

Laboratory Practice Report

Practice 2

September 13, 2023

Computer Systems Engineering

Cloud Architecture

Prof. M.S. Rodolfo Luthe Ríos

Marco Ricardo Cordero Hernández

is727272@iteso.mx



ITESO

Universidad Jesuita
de Guadalajara

Abstract

Following the research and practical line of cloud services discovery, the current investigative and implementation path aims to explore remote storage technology for both block and object handling.

Through the path of development that can be found within this document, description and in-depth analysis of previously mentioned services will be thoroughly explained and looked upon to implement emergent solutions for classic and recurrent problems involving means of digital storage for unspecified and general purposes.

The exploration of the current topics turns out relevant as these concepts make just a small part of what can become a life-saving service or retail product that interfaces with any type of remote service technologies. In fact, as of today, forensics, food, medical, chemical, pharmaceutical, among many other industries are using these operations means to achieve greater goals.

Finally, the main goal by conducting these activities is not to start thinking in the next revolutionary file system that would ultimately substitute current architectures, but to be able to implement, manipulate and communicate with a support tool that, as many other cloud services, would save infrastructure costs and would facilitate physical hardware substitution (but not necessarily replacement).

State of the Art

Over the years and since the last century, file managing has been one of the cornerstones of modern information systems. In any kind of development aimed for any kind of industry, record creation, modification, saving and access is a critical part of their operations, as information calculated or produced in any sort of way needs to be kept for later usage. The most primitive way of achieving this (or at least the step before transitioning into digital systems) was physical folders and organizers with related files inside them, mashed inside cabinets containing thousands of information bits. Nowadays, these file systems are still used as legacy support means for rigid scheme sectors of society such as law or any other governmental activity.

If it was to be analyzed, physical storage has the benefit of fast access most of the time and not needing internet to retrieve desired data, and, evidently, owning the information stored, also accessible without any kind of fees. When abstracting this concept out of files into general information, this method of storage was the way to go for a long time, but, as many other solutions now overtaken by cloud services, physical infrastructure comes with a high cost of physical allocation and physical maintenance.

Most recent storage technology, such as SATA and NVMe for Solid-state drives [SSD] [1] are very expensive additions when talking about either horizontal or vertical scaling [2], and this factor was the decisive aspect of preferring remotely hosted storage, or in simple terms, *cloud storage*.

When talking about the possibility of going with cloud based infrastructure for new developments or migration planning, inevitably storage will be involved in a greater or lesser extent, nonetheless, one key factor among several others [3] to take into account is its *security*, as not only simple text files could potentially be stored remotely, but also corporate documents with a whole different

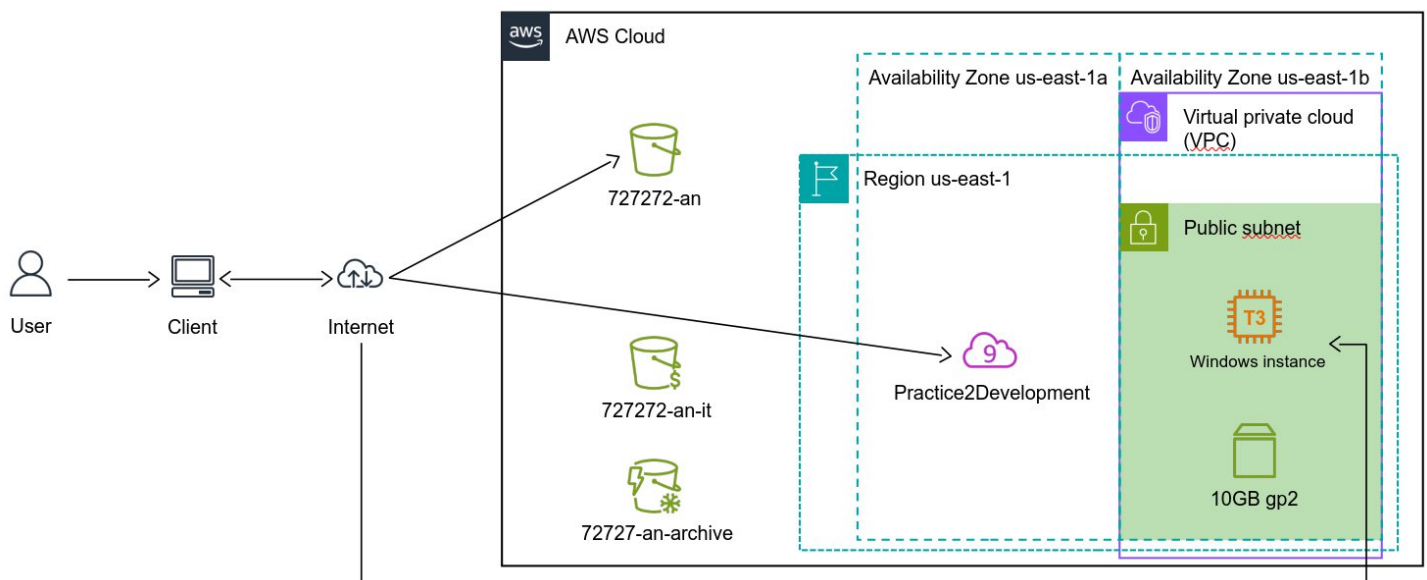
extension of their own, containing sensible information regarding to different business related activities. Among other criteria, this aspect has to be considered thoroughly, as a data leakage would be catastrophic, damaging company reputation and potentially involving it in several severe legal disputes, such as the case of the “Instituto Nacional Electoral” [INE] voters data leakage back in 2016 [4].

Besides security, which is already a whole domain on its own and not to be taken slightly, one should ponder about other involvements, such as: data storage location for fast information retrieval (after all, cloud storage is simply taking physical storage devices and placing them around the world); performance, to determine the bitrate regardless of physical location (but without overlooking it); integration with development cycle (to ensure ease of data access through applications); and cost (because it’s still a service) [5]. All these components aid in decision making for the best provider evaluation, and also, translate to primitive storage devices, where fast data access could be seen as someone turning around their desk to grab with their hands an object stored in a cabinet, performance to how far of them was said object, and the convenience of having these cabinets near by comparable to performance. As for the costs, well, this could be seen as buying more paper, more cabinets, more offices, etc.

The purpose of the current work is to demonstrate the usage of a cloud service solely dedicated to storage: S3, an Amazon Web Services’ [AWS] product aimed for object storage and data retrieval for anywhere in the world [6]. Also, a previously created EC2 instance [7] it’s going to be modified in order to extend its storage capacity.

Diagram

The following architecture it’s proposed as a graphic for solution for the stated goals.



Practice Development

Context inside AWS Console:

- Region: US East (North Virginia, us-east-1)

Existent volume increasing on EC2 instances

First, storage capacity increase will be demonstrated for a previously created EC2 instance with Windows OS. Availability zone has to be kept on sight, as it becomes relevant in the next steps.

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone
Win admin	i-0459a87fd900326c3	Running	t3.small	2/2 checks passed	0 in alarm	us-east-1b

In its current state, the storage capacity is of 30 GB. This information can be found within the “Elastic Block Store” section inside the EC2 Service panel, in “Volumes” option.

Volume ID: vol-0778232b4c9ede69d

Details	Status checks	Monitoring	Tags
Volume ID vol-0778232b4c9ede69d AWS Compute Optimizer finding This user is not authorized to call AWS Compute Optimizer. Retry Encryption Not encrypted Fast snapshot restored No Multi-Attach enabled No	Size 30 GiB Volume state In-use KMS key ID - Snapshot snap-0656dce7eb222831f Attached Instances i-0459a87fd900326c3 (Win admin): /dev/sda1 (attached)	Type gp2 IOPS 100 KMS key alias - Availability Zone us-east-1b Outposts ARN -	

The goal is to create a new 10GB volume in the same availability zone than the instance, this is for data access velocity and because the next step won't work if it isn't like this. The specifications of the new volume are the following:

- Volume type: General Purpose SSD (gp2)
- Size: 10GB
- Availability zone: us-east-1b
- Encryption: none (omitted for this exercise, but relevant on productive environments)

After the volume has been created *and its volume state is “Available”*, it should be listed along the previously available volumes. The next step is to attach it to the instance, which can be running through this process.

Volume ID	Type	Size	IOPS	KMS Key ID	Availability Zone	Created At
vol-01caa650e78f9fcfd	gp2	10 GiB	100	-	us-east-1b	2023/09/13 08:15 GMT-6

Attach volume [Info](#)

Attach a volume to an instance to use it as you would a regular physical hard disk drive.

Basic details

Volume ID
vol-01caa650e78f9cf4

Availability Zone
us-east-1b

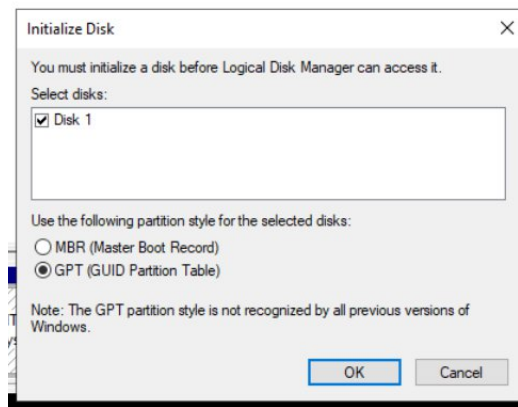
Instance [Info](#)
i-0459a87fd900326c3

Device name [Info](#)
xvdf

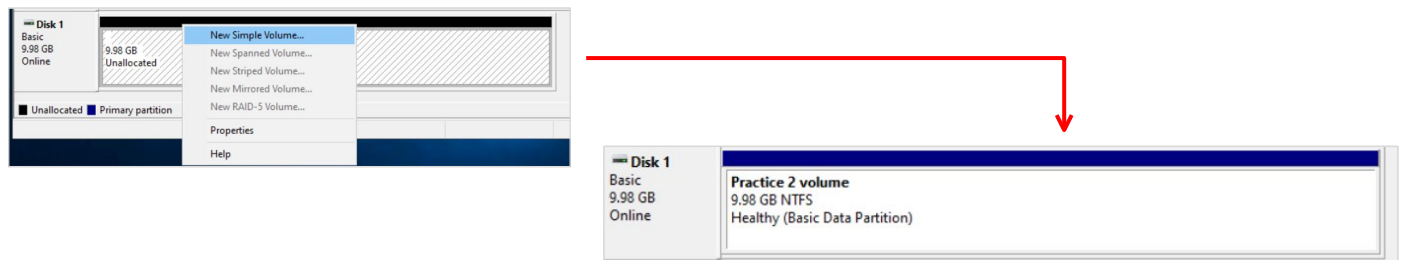
Recommended device names for Windows: /dev/sda1 for root volume, xvdf(p) for data volumes.

As the console suggest, some name conventions should be followed when differentiating between root volumes (for OS and other system files) and data volumes (general purpose storage).

After the previous step, the volume it's not yet ready to be used. Inside the instance, accessed through RDP, a mounting process has to be conducted. This can be done through different methods, however, the "Disk Management utility" tool is the easiest and more direct way to do so. When accessing the tool, the first prompt indicates that the newly found volume has to be initialized. Corresponding to the way it was created, GPT option should be selected.



The next step is also automated, as the same tool provides a wizard facilitator for new volume creation. After completion, the volume should be listed as online.



The next exercise consists in modifying the previously created volume, adding 1GB to its capacity and allocating the new space in the same tool used before. To do this, first, the volume capacity has to be increased.

Name	Volume ID	Type	Size	IOPS	Throughput	Snapshot
-	vol-01caa650e78f9fcf4	gp2	10 GiB	100		
-	vol-0778232b4c9ede69d	gp2	30 GiB	100		
-	vol-0f21d9a561346c3a6	gp2	8 GiB	100		
-	vol-0bcf733b12705eaba	gp3	8 GiB	3000		

Create volume

Modify volume

Create snapshot

Create snapshot lifecycle policy

Delete volume

Attach volume

Detach volume

Force detach volume

Manage auto-enabled I/O

Manage tags

Fault injection

Size (GiB) Info

11

Min: 1 GiB, Max: 16384 GiB. The value must be an integer.

Requested volume modification for volume [vol-01caa650e78f9fcf4](#).
The volume is being modified.

This process might take a few a minutes. Another thing that might be overlooked it's the fact that the new space allocation will be charged accordingly.

Inside the instance which the volume has been attached, new space has to be claimed to extend the corresponding disk.

The screenshot shows the Windows Disk Management console. On the left, a list of disks is shown: Disk 0 (Basic, 30.00 GB, Online) and Disk 1 (Basic, 10.98 GB, Online). The main area displays the details for Disk 1, which contains 'Practice 2 volume (D:)' (9.98 GB NTFS, Healthy (Basic Data Partition)) and a '1.00 GB Unallocated' space. A right-click context menu is open over the 'Practice 2 volume (D:)', with 'Extend Volume...' selected. A red arrow points from the 'Unallocated' space to the 'Extend Volume...' option in the menu.

Bucket creation with S3

Now, another way of storage it's going to be demonstrated, specifically, object storage. An S3 bucket should be made for this.

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](#)

The specifications of this new bucket are the following:

- Name: 727272-an
- 30-day deletion rule (inside lifecycle)

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

Name	AWS Region	Access	Creation date
<input type="radio"/> 727272-an	US East (N. Virginia) us-east-1	Bucket and objects not public	September 13, 2023, 09:22:28 (UTC-06:00)

Lifecycle rules (1)
Use lifecycle rules to define actions you want Amazon S3 to take during an object's lifetime such as transitioning objects to another storage class, archiving them, or deleting them after a specified period of time. [Learn more](#)

Lifecycle rule name	Status	Scope	Current version actions	Noncurrent versions actions	Expired object delete markers	Incomplete multipart uploads
<input type="radio"/> 30-rule file deletion	Enabled	Entire bucket	-	Permanently delete	Permanently delete	Permanently delete

Then, another bucket should be created for comparison reasons. The specifications are:

- Name: 727272-an-archive
- 30-day glacier transition rule (inside lifecycle)

Buckets (2) [Info](#)
Buckets are containers for data stored in S3. [Learn more](#)

Name	AWS Region	Access	Creation date
<input type="radio"/> 727272-an-archive	US East (N. Virginia) us-east-1	Bucket and objects not public	September 13, 2023, 09:50:31 (UTC-06:00)
<input type="radio"/> 727272-an	US East (N. Virginia) us-east-1	Bucket and objects not public	September 13, 2023, 09:22:28 (UTC-06:00)

Lifecycle rules (1)
Use lifecycle rules to define actions you want Amazon S3 to take during an object's lifetime such as transitioning objects to another storage class, archiving them, or deleting them after a specified period of time. [Learn more](#)

Lifecycle rule name	Status	Scope	Current version actions	Noncurrent versions actions	Expired object delete markers	Incomplete multipart uploads
<input type="radio"/> Move to Glacier after 30 days	Enabled	Entire bucket	Transition to Glacier Flexible Retrieval (formerly Glacier)	-	-	-

Before file uploading it's demonstrated, yet another bucket will be created with the following specifications:

- Name: 727272-an-it
- S3 Intelligent-Tiering instant transition (inside lifecycle)

Buckets (3) [Info](#)
Buckets are containers for data stored in S3. [Learn more](#)

Name	AWS Region	Access	Creation date
<input type="radio"/> 727272-an-it	US East (N. Virginia) us-east-1	Bucket and objects not public	September 13, 2023, 11:34:49 (UTC-06:00)
<input type="radio"/> 727272-an-archive	US East (N. Virginia) us-east-1	Bucket and objects not public	September 13, 2023, 09:50:31 (UTC-06:00)
<input type="radio"/> 727272-an	US East (N. Virginia) us-east-1	Bucket and objects not public	September 13, 2023, 09:22:28 (UTC-06:00)

Lifecycle rules (1)
Use lifecycle rules to define actions you want Amazon S3 to take during an object's lifetime such as transitioning objects to another storage class, archiving them, or deleting them after a specified period of time. [Learn more](#)

Lifecycle rule name	Status	Scope	Current version actions	Noncurrent versions actions	Expired object delete markers	Incomplete multipart uploads
<input type="radio"/> Transition to Intelligent-Tiering	Enabled	Entire bucket	Transition to Intelligent-Tiering	-	-	-

Running a summary to what's been created:

- 727272-an: bucket to delete objects inside it after 30 days of original uploading.
- 727272-an-archive: bucket to archive relatively old objects after a 30 day time period.
- 727272-an-it: bucket to categorize objects based on their accesses and allocating them accordingly.
 - S3 Intelligent-Tiering works by automatically storing objects in three access tiers, one optimized for frequent access, one lower-cost tier for infrequent access, and another very low-cost tier optimized for rarely accessed data [8]. This works by monitoring access patterns and automatically moving objects to their corresponding tiers (0/30/90 days pattern).


Static website hosting

Now, to demonstrate one of many S3 capabilities, an static web site entry point will be uploaded to the first bucket created. To achieve this, bucket access has to be modified to allow public access.

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

To view the static file, an index HTML file has to be uploaded. This can be done through the same console without the need of API's or other similar mechanics. After the upload has been completed (shouldn't take more than a few seconds), the bucket has to be modified to include ACL's; this option is automatically prompt for enabling. After that, access permissions for the HTML file can be modified to grant read permission *to anyone*.

Access control list (ACL)

Grant basic read/write permissions to AWS accounts. [Learn more](#)

Grantee	Objects	Object ACL
Object owner (your AWS account) Canonical ID:  044ca9922d44ee6aa6ac3 920fa9c1c24835545d7ac6779d 53a1dd43041badcfb	<input checked="" type="checkbox"/> Read	<input checked="" type="checkbox"/> Read <input checked="" type="checkbox"/> Write
Everyone (public access) Group:  http://acs.amazonaws.co m/groups/global/AllUsers	<input checked="" type="checkbox"/>  Read	<input type="checkbox"/> Read <input type="checkbox"/> Write

Once proper permissions has been granted for the desired static file, the bucket has to be modified again for static website hosting enabling. This can be found at the bottom of the properties section inside bucket configuration. The entry point should have same name as the uploaded file.

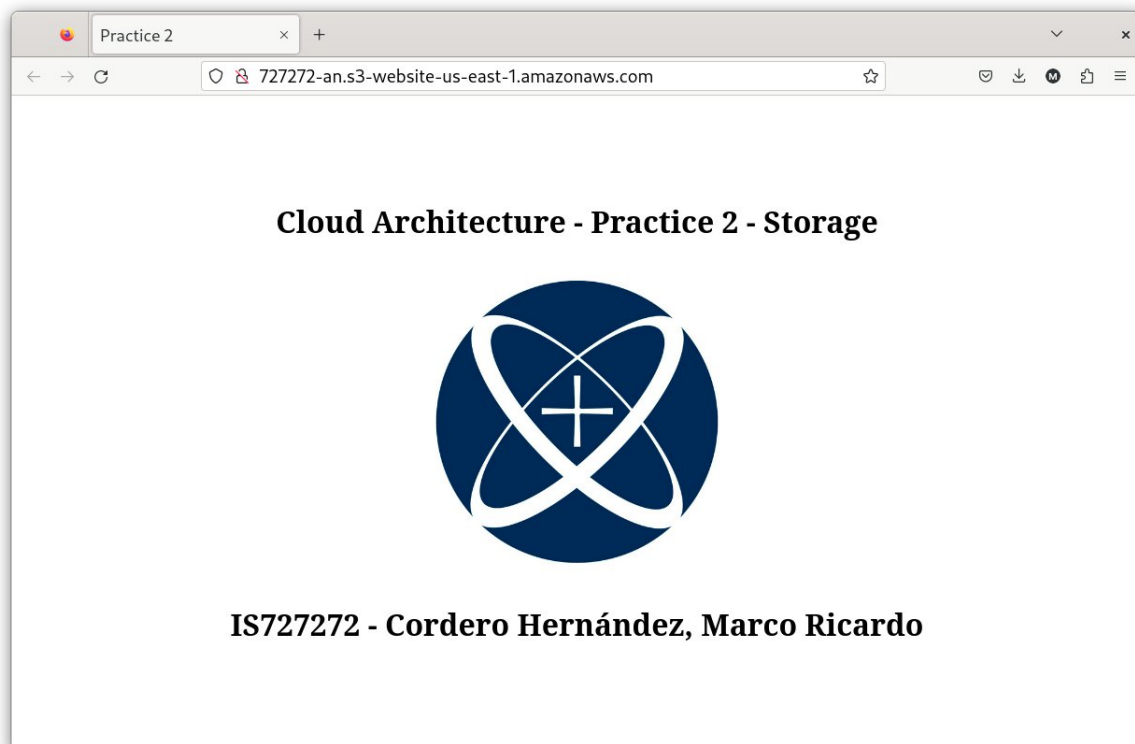
Once again, a final modification has to be made. Inside permissions and bucket policy, the next policy will be created. **Note:** inside “Resource” value, proper bucket name should be input instead of the found in this demonstration.

Bucket policy

The bucket policy, written in JSON, provides access to the objects stored in the bucket. Bucket policies don't apply to objects owned by other accounts. [Learn more](#)

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "PublicReadGetObject",
      "Effect": "Allow",
      "Principal": "*",
      "Action": "s3:GetObject",
      "Resource": "arn:aws:s3:::727272-an/*"
    }
  ]
}
```

Inside the same section where static website hosting was enabled, the access endpoint can be found. This time, the URL provided is <http://727272-an.s3-website-us-east-1.amazonaws.com>.



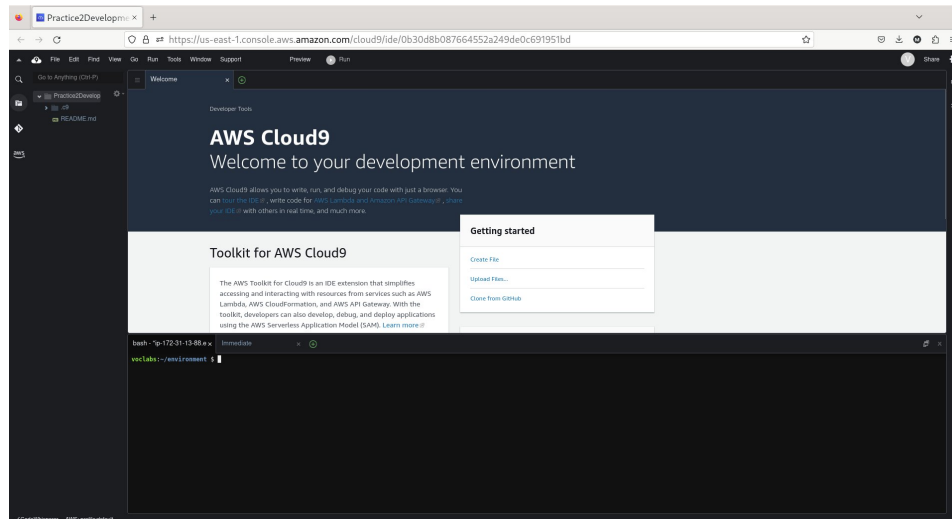
This endpoint works fine, but it exposes several information such as bucket name and availability zone. This can be solved through Domain Name Server [DNS] Services, such as AWS Route 53 [9]. Unfortunately, this can't be conducted because of permission lacking inside the account these procedures are made.

Cloud9

Cloud9 is another AWS service that provides an integrated development environment (IDE) via cloud [10]. In this section, another static website will be deployed to demonstrate its capabilities.

Name	Cloud9 IDE	Environment type	Connection	Permission	Owner ARN
Practice2Development	Open	EC2 Instance	Secure Shell (SSH)	Owner	arn:aws:sts:042979533702:assumed-role/voclabs/user1562810=its727272@iteso.mx

As any other service, a Cloud9 instance can be initialized through the exact same console. Once the instance is ready, the IDE interface should look like this.



Once inside it, a series of commands need to be executed to gain access to a brand new set of files for an static website. These commands, executed inside the instance, are:

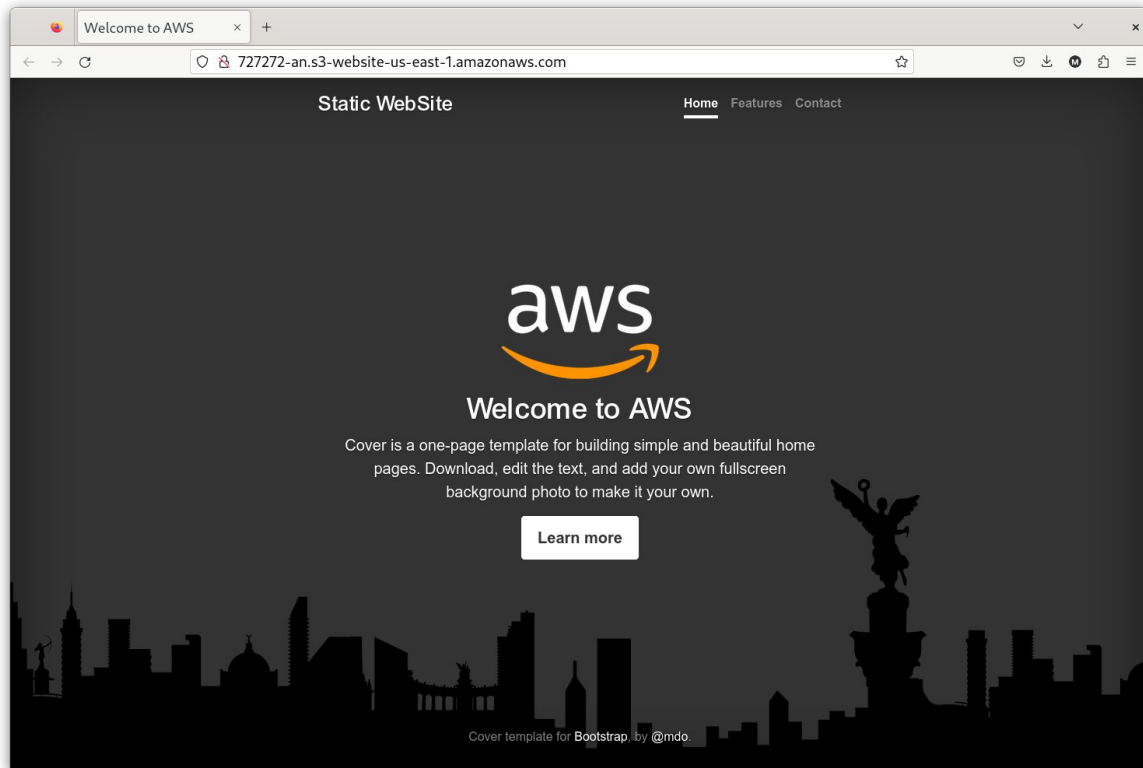
```
git clone https://github.com/aurbac/static-website.git
```

```
aws s3 cp static-website/ s3://727272-an/ --recursive --exclude ".git/*" --acl public-read
```

In simple terms, these actions will download files from a Github repository and then upload them to the previously created bucket (727272-an).

727272-an info						
Publicly accessible						
Objects	Properties	Permissions	Metrics	Management	Access Points	
Objects (6) 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						
<input type="button" value="Copy S3 URI"/> <input type="button" value="Copy URL"/> <input type="button" value="Download"/> <input type="button" value="Open"/> <input type="button" value="Delete"/> <input type="button" value="Actions"/> <input type="button" value="Create folder"/>						
<input type="button" value="Upload"/>						
<input type="text" value="Find objects by prefix"/>						
<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class	
<input type="checkbox"/>	.gitignore	gitignore	September 13, 2023, 13:33:50 (UTC-06:00)	93.0 B	Standard	
<input type="checkbox"/>	css/	Folder	-	-	-	
<input type="checkbox"/>	error.html	html	September 13, 2023, 13:33:51 (UTC-06:00)	1.3 KB	Standard	
<input type="checkbox"/>	images/	Folder	-	-	-	
<input type="checkbox"/>	index.html	html	September 13, 2023, 13:33:51 (UTC-06:00)	2.3 KB	Standard	
<input type="checkbox"/>	README.md	md	September 13, 2023, 13:33:51 (UTC-06:00)	204.0 B	Standard	

Static website hosting configurations doesn't need to be done again, as this was already made in previous sections. And, as such, the same URL it's still valid to visualize the results.



Evidently, the conducted actions overwrote the original index.html, so this could be either troublesome or helpful in a productive context.

Problems and Solutions

Although it's not a problem per se, bucket creation and utilization, along with computational instances such as Cloud9, will still generate expenses. At the end, they're still services, ignoring the educational purpose of these activities.

To avoid unexpected charges, all buckets should be terminated, along with Cloud9 instance. Extra disk space allocated won't be removed, but it'll still be charged for its usage. These actions will render the bucket files unavailable forever, and Cloud9 access no longer possible.

Note: Buckets containing any object in them need to be emptied before deletion.

Experiments and Results

In this occasion, no experiments were conducted whatsoever, but, there was an attempt to utilize AWS Route 53 service in order to avoid full bucket data exposure. However, the account used to make these activities is limited to what it can do, being this action one of the several restricted operations.

Theory

- What's the difference between EBS and Ephemeral?
 - Elastic Block Store [EBS] provides block level storage volumes for EC2 instances, behaving like raw, unformatted block devices [11]; Ephemeral references to a specific type of storage that self-describes its purpose, an ephemeral/temporary storage that would be attached to an EC2 instance root volume to optimize said instance performance [12].
 - The main difference resides in data persistence: EBS stores static data and Ephemeral storage works with dynamic/temporary data.
- What's the difference between EBS and S3?
 - Both of these services have been defined as *block storage* and *object storage* accordingly. Although both kinds of storage can be seen as general purpose storage, the main difference between them resides in that S3 data can be accessed more easily via Internet, which at the same time, comes with fine grained level of security for undesired data access prevention. Also, S3 will provide redundancy by replicating bucket contents across multiple availability zones.
- What's the difference between S3 and Glacier?
 - Glacier is a sub-branch of S3 services, aimed to archive of rarely accessed data.
 - Normal S3 buckets are intended for relatively fast data retrieval, and that's where the main difference between these storage types is; S3 can return desired data in milliseconds, whereas Glacier could take even days. As of this, Glacier will be a much cheaper storage alternative, but it comes with fees per GB when data retrieval is needed.

Budget Justification

Taking into consideration:

- EC2 Windows instance with extra disk space
- S3 Buckets (normal, archive and Intelligent-Tiering)
- Cloud9 instance

... With 100% utilization per month On-Demand for instances and around 5GB usage for buckets, the costs would be the following:

Service Name	Status	Upfront cost	Monthly cost	Description	Region	Config Summary
Amazon EC2	-	0.00 USD	32.72 USD	Windows instance (+10GB)	US East (N. Virginia)	Tenancy (Shared Instances), Operating system (W...
Amazon Simple Storage Service (S3)	-	0.00 USD	0.12 USD	S3 Normal	US East (N. Virginia)	S3 Standard storage (5 GB per month) DT Inboun...
Amazon Simple Storage Service (S3)	-	0.00 USD	0.12 USD	S3 archive	US East (N. Virginia)	S3 Standard storage (5 GB per month)
Amazon Simple Storage Service (S3)	-	0.00 USD	0.12 USD	S3 IT	US East (N. Virginia)	S3 INT Average Object Size (16 MB), Percentage ...
Amazon EC2	-	0.00 USD	8.47 USD	Cloud9	US East (N. Virginia)	Tenancy (Shared Instances), Operating system (Li...

These costs would be similar to those of a file server for personal use.

Conclusions

The exploration and implementation of several storage means has resulted in a very productive and informative activity, because in modern times, cloud storage providers such as Google Drive or Dropbox make it seem so easy to upload and download files remotely, that most people won't even stop to think about how that process works.

At a highly level of abstraction, static web sites consist of a browser interpretation of a remotely located resource, a resource that needs to be stored somewhere in the world, inside a file system. The opportunity of conducting the deployment of a static web page makes this an alternative for free hosting services, making it an interesting technology for web hosting.

The activities found within this development will surely provide support for future implementations, as storage it's always needed, either by storing analytic results or simply by holding a shared resource.

Bibliography

- [1] Kingston Technology, '2 Types of M.2 SSDs: SATA and NVMe'. [Online]. Available: <https://www.kingston.com/en/blog/pc-performance/two-types-m2-vs-ssd>.
- [2] nOps. 'Horizontal vs Vertical scaling: An in-depth Guide'. [Online]. Available: <https://www.nops.io/blog/horizontal-vs-vertical-scaling/>.
- [3] Cloud Industry Home. '8 criteria to ensure you select the right cloud service provider'. [Online]. Available: <https://cloudindustryforum.org/8-criteria-to-ensure-you-select-the-right-cloud-service-provider/>.
- [4] C. Velasco. 'Filtración de la Lista Nominal del Padrón Electoral del INE en Servidores de Amazon'. [Online]. Available: <https://protecciondatos.mx/2016/04/esfiltracin-de-la-lista-nominal-del-padrn-electoral-del-ine-en-servidores-de-amazonendata-breach-mexicos-national-voters-list-national-electoral-institute-ine-amazon-servers/>.
- [5] S. Naraharisetty. 'How to choose the right cloud storage service'. [Online]. Available: <https://www.2brightsparks.com/resources/articles/guidelines-to-select-a-cloud-storage-service.html>.
- [6] Aws.amazon.com, 'Amazon S3'. [Online]. Available: <https://aws.amazon.com/es/s3/>.
- [7] Aws.amazon.com, 'Amazon EC2'. [Online]. Available: <https://aws.amazon.com/es/ec2/>.
- [8] Aws.amazon.com, 'How S3 Intelligent-Tiering works'. [Online]. Available: <https://docs.aws.amazon.com/AmazonS3/latest/userguide/intelligent-tiering-overview.html>.
- [9] Aws.amazon.com, 'Amazon Route 53'. [Online]. Available: <https://aws.amazon.com/route53/>.
- [10] Aws.amazon.com, 'AWS Cloud9', [Online]. Available: <https://aws.amazon.com/cloud9/>.
- [11] Aws.amazon.com, 'Amazon Elastic Block Store (Amazon EBS'. [Online]. Available: <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/AmazonEBS.html>.
- [12] T. Mahmood. 'Amazon EBS vs. Ephemeral Storage'. [Online]. Available: <https://linuxhint.com/amazon-ebs-vs-ephemeral-storage/>.