Final Project Group Proposal – Cloud Computing
Dr. Najafirad, Spring 2022

Members:

Shejan Shuza, Ariel Guerrero, Viswa Bhargavi, Hunter Long, Aaron Gingrich

Project Overview:

Construct a cloud-enabled web-based application that provides a YouTube video archiving and distribution service (like that of Internet Archive's Wayback Machine but for videos) using Python utilities like youtube-dl or yt-dlp (a fork of youtube-dl with better performance) and cloud services such as Cloud Functions, Pub/Sub, Storage, and BigQuery for analytics.

The service can allow a user to submit a YouTube video link to the service, and the service will create a clone on Cloud Storage (including video metadata like descriptions, subtitles, thumbnails, etc.), the user will then receive a "master link" that allows modification of the permissions to their sharable link to the video on the service (count of usable clicks, duration, login requirements, etc.)

Goals:

1. Construct APIs for pulling various parts of a YouTube video, including the video itself, descriptions, subtitles, thumbnails, etc. based on yt-dlp
2. Understand and produce a Pub/Sub workflow for accepting user YouTube links for concurrent downloading.
3. Format a Cloud Storage Bucket to support avoiding duplicates and storing potentially large data.
4. Integrate the service workflow to support Google's BigQuery to gather statistics and analytics of the service itself.

Google Cloud Services

1. Cloud Functions (API environment)
2. Pub/Sub (function organization and concurrency)
3. Cloud Storage (data storage)
4. BigQuery (analytics)

Software Tools

1. youtube-dl or yt-dlp (a fork of youtube-dl with better performance)
2. ffmpeg (used as part of youtube-dl, used for combining and converting media files, could be used for creating separate versions of a video)
3. Python 3.9 (used as the basis of youtube-dl and used as the language of the APIs)