Akash Tanwani **A20448831**

CSP 554—Big Data Technologies

Assignment #1 (Modules 01a & 01b, 12 points)

Que3 (2 points) Submit very brief answers (or bullet points) to the following questions:

- What location or time zone are you in when you attend the course?
  Ans: Chicago, IL Central Daylight Time(CDT) / GMT -5:00 hours.

- Describe any prior experience you might have with use of public cloud, data mining, machine learning, statistics, data science and big data.
  Ans: I have completed courses like Data Mining(CS422), Data preparation and Analysis(CSP-571), Machine Learning(CS-584) and Advance Data Mining(CS522) during my Fall 2019 and Spring 2020 semester. Also, during my summer 2020 Internship, I worked on Azure Cloud where my major tasks include creating web applications and deploying it on Azure cloud, creating database on Azure SQL Database and creating Chatbot using Azure bot and cognitive services.

- Share any big data interests and personal learning goals for the course.
  Ans: As I already worked on different Microsoft Azure cloud services, now I would like to work on Microsoft HDInsight by getting some experience on frameworks like Hadoop, Spark, Hive and Kafka.

- Indicate if there are additional topics in the scope of the course of special interest to you.
  Ans: I would like to learn about technologies involving Social Network Analysis by using Machine Learning techniques.

- Do you have any anticipated personal issues such as expected absences or other necessary accommodations with course impact? (Of course, these will be held in strictest confidence.)
  Ans: No

Que5 (5 points) Answer each of the following questions about the article in just one to three sentences each:

- What was the problem with the Google flu detection algorithm?
  Ans: The problem with the Google flu detection algorithm was that it overestimated the number of flu cases. which is, it predicted more than double the proportion of doctor visited for influenza-like illness (ILI) than the CDC reports.

- What is big data hubris?
  Ans: Big data hubris is the assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis.

- What approach could have been used to improve the Google flu detection algorithm?
  Ans: By combining GFT and lagged CDC data, as well as dynamically recalibrating GFT, we can substantially improve the Google flu detection algorithm.

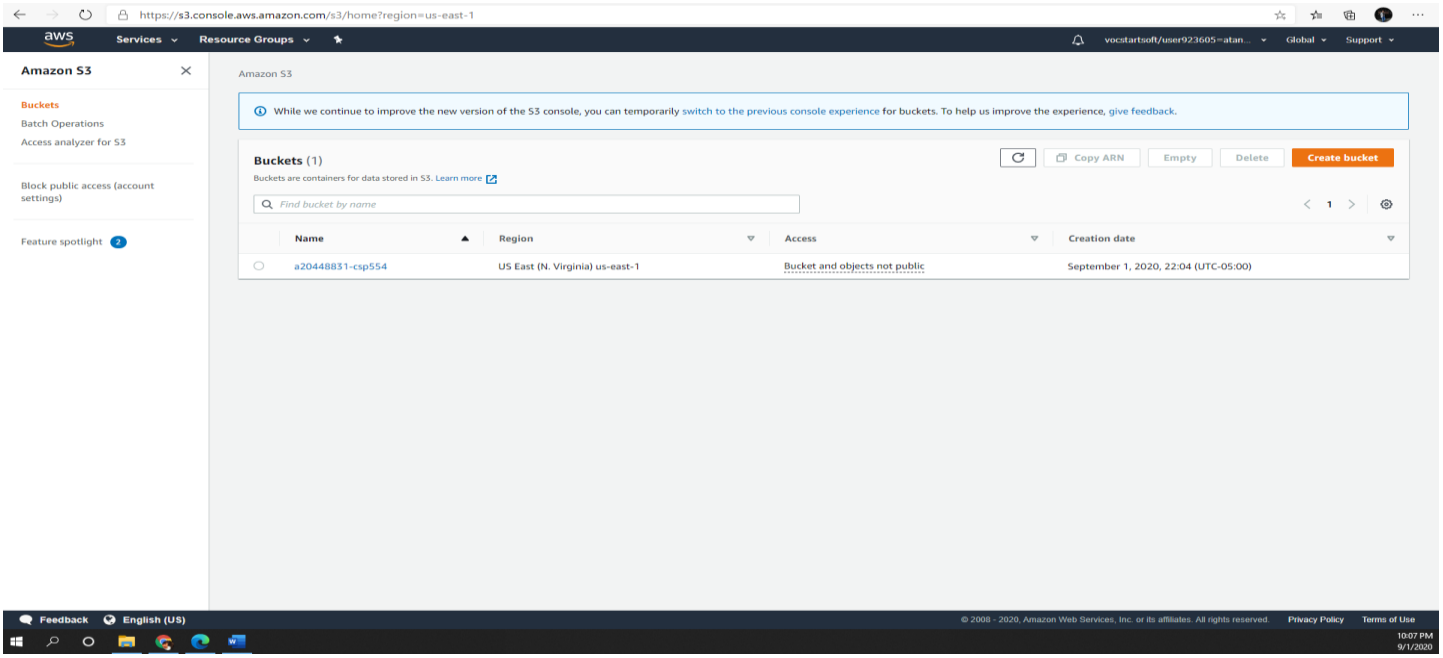
- What is “algorithm dynamics”?

Ans: Algorithm dynamics are the changes made by engineers to improve the commercial service and by consumers in using that service.`

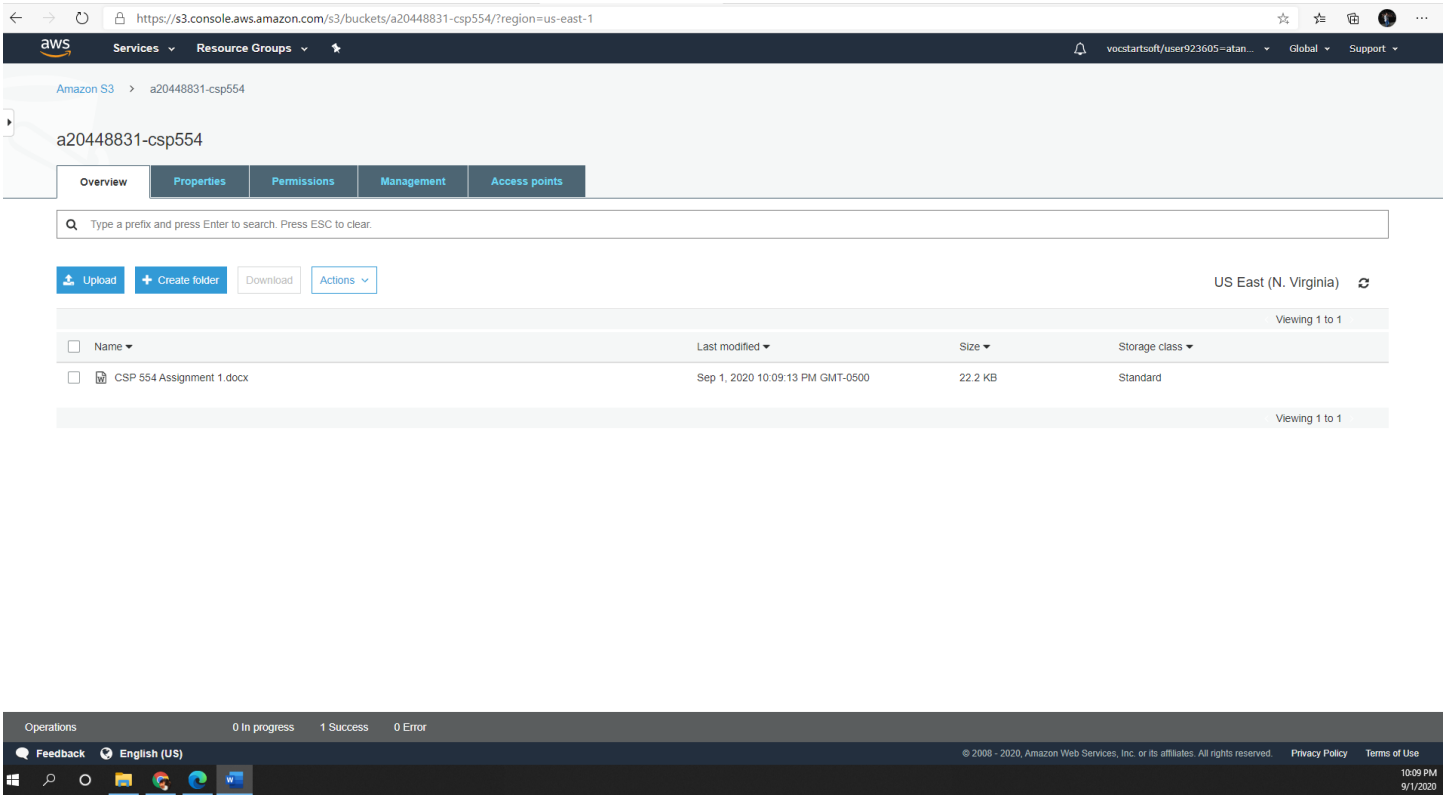
- What aspect of algorithm dynamics impacted the Google flu detection algorithm?
  Ans: Several changes in Google's search algorithm and user behavior probably affected the tracking of GFT. The programmers are constantly testing and improving the search algorithms to improvise user experience which ultimately affects the GFT's tracking.

Que6 (5 points) Set up an Amazon Web Services (AWS) cloud account.

- To receive credit for this question, provide a screen shot showing the S3 bucket you have created. The bucket name should be named something like "YourIITId-CSP554"



- When asked to upload an object to the S3 bucket you have created, just use any text file you have handy (even this one).
- Now also provide a screen shot showing some named object is in the bucket.


https://s3.console.aws.amazon.com/s3/buckets/a20448831-csp554/?region=us-east-1
aws
Services
Resource Groups
vocstartsoft/user923605=atan...
Global
Support
Amazon S3 > a20448831-csp554
a20448831-csp554
Overview
Properties
Permissions
Management
Access points
Type a prefix and press Enter to search. Press ESC to clear.
Upload
Create folder
Download
Actions
US East (N. Virginia)
Viewing 1 to 1
Name
Last modified
Size
Storage class
CSP 554 Assignment 1.docx
Sep 1, 2020 10:09:13 PM GMT-0500
22.2 KB
Standard
Viewing 1 to 1

Operations
0 In progress
1 Success
0 Error
Feedback
English (US)
© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
10:09 PM
9/1/2020