

SWIN
BUR
NE

SWINBURNE
UNIVERSITY OF
TECHNOLOGY

COS80001

Software Development for Cloud Computing

Lecture 03 Storage Services



Reminders



- Assignment 1A demonstrations in next week's lab. You must demo to get a mark
- Assignment 1B out now!!!
- Don't forget to do the online AWS MCQ quizzes.

Last week



- **Virtualisation of Computation**

- ☐ Virtual machines
- ☐ Containers
- ☐ Serverless Computing

- **AWS Compute services (ACF Module 2.1)**

- ☐ 1: Compute Services Overview
- ☐ 2: Introduction to Amazon Elastic Compute Cloud (Amazon EC2)
- ☐ 3: Amazon EC2 Cost Optimization
- ☐ 4: Introduction to AWS Lambda
- ☐ 5: Introduction to AWS Elastic Beanstalk

This week



■ Storage in the Cloud

- ☐ Big Data
- ☐ NFS
- ☐ Distributed File Systems (databases next week)

■ AWS Storage services (ACF Module 2.2)

- ☐ 1: Amazon Elastic Block Store (Amazon EBS)
 - ☐ Plus some extra notes on Instance Storage
- ☐ 2: Amazon Simple Storage Service (Amazon S3)
- ☐ 3: Amazon Elastic File System (Amazon EFS)
- ☐ 4: Amazon Glacier

Big Data – Centre of the Universe



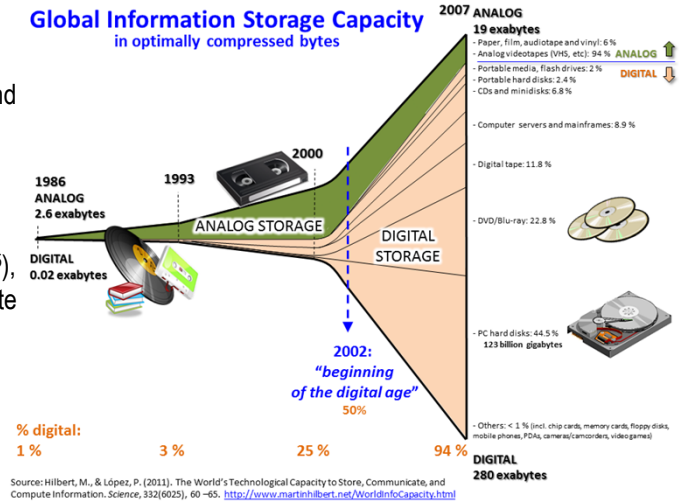
■ Drivers

- Internet commerce, Mobile, Social media, IoT and sensors, Science (e.g. biology, astronomy, meteorology), Health, Spooks, ...

■ Data sizes

- KB (10^3), MB (10^6), GB (10^9), TB (10^{12}), PB(10^{15}), exabyte (EB, 10^{18}), zettabyte (ZB, 10^{21}), yottabyte (YB, 10^{24})
- Everyday: Facebook 10T, Twitter 7T, Youtube 4.5T+

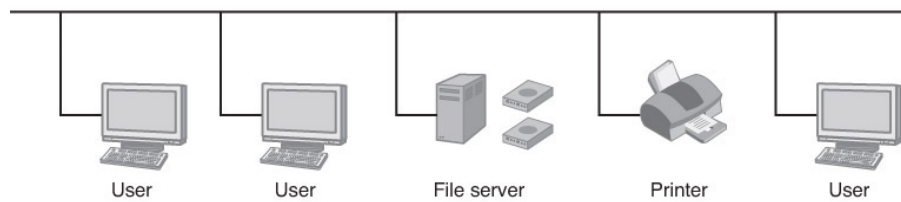
■ 4Vs: volume, variety, velocity, and veracity



Network Storage Began with File Servers



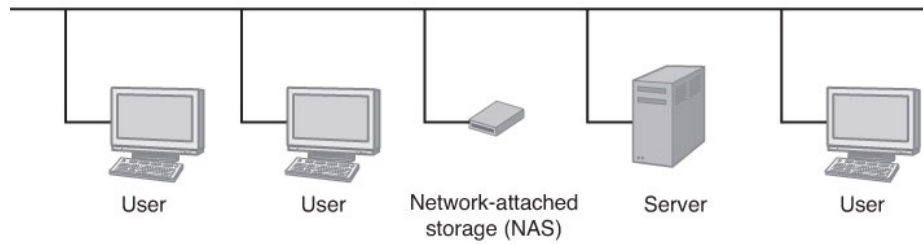
Years ago, local-area networks used special servers, called file servers, to support file sharing, file replication, and storage for large files.



Network-Attached Storage (NAS)



Plug directly into the network.





Advantages of NAS

- **Reliability:** A NAS device typically provides advanced data striping across multiple volumes within the device. If one (or more) volumes fail, the data striping would maintain the data and allow reconstruction of the file contents.
- **Performance:** Because the NAS device did not run a complete operating system, the hardware had less system overhead, which allowed it to outperform a file server.
- **Compatibility:** NAS devices normally support common file systems, which, in turn, make them fully compatible with common operating systems.
- **Ease of performing backups:** NAS devices are commonly used for backup devices. Within a home, for example, all devices can easily access and back up files to a NAS device.

Cloud-Based Storage



- Cloud-based data storage is the next step in the evolution of NAS devices.
- Across the web (the cloud), many providers offer data storage that resides in the cloud.
- Data may be accessible as follows:
 - ☐ Through a web browser interface
 - ☐ Through a mounted disk drive
 - ☐ Through a set of API (application program interface) calls



Advantages of Cloud-Based Storage

- **Scalability:** Most cloud-based data storage providers let users scale their storage capacity (up or down) to align with their storage needs.
- **Pay for use:** With most cloud-based data storage facilities, users pay only for the storage (within a range) that they need.
- **Reliability:** Many cloud-based data storage facilities provide transparent data replication.
- **Ease of access:** Most cloud-based data storage facilities support web-based access to files from any place, at any time, using a variety of devices.
- **Ease of use:** Many cloud-based data storage solutions let users map a drive letter to the remote file storage area and then access the files through the use of a logical drive.

Disadvantages of Cloud-Based Storage



- **Performance:** Because the cloud-based disk storage devices are accessed over the Internet, they will never be as fast as local drives.
- **Security:** Some users will never feel comfortable with their data in the cloud.
- **Data orphans:** Users may abandon data in cloud storage facilities, leaving confidential private or company data at risk.

Cloud-Based Block Storage



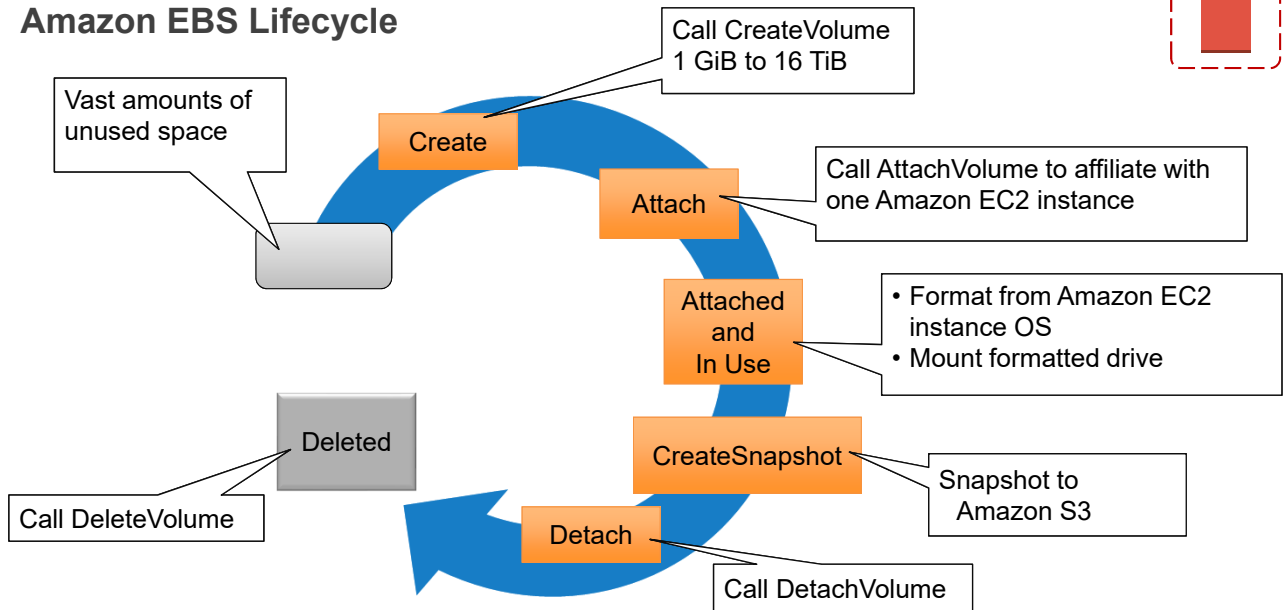
- In the simplest sense, a block of data storage is a fixed-sized sequence of bits. The size of the block normally corresponds to an underlying unit of storage on the **cloud-based block storage device**.
- Some applications work with very large blocks of data, the format of which has meaning only to the application itself—meaning that the data may not map well to storage within a file system or database.

File Systems



- Operating systems exist to allow users to run programs and to store and retrieve data (files) from one user session to the next.
- Within the operating system, special software, called the file system, oversees the storage and retrieval of files to and from a disk.
- When you copy a file, delete a file, or create and move files between folders, the file system is performing the work.
- Initially, file systems allowed users to manipulate only local files that reside on one of the PC's disk drives.
- As networks became more prevalent, so too did network operating systems, which allow users and programs to manipulate files residing on a device across the network.

Amazon EBS Lifecycle



Amazon EBS provides block-level storage volumes for use with Amazon EC2 instances. Amazon EBS volumes are highly available and reliable storage volumes that can be attached to any running instance in the same Availability Zone. The Amazon EBS volumes attached to an Amazon EC2 instance are exposed as storage volumes that persist independently from the life of the instance. When the volumes are not attached to an EC2 instance, you pay only for the cost of storage.

Real World: Hadoop Distributed File System

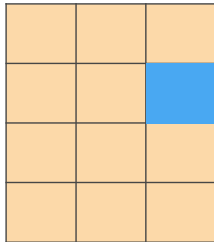


- Apache Hadoop is an open source project, the goal of which is to support reliable, scalable distributed computing.
- Part of the project includes the Hadoop Distributed File System (HDFS), a Java-based file system that is well suited for cloud-based storage.
- HDFS is designed to be highly fault tolerant and robust to maintain operation in the event of a device failure.

AWS Storage Options: Block vs. Object Storage

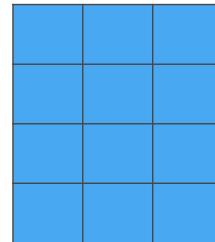


What if you want to **change** one character in a 1-GB file?



Block Storage

Change one block (piece of the file)
that contains the character



Object Storage

Entire file must be updated



One of the critical concepts to understanding the differences between some storage types is whether they offer "block-level" storage or "object-level" storage. This difference has a major impact on the throughput, latency, and cost of your storage solution: block storage solutions are typically faster and use less bandwidth but cost more than object-level storage.



Module 2, Section 2: AWS Core Services - Storage








© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Welcome to Module 2, Section 2 – AWS Core Services - Storage.

What's In This Module



Module 2, Section 2 – Core Services - Storage:

-  Part 1: Amazon Elastic Block Store (Amazon EBS)
 -  Plus some extra notes on Instance Storage
-  Part 2: Amazon Simple Storage Service (Amazon S3)
-  Part 3: Amazon Elastic File System (Amazon EFS)
-  Part 4: Amazon Glacier

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Cloud storage is a critical component of cloud computing, holding the information used by applications. Big data analytics, data warehouses, Internet of Things (IoT), databases, and backup and archive applications all rely on some form of data storage architecture. Cloud storage is typically more reliable, scalable, and secure than traditional on-premises storage systems.

Module Objectives



Goal: Discuss key concepts related to storage

- 📦 Understand the differences between different types of storage
- 📦 Review basic pricing that differentiates the storage solutions

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

The goal of this module is to help you understand the storage resources that are available. We review the different pricing options so you understand how different choices impact your solution cost.

Section 2: Introduction to Storage Services

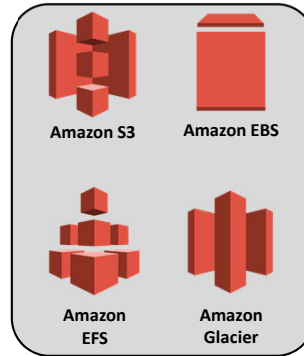
Core AWS Services



Amazon
VPC



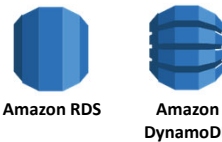
Amazon EC2



Storage



AWS IAM



Database

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

It is not surprising that storage is another AWS core service. There are three broad categories of storage: instance store ("ephemeral"), Amazon EBS, and Amazon S3.

Instance store, or *ephemeral storage*, is a **temporary storage** that it is added to your Amazon EC2 instance. Amazon EBS is **persistent, mountable storage**; it can be mounted as a device to an Amazon EC2 instance. Amazon EBS can only be mounted to an Amazon EC2 instance in the same Availability Zone. Like Amazon EBS, Amazon S3 is persistent storage but it can be accessed from anywhere.

Part 1:

Amazon Elastic Block Store (Amazon EBS)

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

We just finished discussing compute options to power your solution. Amazon Elastic Block Store (Amazon EBS) is an AWS block storage system that is best used for storing persistent data. Amazon EBS provides highly available block level storage volumes for use with Amazon EC2 instances.



Amazon Elastic Block Store (Amazon EBS)

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon EBS provides persistent block storage volumes for use with Amazon EC2 instances in the cloud. Persistent storage is any data storage device that retains data after power to that device is shut off. It is also sometimes referred to as *non-volatile storage*.

Each Amazon EBS volume is automatically replicated *within* its Availability Zone to protect you from component failure, offering high availability and durability. Amazon EBS volumes offer the consistent and low-latency performance needed to run your workloads. With Amazon EBS, you can scale your usage up or down within minutes – all while paying a low price for only what you provision.

Amazon EBS Review



*Amazon EBS allows you to **create individual storage volumes** and **attach them** to an Amazon EC2 instance.*

- 📦 Amazon EBS offers block-level storage
- 📦 Volumes are automatically replicated within its Availability Zone
- 📦 Can be backed up automatically to Amazon S3
- 📦 Uses:
 - 📦 Boot volumes and storage for Amazon EC2 instances
 - 📦 Data storage with a file system
 - 📦 Database hosts
 - 📦 Enterprise applications

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Amazon EBS volumes provide durable, detachable, block-level storage (like an external hard drive) for your Amazon EC2 instances. Because they are directly attached to the instances, they can provide extremely low latency between where the data is stored and where it might be used on the instance. For this reason, they can be used to run a database with an Amazon EC2 instance. Amazon EBS volumes can also be used to back up your instances into Amazon Machine Images (AMI), which are stored in Amazon S3 and can be reused to create new Amazon EC2 instances later.

Amazon EBS Volume Types



	Solid-State Drives (SSD)		Hard Disk Drives (HDD)	
	General Purpose	Provisioned IOPS	Throughput-Optimized	Cold
Max volume size	16 TiB	16 TiB	16 TiB	16 TiB
Max IOPS/volume	10,000	32,000	500	250
Max throughput/volume	160 MiB/s	500 MiB/s	500 MiB/s	250 MiB/s

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Matching the correct technology to your workload is a key best practice for reducing storage costs. Provisioned IOPS SSD-backed Amazon EBS volumes can give you the highest performance, but if your application doesn't require or won't use performance that high, one of the lower-cost options might be a better solution.

For more information, see

<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html>.

Amazon EBS Volume Types



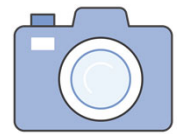
Use Cases

Solid-State Drives (SSD)		Hard Disk Drives (HDD)	
General Purpose	Provisioned IOPS	Throughput-Optimized	Cold
<ul style="list-style-type: none">• Recommended for most workloads• System boot volumes• Virtual desktops• Low-latency interactive apps• Development and test environments	<ul style="list-style-type: none">• I/O-intensive workloads• Relational DBs• NoSQL DBs	<ul style="list-style-type: none">• Streaming workloads requiring consistent, fast throughput at a low price• Big data• Data warehouses• Log processing• Cannot be a boot volume	<ul style="list-style-type: none">• Throughput-oriented storage for large volumes of data that is infrequently accessed• Scenarios where the lowest storage cost is important• Cannot be a boot volume



Snapshots

- Point-in-time snapshots
- Recreate a new volume at any time



Encryption

- Encrypted Amazon EBS volumes
- No additional cost



Elasticity

- Increase capacity
- Change to different types



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

To provide an even higher level of data durability, Amazon EBS gives you the ability to create point-in-time snapshots of your volumes, and AWS allows you to recreate a new volume from a snapshot at any time. Share snapshots or even copy snapshots to different AWS Regions for even greater disaster recovery (DR) protection. You can, for example, encrypt and share your snapshots from Virginia to Tokyo.

You could also have encrypted Amazon EBS volumes at no additional cost. The encryption occurs on the Amazon EC2 side, so the data moving between the Amazon EC2 instance and the Amazon EBS volume inside AWS data centers will be encrypted in transit.

As your company grows, the amount of data stored on your Amazon EBS volumes will likely also grow. Amazon EBS volumes have the ability to increase capacity and change to different types, meaning that you can change from hard disk to SSD or increase from a 50-gigabyte volume to a 16-terabyte volume. For example, you can do this resize operation on the fly without needing to stop the instances.

1. Volumes

- Amazon EBS volumes persist independently from the instance
- All volume types are charged by the amount provisioned per month

2. Input Output Operations per Second (IOPS)

- General Purpose (SSD)
 - Charged by the amount you provision in GB per month until storage is released
- Magnetic
 - Charged by the number of requests to volume
- Provisioned IOPS (SSD)
 - Charged by the amount you provision in IOPS (by % of day / month used)



When you begin to estimate the cost for Amazon EBS, you need to consider the following:

- Volumes** – Volume storage for all Amazon EBS volume types is charged by the amount you provision in GB per month, until you release the storage.
- Input Output Operations per Second (IOPS)** – I/O is included in the price of General Purpose (SSD) volumes, while for Amazon EBS Magnetic volumes, I/O is charged by the number of requests you make to your volume. With Provisioned IOPS (SSD) volumes, you are also charged by the amount you provision in IOPS (multiplied by the percentage of days you provision for the month).

3. Snapshots

- Added cost of Amazon EBS snapshots to Amazon S3 is per GB-month of data stored

4. Data Transfer

- Inbound data transfer is free
- Outbound data transfer charges are tiered



- 3. Snapshot** – Amazon EBS provides the ability to back up snapshots of your data to Amazon S3 for durable recovery. If you opt for Amazon EBS snapshots, the added cost is per GB-month of data stored.
- 4. Data Transfer** – Take into account the amount of data transferred out of your application. Inbound data transfer is free, and outbound data transfer charges are tiered.

Amazon EBS Features:

- ❏ Persistent and customizable block storage for Amazon EC2
- ❏ HDD and SSD types
- ❏ Replicated in the same Availability Zone
- ❏ Easy and transparent encryption
- ❏ Elastic volumes
- ❏ Back up using snapshots

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Amazon EBS provides block level storage volumes for use with Amazon EC2 instances. Amazon EBS volumes are off-instance storage that persists independently from the life of an instance. They are analogous to virtual disks in the cloud. Amazon EBS provides three volume types: General Purpose (SSD), Provisioned IOPS (SSD), and Magnetic.

The three volume types differ in performance characteristics and cost, so you can choose the right storage performance and price for the needs of your applications.

To learn more about Amazon EBS see, <https://aws.amazon.com/ebs/>.

Module 2, Section 2, Lab 2: Working with Amazon EBS

Lab 2 Scenario



This lab focuses on Amazon EBS, a key underlying storage mechanism for Amazon EC2 instances. In this lab, you will create an Amazon EBS volume, attach it to an instance, apply a file system to the volume, and then take a snapshot backup.



Amazon EBS

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon EBS provides persistent block storage volumes for use with [Amazon EC2](#) instances in the AWS cloud. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure, offering high availability and durability.

After completing this lab, you will be able to:

- Create an Amazon EBS volume
- Attach the volume to an instance
- Configure the instance to use the virtual disk
- Create an Amazon EBS snapshot
- Restore the snapshot

Duration: ~45 minutes

Lab 2: Tasks



Amazon EBS

Create a new **EBS volume**.



Amazon EC2
Instance

Attach the volume to an Amazon **EC2 instance**.



Amazon EBS

Create and configure your **file system**.



snapshot

Create a **snapshot**.

Lab 2: Final Product



In this lab, you:

- Created an Amazon EBS volume
- Attached that volume to an instance
- Configured the instance to use the virtual disk
- Created an Amazon EBS snapshot
- Restored the snapshot



~ 45 minutes

Amazon EC2 Instance Storage

- Local, complimentary **direct attached block storage**
- Includes availability, number of disks, and size **based on EC2 instance type**
- Optimized for **up to 365,000 Read IOPS** and 315,000 First Write IOPS
- SSD or magnetic
- Has **no persistence**
- **Automatically deletes** data when an EC2 instance stops, fails or is terminated

An instance store provides temporary block-level storage for your instance. This storage is located on disks that are physically attached to the host computer. Instance store is ideal for temporary storage of information that changes frequently, such as buffers, caches, scratch data, and other temporary content, or for data that is replicated across a fleet of instances, such as a load-balanced pool of web servers.

For more information, see:

- <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/InstanceStorage.html#instance-store-volumes>
- <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/i2-instances.html>

Amazon EBS vs. Amazon EC2 Instance Store

Amazon EBS

- Data stored on an Amazon EBS volume can persist independently of the life of the instance.
- Storage is **persistent**.

Amazon EC2 Instance Store

- Data stored on a local instance store persists only as long as the instance is alive.
- Storage is **ephemeral**.

Use the local instance store only for temporary data. For data that requires a higher level of durability, use Amazon EBS volumes or back up the data to Amazon S3. If you are using an Amazon EBS volume as a root partition and want your Amazon EBS volume to persist outside of the life of the instance, set the Delete on termination flag to “No”.

Instance Lifecycle – Reboot vs. Stop vs. Terminate

Characteristic	Reboot	Stop/Start (EBS-backed instances only)	Terminate
Host computer	The instance stays on the same host computer	The instance runs on a new host computer	
Public IP address	No change	New address assigned	
Elastic IP addresses (EIP)	EIP remains associated with the instance	EIP remains associated with the instance	EIP is disassociated from the instance
Instance store volumes	Preserved	Erased	Erased
EBS volume	Preserved	Preserved	Boot volume is deleted by default
Billing	Instance billing hour doesn't change	Stops incurring charges as soon as state is changed to <i>stopping</i>	Stops incurring charges as soon as state is changed to <i>shutting-down</i>

The table shows the differences between rebooting, stopping, and terminating your instance.

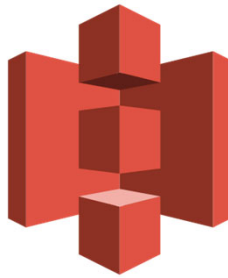
For more information, see:

<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-instance-lifecycle.html>

Part 2: Amazon Simple Storage Service (S3)

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Companies today need the ability to simply and securely collect, store, and analyze their data at a massive scale. Amazon S3 is object storage built to store and retrieve any amount of data from anywhere – web sites and mobile apps, corporate applications, and data from IoT sensors or devices.



Amazon Simple Storage Service (Amazon S3)

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon S3 is object-level storage, which means that if you want to change a part of a file, you have to make the change and then re-upload the entire modified file. Amazon S3 stores data as objects within resources called *buckets*.

Let's take a closer look Amazon S3.

Amazon S3 Review



Managed cloud storage solution designed to scale seamlessly and provide 99.99999999% durability.

- Store as many objects as you want.
- Bucket names must be unique across all existing bucket names in Amazon S3.
- Amazon S3 cannot be used as a bootable drive.
- Data is stored redundantly.
- Access Amazon S3 with the AWS Management Console, one of the AWS SDKs, or a third-party solution.
- Object uploads or deletes can trigger notifications, workflows, or even scripts.
- Data in transit and at rest can be encrypted automatically.
- Storage class analysis (Amazon S3 Analytics) to analyze storage access patterns and transition the right data to the right storage class.

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



You can store as many objects as you want within a bucket, and write, read, and delete objects in your bucket. Bucket names are universal and must be unique across all existing bucket names in Amazon S3. Objects can be up to 5 terabytes in size. By default, data in Amazon S3 is stored redundantly across multiple facilities and multiple devices in each facility.

Amazon S3 is a fully managed storage service that provides a simple API for storing and retrieving data. This means that the data you store in Amazon S3 isn't associated with any particular server, and you don't have to manage any infrastructure yourself. You can put as many objects into Amazon S3 as you want. Amazon S3 holds trillions of objects and regularly peaks at millions of requests per second. Objects can be almost any data file, such as images, videos, or server logs. Since Amazon S3 supports objects as large as several terabytes in size, you could even store database snapshots as objects. Amazon S3 also provides low-latency access to the data over the internet by HTTP or HTTPS, so you can retrieve data anytime from anywhere. You can also access Amazon S3 privately through a virtual private cloud endpoint. You get fine-grained control over who can access your data using identity and access management policies, S3 bucket policies, and even per-object access control lists.

By default, none of your data is shared publicly. You can also encrypt your data in transit and choose to enable server-side encryption on your objects.

Amazon S3 can be accessed via the web-based AWS Management Console, programmatically via the API and SDKs, or with third-party solutions (which use the API/SDKs).

Amazon S3 includes event notifications that allow you to set up automatic notifications when certain events occur, such as an object being uploaded to or deleted from a specific bucket. Those notifications can be sent to you, or they can be used to trigger other processes, such as AWS Lambda scripts.

With storage class analysis, you can analyze storage access patterns and transition the right data to the right storage class. This new Amazon S3 Analytics feature automatically identifies the optimal lifecycle policy to transition less frequently accessed storage to Amazon S3 Standard – Infrequent Access (S3 Standard-IA). You can configure a storage class analysis policy to monitor an entire bucket, a prefix, or object tag. Once an infrequent access pattern is observed, you can easily create a new lifecycle age policy based on the results. Storage class analysis also provides daily visualizations of your storage usage in the AWS Management Console. You can export these to an S3 bucket to analyze using the business intelligence tools of your choice, such as Amazon QuickSight.

Amazon S3 Storage Classes



📦 Amazon S3 provides four classes of object-level storage.

- 📦 Amazon S3 Standard
- 📦 Amazon S3 Standard-IA
- 📦 Amazon S3 One Zone-IA
- 📦 Amazon Glacier

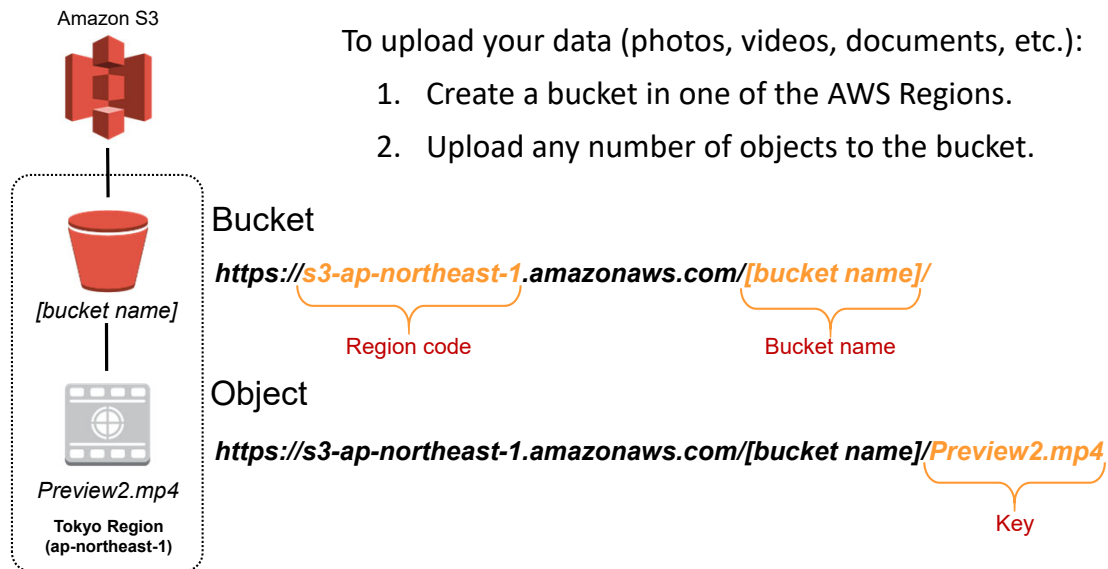


© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

You can select from four different storage classes to store your data in Amazon S3:

- **Amazon S3 Standard:** Amazon S3 Standard offers high durability, availability, and performance object storage for frequently accessed data. Because it delivers low latency and high throughput, Amazon S3 Standard is perfect for a wide variety of use cases including cloud applications, dynamic websites, content distribution, mobile and gaming applications, and Big Data analytics.
- **Amazon S3 Standard-IA:** Amazon S3 Standard-Infrequent Access (Amazon S3 Standard-IA) is an Amazon S3 storage class for data that is accessed less frequently, but requires rapid access when needed. Amazon S3 Standard-IA offers the high durability, high throughput, and low latency of Amazon S3 Standard, with a low per GB storage price and per GB retrieval fee. This combination of low cost and high performance make Amazon S3 Standard-IA ideal for long-term storage, backups, and as a data store for disaster recovery.
- **Amazon S3 One Zone-IA:** Amazon S3 One Zone-Infrequent Access (Amazon S3 One Zone-IA) is an Amazon S3 storage class for data that is accessed less frequently, but requires rapid access when needed. Unlike other Amazon object storage classes, which store data in a minimum of three Availability Zones (AZs), Amazon S3 One Zone-IA stores data in a single AZ.
- **Amazon Glacier:** Amazon Glacier is a secure, durable, and extremely low-cost storage service for data archiving. You can reliably store any amount of data at costs that are competitive with or cheaper than on-premises solutions.

Amazon S3 Review



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



To get the most out of Amazon S3, you need to understand a few simple concepts. First, Amazon S3 stores data inside *buckets*. Buckets are essentially the prefix for a set of files, and as such must be uniquely named across all of Amazon S3. Buckets are logical containers for objects. You can have one or more buckets in your account. For each bucket, you can control access, in other words, who can create, delete and list objects in the bucket. You can also view access logs for the bucket, and its objects, and choose the geographical region where Amazon S3 will store the bucket and its contents.

In the example, we've used Amazon S3 to create a bucket in the Tokyo region, which is identified within AWS formally by its region code: "ap-northeast-1").

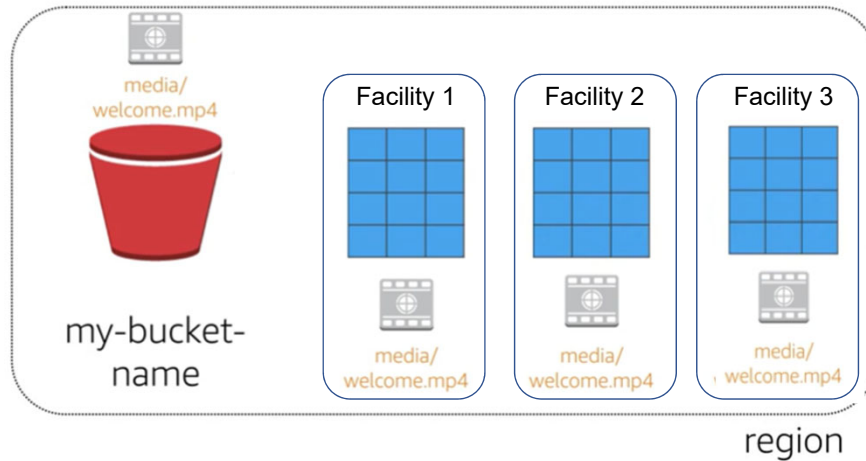
The URL for a bucket is structured as displayed here, with the region code first, followed by amazonaws.com, followed by the bucket name.

Amazon S3 refers to files as *objects*. Once you have a bucket, you can store any number of objects inside of it. An object is composed of data, and any metadata that describes that file. To store an object in Amazon S3, you upload the file you want to store into a bucket.

When you upload a file, you can set permission on the data as well as any metadata. In this example, we're storing the object "Preview2.mp4" inside of our bucket.

The URL for the file includes the object name at the end.

Data Redundantly Stored in Region

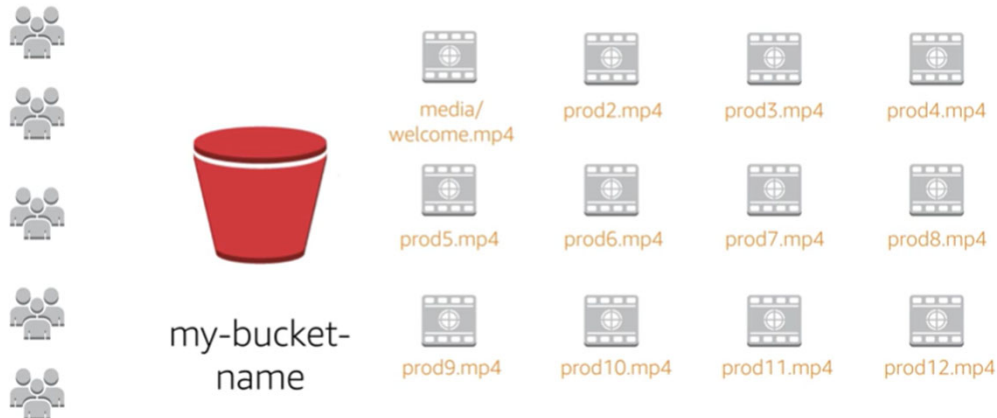


© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



When you create a bucket in Amazon S3, it's associated with a particular AWS Region. Whenever you store data in the bucket, it is redundantly stored across multiple AWS facilities within your selected region. Amazon S3 is designed to durably store your data, even in the case of concurrent data loss in two AWS facilities.

Designed for Seamless Scaling



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Amazon S3 will automatically manage the storage behind your bucket even as your data grows. This allows you to get started immediately and to have your data storage grow with your application needs. Amazon S3 will also scale to handle a high volume of requests. You don't have to provision the storage or throughput, and you'll only be billed for what you use.

Access the Data Anywhere



AWS Management
Console



AWS CLI



AWS SDKs

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



You can access Amazon S3 via the console, AWS CLI, or AWS SDK. Additionally, you can also access the data in your bucket directly via the rest endpoints. These support HTTP or HTTPS access. To support this type of URL-based access, S3 bucket names must be globally unique and DNS-compliant. Also, object keys should be using characters that are safe for URLs.

Common Use Cases



- 📦 Storing application assets
- 📦 Static web hosting
- 📦 Backup and disaster recovery (DR)
- 📦 Staging area for big data
- 📦 *Many more....*



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

This flexibility to store a virtually unlimited amount of data and access that data from anywhere makes Amazon S3 suitable for a wide range of scenarios. Let's look at some use cases for Amazon S3:

- As a location for any application data, Amazon S3 buckets provide that shared location for storing objects that any instances of your application can access, including applications on Amazon EC2 or even traditional servers. This can be useful for user-generated media files, server logs, or other files your application needs to store in a common location. Also, because the content can be fetched directly over the web, you can offload serving of that content from your application and allow clients to directly fetch the data themselves from Amazon S3.
- For static web hosting, Amazon S3 buckets can serve up the static contents of your website, including HTML, CSS, JavaScript, and other files.
- The high durability of Amazon S3 makes it a good candidate to store backups of your data. For even greater availability and disaster recovery capability, Amazon S3 can even be configured to support cross-region replication such that data put into an Amazon S3 bucket in one region can be automatically replicated to another Amazon S3 region. The scalable storage and performance of Amazon S3 make it a great candidate for staging or long-term storage of data you plan to analyze using a variety of big data tools. Given how simple it is to store and access data with Amazon S3, you'll find yourself using it frequently with AWS services and for other parts of your application.

Amazon S3 Pricing



- 📦 Pay only for what you use, including:
 - 📦 GBs per month
 - 📦 Transfer OUT to other regions
 - 📦 PUT, COPY, POST, LIST, and GET requests

- 📦 You do NOT have to pay for:
 - 📦 Transfer IN to Amazon S3
 - 📦 Transfer OUT from Amazon S3 to Amazon CloudFront or Amazon EC2 in the same region.



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Specific costs may vary depending on region and the specific requests made.
As a general rule, you only pay for transfers that cross the boundary of your region, which means you do not pay for transfers to Amazon CloudFront's edge locations within that same region.

Amazon S3: Storage Pricing



To estimate Amazon S3 costs, consider the following:

1. Types of storage classes

- 📦 Standard Storage
 - 📦 99.999999999% durability
 - 📦 99.99% availability
- 📦 Standard-Infrequent Access (SIA)
 - 📦 99.999999999% durability
 - 📦 99.9% availability

2. Amount of storage

- 📦 The number and size of objects
- 📦 Type of storage



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

When you begin to estimate the costs of Amazon S3, you need to consider the following:

1. Storage Class:

- Standard Storage is designed to provide 99.999999999% durability and 99.99% availability.
- Standard – Infrequent Access (SIA) is a storage option within Amazon S3 that you can use to reduce your costs by storing less frequently accessed data at slightly lower levels of redundancy than Amazon S3's standard storage. Standard – Infrequent Access is designed to provide the same 99.999999999% durability as Amazon S3 with 99.9% availability in a given year. It's important to note that each class has different rates.

2. Storage – The number and size of objects stored in your Amazon S3 buckets as well as type of storage.

Amazon S3: Storage Pricing



3. Requests:

- 📦 The number of requests (**GET, PUT, COPY**):
- 📦 Type of requests
 - 📦 Different rates for GET requests than other requests

4. Data Transfer:

- 📦 Pricing based on the amount of data transferred out of the Amazon S3 region
 - 📦 Data transfer in is free, charge for data transfer out



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

- 3. Requests** – The number and type of requests. GET requests incur charges at different rates than other requests, such as PUT and COPY requests.
 - **Get:** Retrieves an object from Amazon S3. You must have READ access to use this operation.
 - **Put:** Adds an object to a bucket. You must have WRITE permissions on a bucket to add an object to it.
 - **Copy:** Creates a copy of an object that is already stored in Amazon S3. A PUT copy operation is the same as performing a GET and then a PUT.
- 4. Data Transfer** – The amount of data transferred out of the Amazon S3 region. Remember that data transfer in is free, but there is a charge for data transfer out.

In Review



- 📦 Amazon S3 is a fully managed cloud storage service
- 📦 Store a virtually unlimited number of objects
- 📦 Pay for only what you use
- 📦 Access at any time, from anywhere
- 📦 Amazon S3 offers rich security controls



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



We've covered an introduction to Amazon S3 including key features and some common use cases.

For more information about Amazon S3, see <https://aws.amazon.com/s3/>.

Amazon EBS and Amazon S3

	 Amazon EBS	 Amazon S3
Paradigm	Block storage with file system	Object store
Performance	Very fast	Fast
Redundancy	Across multiple servers in an Availability Zone	Across multiple facilities in a Region
Security	EBS Encryption – Data volumes and Snapshots	Encryption
Access from the Internet?	No ⁽¹⁾	Yes ⁽²⁾
Typical use case	It is a disk drive	Online storage

(1) Accessible from the Internet if mounted to server and set up as FTP, etc.

(2) Only with proper credentials, unless ACLs are world-readable

This table illustrates the significant differences between Amazon S3 and Amazon EBS. Amazon EBS volumes are network-attached hard drives that can be written to or read from at a block level. Amazon S3 is an object-level storage medium.

This means that you must write whole objects at a time. If you change one small part of a file, you must still rewrite the entire file in order to commit the change to Amazon S3. This can be very time-consuming if you have frequent writes to the same object.

Amazon S3 is optimized for write-once/read-many use cases. The other major difference is cost. With Amazon S3 you pay for what you use, and with Amazon EBS you pay for what you provision.

Amazon S3 Demo

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Please review the Amazon S3 demonstration: M2_S3_S3 v2.0.mp4.

This video demonstration can be found in the learning management system.

Part 3: Amazon Elastic File System (Amazon EFS)

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon Elastic File System (Amazon EFS) provides simple, scalable file storage for use with Amazon EC2 instances in the AWS Cloud. Amazon EFS is easy to use and offers a simple interface that allows you to create and configure file systems quickly and easily.



Amazon EFS

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon EFS provides simple, scalable, elastic file storage for use with AWS services and on-premises resources. It is easy to use and offers a simple interface that allows you to create and configure file systems quickly and easily. Amazon EFS is built to elastically scale on demand without disrupting applications, growing and shrinking automatically as you add and remove files, so your applications have the storage they need, when they need it.

Amazon EFS Features



- 📦 File storage in the AWS cloud
- 📦 Perfect for big data and analytics, media processing workflows, content management, web serving and home directories
- 📦 Petabyte-scale, low latency file system
- 📦 Shared storage
- 📦 Elastic capacity
- 📦 Supports the Network File System versions 4.0 and 4.1 (NFSv4) protocol
- 📦 Compatible with all Linux-based AMIs for Amazon EC2

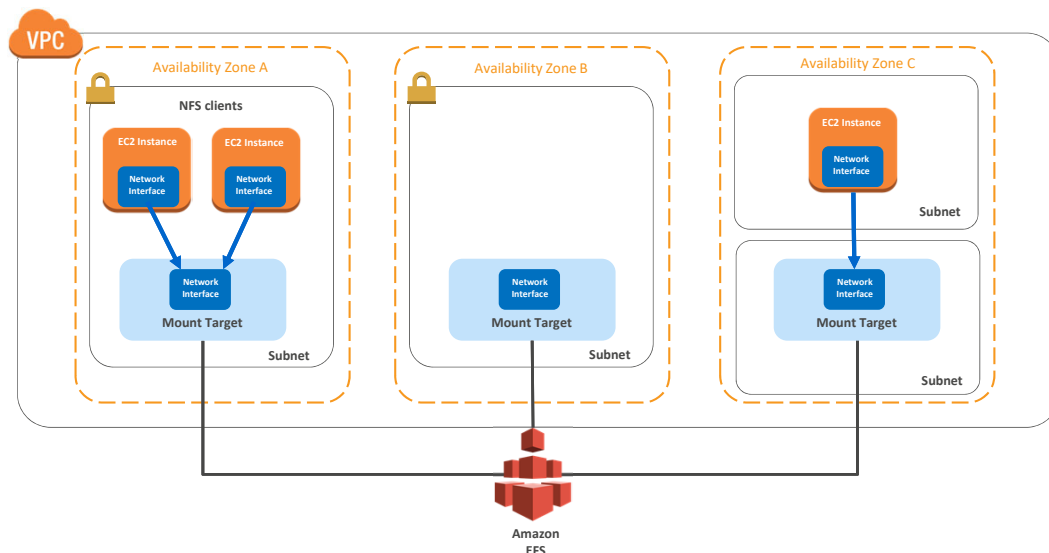


© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon EFS is a fully managed service that makes it easy to set up and scale file storage in the AWS cloud. It is the easiest way to build a file system for big data and analytics, media processing workflows, content management, web serving and home directories. You can create file systems that are accessible to Amazon EC2 instances via a file system interface (using standard operating system file I/O APIs) and that support full file system access semantics (such as strong consistency and file locking).

Amazon EFS file systems can automatically scale from gigabytes to petabytes of data without needing to provision storage. Thousands of Amazon EC2 instances can access an Amazon EFS file system at the same time, and Amazon EFS provides consistent performance to each Amazon EC2 instance. Amazon EFS is designed to be highly durable and highly available. With Amazon EFS, there is no minimum fee or setup costs, and you pay only for the storage you use.

Amazon EFS Architecture



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon EFS provides file storage in the cloud. With Amazon EFS, you can create a file system, mount the file system on an Amazon EC2 instance, and then read and write data from to and from your file system. You can mount an Amazon EFS file system in your VPC, through the Network File System versions 4.0 and 4.1 (NFSv4) protocol.

You can access your Amazon EFS file system concurrently from Amazon EC2 instances in your Amazon VPC, so applications that scale beyond a single connection can access a file system. Amazon EC2 instances running in multiple Availability Zones within the same AWS Region can access the file system, so that many users can access and share a common data source.

In this illustration, the VPC has three Availability Zones, and each has one mount target created in it. We recommend that you access the file system from a mount target within the same Availability Zone. Note that one of the Availability Zones has two subnets. However, a mount target is created in only one of the subnets.

Amazon EFS Implementation



- ① Create your Amazon EC2 resources and launch your Amazon EC2 instance.
- ② Create your Amazon EFS file system.
- ③ Create your target mounts in appropriate subnets.
- ④ Connect your Amazon EC2 instances to target mounts.
- ⑤ Clean up resources and protect your AWS account.

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



There are five steps you need to perform to create and use your first Amazon EFS file system, mount it on an Amazon EC2 instance in your VPC, and test the end-to-end setup.

1. Create your Amazon EC2 resources and launch your instance. (Before you can launch and connect to an Amazon EC2 instance, you need to create a key pair, unless you already have one.)
2. Create your Amazon EFS file system.
3. In the appropriate subnet, create your target mounts.
4. Next, connect to your Amazon EC2 instance and mount the Amazon EFS file system.
5. Finally, clean up your resources and protect your AWS account.

File system

📁 Mount target

- 📁 Subnet ID
- 📁 Security groups
- 📁 One or more per file system
- 📁 Create in a VPC subnet
- 📁 One per Availability Zone
- 📁 Must be in the same VPC

📁 Tags

- 📁 Key-value pairs



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

In Amazon EFS, a file system is the primary resource. Each file system has properties such as ID, creation token, creation time, file system size in bytes, number of mount targets created for the file system, and the file system state.

Amazon EFS also supports other resources to configure the primary resource. These include mount targets and tags.

Mount target: To access your file system, you must create mount targets in your VPC. Each mount target has the following properties:

- The mount target ID
- The subnet ID in which it is created
- The file system ID for which it is created
- An IP address at which the file system may be mounted
- The mount target state.

You can use the IP address or the DNS name in your mount command. Each file system has a DNS name of the following form.

Tags: To help organize your file systems, you can assign your own metadata to each of the file systems you create. Each tag is a key-value pair.

Think of mount targets and tags as subresources that don't exist without being associated with a file system.

In Review



- 📦 Amazon EFS provides file storage over a network
- 📦 Perfect for big data and analytics, media processing workflows, content management, web serving and home directories
- 📦 Fully managed service that eliminates storage administration tasks
- 📦 Accessible from the console, an API, or the CLI
- 📦 Scales up or down as files are added or removed and you pay for what you use.



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

We've covered an introduction to Amazon EFS, including key features and key resources. It provides file storage in the cloud that is perfect for big data and analytics, media processing workflows, content management, web serving and home directories. Amazon EFS scales up or down as files are added or removed and you pay for only what you are using.

Amazon EFS is a fully managed service that is accessible from the console, an API, or the CLI.

For more information about Amazon S3, see <https://aws.amazon.com/efs/>.

Amazon EFS Demo

Please review the Amazon EFS demonstration: M2_S3_EFS v2.0.mp4.

This video demonstration can be found in the learning management system.

Part 4: Amazon Glacier

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon Glacier is a secure, durable, and extremely low-cost cloud storage service for data archiving and long-term backup.



Amazon Glacier

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Let's take a closer look at Amazon Glacier.

Amazon Glacier Review



*Amazon Glacier is a **data archiving service** designed for **security, durability,** and an **extremely low cost**.*

- 📦 Designed for durability of 99.999999999% of objects
- 📦 Supports SSL/TLS encryption of data in transit and at rest
- 📦 The Vault Lock feature enforces compliance via a lockable policy
- 📦 Extremely low-cost design is ideal for long-term archiving
 - 📦 Provides three options for access to archives (Expedited, Standard, and Bulk) from a few minutes to several hours

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Amazon Glacier's data archiving means that although you can store your data at an extremely low cost (even in comparison to Amazon S3), you cannot retrieve your data immediately when you want it. Data stored in Amazon Glacier takes several hours to retrieve, which is why it's ideal for archiving.

There are three key Amazon Glacier terms that you should be familiar with:

- **Archive:** Any object such as a photo, video, file, or document that you store in Amazon Glacier. It is the base unit of storage in Amazon Glacier. Each archive has its own unique ID and can also have a description.
- **Vault:** A container for storing archives. When you create a vault, you specify the vault name and the region in which you would like the vault located.
- **Vault Access Policy:** Determine who can and cannot access the data stored in the vault as well as what operations users can and cannot perform. One vault access permissions policy can be created for each vault to manage access permissions for that vault. You can also use a vault lock policy to make sure a vault cannot be altered. Each vault can have one vault access policy and one vault lock policy attached to it.

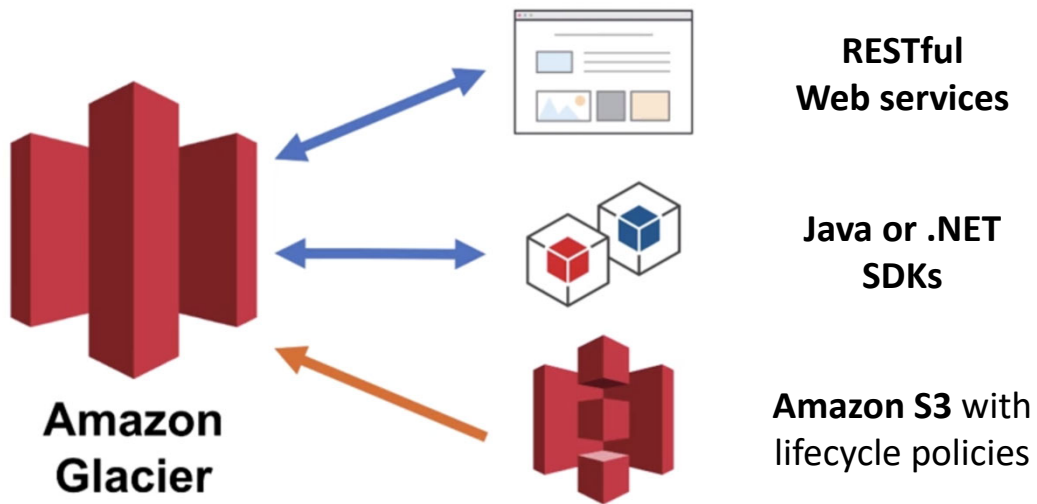
There are three options for retrieving data with varying access times and cost: Expedited, Standard, and Bulk retrievals, as follows:

- Expedited retrievals are typically made available within 1 – 5 minutes (highest cost).
- Standard retrievals typically complete within 3 – 5 hours (less than expedited, more

than bulk).

- Bulk retrievals typically complete within 5 – 12 hours (lowest cost).
Think of it as being like choosing the cost to ship a package.

Using Amazon Glacier



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

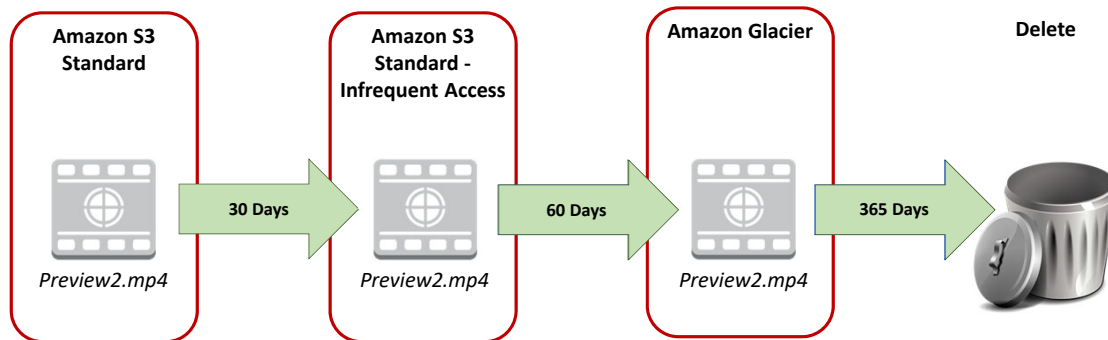
To store and access data in Amazon Glacier, you can use the AWS Management Console; however only a few operations, such as creating and deleting vaults and creating and managing archive policies, are available in the console. Almost all other operations require that you use either the Amazon Glacier REST API, or AWS Java or .NET SDKs to interact with Amazon Glacier via the CLI.

You can also archive data into Amazon Glacier using lifecycle policies. Let's take a closer look at what that means.

Lifecycle Policies



Amazon S3 lifecycle policies allow you to delete or move objects based on age.



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



You should automate the lifecycle of your data stored in Amazon S3. Using lifecycle policies, you can have data cycled at regular intervals between different Amazon S3 storage types. This reduces your overall cost, because you are paying less for data as it becomes less important with time.

In addition to being able to set lifecycle rules per object, you can also set lifecycle rules per bucket.

Let's take a look at an example of a lifecycle policy that moves data as it ages from Amazon S3 Standard to Amazon S3 Standard – Infrequent Access and, finally, into Amazon Glacier before it is deleted. Let's say that the user uploads a video to your application and your application generates a thumbnail preview of the video. This video preview is stored to Amazon S3 Standard, because it is likely that the user will want to access it right away.

Your usage data indicates that most thumbnail previews are not accessed after 30 days. So, your lifecycle policy will take this previews and move them to Amazon S3 infrequent access after 30 days. Once another 30 days have lapsed, it is highly unlikely that it the preview will be accessed again, so it is moved to Amazon Glacier where it remains for 1 year. After one year, the preview is deleted. The important thing to note is that the lifecycle policy manages all of this movement automatically.

For more information, see <http://docs.aws.amazon.com/AmazonS3/latest/dev/object-lifecycle-mgmt.html>.

Storage Comparison



	Amazon S3	Amazon Glacier
Data volume	No limit	No limit
Average latency	ms	min/hrs
Item size	5 TB max	40 TB max
Cost/GB per month	¢¢	¢
Billed requests	PUT, COPY, POST, LIST, and GET	UPLOAD and retrieval
Retrieval pricing	¢ Per request	¢¢ Per request and per GB

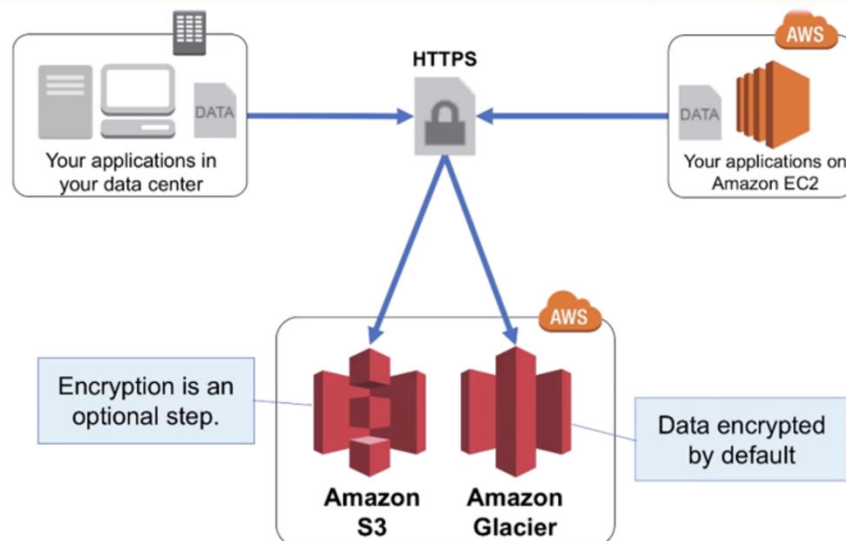
© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



While Amazon S3 and Amazon Glacier are both object storage solutions that allow you to store an unlimited amount of data, there are some critical differences between them that are outlined in this chart.

1. Be careful when deciding which storage solution is correct for your needs. These are two very different services for storage needs. Amazon S3 is designed for frequent, low-latency access to your data, while Amazon Glacier is designed for low-cost, long-term storage of infrequently accessed data.
2. The maximum item size in Amazon S3 is 5 TB, whereas Amazon Glacier can store items up to 40 TB in size.
3. Because Amazon S3 gives you faster access to your data, the storage cost per gigabyte is higher than it is with Amazon Glacier.
4. While both services have per request charges, Amazon S3 charges for PUT, COPY, POST, LIST, GET while Amazon Glacier charges for UPLOAD and retrieval.
5. Because Amazon Glacier was designed for less frequent access to data, it costs more for each retrieval request than Amazon S3. Both the cost per retrieval and the cost per GB are higher for Amazon Glacier.

Server-Side Encryption



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Another important difference between Amazon S3 and Amazon Glacier is how data is encrypted. Server-side encryption is about protecting data at rest. With both solutions, you can securely transfer your data over HTTPS. Any data archived in Amazon Glacier is encrypted by default. With Amazon S3, your application must initiate server-side encryption. There are several ways to accomplish this:

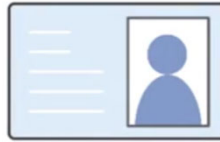
- Server-side encryption with Amazon S3-managed encryption keys (**SSE-S3**) employs strong multi-factor encryption. Amazon S3 encrypts each object with a unique key. As an additional safeguard, it encrypts the key itself with a master key that it regularly rotates. Amazon S3 server-side encryption uses one of the strongest block ciphers available, 256-bit Advanced Encryption Standard (AES-256), to encrypt your data.
- AWS Key Management Service (**AWS KMS**) is a service that combines secure, highly available hardware and software to provide a key management system scaled for the cloud. AWS KMS uses customer master keys (CMKs) to encrypt your Amazon S3 objects. You use AWS KMS via the Encryption Keys section in the IAM console or via AWS KMS APIs to centrally create encryption keys, define the policies that control how keys can be used, and audit key usage to prove they are being used correctly. You can use these keys to protect your data in Amazon S3 buckets.

- Using server-side encryption with customer-provided encryption keys (**SSE-C**) allows you to set your own encryption keys. With the encryption key you provide as part of your request, Amazon S3 manages both encryption (as it writes to disks) and decryption (when you access your objects).

Security with Amazon Glacier



**Amazon
Glacier**



**Control access with
AWS IAM**



**Amazon Glacier encrypts
your data with AES-256**



**Amazon Glacier manages
your keys for you**

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



By default, only you can access your data. You can enable and control access to your data in Amazon Glacier by using AWS IAM. You just set up an AWS IAM policy that specifies user access.

In Review



- ❏ Amazon Glacier is a data archiving service designed for security, durability, and an extremely low cost.
- ❏ Amazon Glacier pricing is region-based.
- ❏ Extremely low-cost design is ideal for long-term archiving.
- ❏ The service is designed for durability of 99.999999999% of objects.

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



We've covered an introduction to Amazon Glacier including key differences between Amazon S3 and Amazon Glacier.

For more information about Amazon Glacier, see <https://aws.amazon.com/glacier/>.

Amazon Glacier Demo

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Please review the Amazon Glacier demonstration: M2_S3_Glacier v2.0.mp4.

This video demonstration can be found in the learning management system.

Section 2.0.2 Review:



- 📦 Reviewed the characteristics of Amazon EBS, Amazon S3, Amazon EFS, and Amazon Glacier
- 📦 Identified appropriate uses for each storage options
- 📦 Briefly looked at the pricing difference for each storage option

To finish this module:

- 📦 Complete:  **Knowledge Assessment**

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

In review, we:

- Discussed storage services including Amazon EBS, Amazon S3, Amazon EFS, and Amazon Glacier
- Reviewed use cases for storage options
- Reviewed storage pricing

To finish this module, please complete the lab and the corresponding knowledge assessment.

Up Next: Unit 2.03 – AWS Core Services - VPC

Now that we have a better understanding some of the storage services offered by AWS, in unit 2.03 we look at Amazon Virtual Private Cloud (**Amazon VPC**). It lets you provision a logically isolated section of the AWS Cloud where you can launch AWS resources in a virtual network that you define.



Thanks for participating!

© 2018 Amazon Web Services, Inc. or its affiliates. All rights reserved. This work may not be reproduced or redistributed, in whole or in part, without prior written permission from Amazon Web Services, Inc. Commercial copying, lending, or selling is prohibited. Corrections or feedback on the course, please email us at: aws-course-feedback@amazon.com. For all other questions, contact us at: <https://aws.amazon.com/contact-us/aws-training/>. All trademarks are the property of their owners.

