# Introduction to Big Data
# Graded Assignment 2

Name - Avijeet Palit
Roll - 21f1005675
Date - 13/10/2024

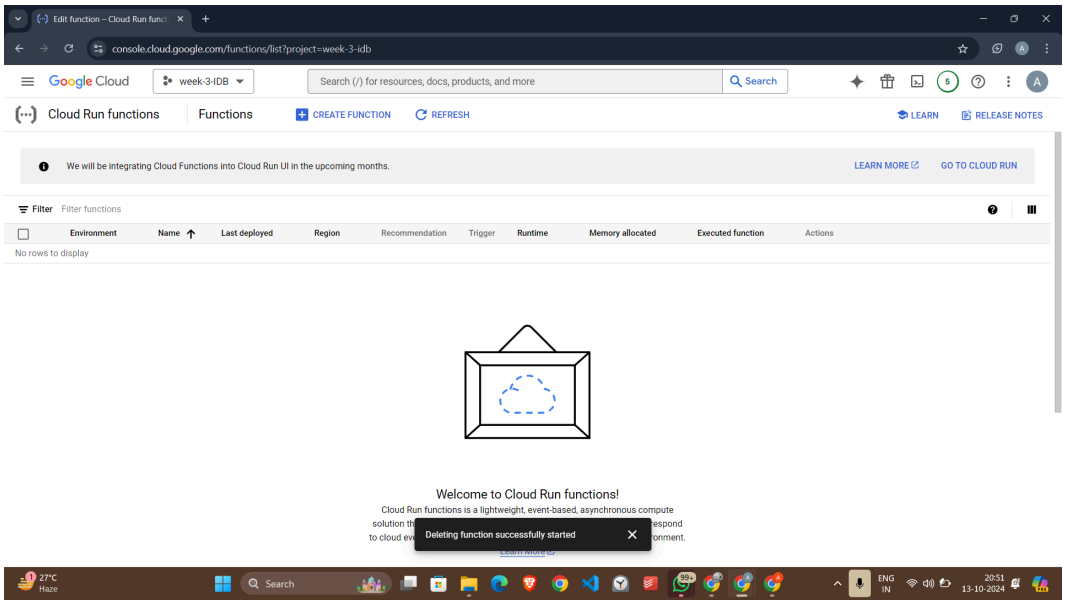
### Week 3: Automated Line Counting in Google Cloud Storage Files

### Introduction

The objective of this project was to implement an automated system to count lines in text files stored in a Google Cloud Storage (GCS) bucket. Utilizing Google Cloud Functions, the system adopts a serverless approach, ensuring efficient and scalable processing of file uploads. This report details the processes involved in project setup, API activation, Python code development, deployment, and testing.

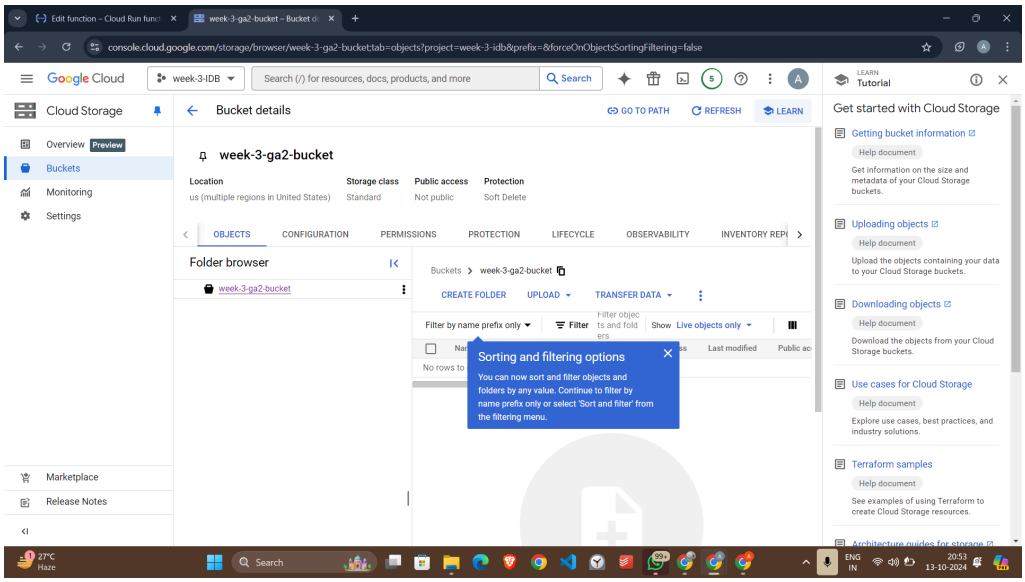
### Project Setup

1. **Creating a Google Cloud Project**: A new project, titled "Week-3," was established in the Google Cloud Console, providing a dedicated environment for all related activities and code implementations.



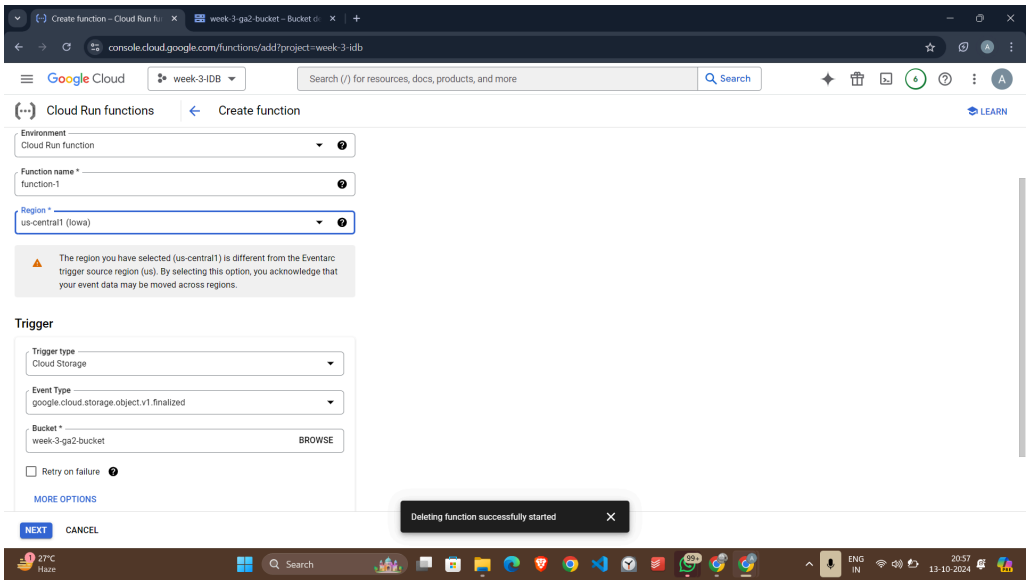

2. **Activating Essential APIs**: Key APIs were enabled to allow interactions with Google Cloud Storage and the deployment of cloud functions:

- Google Cloud Functions API: Facilitates the creation, deployment, and management of cloud functions.
- Google Cloud Storage API: Enables file storage and retrieval operations.

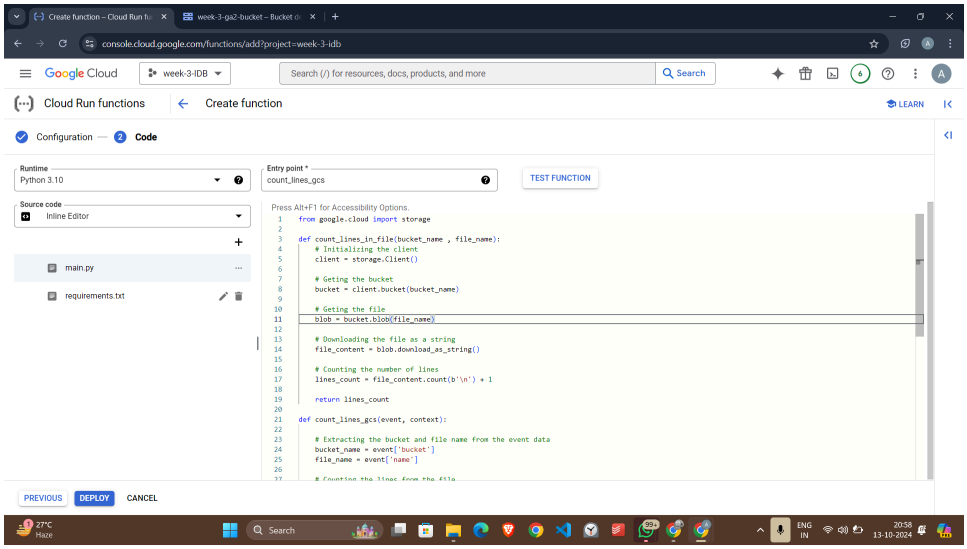


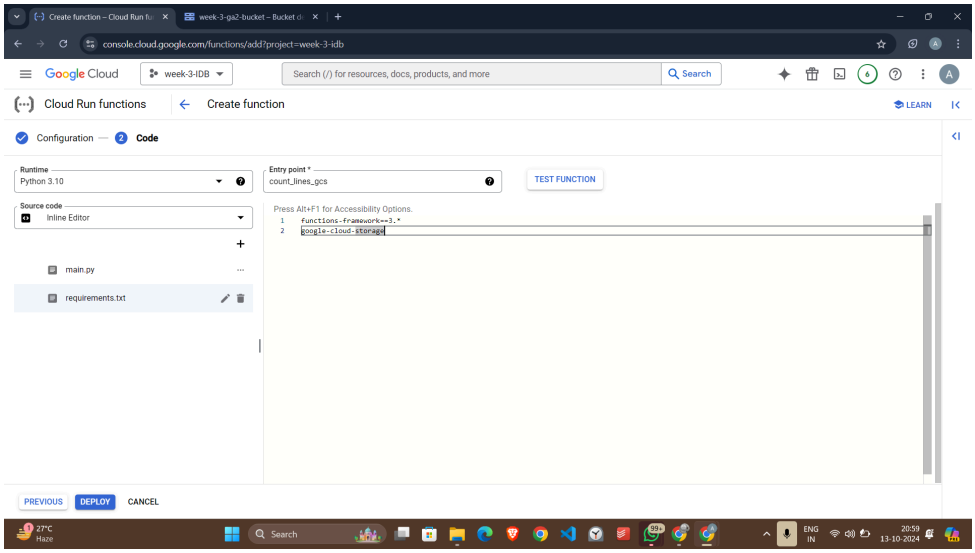

## Development and Deployment

3. **Python Code Development**: A Python script was written to count lines in any file uploaded to the GCS bucket. The script uses the `google-cloud-storage` library to access and manage files.



4. **Cloud Function Deployment**: The cloud function was deployed via the Google Cloud Console with specific configurations:
    - **Trigger**: Set to activate when a file is uploaded to the designated GCS bucket.
    - **Dependencies**: Dependencies, including the `google-cloud-storage` library, were listed in the `requirements.txt` file to ensure proper installation during deployment.

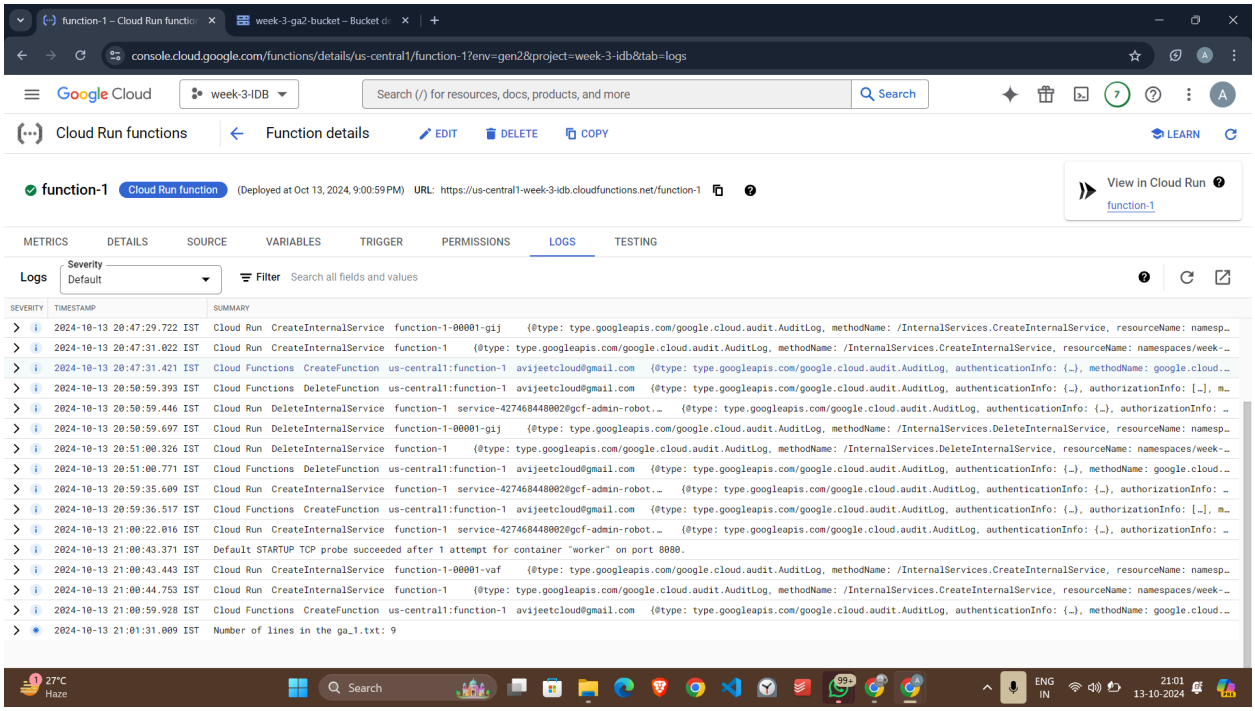



## Testing the Functionality

5. **Execution Test**: The cloud function's performance was tested by uploading a file named `ga_1.txt` to the " week-3-ga2-bucket" GCS bucket. The function

triggered as expected, and the number of lines counted was logged, confirming the functionality of the system.




## Conclusion

This initiative successfully demonstrated the design and implementation of a Google Cloud Function to automatically count lines in text files uploaded to a GCS bucket. The integration of Google Cloud Storage and Cloud Functions, along with Python scripting, highlighted the system's capability for efficient and scalable serverless operations in cloud settings.