pabloUrizar /
Labos_IST

Code Issues Pull requests Actions Projects Wiki Security Insights

Labos_IST / Labo3 / README.md

pabloUrizar Mise en page 16 minutes ago

542 lines (491 loc) · 24.3 KB

Preview Code Blame Raw

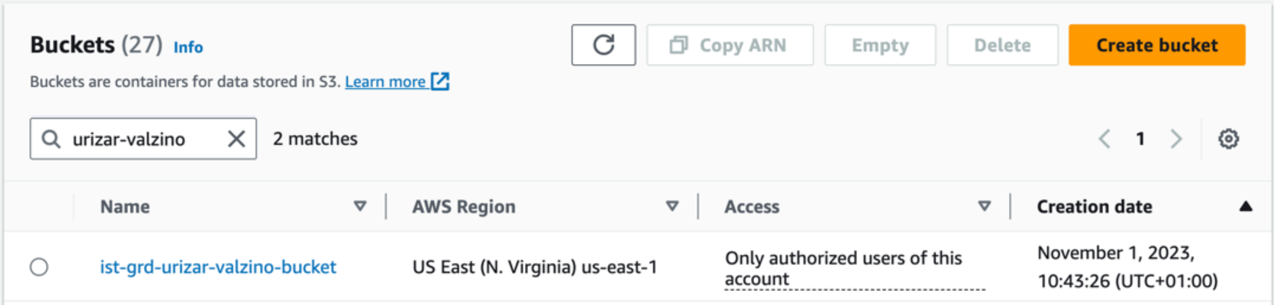
IST_labo3

Authors : Valzino Benjamin, Urizar Pablo

TASK 1: USE THE WEB CONSOLE TO CREATE AN S3 BUCKET AND UPLOAD AND DOWNLOAD OBJECTS (FILES)

1. Navigate to the S3 console.

2.Create a new bucket



Buckets (27) Info

Buckets are containers for data stored in S3. Learn more

urizar-valzino 2 matches 1

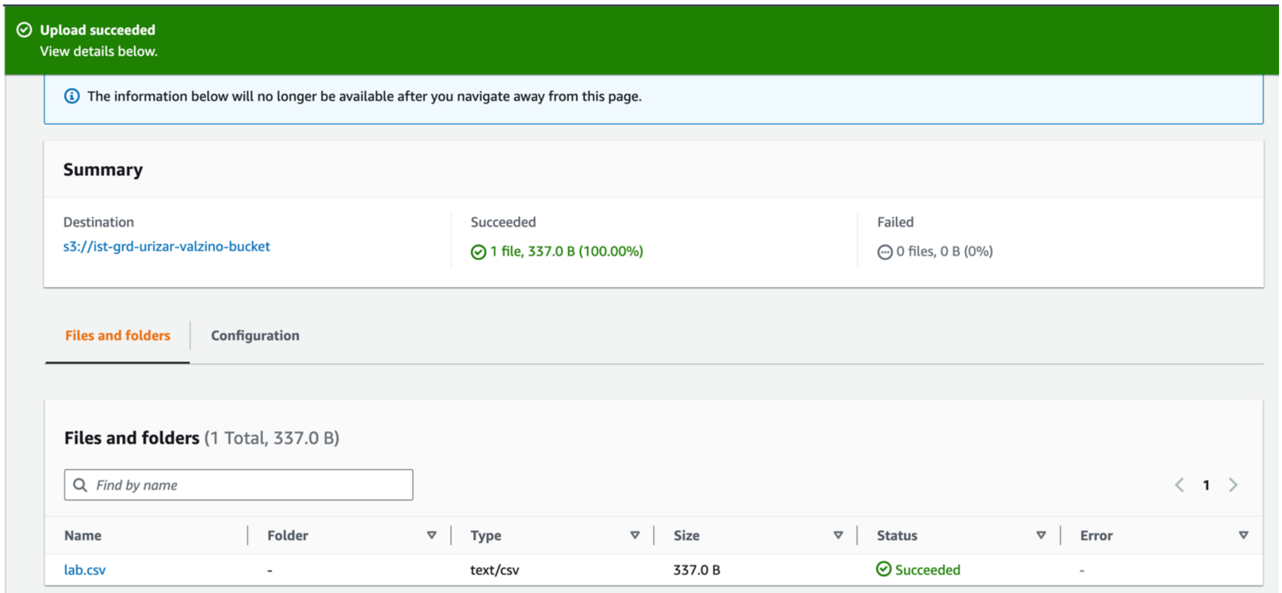
	Name	AWS Region	Access	Creation date
○	ist-grd-urizar-valzino-bucket	US East (N. Virginia) us-east-1	Only authorized users of this account	November 1, 2023, 10:43:26 (UTC+01:00)

3.Create on your local machine a text file named lab.csv with the following content:

```
CustomerID,First Name,Last Name,Join Date,Street Address,City,State,Phone
001,Alejandro,Rosalez,12/12/2013,123 Main St.,Baltimore,MD,765-234-2349
002,Jane,Doe,10/5/2014,456 State St.,Seattle,WA,415-889-4932
003,John,Stiles,9/20/20016,1980 8th St.,Brooklyn,NY,917-123-9308
004,Li,Juan,6/29/2011,1323 22nd Ave.,Albany,NY,917-332-3432
```

```
pablo@Macbook-Pro-M1 Labo3 % cat lab.csv
 CustomerID,First Name,Last Name,Join Date,Street Address,City,State,Phone
 001,Alejandro,Rosalez,12/12/2013,123 Main St.,Baltimore,MD,765-234-2349
 002,Jane,Doe,10/5/2014,456 State St.,Seattle,WA,415-889-4932
 003,John,Stiles,9/20/20016,1980 8th St.,Brooklyn,NY,917-123-9308
 004,Li,Juan,6/29/2011,1323 22nd Ave.,Albany,NY,917-332-3432
```

Upload the file to the bucket.



Upload succeeded
View details below.

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

Summary

Destination	Succeeded	Failed
s3://ist-grd-urizar-valzino-bucket	✓ 1 file, 337.0 B (100.00%)	0 files, 0 B (0%)

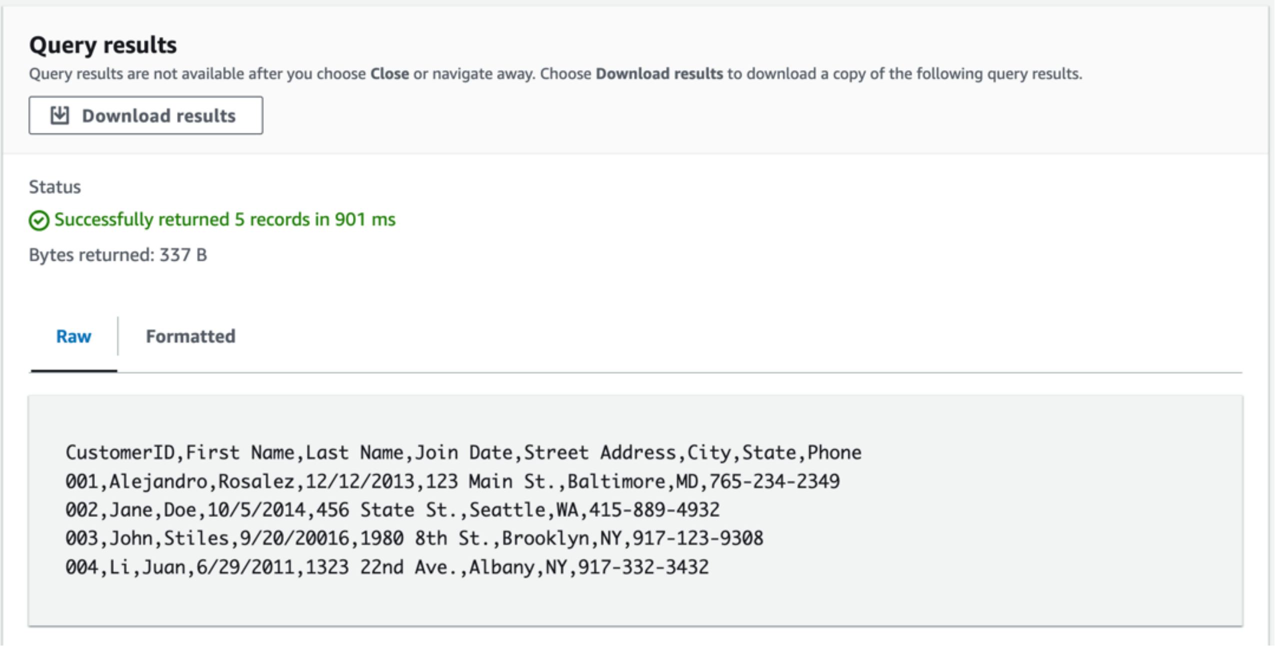
Files and folders | Configuration

Files and folders (1 Total, 337.0 B)

Find by name

Name	Folder	Type	Size	Status	Error
lab.csv	-	text/csv	337.0 B	✓ Succeeded	-

4.To verify that it was uploaded successfully, we will do an SQL query on the file, and at the same time learn that S3 has a buil-in SQL engine that works directly on CSV files.



Query results

Query results are not available after you choose **Close** or navigate away. Choose **Download results** to download a copy of the following query results.

Download results

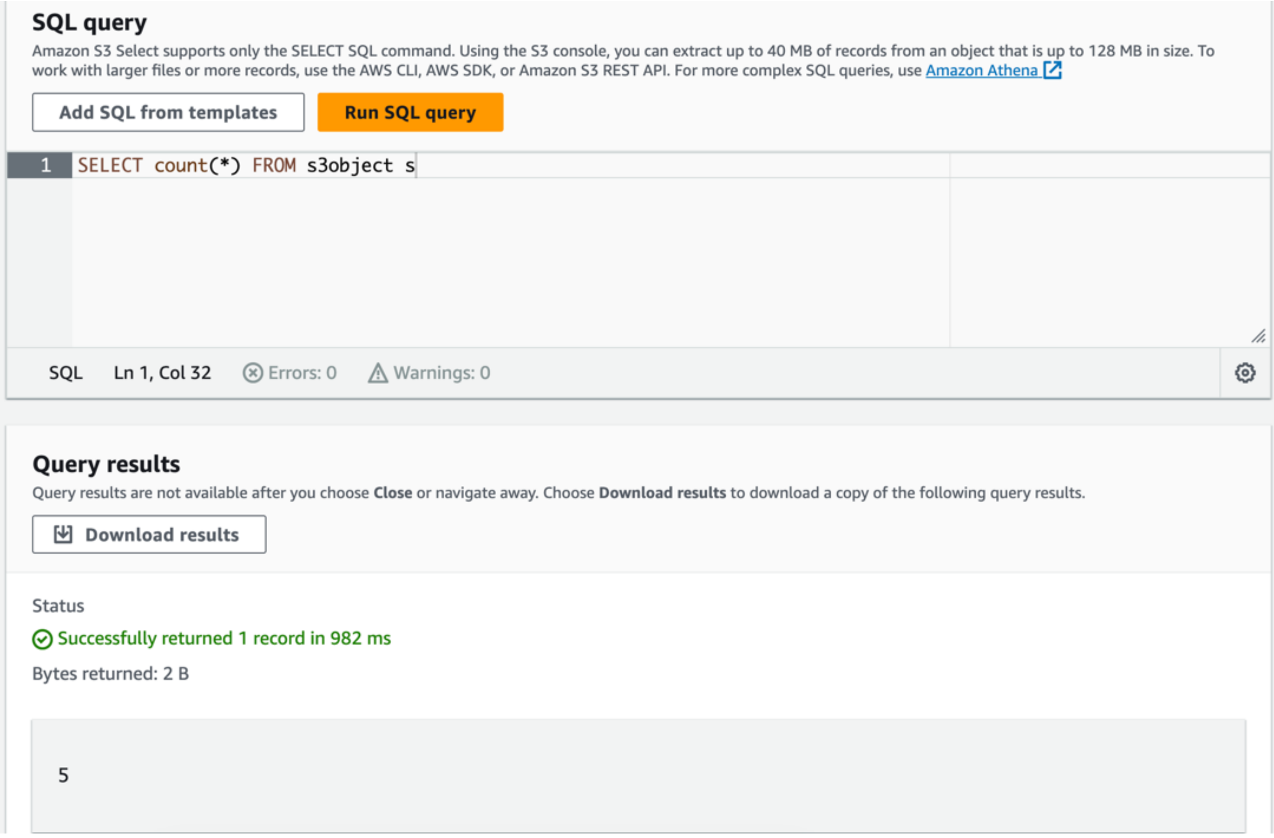
Status

✓ Successfully returned 5 records in 901 ms

Bytes returned: 337 B

Raw | Formatted

```
CustomerID,First Name,Last Name,Join Date,Street Address,City,State,Phone
001,Alejandro,Rosalez,12/12/2013,123 Main St.,Baltimore,MD,765-234-2349
002,Jane,Doe,10/5/2014,456 State St.,Seattle,WA,415-889-4932
003,John,Stiles,9/20/20016,1980 8th St.,Brooklyn,NY,917-123-9308
004,Li,Juan,6/29/2011,1323 22nd Ave.,Albany,NY,917-332-3432
```



SQL query

Amazon S3 Select supports only the SELECT SQL command. Using the S3 console, you can extract up to 40 MB of records from an object that is up to 128 MB in size. To work with larger files or more records, use the AWS CLI, AWS SDK, or Amazon S3 REST API. For more complex SQL queries, use Amazon Athena

```
SELECT count(*) FROM s3object s
```

Query results

Query results are not available after you choose **Close** or navigate away. Choose **Download results** to download a copy of the following query results.

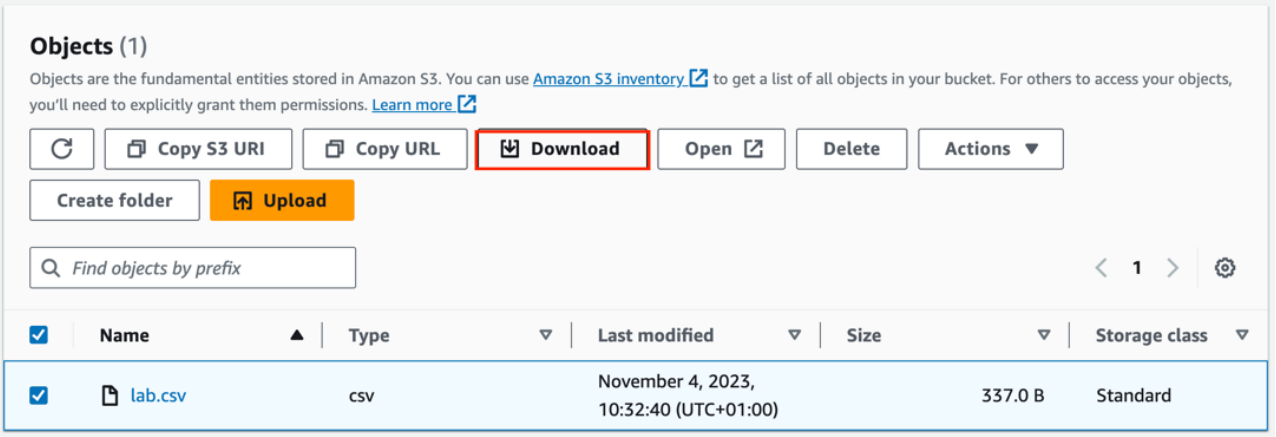
Status

Successfully returned 1 record in 982 ms

Bytes returned: 2 B

```
5
```

5.Use the web console to download the file to your local machine.



Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

☑	Name	Type	Last modified	Size	Storage class
☑	lab.csv	csv	November 4, 2023, 10:32:40 (UTC+01:00)	337.0 B	Standard

TASK 2: USE THE AWS COMMAND-LINE INTERFACE TO MANAGE BUCKETS AND OBJECTS

1.Install the AWS CLI by following the instructions in Installing or updating the latest version of the AWS CLI. Note for macOS users: The AWS CLI is available in homebrew as awscli.

```
pablo@Macbook-Pro-M1 Labo4 % brew install awscli
```

2.Set up your security credentials, default Region and default output format by following the instructions in Quick setup.

```
pablo@Macbook-Pro-M1 Labo4 % aws configure
AWS Access Key ID [****************HZEW]: AKIAQ3ED337FZTQ5HZEW
AWS Secret Access Key [****************aclg]: 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
Default region name [us-east-1]:
Default output format [json]:
```

3.Verify that the tool is configured correctly.

List all available regions:

```
pablo@Macbook-Pro-M1 Labos_IST % aws ec2 describe-regions
{
    "Regions": [
        {
            "Endpoint": "ec2.ap-south-1.amazonaws.com",
            "RegionName": "ap-south-1",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.eu-north-1.amazonaws.com",
            "RegionName": "eu-north-1",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.eu-west-3.amazonaws.com",
            "RegionName": "eu-west-3",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.eu-west-2.amazonaws.com",
            "RegionName": "eu-west-2",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.eu-west-1.amazonaws.com",
            "RegionName": "eu-west-1",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.ap-northeast-3.amazonaws.com",
            "RegionName": "ap-northeast-3",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.ap-northeast-2.amazonaws.com",
            "RegionName": "ap-northeast-2",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.ap-northeast-1.amazonaws.com",
            "RegionName": "ap-northeast-1",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.ca-central-1.amazonaws.com",
            "RegionName": "ca-central-1",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.sa-east-1.amazonaws.com",
            "RegionName": "sa-east-1",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.ap-southeast-1.amazonaws.com",
            "RegionName": "ap-southeast-1",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.ap-southeast-2.amazonaws.com",
            "RegionName": "ap-southeast-2",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.eu-central-1.amazonaws.com",
            "RegionName": "eu-central-1",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.us-east-1.amazonaws.com",
            "RegionName": "us-east-1",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.us-east-2.amazonaws.com",
            "RegionName": "us-east-2",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.us-west-1.amazonaws.com",
            "RegionName": "us-west-1",
            "OptInStatus": "opt-in-not-required"
        },
        {
            "Endpoint": "ec2.us-west-2.amazonaws.com",
            "RegionName": "us-west-2",
            "OptInStatus": "opt-in-not-required"
        }
    ]
}
```

Display account attributes:

```
pablo@Macbook-Pro-M1 Labos_IST %  aws ec2 describe-account-attributes
{
    "AccountAttributes": [
        {
            "AttributeName": "supported-platforms",
            "AttributeValues": [
                {
                    "AttributeValue": "VPC"
                }
            ]
        },
        {
            "AttributeName": "vpc-max-security-groups-per-interface",
            "AttributeValues": [
                {
                    "AttributeValue": "5"
                }
            ]
        },
        {
            "AttributeName": "max-elastic-ips",
            "AttributeValues": [
                {
                    "AttributeValue": "5"
                }
            ]
        },
        {
            "AttributeName": "max-instances",
            "AttributeValues": [
                {
                    "AttributeValue": "20"
                }
            ]
        },
        {
            "AttributeName": "vpc-max-elastic-ips",
            "AttributeValues": [
                {
                    "AttributeValue": "20"
                }
            ]
        },
        {
            "AttributeName": "default-vpc",
            "AttributeValues": [
                {
                    "AttributeValue": "vpc-04d7055d4dd0b93d2"
                }
            ]
        }
    ]
}
```

Display available EC2 Instance types (a filter was applied to only display t2, m5, c5 and t3 instances):

```
pablo@Macbook-Pro-M1 Labos_IST % aws ec2 describe-instance-type-offerings > instance-t

pablo@Macbook-Pro-M1 Labos_IST % cat instance-types.json | jq -c '.InstanceTypeOfferin
{"InstanceType":"m5ad.12xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5.18xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t2.micro","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5.9xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5ad.2xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5d.18xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t2.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t3.2xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5n.24xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5.24xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5a.8xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5n.9xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5dn.large","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5.2xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5a.16xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5n.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5ad.large","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5a.24xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t3a.2xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5.8xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5d.12xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5ad.24xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5ad.2xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t2.medium","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t3a.large","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5ad.4xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5n.2xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5.24xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5n.2xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5a.2xlarge","LocationType":"region","Location":"us-east-1"}
```

```
{"InstanceType":"m5dn.2xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5zn.2xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5n.18xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5.4xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5a.12xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5ad.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t3.small","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t2.large","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5d.large","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5n.16xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5a.16xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5n.large","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5n.4xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5.4xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5zn.3xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t2.nano","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5zn.12xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5.2xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t3.medium","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5d.large","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5dn.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t3a.micro","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t2.small","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t3a.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5d.metal","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5a.large","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5d.2xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t3.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5ad.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5ad.8xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5.large","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5zn.large","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5.large","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5ad.8xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5d.4xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5ad.large","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5n.8xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5d.9xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5dn.12xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t3a.medium","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5.16xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5d.4xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5d.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5n.12xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5d.2xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5d.12xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5a.4xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5a.2xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5n.metal","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5d.metal","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5n.metal","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5zn.6xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5a.8xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5a.12xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5d.24xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5zn.metal","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5.12xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t3a.small","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5.12xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5.metal","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5zn.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5a.24xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5ad.16xlarge","LocationType":"region","Location":"us-east-1"}
```

```
{"InstanceType":"m5n.4xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5dn.24xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5a.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5.metal","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5d.8xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t3.large","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5a.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5d.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5a.large","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5dn.8xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5d.16xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t3.micro","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5dn.metal","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t3.nano","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5ad.24xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5ad.4xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t2.2xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5ad.12xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5dn.4xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5dn.16xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5n.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"m5ad.16xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5a.4xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5.xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"t3a.nano","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5d.24xlarge","LocationType":"region","Location":"us-east-1"}
{"InstanceType":"c5n.large","LocationType":"region","Location":"us-east-1"}
```

List all S3 buckets in the account:

```
pablo@Macbook-Pro-M1 Labo4 % aws s3 ls
2022-02-19 11:41:28 aws-cloudtrail-logs-058258612171-1871bacf
2022-10-31 10:17:21 bucket1-b83c83db-644f-4506-8152-783621c47980
2022-10-31 10:18:14 bucket2-cd834d37-7b45-475e-998a-63dc8a5a4020
2022-10-31 10:18:34 bucket3-cd668ce6-4839-4823-98df-be312db0038c
2023-06-01 17:23:55 custom-labels-console-us-east-1-6aa67c663d
2022-10-06 15:13:32 heigvd-ist
2022-11-24 10:37:02 heigvd-ist-gra
2022-11-10 13:53:32 heigvd-ist-grb
2022-11-24 08:45:31 heigvd-ist-lab5-grc
2023-10-20 13:04:51 heigvd-ist-rms
2023-11-01 10:39:03 ist-gra-ernst-bucket
2023-11-01 11:00:46 ist-gra-ernst-bucketcli
2023-11-01 10:41:40 ist-grb-centeno-merk-bucket
2023-11-01 11:06:04 ist-grb-centeno-merk-bucket2
2023-11-01 11:03:08 ist-grc-rausis-bucket
2023-11-01 10:29:56 ist-grc-reymermet-bucket
2023-11-01 10:43:26 ist-grd-urizar-valzino-bucket
2023-11-01 10:45:27 ist-gre-amos-rosat-bucket
2023-11-01 11:01:21 ist-gre-amos-rosat-bucket-cmd
2023-11-01 10:39:39 ist-grf-conti-bucket
2023-11-01 10:43:35 ist-grf-gogniat-bucket1
2023-11-01 10:40:00 ist-gri-decoppet-bucket1
2023-11-01 10:41:34 ist-gri-lopesgouveia-bucket

pablo@Macbook-Pro-M1 Labo3 % aws s3 ls | grep urizar
2023-11-01 10:43:26 ist-grd-urizar-valzino-bucket
```

4.Manipulate buckets and objects. Use the documentation Using high-level (S3) commands with the AWS CLI as a reference.

Create a new bucket:

```
pablo@Macbook-Pro-M1 Labo4 % aws s3 mb s3://ist-grd-urizar-valzino-bucket-cli --region
make_bucket: ist-grd-urizar-valzino-bucket-cli
```

Upload an object:

```
pablo@Macbook-Pro-M1 Labo4 % aws s3 cp /Users/pablo/Documents/BA5/IST/Labos_IST/Labo4/
upload: ./lab.csv to s3://ist-grd-urizar-valzino-bucket-cli/lab.csv
```

List the objects in the bucket:

```
pablo@Macbook-Pro-M1 Labo4 % aws s3 ls s3://ist-grd-urizar-valzino-bucket-cli/
2023-11-01 11:19:24        337 lab.csv
```

Copy the object:

```
pablo@Macbook-Pro-M1 Labo4 % aws s3 cp s3://ist-grd-urizar-valzino-bucket-cli/lab.csv
copy: s3://ist-grd-urizar-valzino-bucket-cli/lab.csv to s3://ist-grd-urizar-valzino-bu
```

Verify the copy:

```
pablo@Macbook-Pro-M1 Labo4 % aws s3 ls s3://ist-grd-urizar-valzino-bucket-cli/
2023-11-01 11:20:52        337 copied-lab.csv
2023-11-01 11:19:24        337 lab.csv
```

Delete the copied object:

```
pablo@Macbook-Pro-M1 Labo4 % aws s3 rm s3://ist-grd-urizar-valzino-bucket-cli/copied-l
delete: s3://ist-grd-urizar-valzino-bucket-cli/copied-lab.csv
```

5.S3 folders behave differently from file system folders. In fact a folder in S3 is a 0-byte object whose name ends with a slash (/). Read the introduction on the page Organizing objects in the Amazon S3 console using folders.

Create a folder (using the Management Console). Verification that the folder was created:

```
pablo@Macbook-Pro-M1 Labo4 % aws s3 ls s3://ist-grd-urizar-valzino-bucket-cli/
                           PRE Folder/
2023-11-01 11:19:24        337 lab.csv
```

Make a copy of the object and move the copy into the folder:

```
pablo@Macbook-Pro-M1 Labo4 % aws s3 cp s3://ist-grd-urizar-valzino-bucket-cli/lab.csv
copy: s3://ist-grd-urizar-valzino-bucket-cli/lab.csv to s3://ist-grd-urizar-valzino-bu

pablo@Macbook-Pro-M1 Labo4 % aws s3 mv s3://ist-grd-urizar-valzino-bucket-cli/lab_copy
move: s3://ist-grd-urizar-valzino-bucket-cli/lab_copy.csv to s3://ist-grd-urizar-valzi
```

Verification:

```
pablo@Macbook-Pro-M1 Labo4 % aws s3 ls s3://ist-grd-urizar-valzino-bucket-cli/
                           PRE Folder/
2023-11-01 11:19:24        337 lab.csv

pablo@Macbook-Pro-M1 Labo4 % aws s3 ls s3://ist-grd-urizar-valzino-bucket-cli/Folder/
2023-11-01 11:24:44          0
2023-11-01 11:27:52        337 lab_copy.csv
```

What happens if you move an object to a folder that does not exist?

```
pablo@Macbook-Pro-M1 Labo4 % aws s3 mv s3://ist-grd-urizar-valzino-bucket-cli/lab.csv
move: s3://ist-grd-urizar-valzino-bucket-cli/lab.csv to s3://ist-grd-urizar-valzino-bu
```

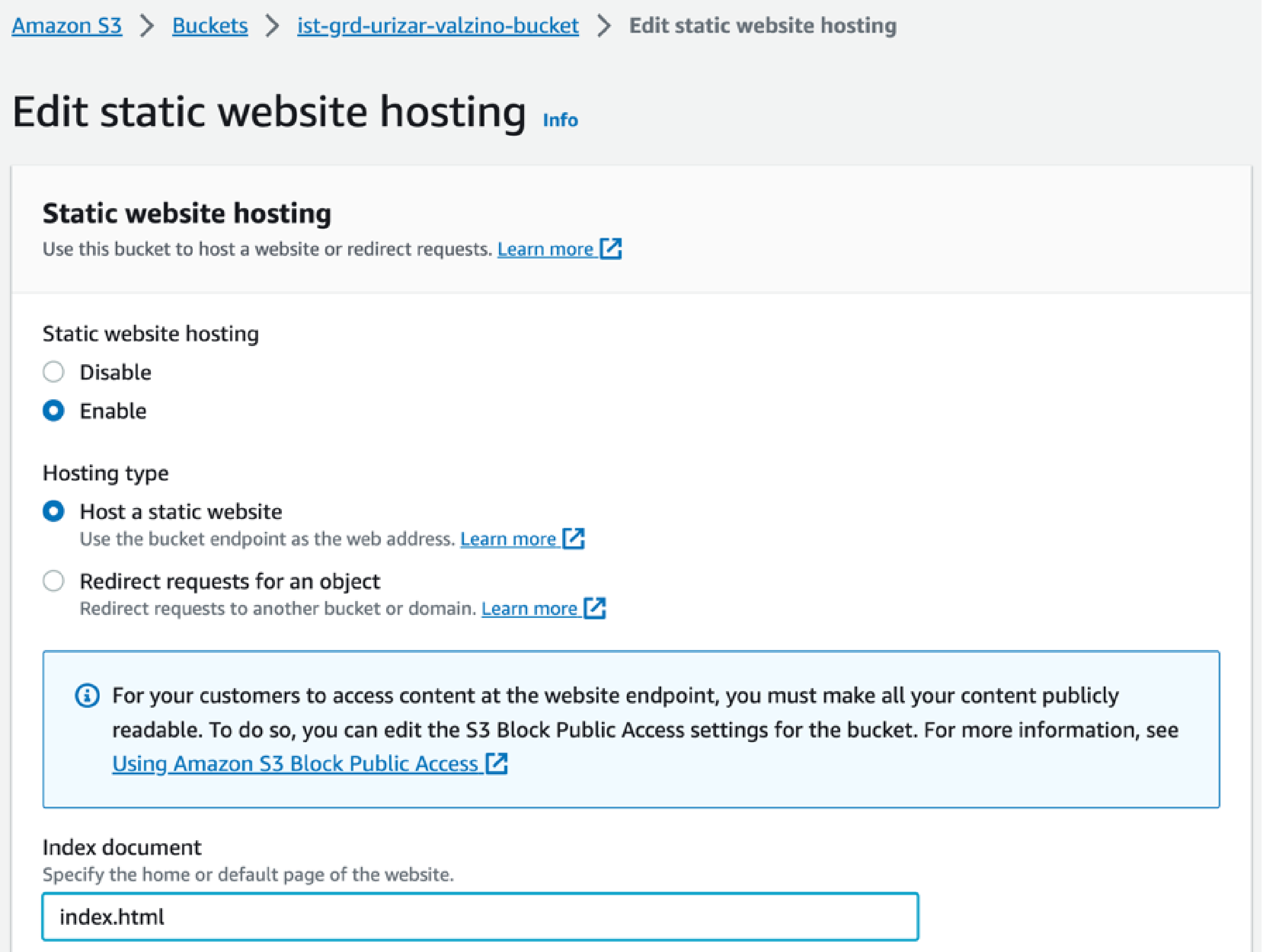
The destination folder that did not exist before was created as specified in the destination path:

```
pablo@Macbook-Pro-M1 Labo4 % aws s3 ls s3://ist-grd-urizar-valzino-bucket-cli/
                           PRE Folder/
                           PRE Folder2/
pablo@Macbook-Pro-M1 Labo4 % aws s3 ls s3://ist-grd-urizar-valzino-bucket-cli/Folder2/
2023-11-01 11:31:01        337 lab.csv
```

TASK 3: CREATE A STATIC WEB SITE

1.Follow the instructions of the Tutorial: Configuring a static website on Amazon S3 to create a new bucket for a static website.

Enable static website hosting:



Amazon S3 > Buckets > ist-grd-urizar-valzino-bucket > Edit static website hosting

Edit static website hosting Info

Static website hosting

Use this bucket to host a website or redirect requests. Learn more

Static website hosting

○ Disable

● Enable

Hosting type

● Host a static website

Use the bucket endpoint as the web address. Learn more

○ Redirect requests for an object

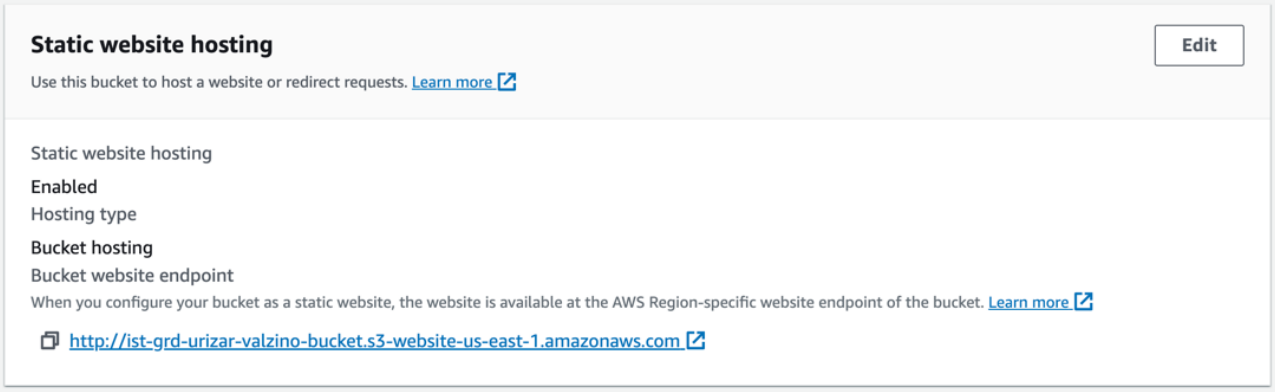
Redirect requests to another bucket or domain. Learn more

For your customers to access content at the website endpoint, you must make all your content publicly readable. To do so, you can edit the S3 Block Public Access settings for the bucket. For more information, see Using Amazon S3 Block Public Access

Index document

Specify the home or default page of the website.

index.html



Static website hosting

Use this bucket to host a website or redirect requests. Learn more

Static website hosting

Enabled

Hosting type

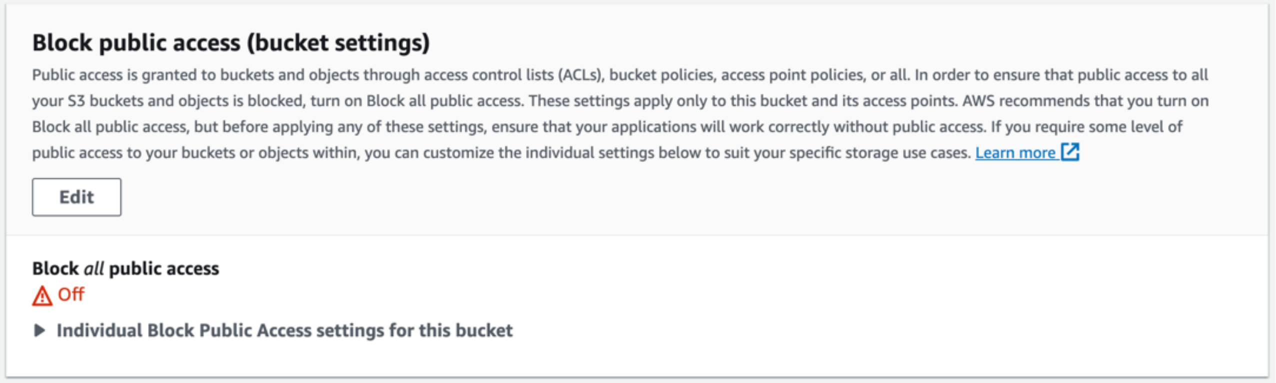
Bucket hosting

Bucket website endpoint

When you configure your bucket as a static website, the website is available at the AWS Region-specific website endpoint of the bucket. Learn more

http://ist-grd-urizar-valzino-bucket.s3-website-us-east-1.amazonaws.com

Edit Block Public Access settings:



Block public access (bucket settings)

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 all your S3 buckets and 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 your buckets or objects within, you can customize the individual settings below to suit your specific storage use cases. Learn more

Block *all* public access

⚠ Off

▶ **Individual Block Public Access settings for this bucket**

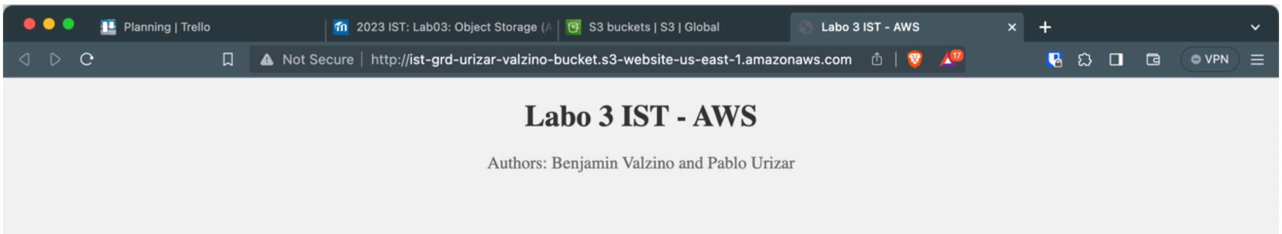
Add a bucket policy that makes your bucket content publicly available

```
{
        "Version": "2012-10-17",
        "Statement": [
                {
                        "Sid": "PublicReadGetObject",
                        "Effect": "Allow",
                        "Principal": "*",
                        "Action": "s3:GetObject",
                        "Resource": "arn:aws:s3:::ist-grd-urizar-valzino-bucket/*"
                }
        ]
}
```

Configure an index document (index.html):

```
<!DOCTYPE html>
<html>
<head>
    <title>Labo 3 IST - AWS</title>
</head>
<body>
    <h1>Labo 3 IST - AWS</h1>
    <p>Authors: Benjamin Valzino and Pablo Urizar</p>
</body>
</html>
```

2.On which URL is your new website reachable? http://ist-grd-urizar-valzino-bucket.s3-website-us-east-1.amazonaws.com



Planning | Trello
2023 IST: Lab03: Object Storage (A
S3 buckets | S3 | Global
Labo 3 IST - AWS
Not Secure
http://ist-grd-urizar-valzino-bucket.s3-website-us-east-1.amazonaws.com
VPN

Labo 3 IST - AWS

Authors: Benjamin Valzino and Pablo Urizar

TASK 4: EXPLORE A PUBLIC BUCKET WITH A LARGE DATASET

1.The data location of the Common Crawl datasets is described on the page Get Started.

- When was the latest crawl? `September/October 2023`
- What is the bucket name? `commoncrawl`
- Under which prefix is the latest crawl stored? `crawl-data/CC-MAIN-2023-40`

2.Log into the AWS S3 Management Console. Replace the browser URL with https://s3.console.aws.amazon.com/s3/buckets// Where you replace with the name of the bucket. You should see a bucket with objects and folders.

https://s3.console.aws.amazon.com/s3/buckets/commoncrawl/?region=us-east-1&tab=objects

3.Navigate to the root folder of the latest crawl. Click on the object index.html. Click the Open button to load it into your browser. What is the URL of this object? https://data.commoncrawl.org/crawl-data/CC-MAIN-2023-40/index.html

4.Explore a bit the objects and folders. What are WARC, WAT and WET files (look at the Get Started guide)?

- "WARC files which store the raw crawl data". WARC files are like web time capsules that save entire websites, including text, images, and more, so we can look back in time and see how they used to be.
- "WAT files which store computed metadata for the data stored in the WARC". WAT files are like labels on the time capsules. They tell us when each web snapshot was taken, what's inside, and where it came from.
- "WET files which store extracted plaintext from the data stored in the WARC". WET files are like books made from the time capsules. They take out the words from websites, so we can easily read and search them.

What is the typical size of a WARC file (ballpark)? "The WARC format allows for more efficient storage and processing of Common Crawl's free multi-billion page web archives, which can be **hundreds of terabytes** in size."

Why is it not sufficient to just store the WARC, WAT and WET files in the bucket? What other type of file is needed? If we only have the WARC, WAT and WET files, the content may not be easily accessible or readable by users without additional tools. These files provide the raw data and metadata but accessing and understanding its content will require other resources such as indexes, checksums, software tools, and documentation.

What storage classes have the Common Crawl developers chosen to store the data? They use Standard Storage for data that is frequently accessed, Intelligent-Tiering to automatically move objects between two access tiers (frequent and infrequent) and Glacier for less frequently accessed data.