

# Big Data on the Cloud

Lecture 1

Dr. Ramesh Shankar

# Uploading your data to Amazon S3

- Create your AWS account with AWS Educate (\$100 free credit) (requires credit card) or AWS Educate Starter Account (third party: qwikLABs, \$75 free credit)

<https://aws.amazon.com/education/awseducate/>

<https://aws.amazon.com/premiumsupport/knowledge-center/educate-starter-account/>

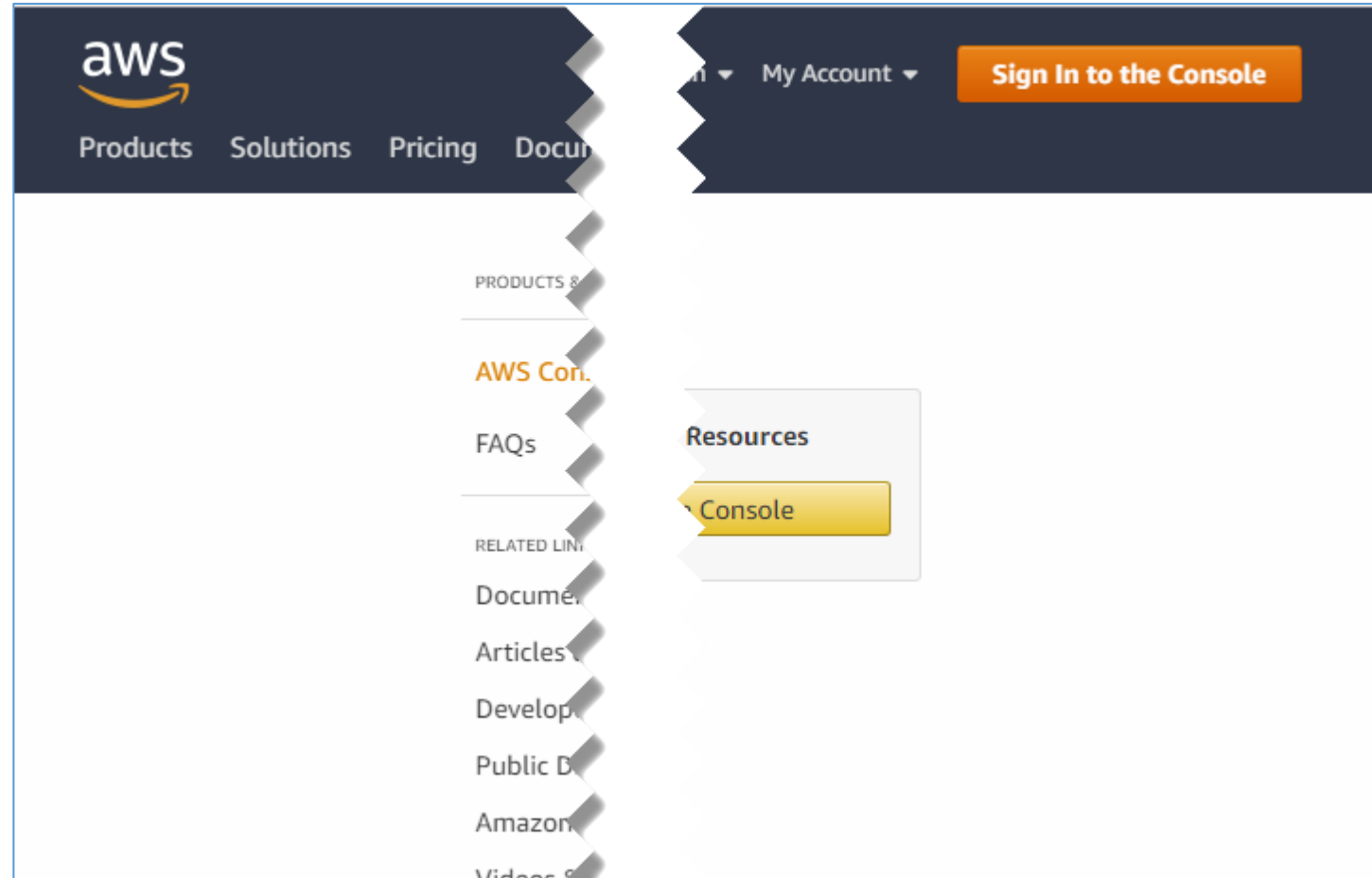
- Request free tier
- Login to AWS
- Create “bucket” (unique name across AWS)
- Upload data to your bucket

Download this file to your computer

[https://www.dropbox.com/s/ir4f2ouerlkg9sx/trip\\_data\\_1.csv?dl=0](https://www.dropbox.com/s/ir4f2ouerlkg9sx/trip_data_1.csv?dl=0)

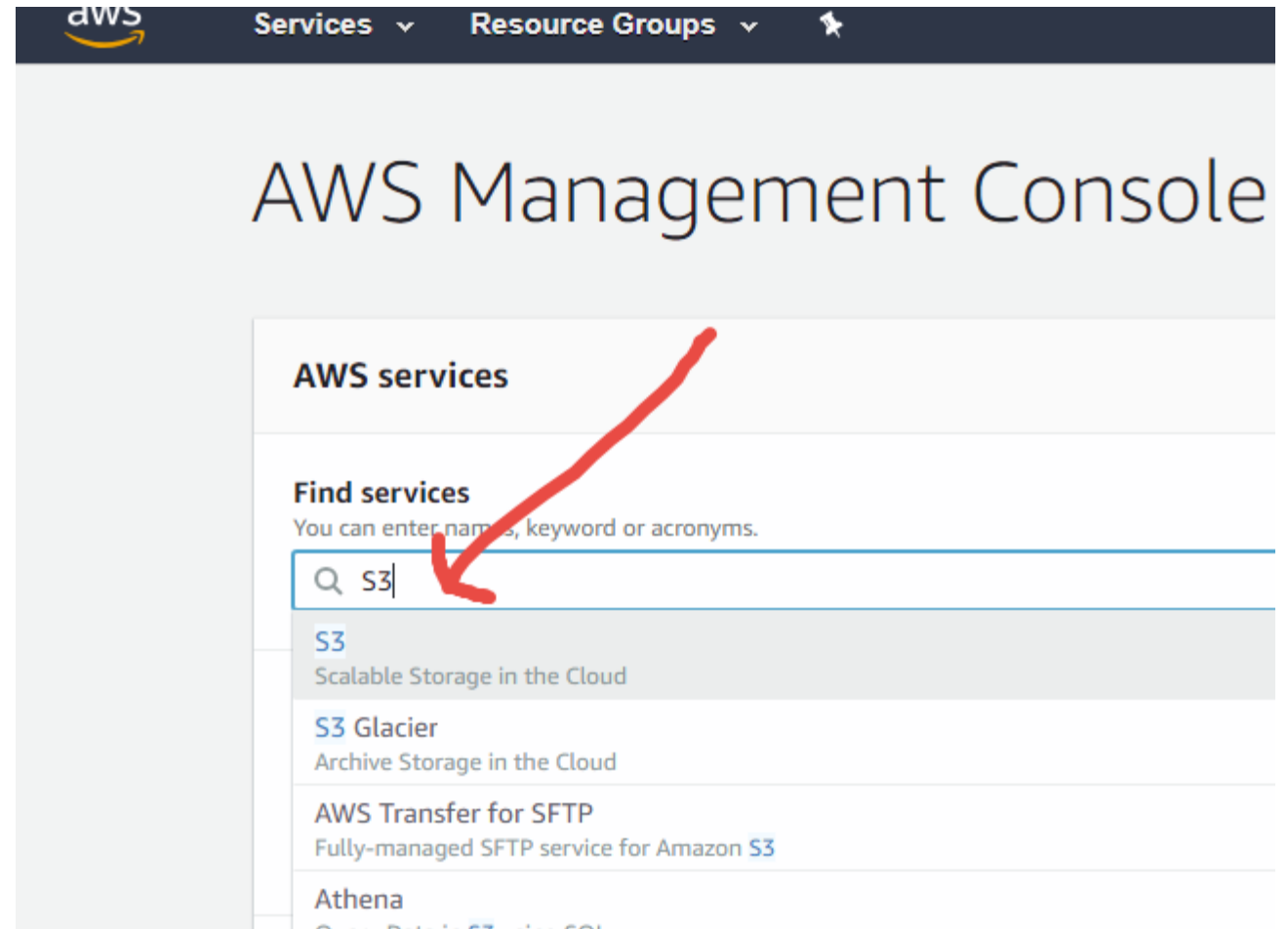
# Login to AWS

<https://aws.amazon.com/console/>



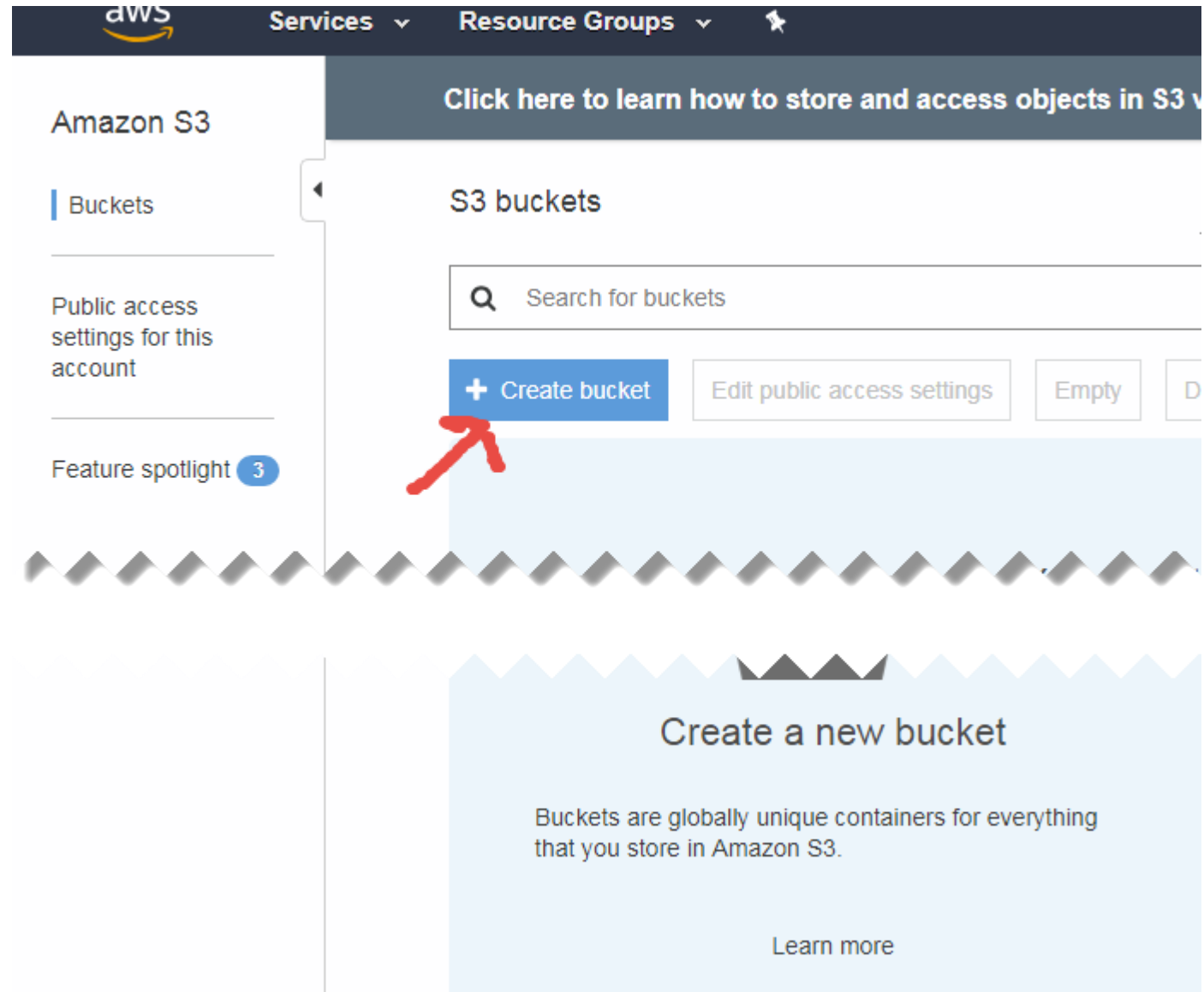
# Type “s3” in search box

- “S3” stands for “Simple Storage Service”
- Cloud based data storage facility on Amazon AWS



# Create a new bucket

- “Bucket” is a public cloud storage resource (basic unit or “folder” in which you can store data)
- You create a bucket in a region close to you (e.g. east, west, etc)



# Choose bucket name, region

- Name is a unique identifier across all of AWS
- Region is the geographic region closest to you

The screenshot shows the AWS S3 'Create bucket' wizard, step 1: Name and region. The interface has a blue header with a close button (X) and two progress indicators: '1 Name and region' and '4 Review'. The main content area is dark gray. Under the heading 'Name and region', there is a 'Bucket name' label with an information icon (i) and a text input field containing 'nyctaxi-jan-2013-trip1'. A yellow arrow points to this input field. Below this is a 'Region' label and a dropdown menu showing 'US East (N. Virginia)' with a blue downward arrow. Further down is a section 'Copy settings from an existing bucket' with a dropdown menu showing 'You have no buckets0 Buckets' and a blue downward arrow. At the bottom, there are three buttons: 'Create' (disabled), 'Cancel', and 'Next' (active). A yellow arrow points to the 'Next' button.

# Use default bucket settings for now

The image displays three overlapping screenshots of the AWS 'Create bucket' wizard, specifically the 'Configure options' step. The wizard is titled 'Create bucket' and has a progress bar at the top with four steps: 1. Name and region (checked), 2. Configure options (active), 3. Set permissions (checked), and 4. Review (checked).

The 'Configure options' step shows the following settings:

- Versioning:** ☐ Keep all versions of an object in the same bucket. [Learn more](#)
- Server access logging:** ☐ Log requests for access to your bucket. [Learn more](#)
- Tags:** You can use tags to track project costs. [Learn more](#)  
Key:  Value:   
[+ Add another](#)
- Object-level logging:** ☐ Record object-level API activity using AWS CloudTrail for an additional cost. See [CloudTrail pricing](#) or [learn more](#)
- Default encryption:** ☐ Automatically encrypt objects when they are stored in S3. [Learn more](#)
- Advanced settings:** [View](#)


At the bottom of the wizard, there are 'Previous' and 'Next' buttons. A yellow arrow points to the 'Next' button in each of the three screenshots, indicating the flow of the wizard.




# And the bucket is created

## S3 buckets

 [Discover the console](#)

 Search for buckets

All access types 

 Create bucket

Edit public a

1 Buckets

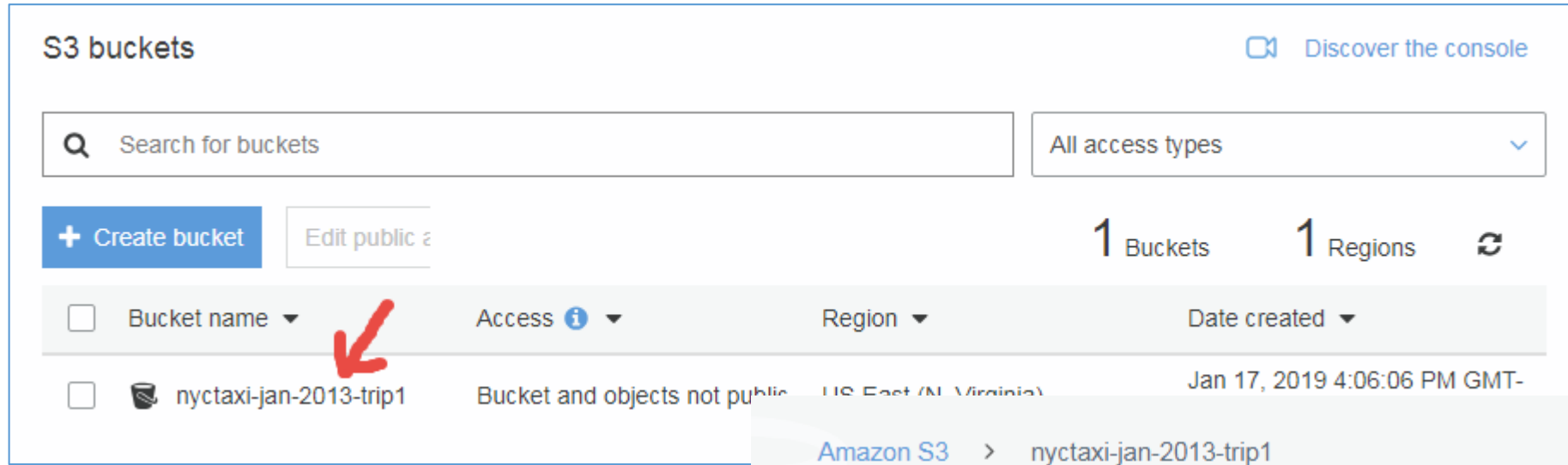
1 Regions



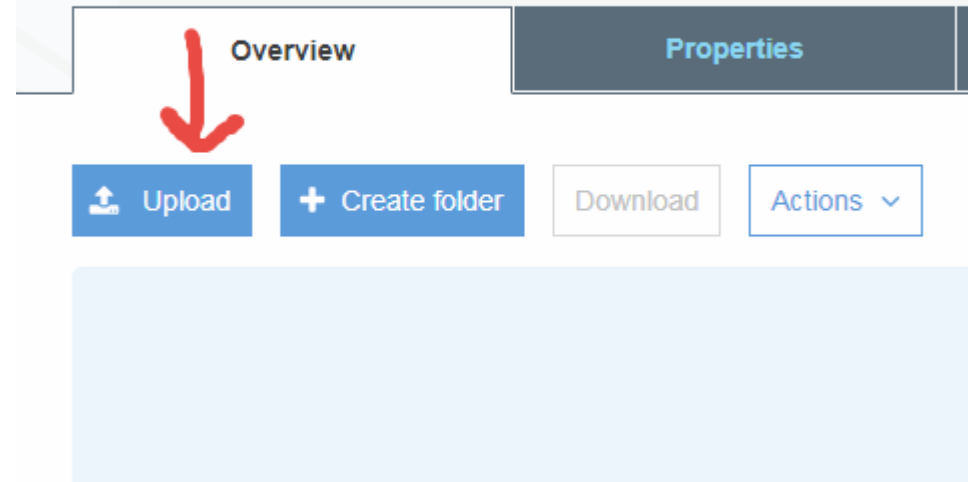
<input type="checkbox"/>	Bucket name ▼	Access  ▼	Region ▼	Date created ▼
<input type="checkbox"/>	 nyctaxi-jan-2013-trip1	Bucket and objects not public	US East (N. Virginia)	Jan 17, 2019 4:06:06 PM GMT-0500

# Upload a file into the bucket

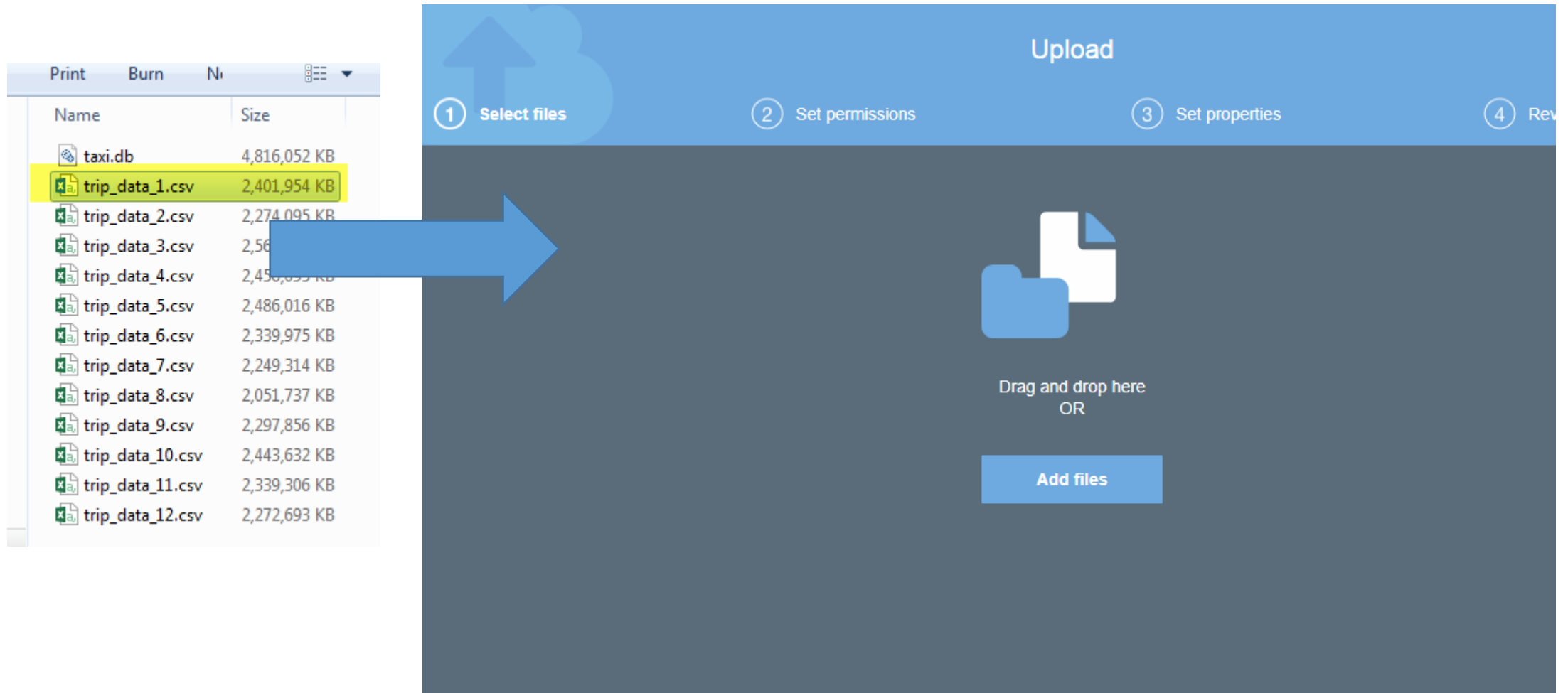
- Click on the bucket:



- Click “upload”:



# Drag-drop your file here (it'll take a while)



(Choose standard options for now)