CCMP 600 – Fundamentals of Cloud Computing

Cloud Data Management

Mayra Samaniego MSc. Ph.D.

Outline

o Learning Outcome

o Technology Terminology

o AWS Service Review

o Amazon EBS

o Amazon S3

o Amazon EFS

o Amazon S3 Glacier

Learning Outcome

LO4. Explain cloud data management.

By the end of this lesson, you should be able to

- Identify the different type of storage
- Explain a cloud data storage service
- Identify the functionalities of a cloud data storage service
- Differentiate between types of cloud data storage services

SASKATCHEWAN
POLYTECHNIC

Technology Terminology

Amazon Elastic Block Store (Amazon EBS)

Amazon Elastic Compute Cloud (Amazon EC2)

Hard disk drive (HDD)

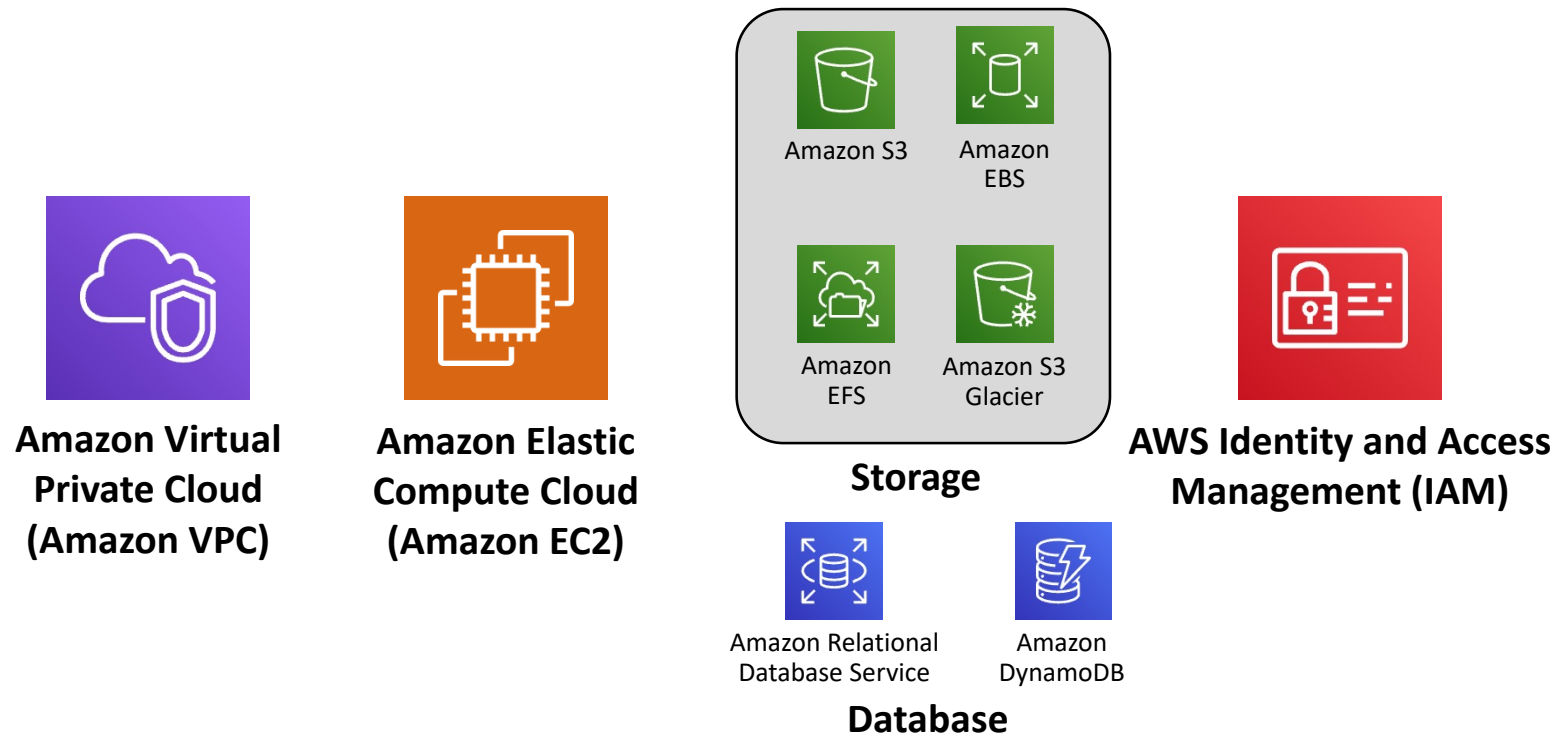Input/Output Operations Per Second (IOPS)

Solid State Drive (SSD)

SASKATCHEWAN POLYTECHNIC

AWS storage options

An HDD is a lower-speed storage option that uses a spinning disk to store data.

An SSD is a high-speed storage option that uses flash memory instead of a spinning disk.

Cloud Computing and Blockchain - Mayra Samaniego MSc. Ph.D.

SASKATCHEWAN POLYTECHNIC

AWS Core Services

Amazon Virtual Private Cloud (Amazon VPC)

Amazon Elastic Compute Cloud (Amazon EC2)

Amazon S3

Amazon EBS

Amazon EFS

Amazon S3 Glacier

Storage

Amazon Relational Database Service

Amazon DynamoDB

Database

AWS Identity and Access Management (IAM)

SASKATCHEWAN POLYTECHNIC
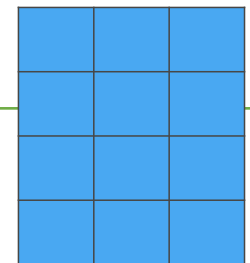
AWS storage options

Block Storage

- What happens if you want to change one character in a 1-GB file?
- With **block** storage, you change only the block that contains the character. With object storage, the entire file must be updated.
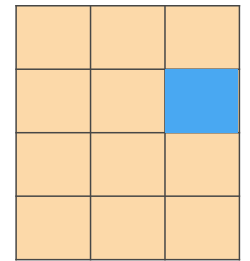
Block storage

Object Storage

- What happens if you want to change one character in a 1-GB file?
- With **object** storage, the entire file must be updated.

Object storage

AWS storage options

One critical difference between some storage types is whether they offer block-level storage or object-level storage.

This difference has a major effect on the throughput, latency, and cost of your storage solution.

Block storage solutions are typically faster and use less bandwidth, but they can cost more than object-level storage.

SASKATCHEWAN
POLYTECHNIC

Amazon Elastic Block Store (Amazon EBS)

Amazon EBS provides persistent block storage volumes for use with Amazon EC2 instances.

- Persistent storage is any data storage device that retains data after power to that device is shut off.
- It is also sometimes called non-volatile storage.

Each Amazon EBS volume is automatically replicated within its Availability Zone

- This is done to protect you from component failure.

Amazon EBS is designed for high availability and durability.

SASKATCHEWAN POLYTECHNIC

Amazon Elastic Block Store (Amazon EBS)

Amazon EBS volumes provide the consistent and low-latency performance that is needed to run your workloads.
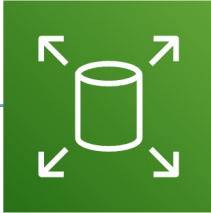
With Amazon EBS, you can scale your usage up or down within minutes, while paying a low price for only what you provision.

Because they are directly attached to the instances, they can provide 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.

SASKATCHEWAN
POLYTECHNIC

Amazon Elastic Block Store (Amazon EBS)

Amazon EBS volumes are included as part of the backup of your instances into Amazon Machine Images (or AMIs).
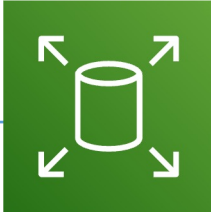
- AMIs are stored in Amazon S3 and can be reused to create new Amazon EC2 instances later.

A backup of an Amazon EBS volume is called a snapshot.

- The first snapshot is called the baseline snapshot.
- Any other snapshot after the baseline captures only what is different from the previous snapshot.

Amazon Elastic Block Store (Amazon EBS)

Amazon EBS volumes uses include:

- Boot volumes and storage for Amazon EC2 instances
- Data storage with a file system
- Database hosts
- Enterprise applications

SASKATCHEWAN POLYTECHNIC

Class Activity

Objective

- The objective of this activity is to learn what Amazon EBS is, and what the benefits, features, and use cases of the four types of EBS volumes are.

Instructions

- Discuss the benefits of EBS.
- Complete the "Class Activity - All About Amazon EBS.docx" sheet.

SASKATCHEWAN POLYTECHNIC

Benefits of Amazon Elastic Block Store (Amazon EBS)

Data availability

- When you create an EBS volume in an Availability Zone, it is automatically replicated within that zone to prevent data loss due to the failure of any single hardware component.

- After you create a volume, you can attach it to any EC2 instance in the same Availability Zone.

- An EBS volume can be attached to only one instance at a time, but multiple volumes can be attached to a single instance.

- An EBS volume and the instance to which it attaches must be in the same Availability Zone.

SASKATCHEWAN POLYTECHNIC

# Benefits of Amazon Elastic Block Store (Amazon EBS)

Data persistence

- An EBS volume is off-instance storage that can persist independently from the life of an instance.
  - You continue to pay for the volume usage if the data persists.

- EBS volumes that are attached to a running instance can automatically detach from the instance with their data intact when the instance is terminated.
  - The volume can then be reattached to a new instance, permitting quick recovery.

SASKATCHEWAN
POLYTECHNIC

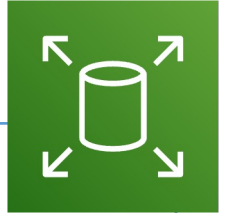# Benefits of Amazon Elastic Block Store (Amazon EBS)

Data encryption
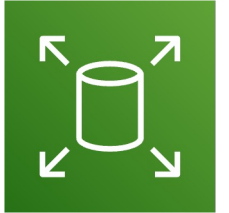
- For simplified data encryption, you can create encrypted EBS volumes with the Amazon EBS encryption feature.

- All EBS volume types support encryption.
  - You can use encrypted EBS volumes to meet a wide range of data-at-rest encryption requirements for regulated or audited data and applications.

SASKATCHEWAN POLYTECHNIC

Benefits of Amazon Elastic Block Store (Amazon EBS)

Snapshots

- Amazon EBS provides the ability to create snapshots (backups) of any EBS volume and write a copy of the data in the volume to Amazon S3, where it is stored redundantly in multiple Availability Zones.

- Snapshots of encrypted EBS volumes are automatically encrypted.

- When you create a new volume from a snapshot, it's an exact copy of the original volume at the time the snapshot was taken.

- EBS volumes that are restored from encrypted snapshots are automatically encrypted.

SASKATCHEWAN POLYTECHNIC

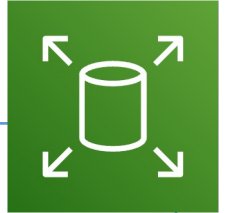# Benefits of Amazon Elastic Block Store (Amazon EBS)

Snapshots

- Amazon EBS provides the ability to create snapshots (backups) of any EBS volume and write a copy of the data in the volume to Amazon S3, where it is stored redundantly in multiple Availability Zones.

- Snapshots of encrypted EBS volumes are automatically encrypted.

- When you create a new volume from a snapshot, it's an exact copy of the original volume at the time the snapshot was taken.

- EBS volumes that are restored from encrypted snapshots are automatically encrypted.

SASKATCHEWAN POLYTECHNIC

Homework

Objective

- The objective of this homework is to Analyze a use case and recommend the best type of virtual storage for a particular situation.

Instructions

- Analyze the use cases presented in the "Homework - AmazonEBSUseCases.docx" sheet.

SASKATCHEWAN POLYTECHNIC

Lab: Attaching an EBS Volume

Objective

- The objective of this lab is to attach an EBS volume to an EC2 instance.

Instructions

- Follow the instructions on the Lab: Attaching an EBS Volume of Module 6 on the AWS Academy Introduction to Cloud: Semester 1 course.

SASKATCHEWAN POLYTECHNIC

Amazon Simple Storage Service (Amazon S3)

Amazon S3 is object-level storage, which means that if you want to change a part of a file, you must make the change and then re-upload the entire modified file.

Amazon S3 stores data as objects within resources that are called buckets.

SASKATCHEWAN
POLYTECHNIC

Amazon Simple Storage Service (Amazon S3)

Amazon S3 is a managed cloud storage solution that is designed to scale seamlessly and provide 11 9s of durability.

- You can store virtually as many objects as you want in a bucket, and you can 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 TB in size.

By default, data in Amazon S3 is stored redundantly across multiple facilities and multiple devices in each facility.

SASKATCHEWAN POLYTECHNIC

Amazon Simple Storage Service (Amazon S3)

The data that you store in Amazon S3 is not associated with any server

- You do not need 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.

Amazon Simple Storage Service (Amazon S3)

Objects can be almost any data file, such as

- images
- videos
- server logs.

Because Amazon S3 supports objects as large as several terabytes in size, you can even store database snapshots as objects.

Amazon S3 also provides low-latency access to the data over the internet by Hypertext Transfer Protocol (HTTP) or Secure HTTP (HTTPS).

- Therefore, you can retrieve data anytime from anywhere.

You can also access Amazon S3 privately through a virtual private cloud (VPC) endpoint.

Amazon Simple Storage Service (Amazon S3)

You get fine-grained control over who can access your data by using

- AWS Identity and Access Management (IAM) policies
- Amazon 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.

You can access Amazon S3 through

- the web-based AWS Management Console
- programmatically through the API and SDKs
- with third-party solutions, which use the API or the SDKs.

SASKATCHEWAN
POLYTECHNIC

Amazon Simple Storage Service (Amazon S3)

Amazon S3 includes event notifications that enable you to set up automatic notifications when certain events occur:

- When an object is uploaded to a bucket or deleted from a specific bucket.
- These notifications can be sent to you or used to trigger other processes, such as AWS Lambda functions.

With storage class analysis, you can analyze storage access patterns and transition the right data to the right storage class.

SASKATCHEWAN
POLYTECHNIC

Amazon Simple Storage Service (Amazon S3)

The Amazon S3 Analytics feature automatically identifies the optimal lifecycle policy to transition less frequently accessed storage to Amazon S3 Standard – Infrequent Access (Amazon S3 Standard-IA).

- You can configure a storage class analysis policy to monitor an entire bucket, a prefix, or an object tag.

Storage class analysis also provides daily visualizations of your storage usage in the AWS Management Console.

- You can export them to an Amazon S3 bucket to analyze by using the business intelligence (BI) tools of your choice, such as Amazon QuickSight.

SASKATCHEWAN POLYTECHNIC

Amazon S3 Storage Classes

Amazon S3 Standard

Amazon S3 Intelligent-Tiering

Amazon S3 Standard-Infrequent Access (Amazon S3 Standard-IA)

Amazon S3 One Zone-Infrequent Access (Amazon S3 One Zone-IA)

Amazon S3 Glacier

Amazon S3 Glacier Deep Archive

Homework

Objective

- The objective of this activity is to analyze the Amazon S3 Storage Classes

Instructions

- Read about Amazon S3 storage classes in the links below
- [Amazon S3 Storage Classes](#)
- [Using Amazon S3 storage classes](#)

SASKATCHEWAN
POLYTECHNIC

Amazon Simple Storage Service (Amazon S3)

To use Amazon S3 effectively, you must understand a few simple concepts.

- First, Amazon S3 stores data inside buckets.

Buckets are essentially the prefix for a set of files and must be **uniquely** named across all of Amazon S3 globally.

- Buckets are logical containers for objects. You can have one or more buckets in your account.
- You can control access for each bucket. Who can
  - create
  - delete
  - and list objects in the bucket.

- You can also view access logs for the bucket and its objects.
- You can choose the geographical region where Amazon S3 stores the bucket and its contents.

SASKATCHEWAN POLYTECHNIC

Amazon S3 bucket URLs (two styles)

Amazon S3 bucket URLs

Amazon S3

[bucket name]

Preview2.mp4

Tokyo Region (ap-northeast-1)
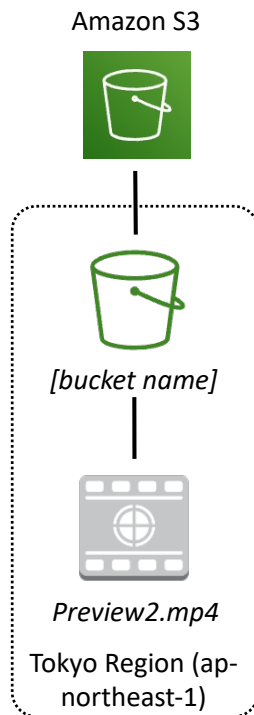
To upload your data:

1. Create a **bucket** in an AWS Region.

2. Upload almost any number of **objects** to the bucket.

Bucket path-style URL endpoint:

https://s3.*ap-northeast-1*.amazonaws.com/*bucket-name*

Region code   Bucket name

Bucket virtual hosted-style URL endpoint:

https:// *bucket-name*.*s3-ap-northeast-1*.amazonaws.com

Bucket name   Region code

SASKATCHEWAN POLYTECHNIC

Class Activity

Objective

- The objective of this activity is to analyze the Amazon S3 URLs

Instructions

- Read about Amazon S3 URLs in the link bellow
- [Methods for accessing a bucket](#)

SASKATCHEWAN POLYTECHNIC

Amazon S3 Redundantly stored in Regions

When you create a bucket in Amazon S3, it is associated with a specific AWS Region.

- When you store data in the bucket, it is redundantly stored across multiple AWS facilities within your selected Region.

media/welcome.mp4

my-bucket-name

Facility 1	Facility 2	Facility 3
media/ welcome.mp4	media/ welcome.mp4	media/ welcome.mp4

Region

SASKATCHEWAN
POLYTECHNIC
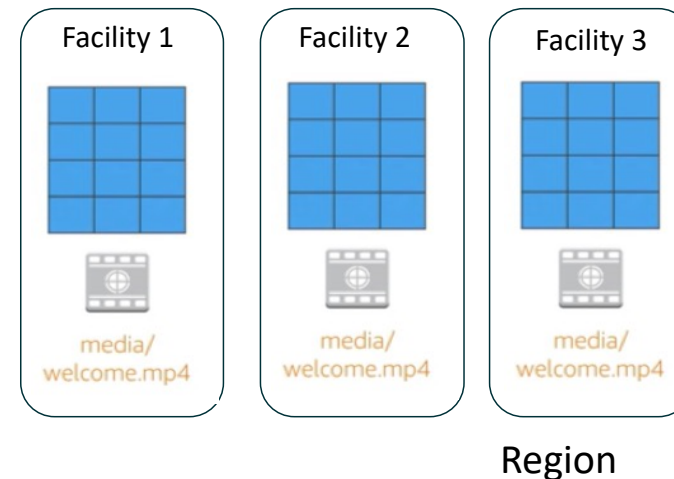
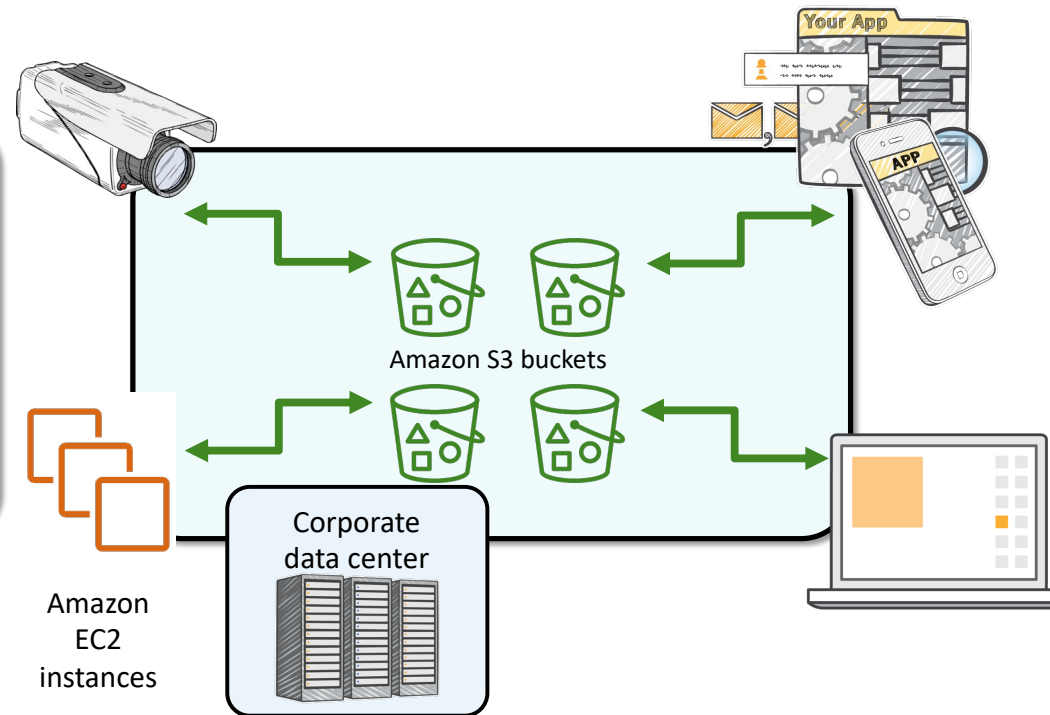# Amazon S3 Common Use Cases

Storing application assets

Static web hosting

Backup and disaster recovery (DR)

Staging area for big data

Amazon S3 Common Scenarios

- Backup and storage
- Application hosting
- Media hosting
- Software delivery

Amazon S3 buckets

Amazon EC2 instances

Corporate data center

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 pay for –

- Transfers IN to Amazon S3
- Transfers OUT from Amazon S3 to Amazon CloudFront or Amazon EC2 in the same Region

Amazon S3 Pricing

To estimate Amazon S3 costs, consider the following:

Storage class type –

- Standard storage is designed for:
  - 11 9s of durability
  - Four 9s of availability
- S3 Standard-Infrequent Access (S-IA) is designed for:
  - 11 9s of durability
  - Three 9s of availability

Amount of storage –

- The number and size of objects

Amazon S3 Pricing

Requests –

- The number and type of requests (GET, PUT, COPY)
- Type of requests:
  - Different rates for GET requests than other requests.

Data transfer –

- Pricing is based on the amount of data that is transferred out of the Amazon S3 Region
  - Data transfer in is free, but you incur charges for data that is transferred out.

Amazon Elastic File System (Amazon EFS)

Amazon Elastic File System (Amazon EFS) provides simple, scalable, elastic file storage for use with AWS services and on-premises resources.

- It offers a simple interface that enables you to create and configure file systems quickly and easily.

Amazon EFS is built to dynamically scale on demand without disrupting applications—it will grow and shrink automatically as you add and remove files.

- It is designed so that your applications have the storage they need when they need it.

SASKATCHEWAN
POLYTECHNIC

Amazon EFS Features

Shared storage

Elastic capacity

Petabyte-scale, low-latency file system

- Amazon EFS file systems can automatically scale from gigabytes to petabytes of data without the need to provision storage.

Amazon EFS Common Use Cases

You can use Amazon EFS to build a file system for:

- big data and analytics

- media processing workflows

- content management

- web serving

- home directories

SASKATCHEWAN POLYTECHNIC

Comparison among Amazon S3, EBS, and EFS

AMAZON S3	AMAZON EBS	AMAZON EFS
Can be publicly accessible	Accessible only via an EC2 Instance	Accessible via several EC2 Instances and AWS services
Web interface	File system interface	Web and file system interface
Object Storage	Block Storage	File storage
Scalable	Hardly scalable	Scalable
Slower than EBS and EFS	Faster than S3 and EFS	Faster than S3, slower than EBS
Good for backups and static data	It is meant to be an EC2 drive	Good for applications and shareable workloads

SASKATCHEWAN POLYTECHNIC

Amazon S3 Glacier

Amazon S3 Glacier is a data archiving service that is designed for security, durability, and an extremely low cost.

Amazon S3 Glacier is designed to provide 11 9s of durability for objects.

It supports the encryption of data in transit and at rest.

Amazon S3 Glacier

When you use Amazon S3 Glacier to archive data, you can store your data at an extremely low cost (even in comparison to Amazon S3)

- but you cannot retrieve your data immediately when you want it.

Data that is stored in Amazon S3 Glacier can take several hours to retrieve, which is why it works well for archiving.

SASKATCHEWAN
POLYTECHNIC

Amazon S3 Glacier Use Cases

Media asset archiving

- Media assets—such as video and news footage—require durable storage and can grow to many petabytes over time.
- Amazon S3 Glacier enables you to archive older media content affordably and then move it to Amazon S3 for distribution when needed.

Healthcare information archiving

- To meet regulatory requirements, hospital systems must retain petabytes of patient records—such as:
  - Low-Income Subsidy (LIS) information
  - Picture archiving and communication system (PACS) data
  - Electronic Health Records (EHR)—for decades.

- Amazon S3 Glacier can help you reliably archive patient record data securely at a very low cost.

Amazon S3 Glacier Use Cases

Regulatory and compliance archiving

- Many enterprises, like those in financial services and healthcare, must retain regulatory and compliance archives for extended durations.
- Amazon S3 Glacier Vault Lock can help you set compliance controls so you can work towards meeting your compliance objectives.

Scientific data archiving

- Research organizations generate, analyze, and archive large amounts of data.
- By using Amazon S3 Glacier, you can reduce the complexities of hardware and facility management and capacity planning.

SASKATCHEWAN POLYTECHNIC

Amazon S3 Glacier Use Cases

Digital preservation

- Libraries and government agencies must handle data integrity challenges in their digital preservation efforts.
- Amazon S3 Glacier performs regular, systematic data integrity checks, and it is designed to be automatically self-healing.

Magnetic tape replacement

- On-premises or offsite tape libraries can lower storage costs, but they can require large upfront investments and specialized maintenance.
- Amazon S3 Glacier has no upfront cost and reduces the cost and burden of maintenance.

SASKATCHEWAN POLYTECHNIC

References

- Amazon EBS

- Amazon EBS Volume Types

- Amazon S3

- Tutorial: Configuring a static website on Amazon S3

- AWS S3 Glacier