

Kajol Tanesh Shah
A20496724
kshah127@hawk.iit.edu
CSP 554 Fall 2022

CSP 554 - Big Data Technologies Assignment #1

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

1. What was the problem with the Google flu detection algorithm?

Ans. The problem with the Google Flu detection algorithm was that it overestimated the prevalence of flu i.e. it was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which based its estimates on surveillance reports from laboratories across the United States. GFT used relative prevalence of search terms in its model and so, any changes in the search algorithm adversely affected GFT's estimates.

Below are the issues that contributed to GFT's mistakes:-

- 1) Big Data Hubris and
- 2) Algorithm Dynamics

2. What is big data hubris?

Ans. Big data hubris is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis. There are enormous scientific possibilities in big data, however, the quantity of data does not mean that we can ignore foundational issues of measurement and construct validity and reliability, and dependencies among data.

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

Ans. Google Flu detection algorithm could have been combined with other near real-time health data for improvement. By combining GFT and lagged CDC data, as well as dynamically recalibrating GFT, the performance of GFT could have been substantially improved. Also, Google developers could have changed the methodology of GFT by focusing on why the trending flu-related keywords were searched for instead of the number of the trending keywords.

4. What is "algorithm dynamics?"

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

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

Ans. Algorithm dynamics affected Google's search algorithm and this is what likely led to GFT becoming an unstable reflection of the prevalence of the flu. Several changes in Google's search

algorithm and user behavior likely affected GFT's tracking. Also, the explanation for changes in the relative search behavior was blue team dynamics where the algorithm producing the data was modified by the service provider in accordance with their business model. Because GFT used the relative prevalence of search terms in its model, improvements in the search algorithm adversely affected GFT's estimates.

Q. (5 points) Set up an Amazon Web Services (AWS) cloud account, if you don't already have one (see below for details), and then follow the tutorial about how to work with a storage service called S3. Since we will do most of our assignments using AWS, this will get you started. In a while we will come to understand S3 as one critical element of a big data processing architecture known as the "data lake."

- a. To receive credit for this question, provide a screenshot showing the S3 bucket you have created. The bucket name should be named something like "YourIITId-CSP554", for example:
"A1234567_CSP554"
- b. When asked to upload an object to the S3 bucket you have created, just use any text file you have handy (even this one).
- c. Now also provide a screenshot showing some named object is in the bucket.
- d. Make sure to follow the instructions in the pdf file for deleting your bucket at the end of the assignment so you do not incur additional costs.

Solution:-

Step 1:- Create bucket

aws

Services

Search for services, features, blogs, docs, and more

[Alt+S]

We're continuing to improve the S3 console to make it faster and easier to use. If you have feedback on the updated experience, choose [Provide feedback](#).

Amazon S3

>

Buckets

>

Create bucket

Create bucket

Info

Buckets are containers for data stored in S3. [Learn more](#)

General configuration

Bucket name

a20496724-sp554

Bucket name must be globally unique and must not contain spaces or uppercase letters. [See rules for bucket naming](#)

AWS Region

US East (Ohio) us-east-2

Copy settings from existing bucket - optional

Only the bucket settings in the following configuration are copied.

Choose bucket

Object Ownership

Info

Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.

☒ **ACLs disabled (recommended)**
All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies.

☐ **ACLs enabled**
Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using ACLs.

Object Ownership

Bucket owner enforced

aws

Services

Search for services, features, blogs, docs, and more

[Alt+S]

We're continuing to improve the S3 console to make it faster and easier to use. If you have feedback on the updated experience, choose [Provide feedback](#).

Block Public Access settings for this bucket

Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, access point policies, or all. In order to ensure that public access to this bucket and its objects is blocked, turn on Block all public access. These settings apply only to this bucket and its access points. AWS recommends that you turn on Block all public access, but before applying any of these settings, ensure that your applications will work correctly without public access. If you require some level of public access to this bucket or objects within, you can customize the individual settings below to suit your specific storage use cases. [Learn more](#)

☒ **Block all public access**
Turning this setting on is the same as turning on all four settings below. Each of the following settings are independent of one another.

☒ **Block public access to buckets and objects granted through new access control lists (ACLs)**
S3 will block public access permissions applied to newly added buckets or objects, and prevent the creation of new public access ACLs for existing buckets and objects. This setting doesn't change any existing permissions that allow public access to S3 resources using ACLs.

☒ **Block public access to buckets and objects granted through any access control lists (ACLs)**
S3 will ignore all ACLs that grant public access to buckets and objects.

☒ **Block public access to buckets and objects granted through new public bucket or access point policies**
S3 will block new bucket and access point policies that grant public access to buckets and objects. This setting doesn't change any existing policies that allow public access to S3 resources.

☒ **Block public and cross-account access to buckets and objects through any public bucket or access point policies**
S3 will ignore public and cross-account access for buckets or access points with policies that grant public access to buckets and objects.

Bucket Versioning

Versioning is a means of keeping multiple variants of an object in the same bucket. You can use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket. With versioning, you can easily recover from both unintended user actions and application failures. [Learn more](#)

Bucket Versioning

☐ Disable

☒ Enable

Tags (0) - optional

Track storage cost or other criteria by tagging your bucket. [Learn more](#)

No tags associated with this bucket.

Add tag

Default encryption

Automatically encrypt new objects stored in this bucket. [Learn more](#)

Server-side encryption

☒ Disable

☐ Enable

► **Advanced settings**

After creating the bucket you can upload files and folders to the bucket, and configure additional bucket settings.

Cancel

Create bucket

Step 2:- Bucket created

aws

Services

Search for services, features, blogs, docs, and more

[Alt+S]

Global

kshah127

Amazon S3

Buckets

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

Access analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards

AWS Organizations settings

Feature spotlight

AWS Marketplace for S3

We're continuing to improve the S3 console to make it faster and easier to use. If you have feedback on the updated experience, choose **Provide feedback**.

Learn how to effectively use the S3 Storage Classes.

Amazon S3 > Buckets

Account snapshot

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

View Storage Lens dashboard

Buckets (1) Info

Refresh

Copy ARN

Empty

Delete

Create bucket

Buckets are containers for data stored in S3. [Learn more](#)

Find buckets by name

Name

AWS Region

Access

Creation date

<input type="radio"/> a20496724-csp554	US East (Ohio) us-east-2	Bucket and objects not public	August 30, 2022, 22:43:45 (UTC-05:00)
--	--------------------------	-------------------------------	---------------------------------------

Step 3:- Upload an object to S3 bucket

The screenshot shows the AWS S3 console interface. At the top, there's a navigation bar with the AWS logo, 'Services' link, a search bar, and a user profile 'kshah127'. Below the navigation bar, a blue banner contains a message about improving the S3 console. A green banner below that says 'Upload succeeded' with a link to 'View details below.' The main content area is titled 'Upload: status' and includes a 'Close' button. A light blue box contains a warning: 'The information below will no longer be available after you navigate away from this page.' The 'Summary' section shows the destination 's3://a20496724-csp554', a success status with a green checkmark, and details: '1 file, 23.3 KB (100.00%)'. It also shows a 'Failed' status with '0 files, 0 B (0%)'. Below the summary, there are tabs for 'Files and folders' (selected) and 'Configuration'. The 'Files and folders' section shows 'Files and folders (1 Total, 23.3 KB)' with a search bar. A table lists the uploaded file:

Name	Folder	Type	Size	Status	Error
CSP 554 Assignment 1 v3.docx	-	application/vnd.openxmlformats-officedocument.wordprocessingml.document	23.3 KB	Succeeded	-

Step 4:- Delete object

The screenshot shows the AWS S3 console interface for deleting an object. The navigation bar is the same as in Step 3. Below the navigation bar, a blue banner contains a message about improving the S3 console. A blue banner below that says 'Learn how to effectively use the S3 Storage Classes.' The main content area is titled 'Delete objects' with an 'Info' link. A warning box states: 'If a folder is selected for deletion, all objects in the folder will be deleted, and any new objects added while the delete action is in progress might also be deleted. If an object is selected for deletion, any new objects with the same name that are uploaded before the delete action is completed will also be deleted.' Below the warning, a blue box contains an information icon and text: 'Deleting the specified objects adds delete markers to them. If you need to undo the delete action, you can delete the delete markers.' The 'Specified objects' section shows a search bar and a table with the object to be deleted:

Name	Type	Last modified	Size
CSP 554 Assignment 1 v3.docx	docx	August 30, 2022, 22:46:45 (UTC-05:00)	23.3 KB

The 'Delete objects?' section contains a text input field with the word 'delete' entered. At the bottom, there are 'Cancel' and 'Delete objects' buttons.

aws Services Search for services, features, blogs, docs, and more [Alt+S]

ⓘ We're continuing to improve the S3 console to make it faster and easier to use. If you have feedback on the updated experience, choose **Provide feedback**.

✔ **Successfully deleted objects**
View details below.

ⓘ Learn how to effectively use the S3 Storage Classes.

Delete objects: status

ⓘ The information below will no longer be available after you navigate away from this page.

Summary		
Source s3://a20496724-csp554	Successfully deleted ✔ 1 object, 23.3 KB	Failed to delete 0 objects

Step 5:- Delete bucket

aws Services Search for services, features, blogs, docs, and more [Alt+S]

ⓘ We're continuing to improve the S3 console to make it faster and easier to use. If you have feedback on the updated experience, choose **Provide feedback**.

Amazon S3 > Buckets > a20496724-csp554 > Empty bucket

Empty bucket Info

⚠

- Emptying the bucket deletes all objects in the bucket and cannot be undone.
- Objects added to the bucket while the empty bucket action is in progress might be deleted.
- To prevent new objects from being added to this bucket while the empty bucket action is in progress, you might need to update your bucket policy to stop objects from being added to the bucket.

[Learn more](#)

ⓘ If your bucket contains a large number of objects, creating a lifecycle rule to delete all objects in the bucket might be a more efficient way of emptying your bucket. [Learn more](#)

[Go to lifecycle rule configuration](#)

Permanently delete all objects in bucket "a20496724-csp554"?

To confirm deletion, type *permanently delete* in the text input field.

Cancel **Empty**

aws

Services

Search for services, features, blogs, docs, and more

[Alt+S]

Global

kshah127

We're continuing to improve the S3 console to make it faster and easier to use. If you have feedback on the updated experience, choose [Provide feedback](#).

Successfully emptied bucket "a20496724-csp554"
View details below. If you want to delete this bucket, use the [delete bucket configuration](#).

Empty bucket: status

Cancel

Exit

The details below are no longer available after you navigate away from this page.

Summary

Source s3://a20496724-csp554	Successfully deleted 2 objects, 25.3 KB	Failed to delete 0 objects
---------------------------------	--	-------------------------------

Failed to delete (0)

Find objects by name

Name	Prefix	Version ID	Type	Last modified	Size	Error
No failed object deletions						

aws

Services

Search for services, features, blogs, docs, and more

[Alt+S]

Global

kshah127

We're continuing to improve the S3 console to make it faster and easier to use. If you have feedback on the updated experience, choose [Provide feedback](#).

[Amazon S3](#) > [Buckets](#) > [a20496724-csp554](#) > Delete bucket

Delete bucket

Info

⚠

- Deleting a bucket cannot be undone.
- Bucket names are unique. If you delete a bucket, another AWS user can use the name.

[Learn more](#)

Delete bucket "a20496724-csp554"?

To confirm deletion, enter the name of the bucket in the text input field.

a20496724-csp554

Cancel

Delete bucket

① We're continuing to improve the S3 console to make it faster and easier to use. If you have feedback on the updated experience, choose [Provide feedback](#).

① Replicate your data to any storage class to save on storage costs.

Amazon S3 > Buckets

► **Account snapshot**

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

[View Storage Lens data](#)

Buckets (0) [Info](#)

Buckets are containers for data stored in S3. [Learn more](#)



Copy ARN

Empty

Delete

Create

Find buckets by name

< 1

Name	AWS Region	Access	Creation date
------	------------	--------	---------------

No buckets

You don't have any buckets.

Create bucket