# AWS Solution Architect Associate Certification Training – Module 10

## 10. Simple Storage Service (S3)

### Introduction to Simple Storage Service (S3)

Amazon Simple Storage Service is storage for the Internet. It is designed to make web-scale computing easier for developers.

Amazon S3 has a simple web services interface that you can use to store and retrieve any amount of data, at any time, from anywhere on the web. It gives any developer access to the same highly scalable, reliable, fast, inexpensive data storage infrastructure that Amazon uses to run its own global network of web sites. The service aims to maximize benefits of scale and to pass those benefits on to developers.

Amazon Simple Storage Service (Amazon S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance. This means customers of all sizes and industries can use it to store and protect any amount of data for a range of use cases, such as websites, mobile applications, backup and restore, archive, enterprise applications, IoT devices, and big data analytics. Amazon S3 provides easy-to-use management features so you can organize your data and configure finely-tuned access controls to meet your specific business, organizational, and compliance requirements. Amazon S3 is designed for 99.999999999% (11 9's) of durability, and stores data for millions of applications for companies all around the world.

**Benefits:**

**Industry-leading performance, scalability, availability, and durability**

Scale your storage resources up and down to meet fluctuating demands, without upfront investments or resource procurement cycles. Amazon S3 is designed for 99.999999999% (11 9’s) of data durability because it automatically creates and stores copies of all S3 objects across multiple systems. This means your data is available when needed and protected against failures, errors, and threats.

**Wide range of cost-effective storage classes**

Save costs without sacrificing performance by storing data across the S3 Storage Classes, which support different data access levels at corresponding rates. You can use S3 Storage Class Analysis to discover data that should move to a lower-cost storage class based on access patterns, and configure an S3 Lifecycle policy to execute the transfer. You can also store data with changing or unknown access patterns in S3 Intelligent-Tiering, which tiers objects based on changing access patterns and automatically delivers cost savings.

**Unmatched security, compliance, and audit capabilities**

Store your data in Amazon S3 and secure it from unauthorized access with encryption features and access management tools. Amazon S3 maintains compliance programs, such as PCI-DSS, HIPAA/HITECH, FedRAMP, EU Data Protection Directive, and FISMA, to help you meet regulatory requirements. AWS also supports numerous auditing capabilities to monitor access requests to your S3 resources.

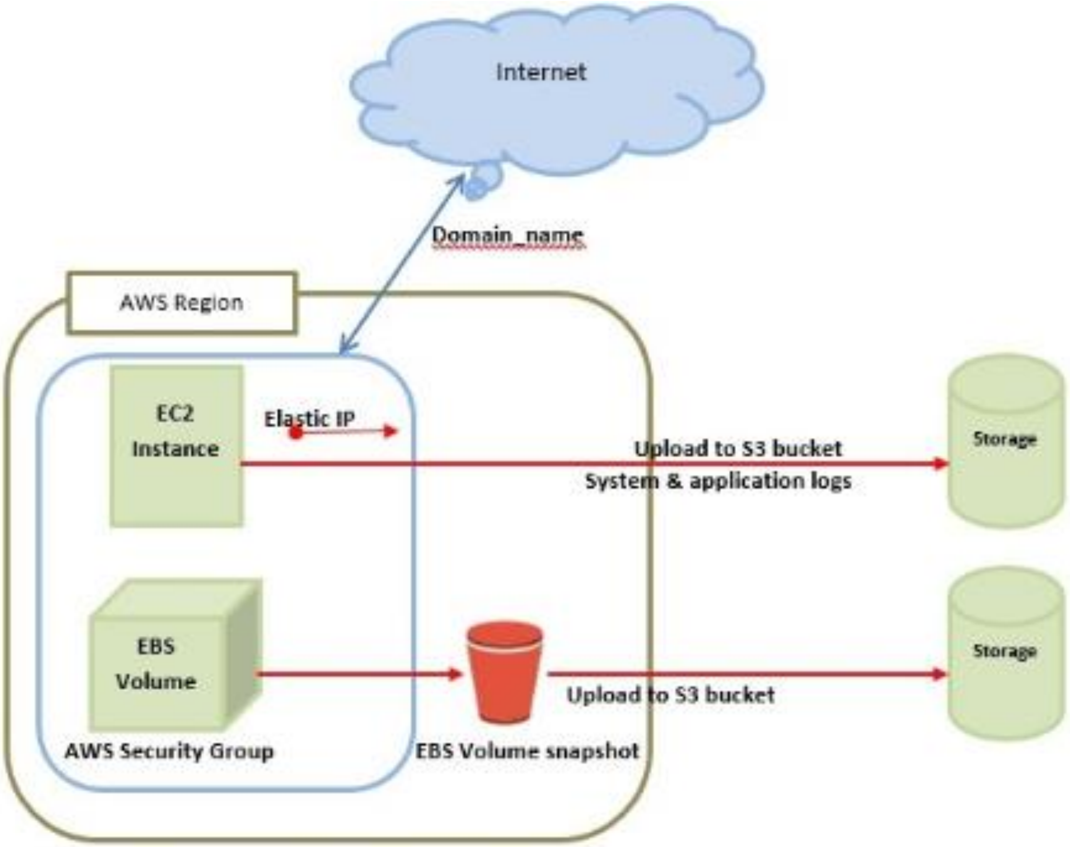
**Management tools for granular data control**

Classify, manage, and report on your data using features, such as: S3 Storage Class Analysis to analyze access patterns; S3 Lifecycle policies to transfer objects to lower-cost storage classes; S3 Cross-Region Replication to replicate data into other regions; S3 Object Lock to apply retention dates to objects and protect them from deletion; and S3 Inventory to get visibility into your stored objects, their metadata, and encryption status. You can also use S3 Batch Operations to change object properties and perform storage management tasks for billions of objects. Since Amazon S3 works with AWS Lambda, you can log activities, define alerts, and automate workflows without managing additional infrastructure.

**Query-in-place services for analytics**

Run big data analytics across your S3 objects (and other data sets in AWS) with our query-in-place services. Use Amazon Athena to query S3 data with standard SQL expressions and Amazon Redshift Spectrum to analyze data that is stored across your AWS data warehouses and S3 resources. You can also use S3 Select to retrieve subsets of object metadata, instead of the entire object, and improve query performance by up to 400%.

**The architecture of Simple Storage Service (S3)**



**S3 buckets and Objects**

An Amazon S3 bucket is a public cloud storage resource available in Amazon Web Services' (AWS) Simple Storage Service (S3), an object storage offering. Amazon S3 buckets, which are similar to file folders, store objects, which consist of data and its descriptive metadata.

To upload your data (photos, videos, documents etc.), you first create a bucket in one of the AWS Regions. You can then upload any number of objects to the bucket.

By default, you can create up to 100 buckets in each of your AWS accounts. If you need more buckets, you can increase your account bucket limit to a maximum of 1,000 buckets by submitting a service limit increase.

### Working with Amazon S3 Objects

Amazon S3 is a simple key, value store designed to store as many objects as you want. You store these objects in one or more buckets. An object consists of the following:

- **Key** – The name that you assign to an object. You use the object key to retrieve the object.
- **Version ID** – Within a bucket, a key and version ID uniquely identify an object.The version ID is a string that Amazon S3 generates when you add an object to a bucket.
- **Value** – The content that you are storing.An object value can be any sequence of bytes. Objects can range in size from zero to 5 TB.
- **Metadata** – A set of name-value pairs with which you can store information regarding the object.You can assign metadata, referred to as user-defined metadata, to your objects in Amazon S3. Amazon S3 also assigns system-metadata to these objects, which it uses for managing objects.
- **Access Control Information** – You can control access to the objects you store in Amazon S3.

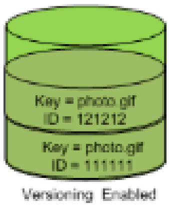
### Features of S3 Buckets

Amazon S3 has various features you can use to organize and manage your data in ways that support specific use cases, enable cost efficiencies, enforce security, and meet compliance requirements. Data is stored as objects within resources called “buckets”, and a single object can be up to 5 terabytes in size. S3 features include capabilities to append metadata tags to objects, move and store data across the S3 Storage Classes, configure and enforce data access controls, secure data against unauthorized users, run big data analytics, and monitor data at the object and bucket levels.

### Versioning, Logging, Access control

Versioning is a means of keeping multiple variants of an object in the same bucket. You can use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket. With versioning, you can easily recover from both unintended user actions and application failures.

In one bucket, for example, you can have the two objects with the same key, but with different version IDs, such as photo.gif(version 111111) and photo.gif (version 121212).



Versioning-enabled buckets enable you to recover objects from accidental deletion or overwrite. For example:

- If you delete an object, instead of removing it permanently, Amazon S3 inserts a delete marker, which becomes the current object version. You can always restore the previous version.
- If you overwrite an object, it results in a new object version in the bucket. You can always restore the previous version.

Buckets can be in one of three states: unversioned (the default), versioning-enabled, or versioning-suspended.

Once you version-enable a bucket, it can never return to an unversioned state. You can, however, suspend versioning on that bucket.

**Logging S3**

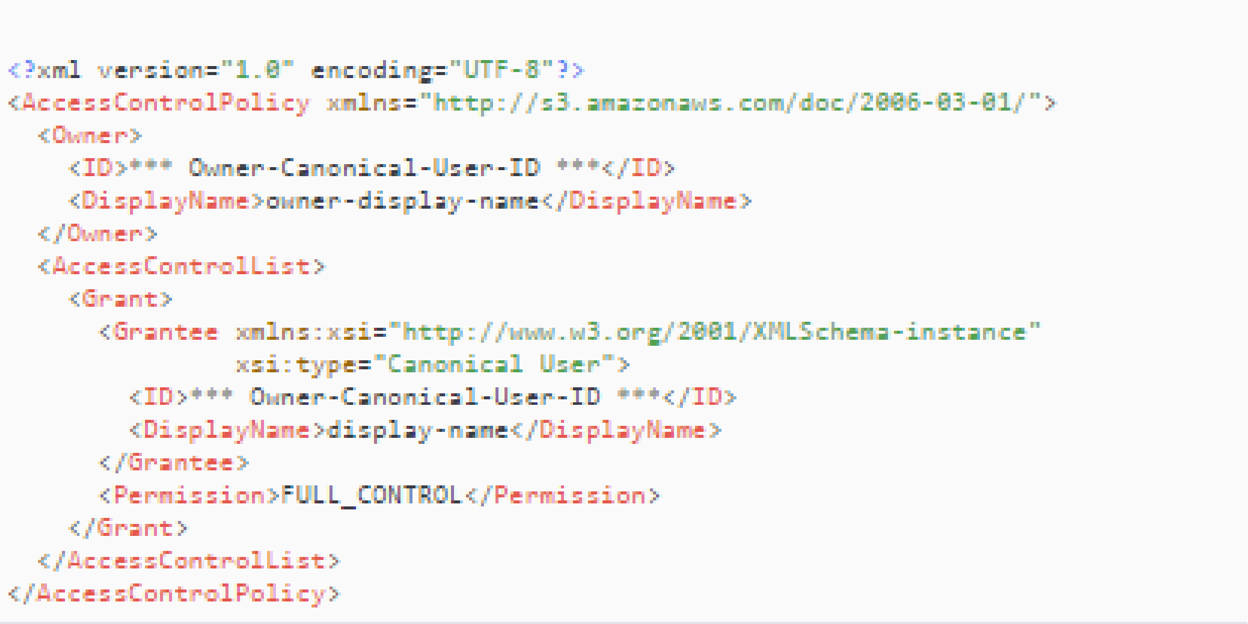
By default, Amazon Simple Storage Service (Amazon S3) doesn't collect server access logs. When you enable logging, Amazon S3 delivers access logs for a source bucket to a target bucket that you choose. The target bucket must be in the same AWS Region as the source bucket. Server access logging provides detailed records for the requests that are made to an S3 bucket. Server access logs are useful for many applications. For example, access log information can be useful in security and access audits. It can also help you learn about your customer base and understand your Amazon S3 bill.

**Important:** There is no extra charge for enabling server access logging on an Amazon S3 bucket. However, any log files that the system delivers to you will accrue the usual charges for storage. (You can delete the log files at any time.) We do not assess data transfer charges for log file delivery, but we do charge the normal data transfer rate for accessing the log files.

**Access control**

Amazon S3 access control lists (ACLs) enable you to manage access to buckets and objects. Each bucket and object has an ACL attached to it as a sub-resource. It defines which AWS accounts or groups are granted access and the type of access. When a request is received against a resource, Amazon S3 checks the corresponding ACL to verify that the requester has the necessary access permissions.

When you create a bucket or an object, Amazon S3 creates a default ACL that grants the resource owner full control over the resource. This is shown in the following sample bucket ACL (the default object ACL has the same structure):



```
<?xml version="1.0" encoding="UTF-8"?>
<AccessControlPolicy xmlns="http://s3.amazonaws.com/doc/2006-03-01/">
  <Owner>
    <ID>*** Owner-Canonical-User-ID ***</ID>
    <DisplayName>owner-display-name</DisplayName>
  </Owner>
  <AccessControlList>
    <Grant>
      <Grantee xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
               xsi:type="Canonical User">
        <ID>*** Owner-Canonical-User-ID ***</ID>
        <DisplayName>display-name</DisplayName>
      </Grantee>
      <Permission>FULL_CONTROL</Permission>
    </Grant>
  </AccessControlList>
</AccessControlPolicy>
```

**Encryption**

Amazon S3 default encryption provides a way to set the default encryption behavior for an S3 bucket. You can set default encryption on a bucket so that all objects are encrypted when they are stored in the bucket. The objects are encrypted using server-side encryption with either Amazon S3-managed keys (SSE-S3) or AWS KMS-managed keys (SSE-KMS).

When you use server-side encryption, Amazon S3 encrypts an object before saving it to disk in its data centers and decrypts it when you download the objects.

**Protecting Data Using Encryption**

- Protecting Data Using Server-Side Encryption
- Protecting Data Using Client-Side Encryption

Data protection refers to protecting data while in-transit (as it travels to and from Amazon S3) and at rest (while it is stored on disks in Amazon S3 data centers).

**Use Server-Side Encryption** – You request Amazon S3 to encrypt your object before saving it on disks in its data centers and decrypt it when you download the objects.

**Use Client-Side Encryption** – You can encrypt data client-side and upload the encrypted data to Amazon S3. In this case, you manage the encryption process, the encryption keys, and related tools.

**Permissions for the Amazon S3 Bucket**

By default, all Amazon S3 buckets and objects are private. Only the resource owner and the AWS account that created the bucket can access that bucket and any objects it contains. The resource owner can, however, choose to grant access permissions to other resources and users. One way to do this is to write an access policy. If Config creates an S3 bucket for you automatically (for example, if you use the *AWS Config console* or use the *aws config subscribe* command to set up your deliver channel) or you choose an existing S3 bucket already existing in your account, these permissions are automatically added to S3 bucket. However, if you specify an existing S3 bucket from another account, you must ensure that the S3 bucket has the correct permissions.

**Static Website Hosting**

You can host a static website on Amazon Simple Storage Service (Amazon S3). On a static website, individual webpages include static content. They might also contain client-side scripts. By contrast, a dynamic website relies on server-side processing, including server-side scripts such as PHP, JSP, or ASP.NET. Amazon S3 does not support server-side scripting. Amazon Web Services (AWS) also has resources for hosting dynamic websites.

**Amazon S3 Transfer Acceleration**

Amazon S3 Transfer Acceleration enables fast, easy, and secure transfers of files over long distances between your client and an S3 bucket. Transfer Acceleration takes advantage of Amazon CloudFront's globally distributed edge locations. As the data arrives at an edge location, data is routed to Amazon S3 over an optimized network path. When using Transfer Acceleration, additional data transfer charges may apply.

**Backup and Archiving using Glacier Storage**

Amazon S3 Glacier is a secure, durable, and extremely low-cost cloud storage service for data archiving and long-term backup. It is designed to deliver 99.999999999% durability, and provides comprehensive security and compliance capabilities that can help meet even the most stringent regulatory requirements. Amazon S3 Glacier provides query-in-place functionality, allowing you to run powerful analytics directly on your archive data at rest. Customers can store data for as little as $0.004 per gigabyte per month, a significant savings compared to on-premises solutions. To keep costs low yet suitable for varying retrieval needs, Amazon S3 Glacier provides three options for access to archives, from a few minutes to several hours.

Glacier is an extremely low-cost storage service that provides durable storage with security features for data archiving and backup. With Glacier, customers can store their data cost effectively for months, years, or even decades. Glacier enables customers to offload the administrative burdens of operating and scaling storage to AWS, so they don't have to worry about capacity planning, hardware provisioning, data replication, hardware failure detection and recovery, or time-consuming hardware migrations.

**Life cycle management of Objects**

To manage your objects so that they are stored cost effectively throughout their lifecycle, configure their lifecycle. A *lifecycle configuration* is a set of rules that define actions that Amazon S3 applies to a group of objects. There are two types of actions:

**Transition actions**—Define when objects transition to another storage class. For example, you might choose to transition objects to the STANDARD_IA storage class 30 days after you created them, or archive objects to the GLACIER storage class one year after creating them.

There are costs associated with the lifecycle transition requests.

**Expiration actions**—Define when objects expire. Amazon S3 deletes expired objects on your behalf. The lifecycle expiration costs depend on when you choose to expire objects.

**When Should I Use Lifecycle Configuration?**

- If you upload periodic logs to a bucket, your application might need them for a week or a month. After that, you might want to delete them.
- Some documents are frequently accessed for a limited period of time. After that, they are infrequently accessed. At some point, you might not need real-time access to them, but your organization or regulations might require you to archive them for a specific period. After that, you can delete them.
- You might upload some types of data to Amazon S3 primarily for archival purposes. For example, you might archive digital media, financial and healthcare records, raw genomics sequence data, long-term database backups, and data that must be retained for regulatory compliance.