

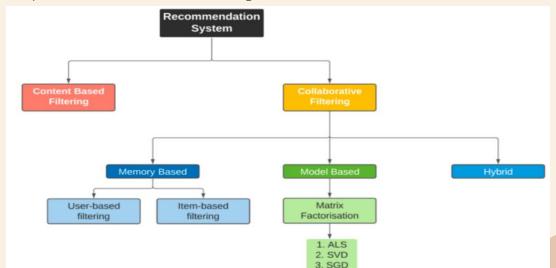
Ashritha Lopelli







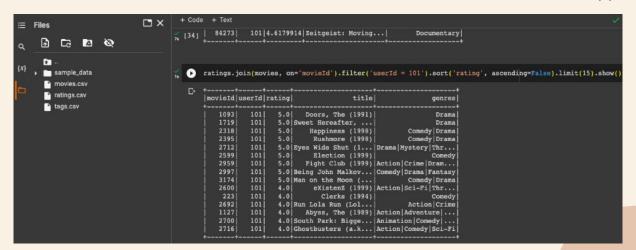
- The MovieLens dataset is widely used as a benchmark for evaluating recommendation systems.
- The program utilizes PySpark, a Python library designed for distributed data processing, to leverage collaborative filtering techniques.
- This Python program showcases the implementation of a recommendation engine that utilizes the ALS (Alternating Least Squares) model on the MovieLens dataset.
- By applying collaborative filtering, the program generates personalized movie recommendations for users based on their preferences and historical ratings.





Steps on Colab:

- Download the recommendation_movielens1.py file, movies.csv, rating.csv, tags.csv files.
- Open and login to your Google colab.
- Upload the recommendation_movielens1 file to the colab.
- Also add the movies.csv, rating.csv, tags.csv files to the folder of the colab.
- Run all the command on the recommendation_movielens1 file and then Download this file as .py file.

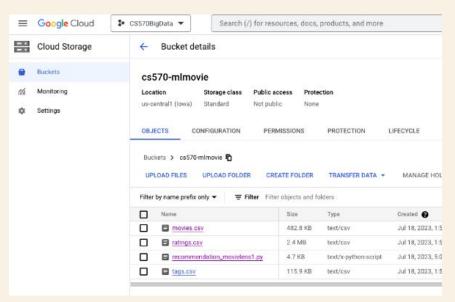






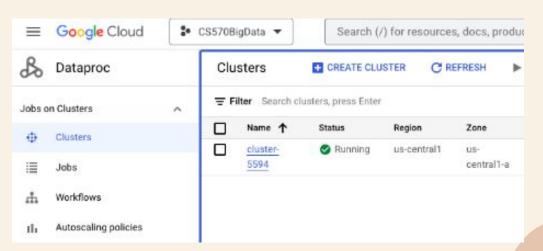
Steps on GCP:

- 1. Create a bucket on the GCP.
 - Upload the movies.csv, rating.csv, tags.csv files to the above created bucket.
 - Also upload the Downloaded .py file (recommendation_movielens1.py) to the bucket.



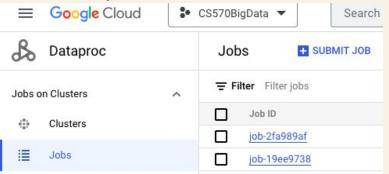


- 2. Now create a dataproc cluster.
 - In the Navigation menu, click on "Dataproc".
 - Click on the "Create cluster" button to create a new cluster.
 - Give the necessary details such as Cluster name, Region, Zone and Cluster type.
 - Create cluster.
 - The cluster is now created and running.





- 3. Now submit the job through the dataproc Jobs.
 - In the Navigation menu, click on "Dataproc".
 - Click on the "Jobs" and then on "submit job" tab to create a new Job.



- Fill out all the required details like Job id, Region, Cluster and Job type.
- Also provide the main file path (here it is recommendation_movielens1.py from the bucket above copy its path and paste it).



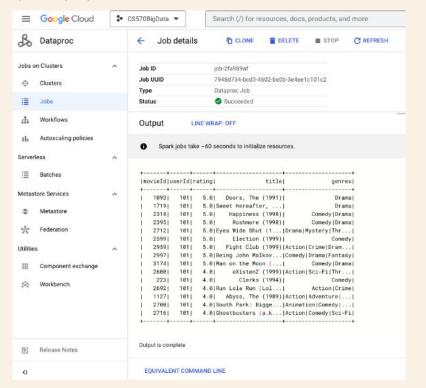
=	Google Cloud	8 0	CS570BigData ▼	Search (/) for resources, docs, pro	
B	Dataproc		← Submit a job		
Jobs on Clusters		Job ID * job-92fbb93c			
•	Clusters		Region *us-central1 Specifies the Cloud Dataproc regional service, which determines what cavailable.		
≣	Jobs				
ф	Workflows				
th	Autoscaling policies		Cluster * cluster-5594		
Serverless		^	Job type *		
≣	Batches		PySpark		
Metastore Services		gs://cs570-mlmovie/recommendation_movielens.py			
Φ	Metastore		Can be a GCS file with the gs:// prefix, an HDFS file on the cluster with the a local file on the cluster with the file:// prefix Additional python files		
్గిర్తీ	Federation				
Utilities		^	<u></u>		

• Now click on "submit" button.





• And we can see the output displayed.







Thank You!!!