CSP 554 BIG DATA TECHNOLOGIES MOHAMMED WASIM R D(A20497053)

ASSIGNMENT 1

- 1. Answer each of the following questions about the article in just one to three sentences each:
 - O What was the problem with the Google flu detection algorithm?

The goal behind Google Flu Trends was to evaluate the vast number of Google search queries obtained to see if there was a presence of flu-like illness in a population by monitoring millions of user's health tracking activities online.

O What is big data hubris?

Big data hubris is a belief that big data should be used as a replacement for rather than an addition to traditional data collecting and analysis

 What approach could have been used to improve the Google flu detection algorithm?

Larger value can be obtained by combining GFT with other real time health data

- a) Combining GFT and lagged CDC data
- b) Dynamically recalibrating GFT

O What is "algorithm dynamics?"

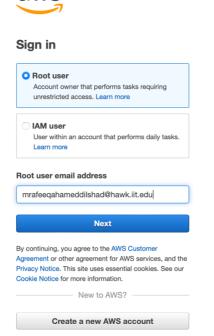
Algorithm dynamics are the changes made by engineers to improve the commercial service and by consumers in using that service. Several changes in Google's search algorithm and user behavior likely affected GFT's tracking.

What aspect of algorithm dynamics impacted the Google flu detection algorithm?

The most common explanation for GFT's error is a media-stoked panic last flu season (1, 15). Although this may have been a factor, it cannot explain why GFT has been missing high by wide margins for more than 2 years. Moreover, the data sources for GFT were not appropriate to produce valid and reliable results compatible for scientific research

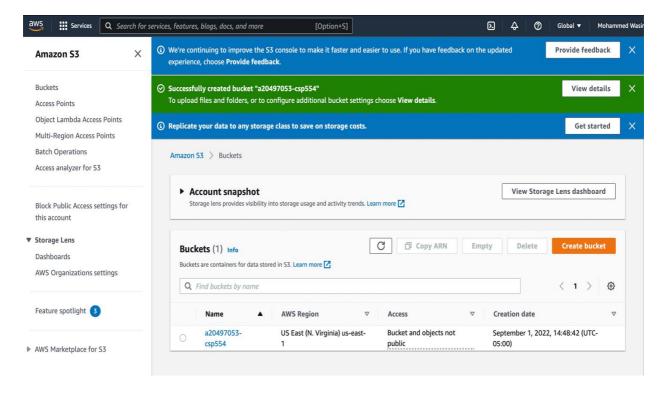
2. Setting up AWS account and uploading files to S3

a) Sign up to AWS Account as root user

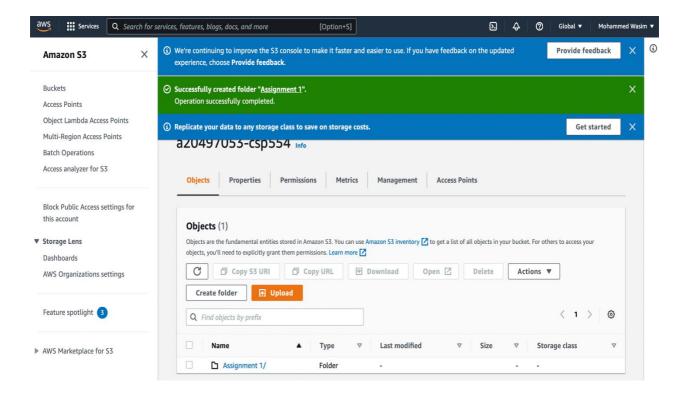




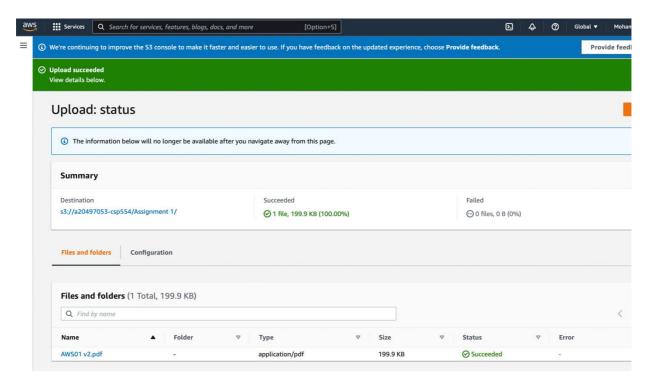
b) Creating bucket with id and course code



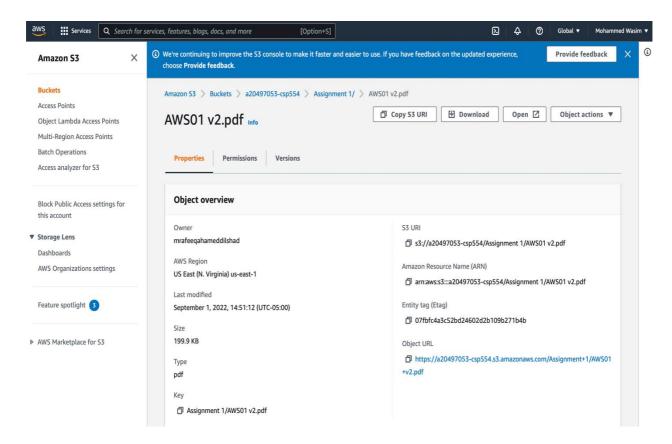
c) Creating folder in the bucket



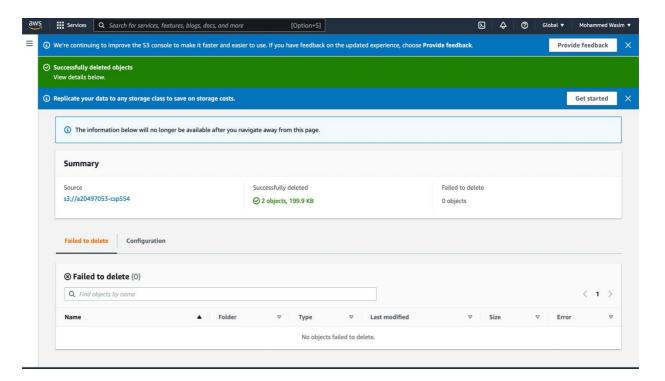
d) Uploading files in folder which are called objects



e) Overview of uploaded file



f) Deleting folder and objects



g) Deleting Bucket

