Name – Rishabh Jain
Mail – rjain35@hawk.iit.edu
Course – CSP 554

CSP 554—Big Data Technologies
Assignment – 1

Q1) What was the problem with the Google flu detection algorithm?

Sol.) The main problem with the Google flu detection algorithm was, it predicts more than double count of doctor visit for the influenza illness than the Centers for Disease Control and Prevention (CDC). This all happened because the algorithm was measuring the flu trend by measuring the count of keywords. It does not analyze why people were searching for those keywords. In this case, the model was making prediction using only the search queries from previous few years.

Also, the two main issue that contributed to this mistake are: -

1. Big Data Hubris
2. Algorithm Dynamics

Q2) What is big data hubris?

Sol.) Big data hubris can be defined as the misunderstanding of making big data as a substitute, rather than a supplement to, traditional data collection and analysis. Also, the quantity of data does not mean that we can ignore foundational issues of measuring and validation among data.

Q3) What approach could have been used to improve the Google flu detection algorithm?

Sol.) Google flu detection algorithm can be improved over time if the developer had changed the methodology of flu detection as “why these trending flu-related keywords are searched for” instead of “creating the data by the count of trending flu keywords

At the same time, the concept of the algorithm dynamics should also be very precise and limited, as it is also one of the main factors which make the algorithm show wrong data contributing to high error %.

Q4) What is “algorithm dynamics?”

Sol.) Algorithm dynamics are the changes made by engineers to improve the commercial service and by consumers in using that service. Algorithm dynamics are important because it helps the algorithm to adapt to the new changes as per the changes happening to the society in a period of time.

Q5) What aspect of algorithm dynamics impacted the Google flu detection algorithm?

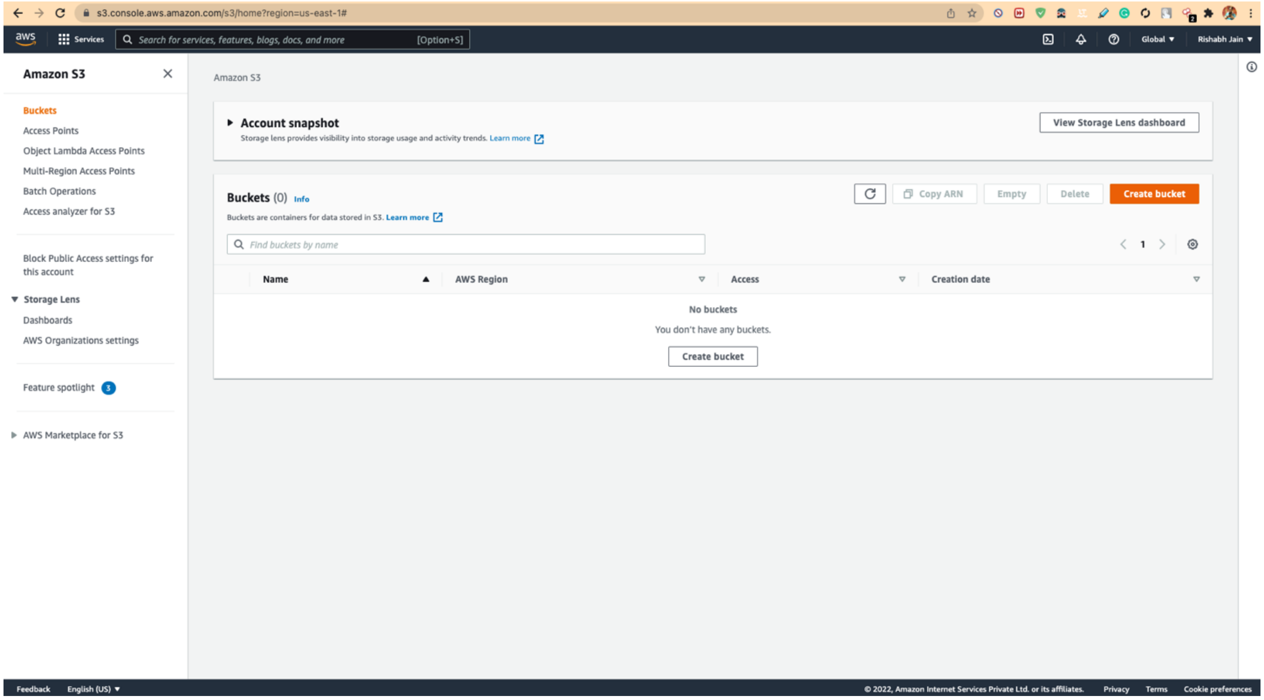
Sol.) Algorithm dynamics leads the GFT to act abnormally, and it starts showing the unstable reflection of the prevalence of the flu. GFT uses the relative frequency of search terms in its model, improvements made in the GFT algorithm adversely affects the GFT estimates

The changes and modifications contributed to biased data collection, which leads to the high % error and showed double the count of actual data.

Name – Rishabh Jain
Mail – rjain35@hawk.iit.edu
Course – CSP 554

Q6)

1. Below screen shot shows the S3 (Simple Storage Service) dashboard of the AWS cloud.


s3.console.aws.amazon.com/s3/home?region=us-east-1#
aws
Services
Search for services, features, blogs, docs, and more
[Option+S]
Global ▼
Rishabh Jain ▼
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 3
▶ AWS Marketplace for S3
Amazon S3
View Storage Lens dashboard
Copy ARN
Empty
Delete
Create bucket
Find buckets by name
1
Create bucket
Feedback
English (US) ▼
© 2022, Amazon Internet Services Private Ltd. or its affiliates.
Privacy
Terms
Cookie preferences

▶ Account snapshot

Storage lens provides visibility into storage usage and activity trends. Learn more

Buckets (0) Info

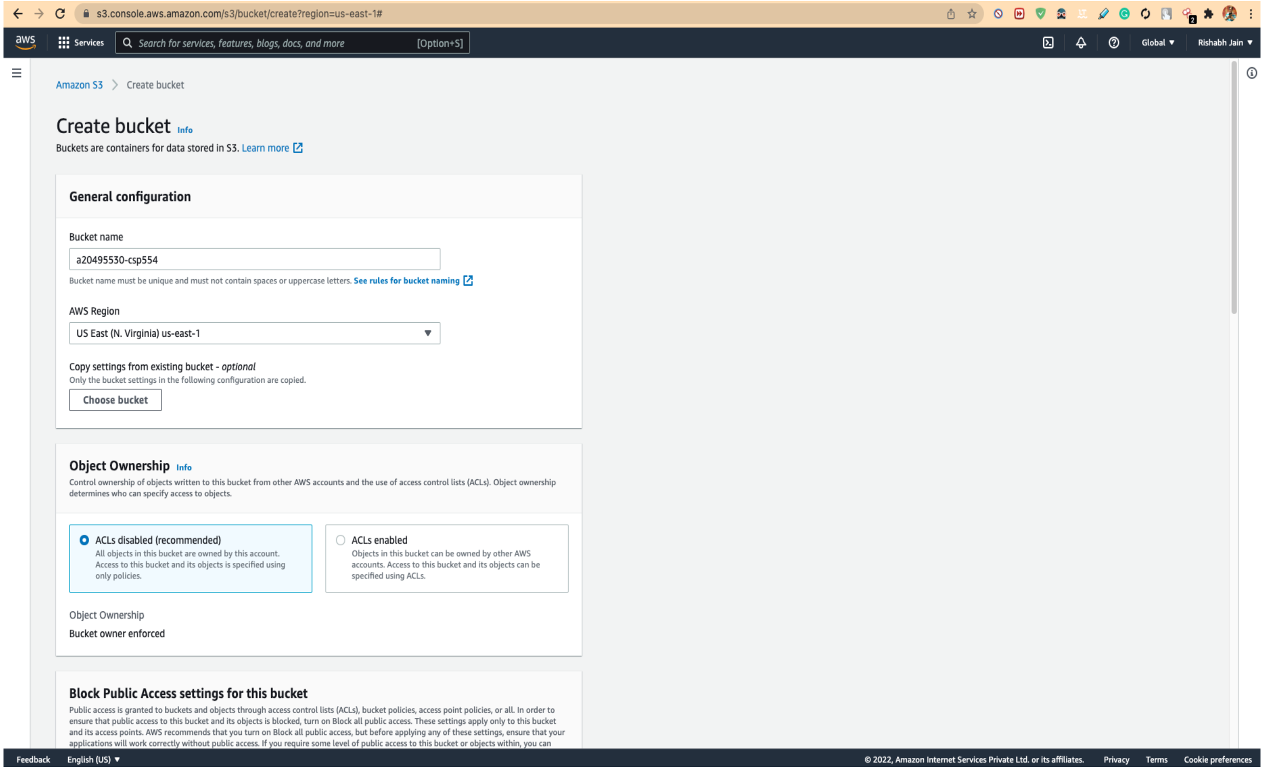
Buckets are containers for data stored in S3. Learn more

Name ▲	AWS Region ▽	Access ▽	Creation date ▽

No buckets

You don't have any buckets.

2. Next step is to create your own S3 bucket. Please find the chain of screen shot for the same.


s3.console.aws.amazon.com/s3/bucket/create?region=us-east-1#
aws
Services
Search for services, features, blogs, docs, and more
[Option+S]
Global ▼
Rishabh Jain ▼
Amazon S3 > Create bucket
Choose bucket
Feedback
English (US) ▼
© 2022, Amazon Internet Services Private Ltd. or its affiliates.
Privacy
Terms
Cookie preferences

Create bucket Info

Buckets are containers for data stored in S3. Learn more

General configuration

Bucket name

a20495530-csp554

Bucket name must be unique and must not contain spaces or uppercase letters. See rules for bucket naming

AWS Region

US East (N. Virginia) us-east-1 ▼

Copy settings from existing bucket - *optional*

Only the bucket settings in the following configuration are copied.

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.

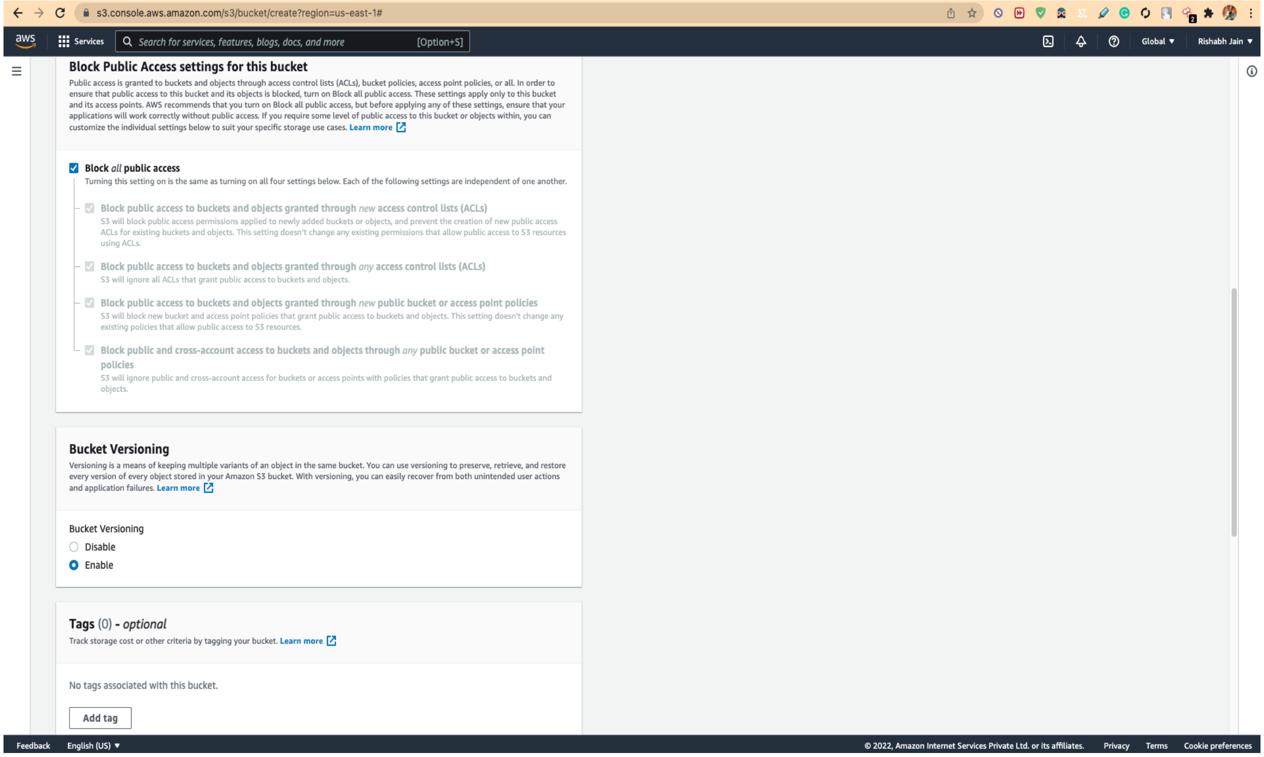
Object Ownership

Bucket owner enforced

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

Name – Rishabh Jain
Mail – rjain35@hawk.iit.edu
Course – CSP 554


s3.console.aws.amazon.com/s3/bucket/create?region=us-east-1#
aws
Services
Search for services, features, blogs, docs, and more
[Option+S]
Global
Rishabh Jain
Feedback
English (US)
© 2022, Amazon Internet Services Private Ltd. or its affiliates.
Privacy
Terms
Cookie preferences

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

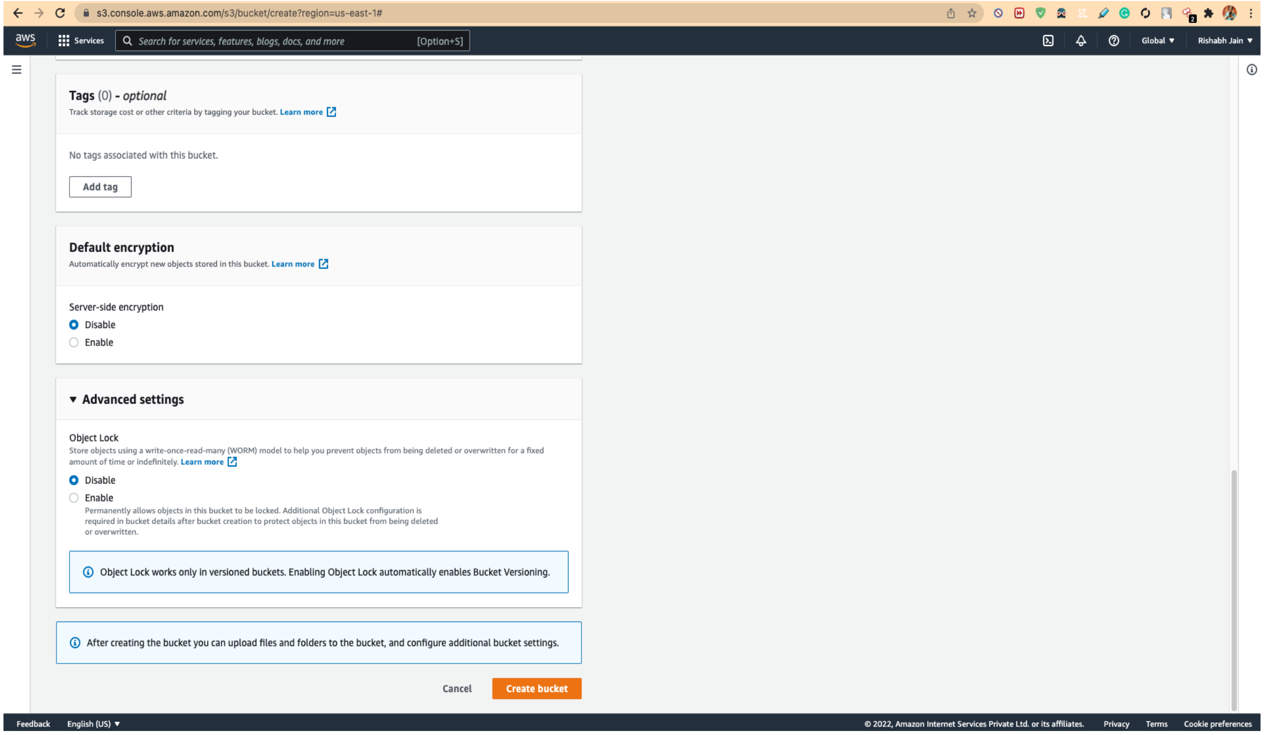
Bucket Versioning
○ Disable
● Enable

Tags (0) - *optional*

Track storage cost or other criteria by tagging your bucket. Learn more

No tags associated with this bucket.

Add tag


s3.console.aws.amazon.com/s3/bucket/create?region=us-east-1#
aws
Services
Search for services, features, blogs, docs, and more
[Option+S]
Global
Rishabh Jain
Feedback
English (US)
© 2022, Amazon Internet Services Private Ltd. or its affiliates.
Privacy
Terms
Cookie preferences

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

Object Lock
Store objects using a write-once-read-many (WORM) model to help you prevent objects from being deleted or overwritten for a fixed amount of time or indefinitely. Learn more

● Disable
○ Enable
Permanently allows objects in this bucket to be locked. Additional Object Lock configuration is required in bucket details after bucket creation to protect objects in this bucket from being deleted or overwritten.

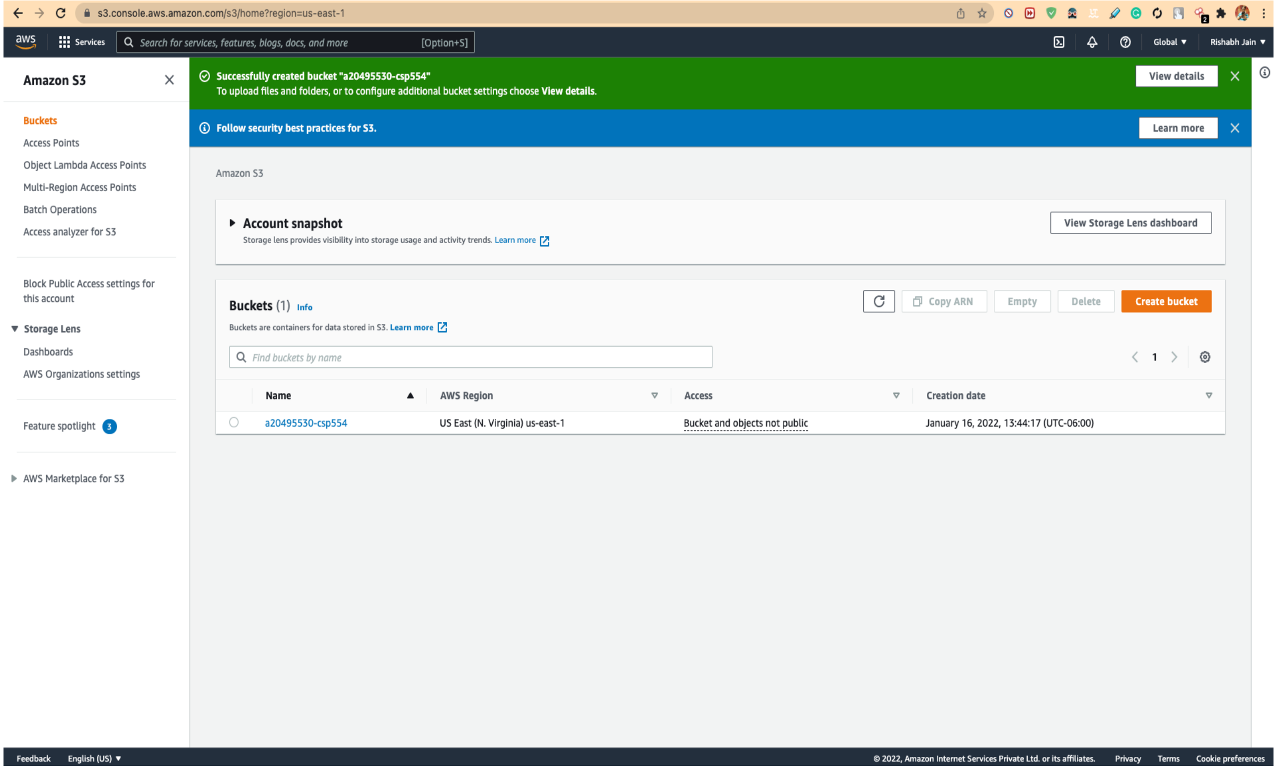
Object Lock works only in versioned buckets. Enabling Object Lock automatically enables Bucket Versioning.

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

Cancel | Create bucket

Name – Rishabh Jain
Mail – rjain35@hawk.iit.edu
Course – CSP 554

3. A bucket is created with name “a20495530-csp554”



s3.console.aws.amazon.com/s3/home?region=us-east-1

Successfully created bucket "a20495530-csp554"
To upload files and folders, or to configure additional bucket settings choose **View details**.

Follow security best practices for S3.

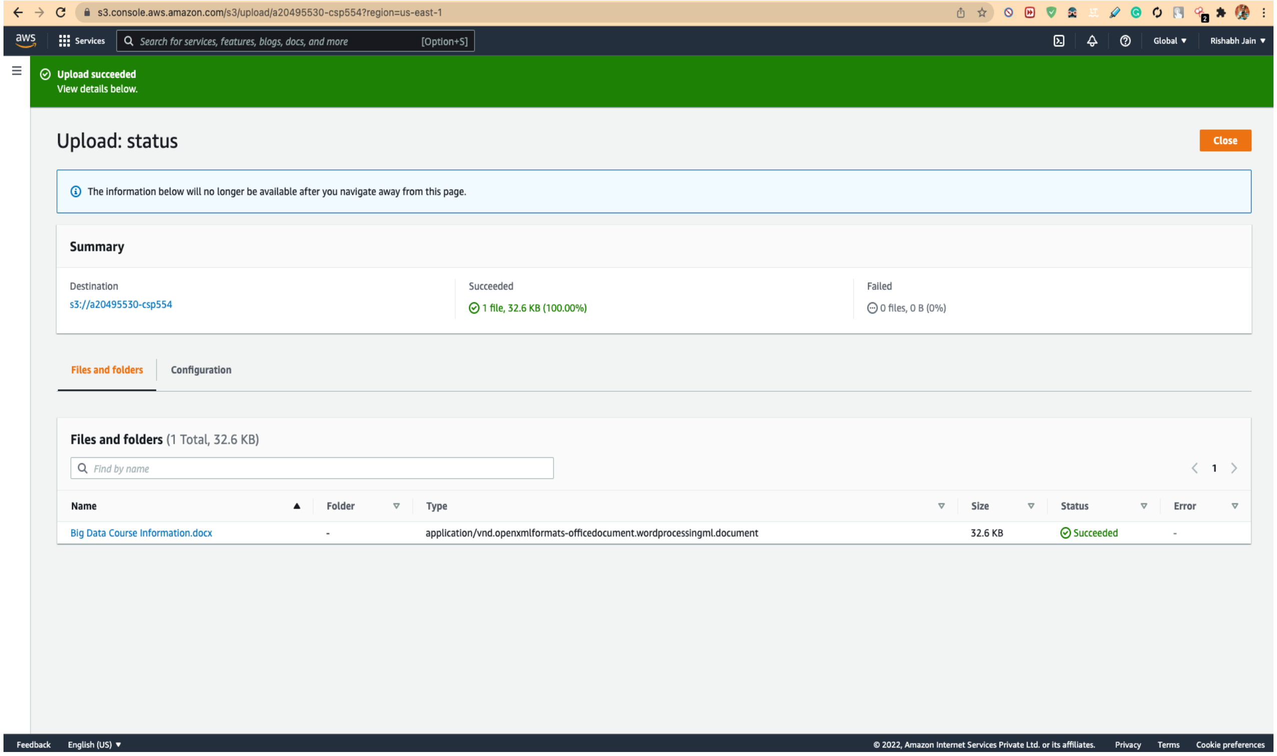
Amazon S3

Account snapshot
Storage lens provides visibility into storage usage and activity trends. Learn more

Buckets (1) Info
Buckets are containers for data stored in S3. Learn more

Name	AWS Region	Access	Creation date
a20495530-csp554	US East (N. Virginia) us-east-1	Bucket and objects not public	January 16, 2022, 13:44:17 (UTC-06:00)

4. Uploaded the sample object successfully in newly created S3 bucket.



s3.console.aws.amazon.com/s3/upload/a20495530-csp554?region=us-east-1

Upload succeeded
View details below.

Upload: status

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

Summary

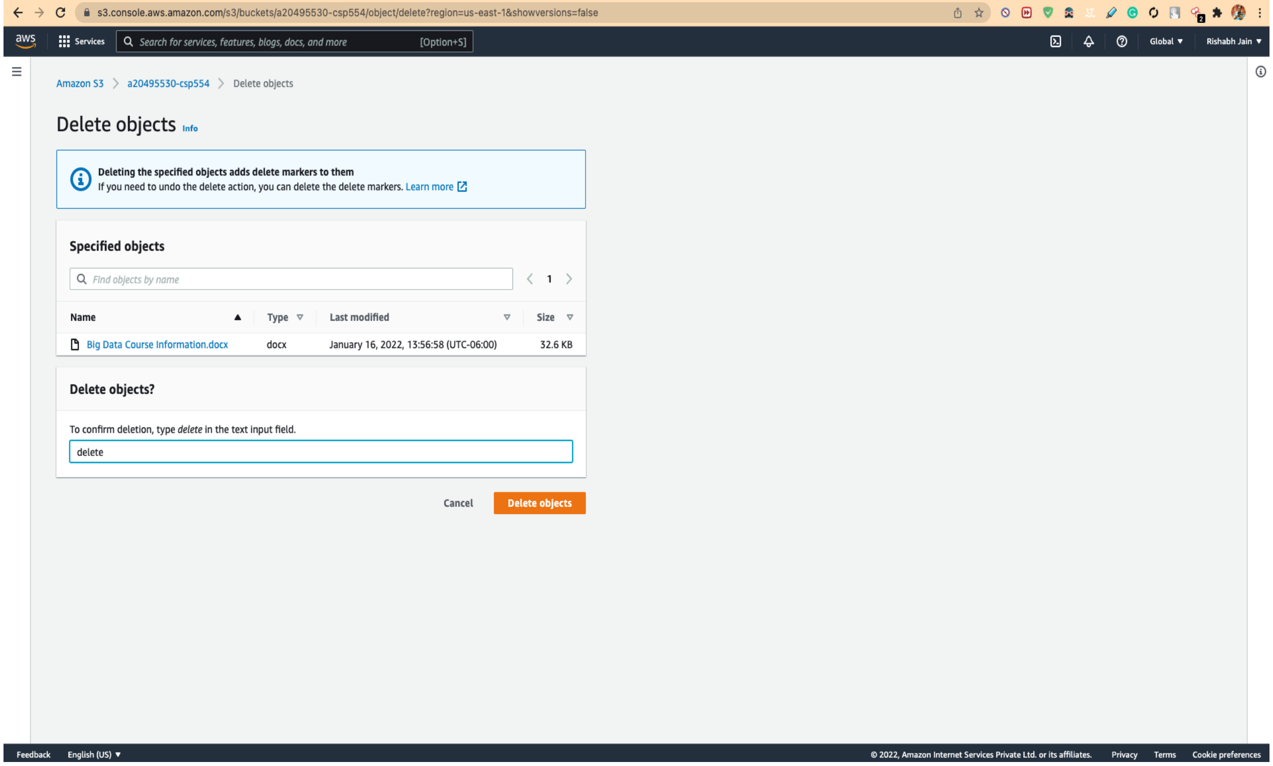
Destination	Succeeded	Failed
s3://a20495530-csp554	1 file, 32.6 KB (100.00%)	0 files, 0 B (0%)

Files and folders (1 Total, 32.6 KB)

Name	Folder	Type	Size	Status	Error
Big Data Course Information.docx	-	application/vnd.openxmlformats-officedocument.wordprocessingml.document	32.6 KB	Succeeded	-

Name – Rishabh Jain
Mail – rjain35@hawk.iit.edu
Course – CSP 554

5. Deleting the newly created bucket to avoid extra costing. To delete a bucket, it is mandatory that it should be empty and there is no object inside it.


s3.console.aws.amazon.com/s3/buckets/a20495530-csp554/object/delete?region=us-east-1&showversions=false
aws
Services
Search for services, features, blogs, docs, and more [Option+S]
Global ▼
Rishabh Jain ▼
Amazon S3 > a20495530-csp554 > Delete objects

Delete objects Info

Deleting the specified objects adds delete markers to them
If you need to undo the delete action, you can delete the delete markers. Learn more

Specified objects

Find objects by name

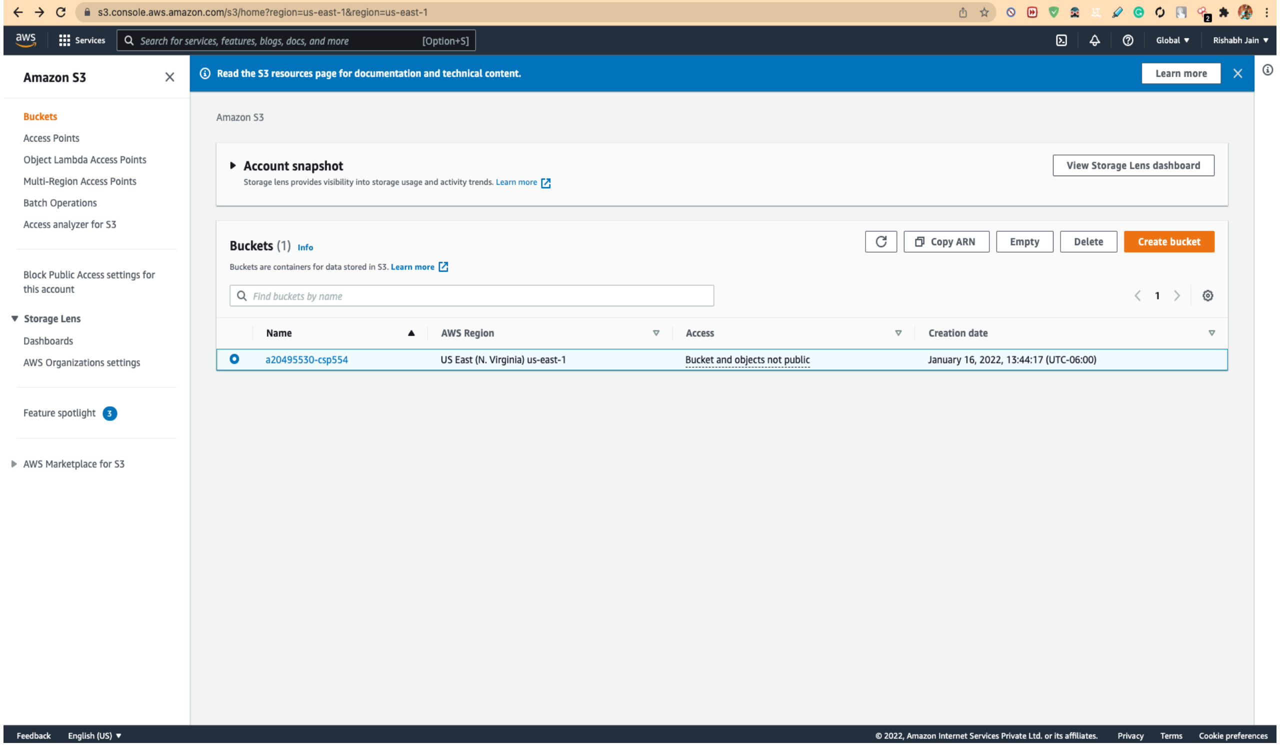
Name	Type	Last modified	Size
Big Data Course Information.docx	docx	January 16, 2022, 13:56:58 (UTC-06:00)	32.6 KB

Delete objects?

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

delete

First, choose the bucket you want to delete and then click on delete button showed on right side.


s3.console.aws.amazon.com/s3/home?region=us-east-1&region=us-east-1
aws
Services
Search for services, features, blogs, docs, and more [Option+S]
Global ▼
Rishabh Jain ▼
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 3
AWS Marketplace for S3

Read the S3 resources page for documentation and technical content.

Account snapshot
Storage lens provides visibility into storage usage and activity trends. Learn more

View Storage Lens dashboard

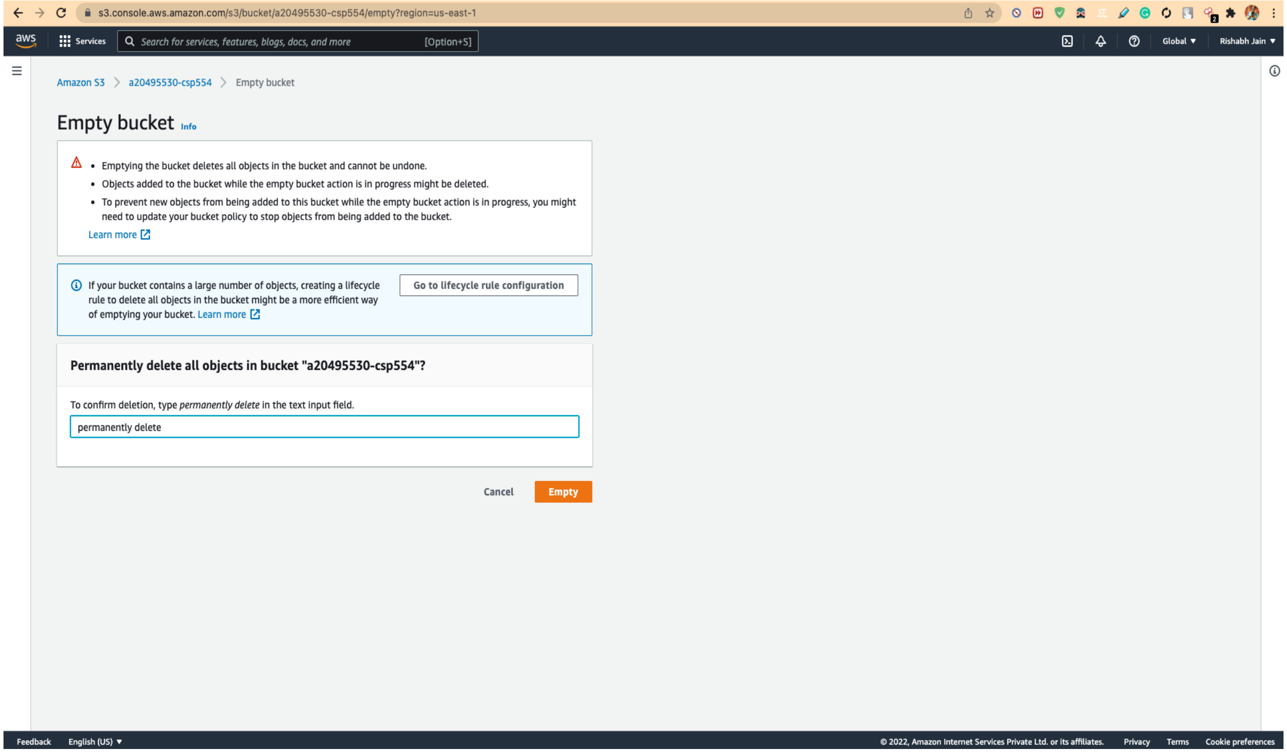
Buckets (1) Info
Buckets are containers for data stored in S3. Learn more

Copy ARN
Empty
Delete
Create bucket
Find buckets by name

	Name	AWS Region	Access	Creation date
●	a20495530-csp554	US East (N. Virginia) us-east-1	Bucket and objects not public	January 16, 2022, 13:44:17 (UTC-06:00)

Feedback
English (US) ▼
© 2022, Amazon Internet Services Private Ltd. or its affiliates.
Privacy
Terms
Cookie preferences

Name – Rishabh Jain
Mail – rjain35@hawk.iit.edu
Course – CSP 554


s3.console.aws.amazon.com/s3/bucket/a20495530-csp554/empty?region=us-east-1

aws
Services
Search for services, features, blogs, docs, and more
[Option+S]
Global ▼
Rishabh Jain ▼
Amazon S3 > a20495530-csp554 > Empty bucket

Empty bucket

- 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

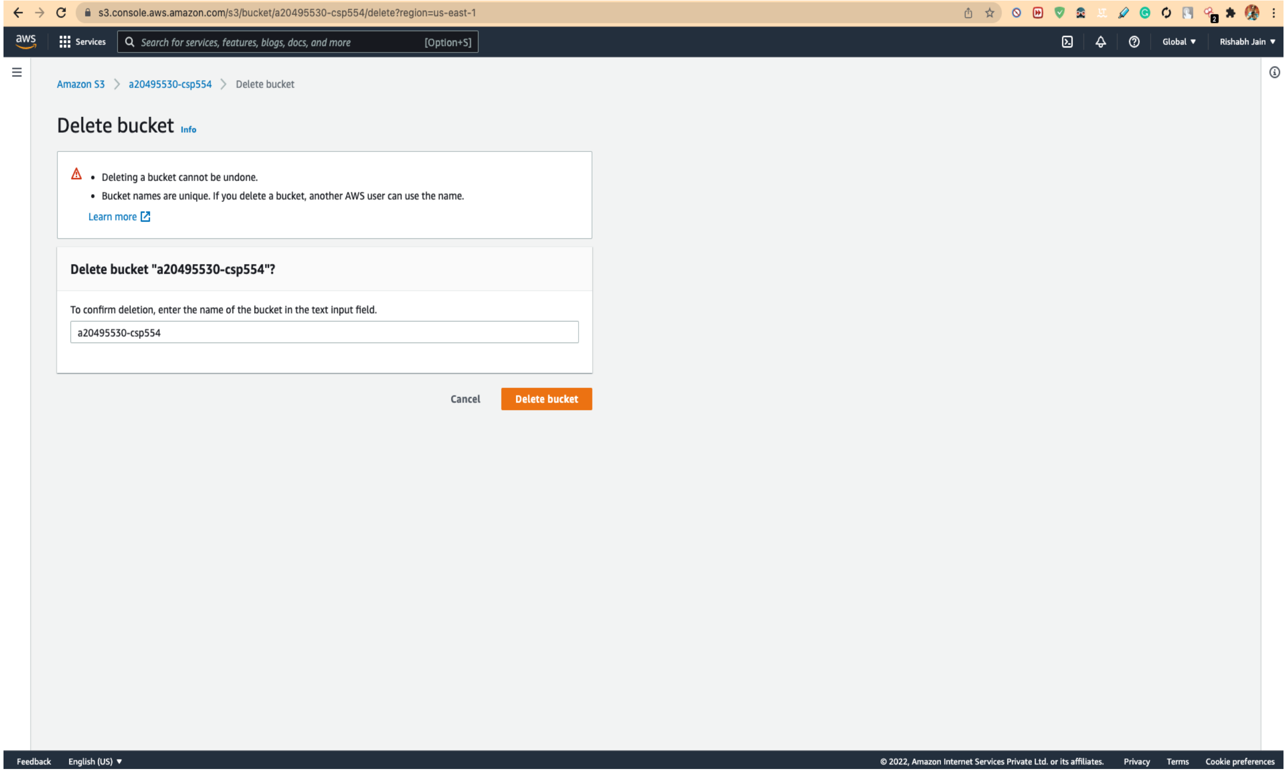
If your bucket contains a large number of objects, creating a lifecycle rule to delete all objects in the bucket might be a more efficient way of emptying your bucket. Learn more

Permanently delete all objects in bucket "a20495530-csp554"?

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

permanently delete

Cancel
Empty
Feedback
English (US) ▼
© 2022, Amazon Internet Services Private Ltd. or its affiliates.
Privacy
Terms
Cookie preferences


s3.console.aws.amazon.com/s3/bucket/a20495530-csp554/delete?region=us-east-1

aws
Services
Search for services, features, blogs, docs, and more
[Option+S]
Global ▼
Rishabh Jain ▼
Amazon S3 > a20495530-csp554 > Delete bucket

Delete bucket

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

Learn more

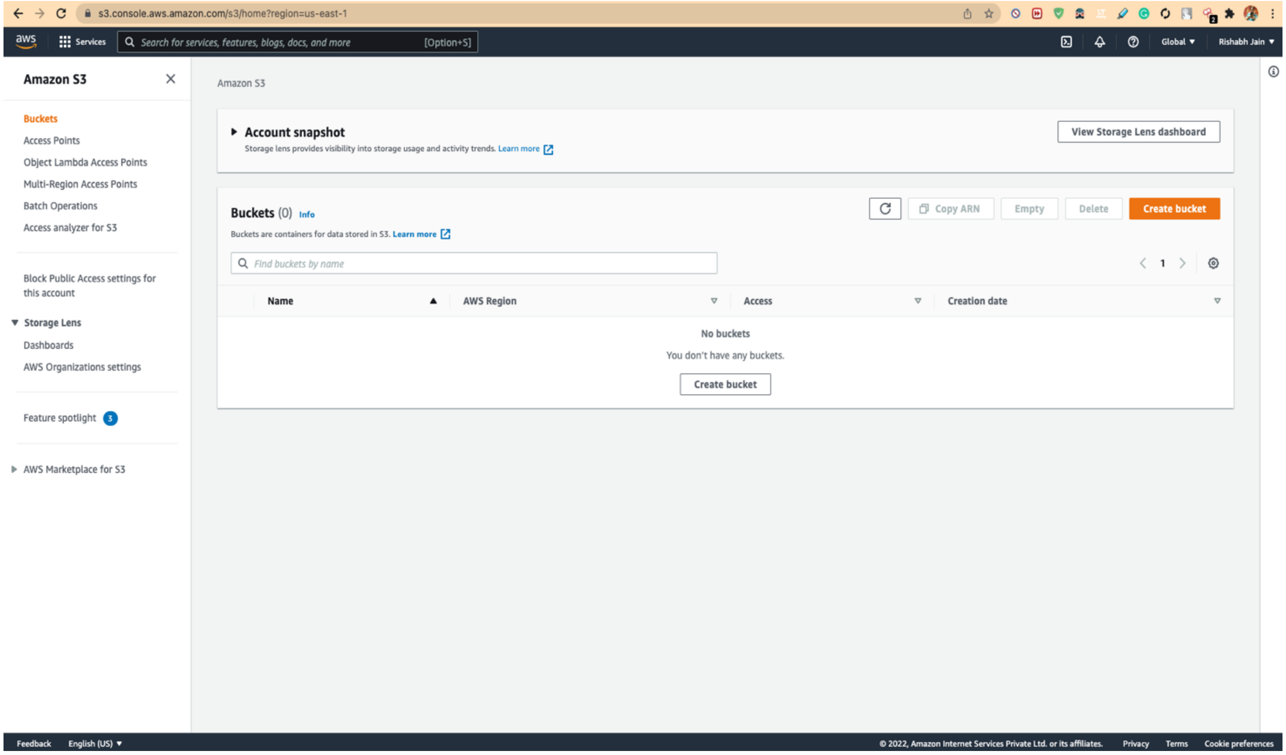
Delete bucket "a20495530-csp554"?

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

a20495530-csp554

Cancel
Delete bucket
Feedback
English (US) ▼
© 2022, Amazon Internet Services Private Ltd. or its affiliates.
Privacy
Terms
Cookie preferences

Name – Rishabh Jain
Mail – rjain35@hawk.iit.edu
Course – CSP 554



Account snapshot

Storage lens provides visibility into storage usage and activity trends. Learn more

Buckets (0) Info

Buckets are containers for data stored in S3. Learn more

Name ▲	AWS Region	Access	Creation date

No buckets

You don't have any buckets.

© 2022, Amazon Internet Services Private Ltd. or its affiliates.