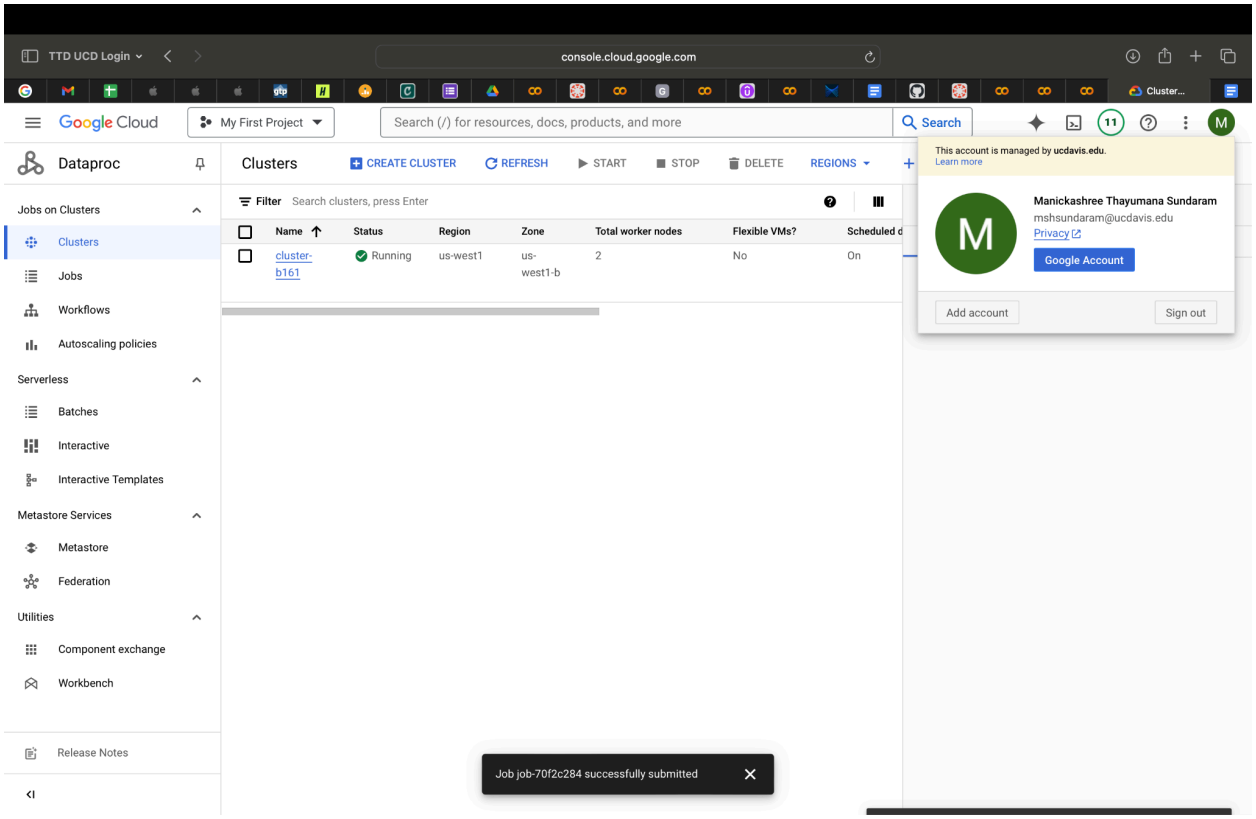
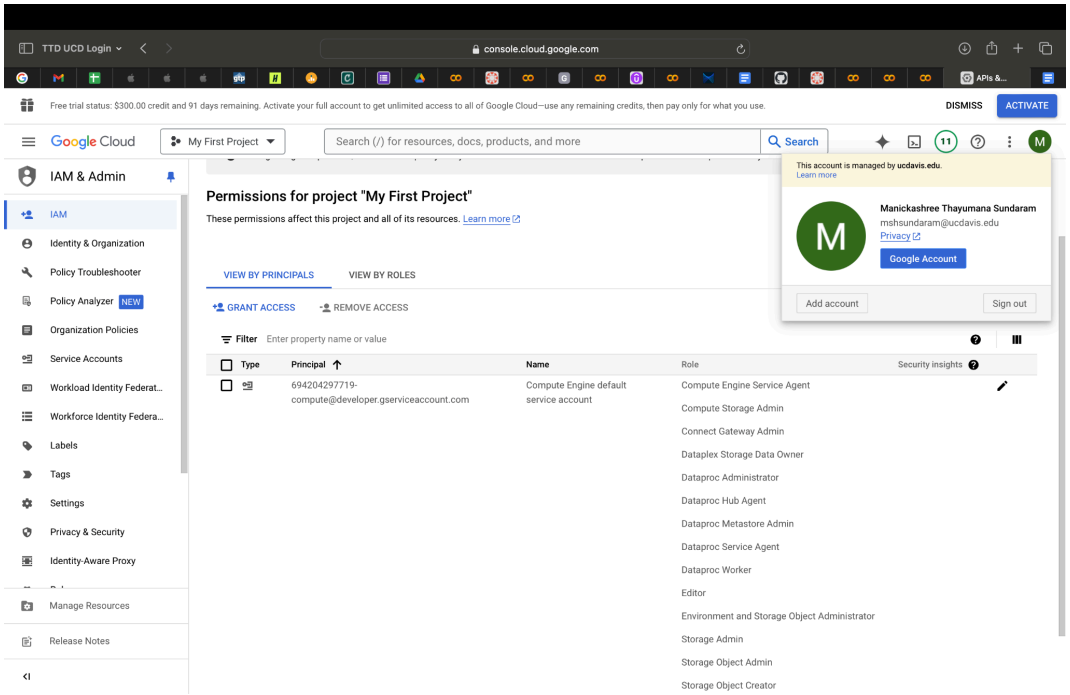
# Homework 2: Getting Started with Spark and Climate Change

## Cluster in GCP Steps

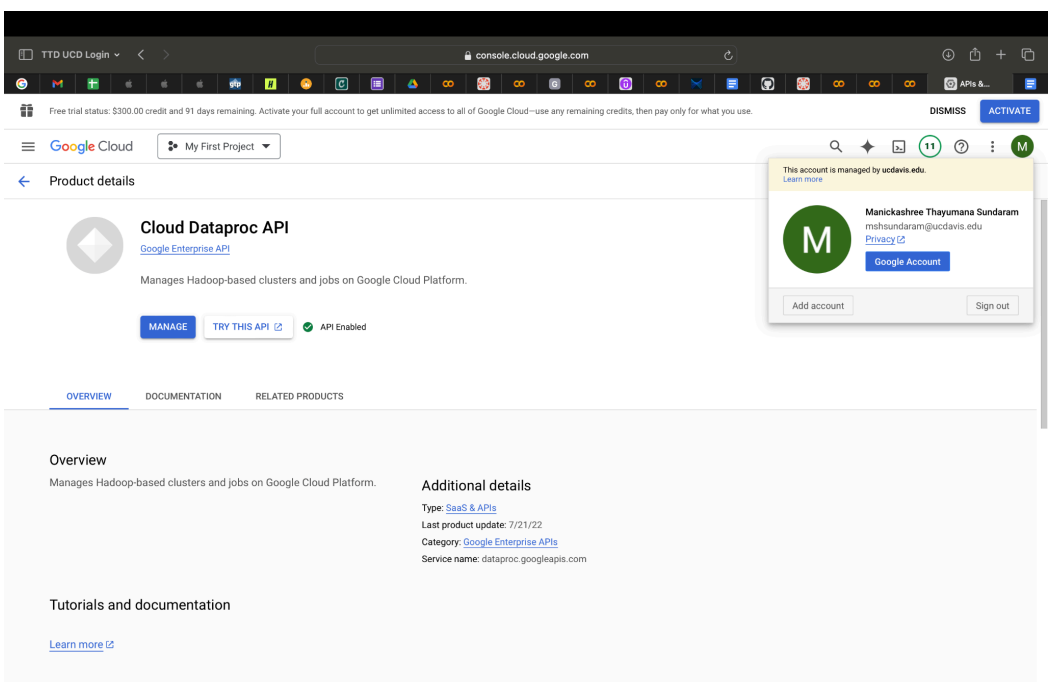


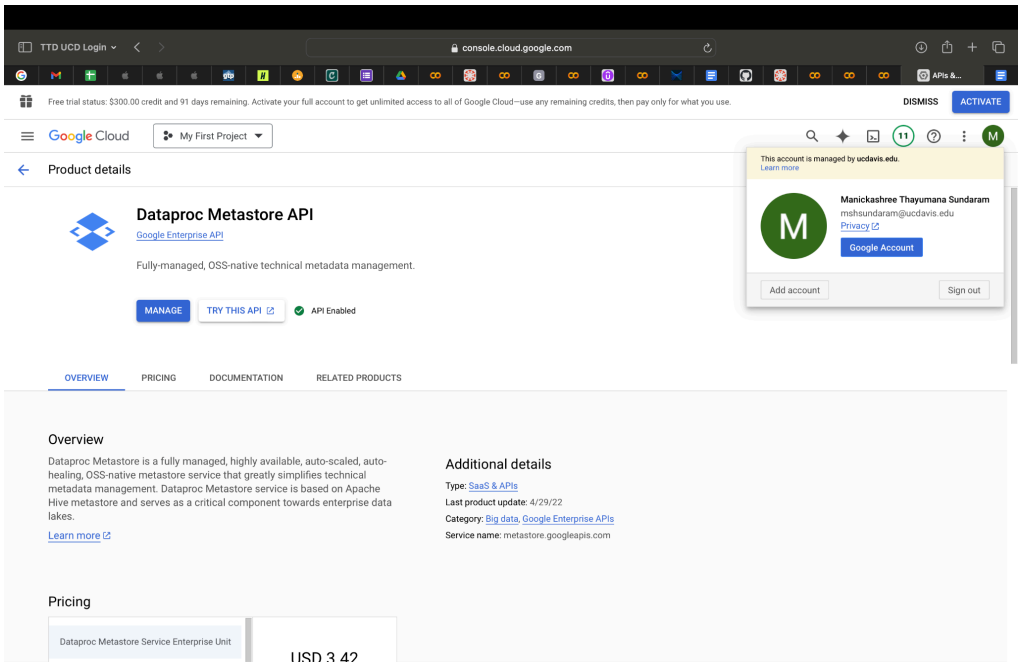

## Steps

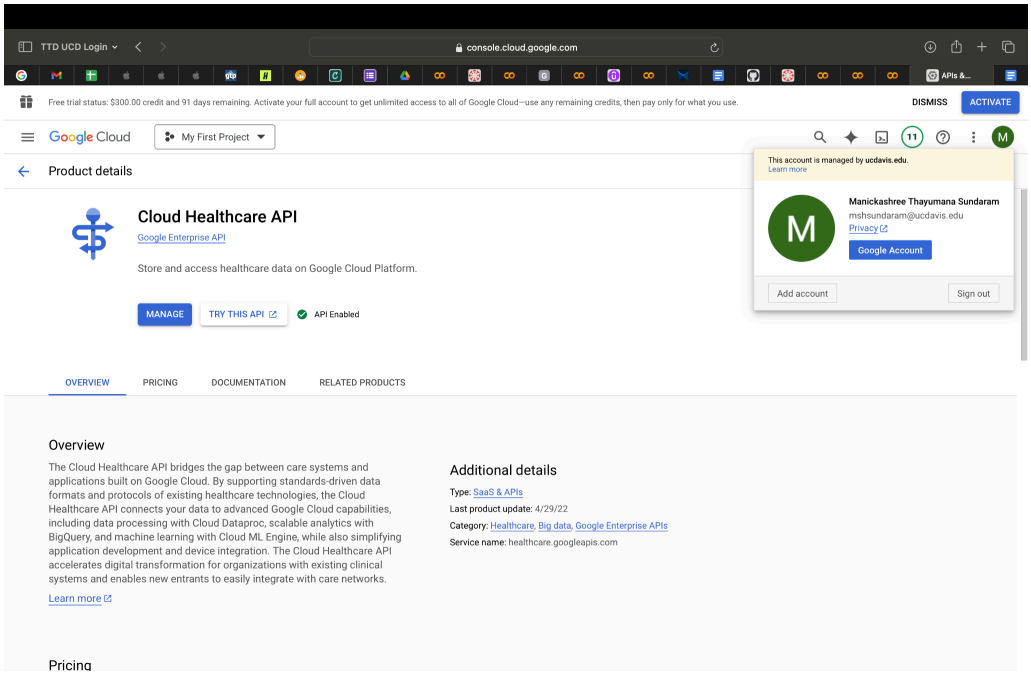
1. Setting up your Google Cloud account.
2. Define necessary IAM roles to users




3. Enabling Google Cloud Dataproc API(Cloud Dataproc API, Dataproc Metastore API, Cloud Resource Manager API)






4. Create a Dataproc Cluster by providing the necessary information for your cluster such as name, region and zone. Proceed to deploy the cluster created.
5. Create a bucket and upload all the required files($CO_2$ emissions per capita per country.csv and GlobalLandTemperatures_GlobalLandTemperaturesByCountry.csv)

6. Upload the PySpark script to Google Cloud Storage bucket and copy the URI of it.
7. Now in the cluster details, select ‘Submit job’ and choose PySpark job type. Then click on submit to run the job.
8. Monitor the job, in case of failure, rectify the error and re-run the job by uploading the latest file.
9. Finally after running the job, clean up the resource.

## Reflect on your everyday life activities. What can you personally do to make a positive change for the environment? Write a short paragraph with your thoughts.

There are many practical ways in which I contribute positively to the environment in my daily life. The first and most important step is to reduce waste. To achieve this, I start by minimizing the use of single-use plastics and opting for reusable alternatives such as water bottles, shopping bags, and food containers. This can significantly reduce the amount of waste that ends up in landfills.

Recycling and composting can also play an essential role in decreasing the amount of waste that ends up in landfills. By separating my trash into recyclables, compostables, and regular waste, I can ensure that items are disposed of in the most environmentally friendly way possible.

Conserving energy is another vital area that I can focus on. By using energy-efficient appliances, I can save money on my electricity bills and reduce my carbon footprint. Turning off lights and electronics when not in use, maximizing natural light, and choosing public transportation, carpooling, or biking instead of driving can also help reduce greenhouse gas emissions.

Lastly, I can enhance energy conservation in my home by switching to energy-efficient appliances and light bulbs and being mindful of turning off lights and electronics when not in use. By integrating these practices into my daily routine, I can help foster a more sustainable and environmentally friendly world for myself and future generations.

## Ideate 3 ideas about new sources of information that can underpin new companies, and type a short paragraph describing your ideas (bullet points are accepted)

### 1. Language Learning Progression Analytics Platform:

The concept is to create a platform that uses detailed user data from language learning apps to provide analytics on user progression, retention rates, and learning styles. The data source would be anonymized user data, similar to the Duolingo dataset, focusing on metrics such as recall accuracy, session activity, and learning preferences. The platform could serve educational institutions and language learning app developers by offering insights into effective curriculum development and personalized learning pathways.

### 2. Adaptive Learning Content Generator:

**Idea:** To develop a service that creates personalized learning material by analyzing user performance data from different educational applications.

**Data Source:** The service will use datasets containing user performance metrics, learning speeds, and content interaction details to tailor educational content dynamically.

**Service Offered:** This service will be beneficial for online educational platforms seeking to optimize their content for better engagement and learning outcomes. It will adapt in real-time to cater to the user's needs.

### 3. Multilingual User Experience Optimization Tool:

**Concept:** Create a tool that assists app developers in comprehending how diverse languages and cultural contexts influence user interaction and app usability.

**Data Source:** The data for the tool would be collected from apps like Duolingo, by analyzing user interface language preferences, performance metrics across various linguistic backgrounds, and user feedback.

**Service Offered:** The tool would be ideal for developers aiming to localize their apps effectively across multiple regions. It would ensure optimal user experience by analyzing and predicting language-based usage patterns.