creating web applications.
* **Purpose**: Used for building an interactive dashboard to explore and visualize the harvested data.

**7. Git and GitHub:**

* **Description:** Version control system and hosting platform.
* **Purpose:** Manages collaborative development, version tracking, and project sharing.

Installation

To run the YouTube Data Harvesting and Warehousing project, follow these steps:

**Install Pakages:**

1. **pip install streamlit**
2. **pip install pandas**
3. **pip install SQLAlchemy**
4. **pip install mysql-connector-python**
5. **pip install google-api-python-client**
6. **pip install pymongo**

Usage

1**. Channel Data Harvesting:**

* + Use the project to extract detailed information about a specific YouTube channel.
  + Retrieve data such as channel name, description, subscriber count, view count, and more.
  + Get insights into the channel's statistics and metadata.

**2. Video Data Harvesting:**

* + Extract comprehensive details about specific YouTube videos.
  + Retrieve data including video title, description, upload date, view count, and likes/dislikes.
  + Obtain information about the video's comments and associated metadata.

**3. Customization and Options:**

* Tailor your data extraction by exploring various command-line options.
* Use additional parameters to specify the type and depth of information you want to collect.
* Refer to the help` command for a list of available options.

Conclusion

In summary, the YouTube Data Harvesting project is a versatile tool for extracting valuable insights from YouTube channels and videos. Its user-friendly command-line interface allows easy customization, and the modular structure caters to both beginners and experienced developers. The project outputs data in a structured JSON format, facilitating seamless integration into various applications and workflows.

As an open-source initiative, collaboration is encouraged. The accompanying documentation provides comprehensive guidance, making the tool accessible to a wide audience. Whether you're a data enthusiast, researcher, or developer, this project empowers you to efficiently explore and analyze YouTube data for valuable insights. Start harvesting today!