



# Assignment 1

1. Read an article on "Blackboard" in "Articles" section

### **The Parable of Google Flu (just 3 pages!)**

(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.** GFT or Google flu detection algorithm was predicting more than double the proportion of doctor visits for influenza-like illness than the Centers for Disease Control and Prevention (CDC), which were based on its estimates on surveillance reports from laboratories across the United States. Basically, GFT was over-predicting the flu by a significant margin and its inaccuracies raised concerns about the reliability of big data analytics for public health surveillance. Two issues that contributed to GFT's mistakes:

- 1) Big data hubris
- 2) Algorithm dynamics

- **What is big data hubris?**

**Ans.** Big data hubris is commonly an implicit assumption that big data is a replacement, rather than a supplement to traditional data collection and analysis. It fails to consider the limitations of big data, such as its susceptibility to bias and noise. Further, the quantity of data does not mean that one can ignore fundamental measurement issues and build validity, reliability, and dependencies between data. Therefore, to avoid big data hubris, it is important for us to be aware of the limitations of big data and to use it in combination with other methods.

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

**Ans.** Google Flu detection algorithms could have been improved if the developers had changed the detection method from simply counting the number of flu-related keywords to understanding the why behind those searches. For example, the search term "high school basketball" is strongly correlated with the CDC data because flu cases tend to peak during the winter, when high school basketball is in season. However, the search term "high school basketball" is not actually related to the flu. It is simply a seasonal term that is correlated with the flu by chance. This should have been a warning that the big data were overfitting the small number of cases. Further, the concept of algorithm dynamics must have been carefully considered, as it is one of the main factors that can lead to an algorithm displaying erroneous data, which can result in a high percentage of error.

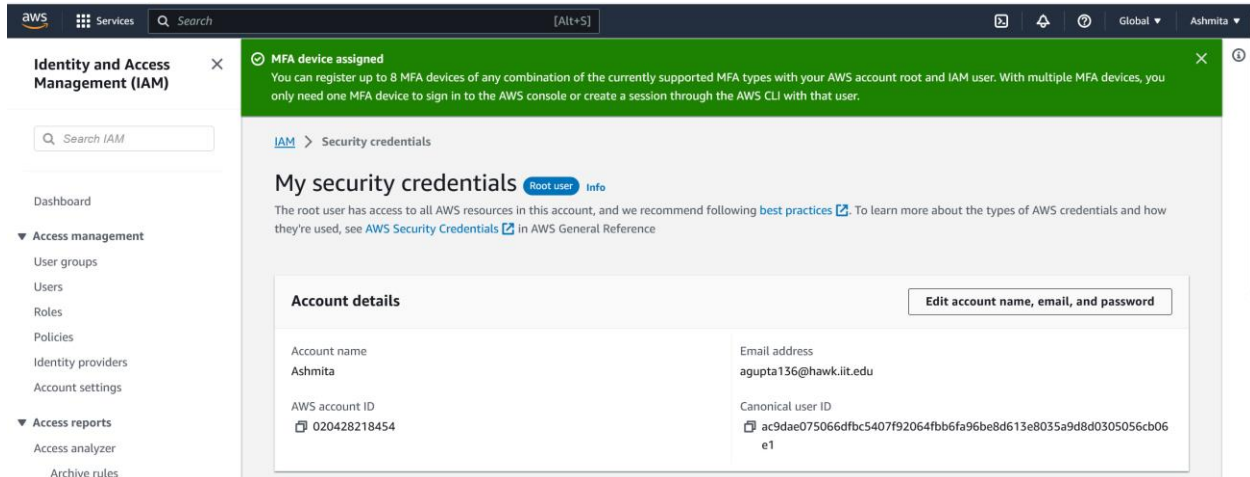
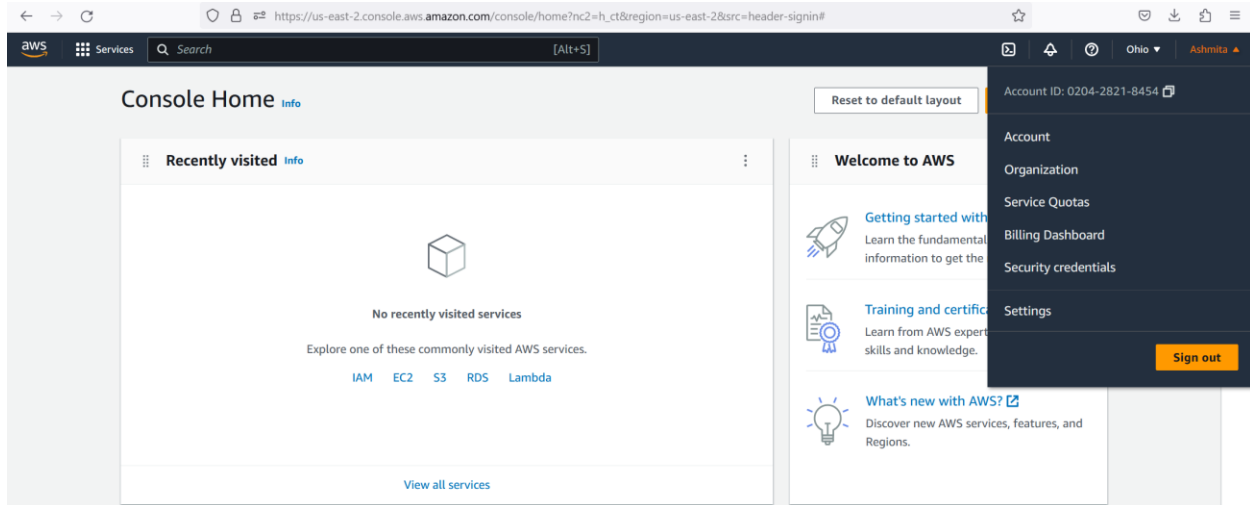
- **What is "algorithm dynamics?"**

**Ans.** Algorithm dynamics refers to the way that an algorithm changes over time in response to new data. It is basically the changes made by engineers to improve the commercial service and by consumers in using that service. Algorithm dynamism is important because it helps the algorithm adapt to new changes based on the changes occurring over a certain period in society

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

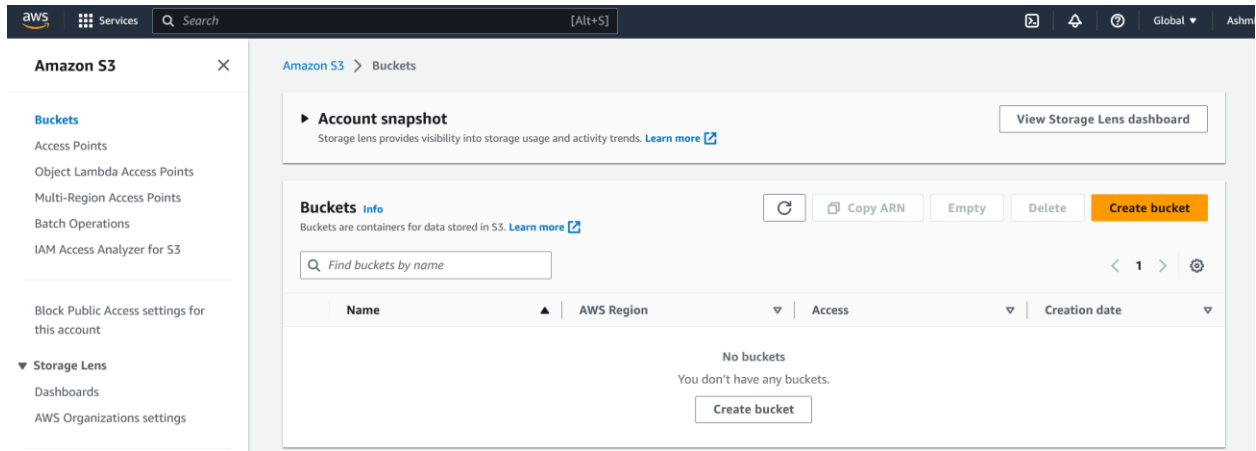
**Ans.** The algorithm was trained on data from previous years, but search behavior can change over time. For instance, people are more likely to search for flu-related terms during a media-stoked panic. This led to the algorithm overestimating the number of flu cases. GFT uses the relative popularity of the search terms in its model and thus impacted the Google flu detection algorithm. Further, it acts abnormally by showing unstable reflection about the prevalence of the flu.

## 2. Amazon account creation:

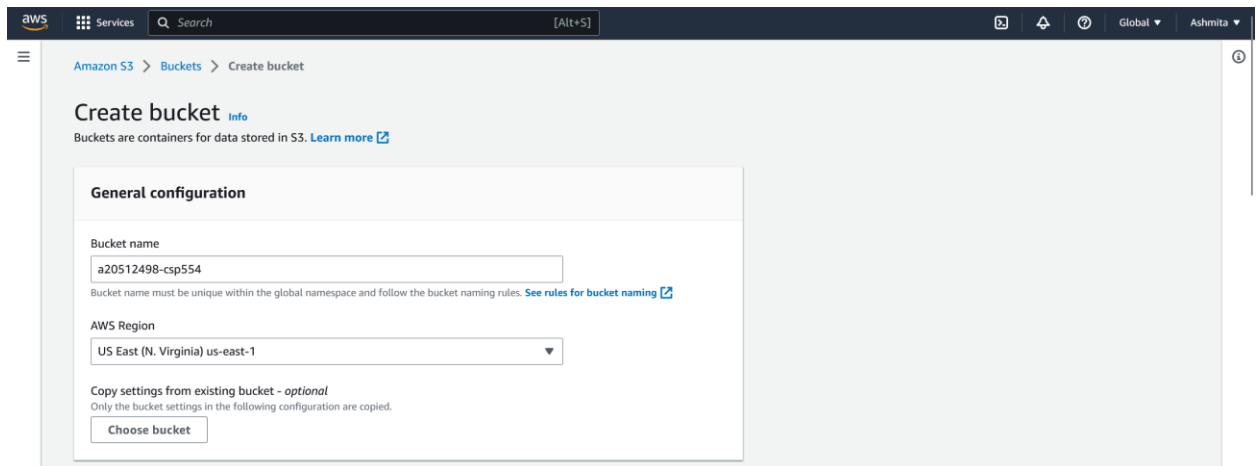


## (a) Amazon S3 Bucket

Below is the screenshot of the Amazon S3:



## (b) Creating an S3 Bucket with the name a20512498-csp554:



Services

Search

[Alt+S]

Global

Ashmita

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 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 is 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.

Bucket Versioning

Disable

Enable

Tags (0) - optional

You can use bucket tags to track storage costs and organize buckets. [Learn more](#)

No tags associated with this bucket.

Add tag

Default encryption

Info

Server-side encryption is automatically applied to new objects stored in this bucket.

Encryption type

Info

Server-side encryption with Amazon S3 managed keys (SSE-S3)

Server-side encryption with AWS Key Management Service keys (SSE-KMS)

Dual-layer server-side encryption with AWS Key Management Service keys (DSSE-KMS)

Secure your objects with two separate layers of encryption. For details on pricing, see [DSSE-KMS pricing](#) on the [Storage](#) tab of the [Amazon S3 pricing page](#).

Bucket Key

Using an S3 Bucket Key for SSE-KMS reduces encryption costs by lowering calls to AWS KMS. S3 Bucket Keys aren't supported for DSSE-KMS. [Learn more](#)

Disable

Enable

Services

Search

[Alt+S]

Global

Ashmita

Successfully created bucket "a20512498-csp554"

To upload files and folders, or to configure additional bucket settings choose [View details](#).

Amazon S3

Buckets

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](#)

Find buckets by name

Name

AWS Region

Access

Creation date

a20512498-csp554

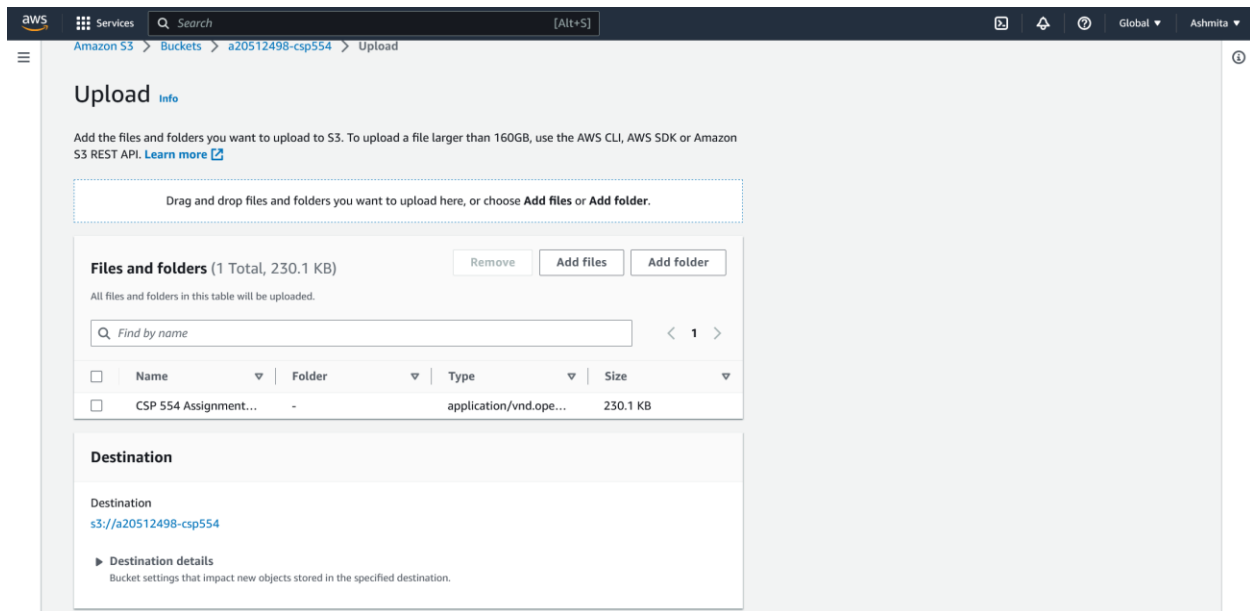
US East (N. Virginia) us-east-1

Bucket and objects not public

August 30, 2023, 22:34:37 (UTC-05:00)

Ashmita Gupta

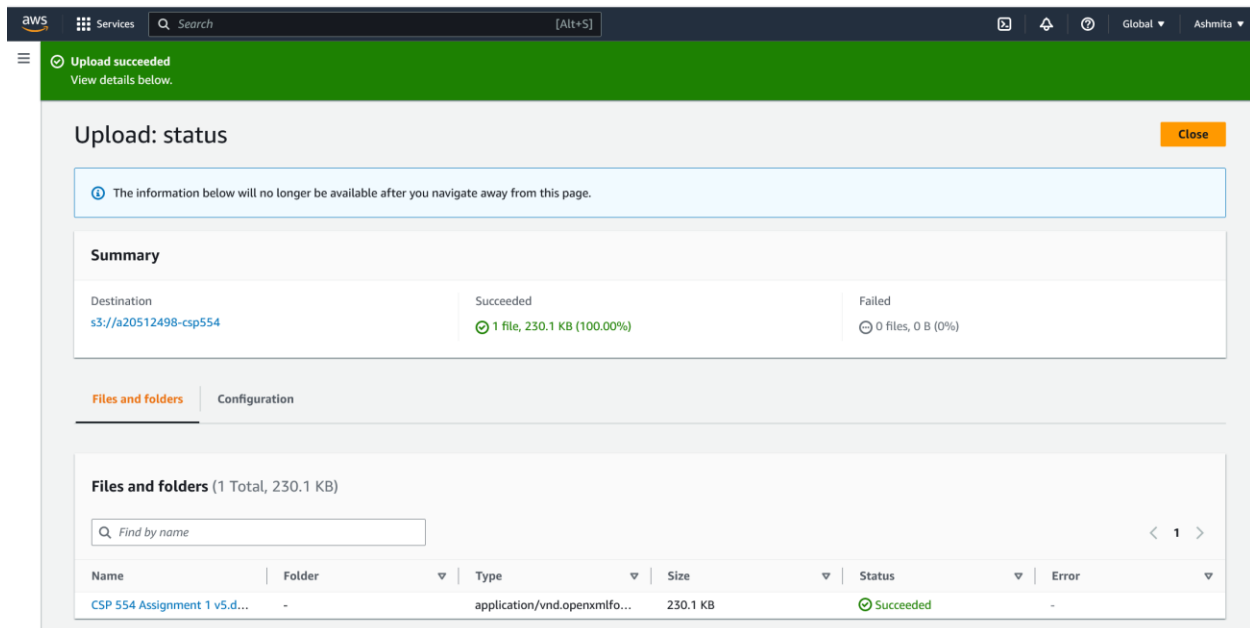
### (c) Creating and uploading an object in the Bucket:



The screenshot shows the AWS S3 'Upload' page for bucket 'a20512498-csp554'. The page has a dark header with the AWS logo, 'Services', a search bar, and user information 'Global' and 'Ashmita'. The breadcrumb trail is 'Amazon S3 > Buckets > a20512498-csp554 > Upload'. The main heading is 'Upload' with an 'Info' link. Below it, a message states: 'Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)'. A dashed box contains the instruction: 'Drag and drop files and folders you want to upload here, or choose **Add files** or **Add folder**.' Below this is a section titled 'Files and folders (1 Total, 230.1 KB)' with 'Remove', 'Add files', and 'Add folder' buttons. A note says 'All files and folders in this table will be uploaded.' There is a search bar 'Find by name' and a pagination control '< 1 >'. A table lists the upload:

<input type="checkbox"/>	Name	Folder	Type	Size
<input type="checkbox"/>	CSP 554 Assignment...	-	application/vnd.ope...	230.1 KB

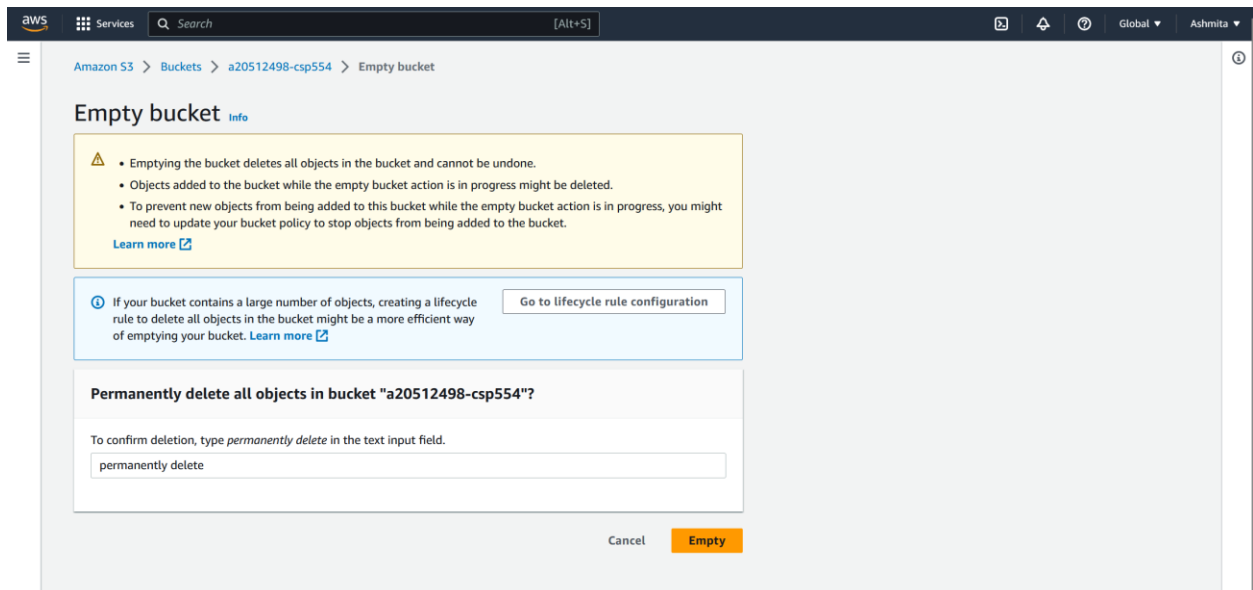
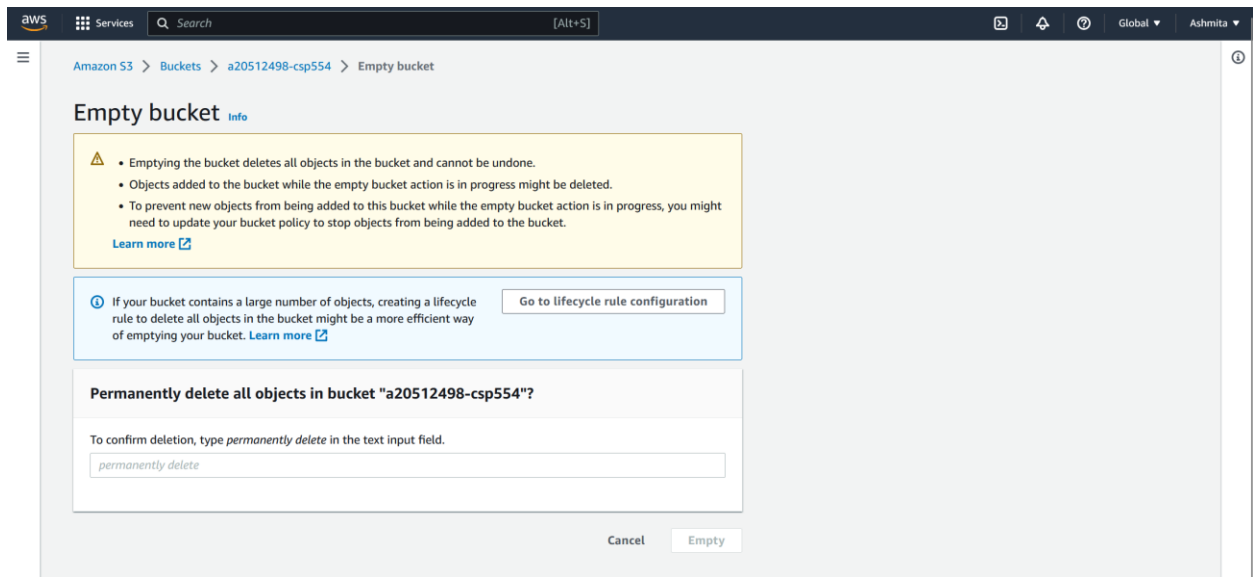
Below the table is a 'Destination' section with the text 'Destination' and the path 's3://a20512498-csp554'. A link 'Destination details' is shown with the subtext 'Bucket settings that impact new objects stored in the specified destination.'



The screenshot shows the 'Upload: status' page. The header is green with a checkmark icon and the text 'Upload succeeded' and 'View details below.'. A 'Close' button is in the top right. Below the header is a light blue box with an information icon and the text: 'The information below will no longer be available after you navigate away from this page.' The main section is titled 'Upload: status' with a 'Close' button. It contains a 'Summary' section with three columns: 'Destination' (s3://a20512498-csp554), 'Succeeded' (1 file, 230.1 KB (100.00%)), and 'Failed' (0 files, 0 B (0%)). Below the summary are two tabs: 'Files and folders' (selected) and 'Configuration'. The 'Files and folders' tab shows a table with the upload details:

Name	Folder	Type	Size	Status	Error
CSP 554 Assignment 1 v5.d...	-	application/vnd.openxmlfo...	230.1 KB	✓ Succeeded	-

## (d) Deleting a bucket and its objects:



aws

Services

Search

[Alt+S]

Global

Ashmita

Successfully emptied bucket "a20512498-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://a20512498-csp554

Successfully deleted

1 object, 230.1 KB

Failed to delete

0 objects

Failed to delete (0)

Find objects by name

< 1 >

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

aws

Services

Search

[Alt+S]

Global

Ashmita

Amazon S3

Buckets

a20512498-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.

If this bucket is used with a Multi-Region Access Point in an external account, initiate failover before deleting the bucket.

If this bucket is used with an access point in an external account, the requests made through those access points will fail after you delete this bucket.

[Learn more](#)

Delete bucket "a20512498-csp554"?

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

a20512498-csp554

Cancel

Delete bucket

aws

Services

Search

[Alt+S]

Global

Ashmita

Successfully deleted bucket "a20512498-csp554"

Storage

Amazon S3

Store and retrieve any amount of data from anywhere

Amazon S3 is an object storage service that offers industry-leading scalability, data availability, security, and performance.

Create a bucket

Every object in S3 is stored in a bucket. To upload files and folders to S3, you'll need to create a bucket where the objects will be stored.

Create bucket

How it works

Introduction to Amazon S3

Copy link

aws

Pricing

With S3, there are no minimum fees. You only pay for what you use. Prices are based on the location of your S3 bucket.

Estimate your monthly bill using the [AWS Simple Monthly Calculator](#)

[View pricing details](#)

Resources