**Movie Recommender System**

(An All-Google Cloud Solution)

Recommender System is a system that seeks to predict or filter preferences according to the user’s choices. Recommender systems are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general. I am going to build a Movie Recommender System.

Data Set: (6.17GB) <https://www.kaggle.com/clementmsika/mubi-sqlite-database-for-movie-lovers>

There are multiple files in the database. I have used only mubi\_movie\_data.csv and mubi\_ratings\_data.csv files.

* I loaded these files in BigQuery using the Google Cloud Console.
* I used BigQuery for fast query execution and retrieval.
* I will use a Cloud Schedular to run a python code weekly that will extract the data for the week from the BigQuery and store it as a separate file in the Cloud Storage(Bucket) named data\_<date>.txt .
* I used a python program to do this for the whole data since I have data from 2008 to 2020. Thus, I used a for loop to loop through all the data weekly.
* Pub-Sub architecture is used. When the python program loads the file in the storage, it sends a message to the Publish topic that a file is loaded to the storage.
* As a part of Pub-Sub architecture, whenever a message is published to the topic, it calls a Cloud Function. The Cloud function reads the file and removes irrelevant columns from the file, converts it into a dictionary and loads it in a Json file in another Cloud Storage. This Json File will then be used for giving recommendations.
* The above part was the data processing part and the pipeline.
* There will be a Flask application that will be running in the Kubernetes Cluster since there is a possibility of huge number of users using the recommendation system.
* There will be a load balancer the balance load around the Kube clusters.
* The flask app will load the data json file as dictionary and process it to give recommendations to Users.
* The recommendations System will be of 2 types:
  + User Based recommendation system
  + Top rated recommendation system
* The pipeline will run every Monday at night (4 pm) local time.
* The Flask API will be up and running and serve several users at a time. There will also be auto-scaling feature since the number of users requesting would be less during the day time and more during the evening time, and even more in weekends.