



Bansilal Ramnath Agarwal Charitable Trust's
Vishwakarma Institute of Information Technology

Department of
Artificial Intelligence and Data Science

Name: Agrawal Yash Pradeep

Class: TY

Division: A

Roll No: 371005

Semester: V

Academic Year: 2022-2023

Subject Name & Code: Cloud Computing and Analytics

Title of Assignment: Study and installation of Storage as Services on cloud.

Date of Performance: 2/09/2022

Date of Submission: 05/11/2022

Aim: Study and installation of Storage as Services on cloud.

Problem Statement: Study and installation of Storage as Services on cloud.

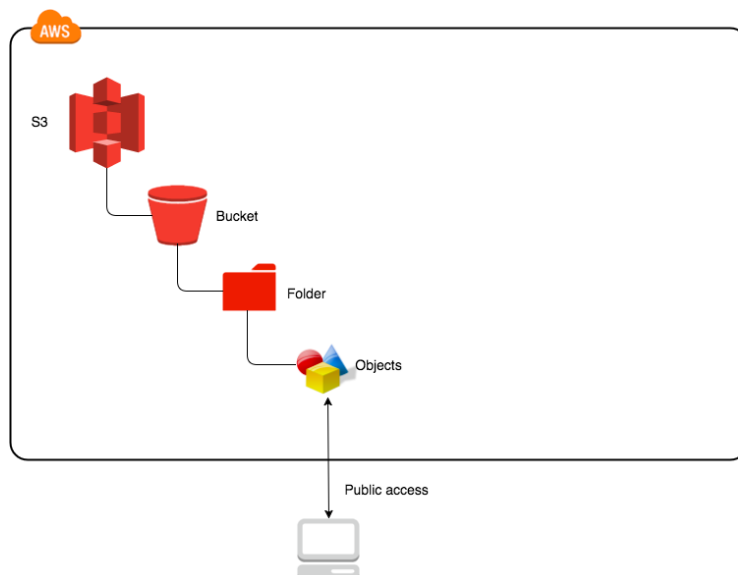
Background Information:

- **AWS Storage Services:**

- AWS offers a wide range of storage services that can be provisioned depending on your project requirements and use case.
- AWS storage services have different provisions for highly confidential data, frequently accessed data, and the not so frequently accessed data.
- Types are: *object storage, file storage, block storage services, backups, and data migration* options. All of which fall under the AWS Storage Services list.

- **AWS Simple Storage Service (S3):**

- From the mentioned list, S3, is the object storage service provided by AWS. It is probably the most commonly used, go-to storage service for AWS users given the features like extremely high availability, security, and simple connection to other AWS Services.
- AWS S3 can be used by people with all kinds of use cases like mobile/web applications, big data, machine learning and many more.



AWS S3 Terminology:

- **Bucket:** Data, in S3, is stored in containers called *buckets*.
 - Each bucket will have its own set of policies and configuration. This enables users to have more control over their data.
 - Bucket Names must be unique.
 - Can be thought of as a parent folder of data.
 - There is a limit of 100 buckets per AWS accounts. But it can be increased if requested from AWS support.
- **Bucket Owner:** The person or organization that owns a particular bucket is its *bucket owner*.
- **Import/Export Station:** A machine that uploads or downloads data to/from S3.
- **Key:** Key, in S3, is a unique identifier for an object in a bucket.
 - It is important to note that 'bucketName + key' is unique for all objects.
 - This also means that there can be only one object for a key in a bucket. If you upload 2 files with the same key. The file uploaded latest will overwrite the previously contained file.
- **Versioning:** Versioning means to always keep a record of previously uploaded files in S3. Points to note:
 - Versioning is not enabled by default. Once enabled, it is enabled for all objects in a bucket.
 - Versioning keeps all the copies of your file, so, it adds cost for storing multiple copies of your data.
 - Versioning is helpful to prevent unintended overwrites and deletions.
 - Note that objects with the same key can be stored in a bucket if versioning is enabled (since they have a unique version ID).
- **Null Object:** Version ID for objects in a bucket where versioning is suspended is null. Such objects may be referred to as null objects.
 - For buckets with versioning enabled, each version of a file has a specific version ID.
- **Object:** Fundamental entity type stored in AWS S3.
- **Access Control Lists (ACL):** A document for verifying the access to S3 buckets from outside your AWS account. Each bucket has its own ACL.
- **Bucket Policies:** A document for verifying the access to S3 buckets from within your AWS account, this controls which services and users have what kind of access to your S3 bucket. Each bucket has its own Bucket Policies.
- **Lifecycle Rules:** This is a cost-saving practice that can move your files to AWS Glacier (The AWS Data Archive Service) or to some other S3 storage class for cheaper storage of old data or completely delete the data after the specified time.

Features of AWS S3:

- **Durability:** AWS claims Amazon S3 to have a possibility of losing your data stored on S3 is one in a billion.
- **Availability:** AWS ensures that the up-time of AWS related to being able to access data and durability is related to losing data altogether.
- **Server-Side-Encryption (SSE):** AWS S3 supports three types of SSE models:
 - **SSE-S3:** AWS S3 manages encryption keys.
 - **SSE-C:** The customer manages encryption keys.
 - **SSE-KMS:** The AWS Key Management Service (KMS) manages the encryption keys.
- **File Size support:** AWS S3 can hold files of size ranging from 0 bytes to 5 terabytes. A 5TB limit on file size should not be a blocker for most of the applications in the world.
- **Infinite storage space:** Theoretically AWS S3 is supposed to have infinite storage space. This makes S3 infinitely scalable for all kinds of use cases.

- **Pay as you use:** The users are charged according to the S3 storage they hold.
- **AWS-S3** is region-specific.

S3 storage classes:

- **Standard:** Suitable for frequently accessed data, that needs to be highly available and durable.
- **Standard Infrequent Access (Standard IA):** This is a cheaper data-storage class and as the name suggests, this class is best suited for storing infrequently accessed data like log files or data archives. Note that there may be a per GB data retrieval fee associated with Standard IA class.
- **Intelligent Tiering:** This service class classifies your files automatically into frequently accessed and infrequently accessed and stores the infrequently accessed data in infrequent access storage to save costs. This is useful for unpredictable data access to an S3 bucket.
- **One Zone Infrequent Access (One Zone IA):** All the files on your S3 have their copies stored in a minimum of 3 Availability Zones. One Zone IA stores this data in a single availability zone. It is only recommended to use this storage class for infrequently accessed, non-essential data. There may be a per GB cost for data retrieval.
- **Reduced Redundancy Storage (RRS):** All the other S3 classes ensure the durability of 99.999999999%. RRS only ensures a 99.99% durability. AWS no longer recommends RRS due to its less durability. However, it can be used to store non-essential data.

Cloud Resource Requirements: AWS Console - S3

Steps:

Step 1: Set up and log into your AWS account

To use Amazon S3, you need an AWS account. If you don't already have one, you'll be prompted to create one when you sign up for Amazon S3. You will not be charged for Amazon S3 until you use it.

Step 2: Create a bucket

Every object in Amazon S3 is stored in a bucket. Before you can store data in Amazon S3, you must create an S3 bucket.

Step 3: Start building with AWS

Now that you've created a bucket, you're ready to add an object to it. An object can be any kind of file: a text file, a photo, a video, and so on. Read the Getting Started Guide to learn more and start building.

Conclusion:

It is Amazon's object storage service that deploys at the click of a mouse. S3 is a safe place to store all of your files (object storage). Object storage manages data and links that data to associated metadata (data about data). An object is a file and potentially any metadata (data about data describing the file). Buckets store objects, objects have keys which are unique identifiers. S3 isn't suitable to install an operating system or database on because it's object-based (you would want to use block storage for these).