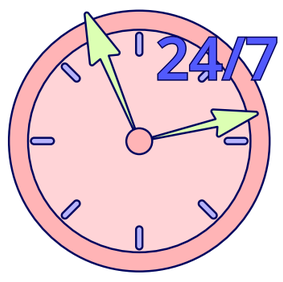
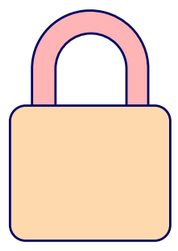
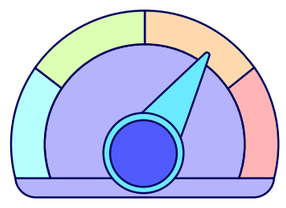
Key Benefits of AWS Storage

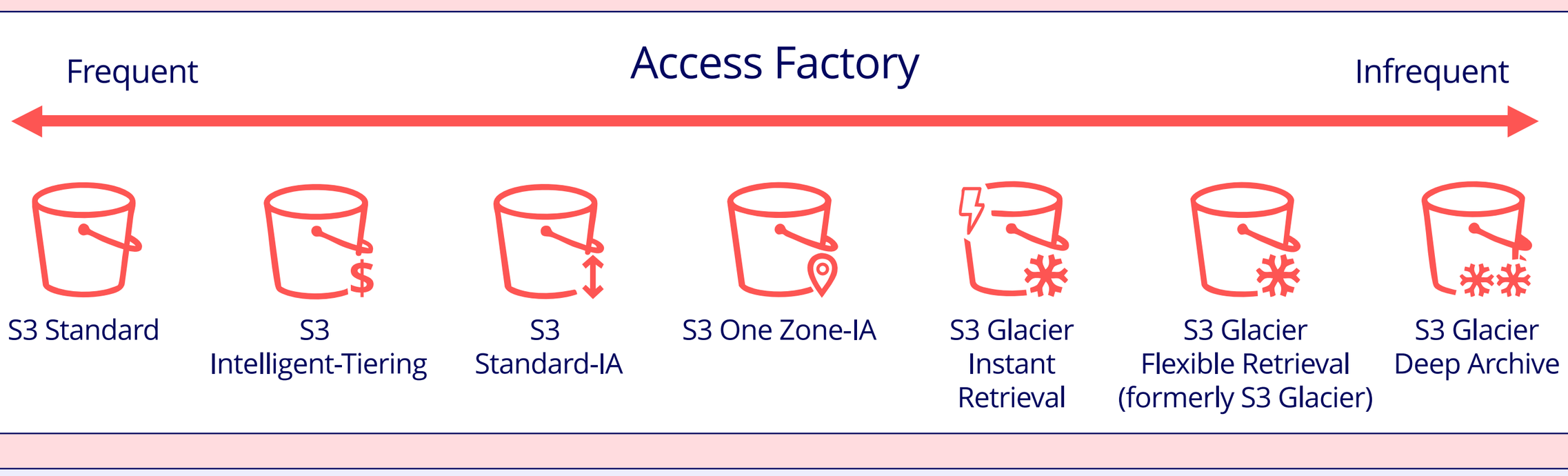
 Availability	 Durability	 Security	 Performance

Popular Services

Simple Storage Service (S3)

- S3 is a scalable storage service that allows us to store any object in a bucket identified through a unique key. A bucket has virtually unlimited storage capacity.
- S3 offers multiple storage classes to optimize storage costs based on access frequency.
- S3 allows us to define life cycle policies to automatically transfer the data to a different storage class based on access patterns to optimize costs.

Storage Class	Purpose	First-Byte Latency
S3 Standard	Used for general purpose and frequently accessed data.	Milliseconds
S3 Express One zone	Stores data in one AZ, only improving access latency.	Single digit millisecond
S3 intelligent Tiering	Automatically moves data between multiple access tiers to save costs.	Milliseconds
S3 Standard Infrequent Access	Stores infrequently accessed data.	Milliseconds
S3 One Zone Infrequent Access	Stores infrequently accessed data in one availability zone.	Milliseconds
S3 Glacier Instant Retrieval	Used for long-term archival, which is instantly retrievable.	Milliseconds
S3 Glacier Flexible Retrieval	Used for long-term archival accessed a few times a year.	Minutes to hours
S3 Glacier Deep Archive	Used for long-term data retention accessed rarely in a year.	Hours

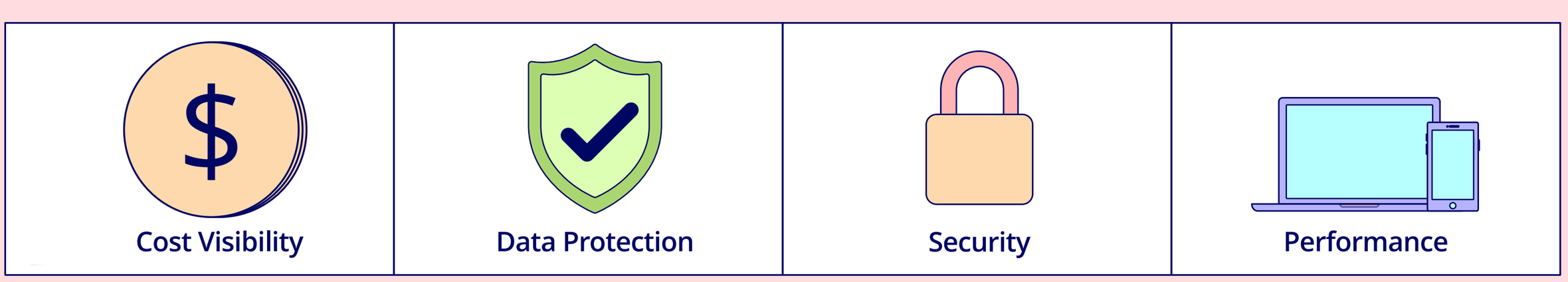


- S3 provides four options to encrypt data.

Type	Description	Use Case
Server-side encryption (SSE)	S3 manages the encryption process.	Data protection in transit.
SSE-KMS (AWS KMS)	Uses AWS Key Management Service for encryption.	Enhanced control and auditing for data in transit.
SSE-C (customer-provided keys)	The customer manages encryption keys.	Full control by the customer.
Client-side encryption (CSE)	Data is encrypted on the client side before upload.	Max control by customers for sensitive data.

S3 Storage Lens

- Helps optimize the storage and performance utilization in S3 buckets.
- Provides a single view of buckets across hundreds of accounts.
- Provides an interactive dashboard to view the metrics.

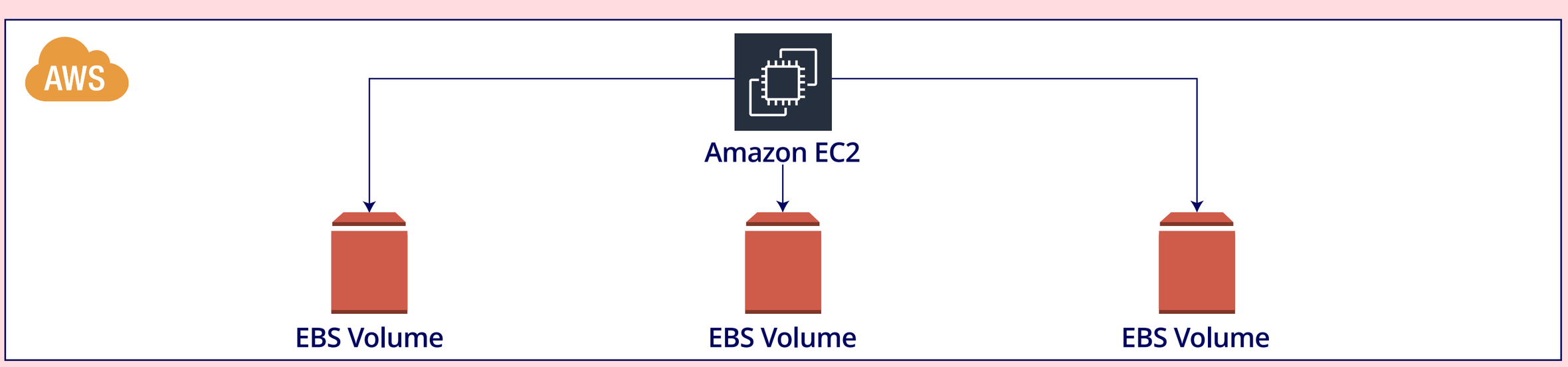


Elastic Block Store (EBS)

- A block storage service that offers volumes that can be mounted to EC2 instances.
- EBS volumes retain data after the termination of the EC2 instance they are mounted to and can be mounted to another EC2 instance later on.
- Typically used as primary storage and boot volumes. We can back up EBS snapshots in S3 buckets.
- To optimize costs, EBS offers multiple types of EBS volumes based on the I/O frequency.

Volume	Purpose	Max IOPS
General purpose SSD (gp2, gp3)	Used as general-purpose storage volumes.	16000
Provisioned IOPS (io1, io2)	Used for workloads with frequent read-write operations such as databases.	64000
Throughput Optimized HDD (st1)	Used for large sequential I/O workloads such as data warehouses.	500
Cold HDD (sc1)	Used for long-term retention of data.	250

- EBS uses KMS to encrypt volumes and snapshots.
- The encryption is done at the EC2 instances to which volumes are attached.

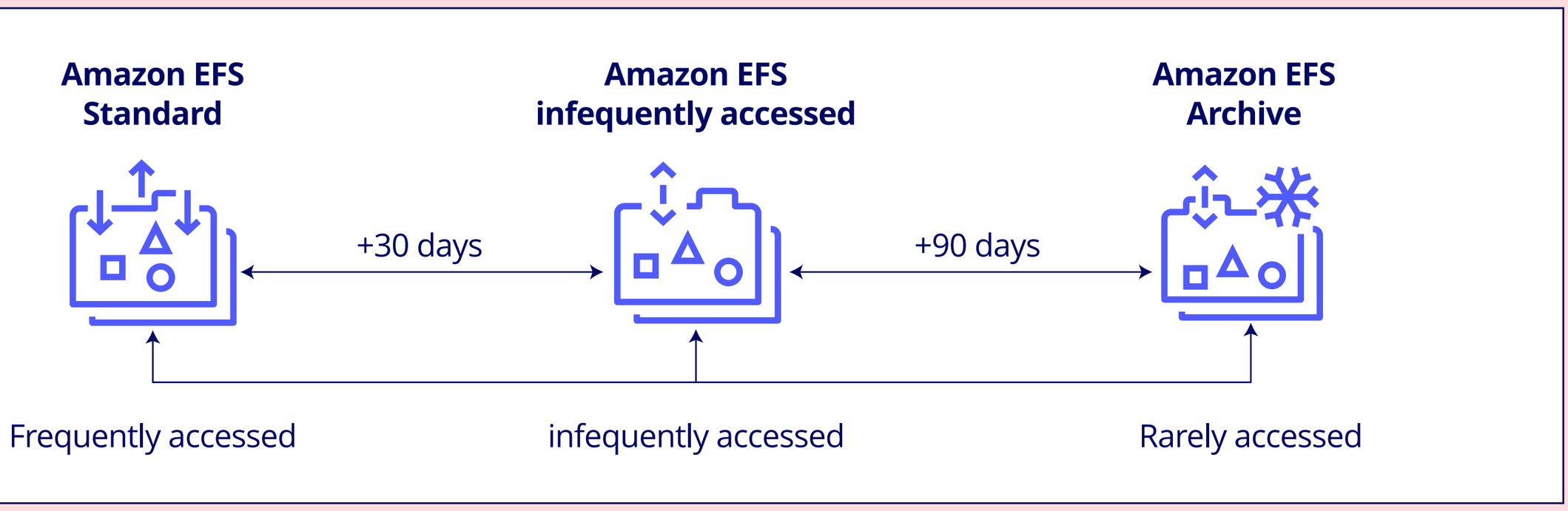


Elastic File System (EFS)

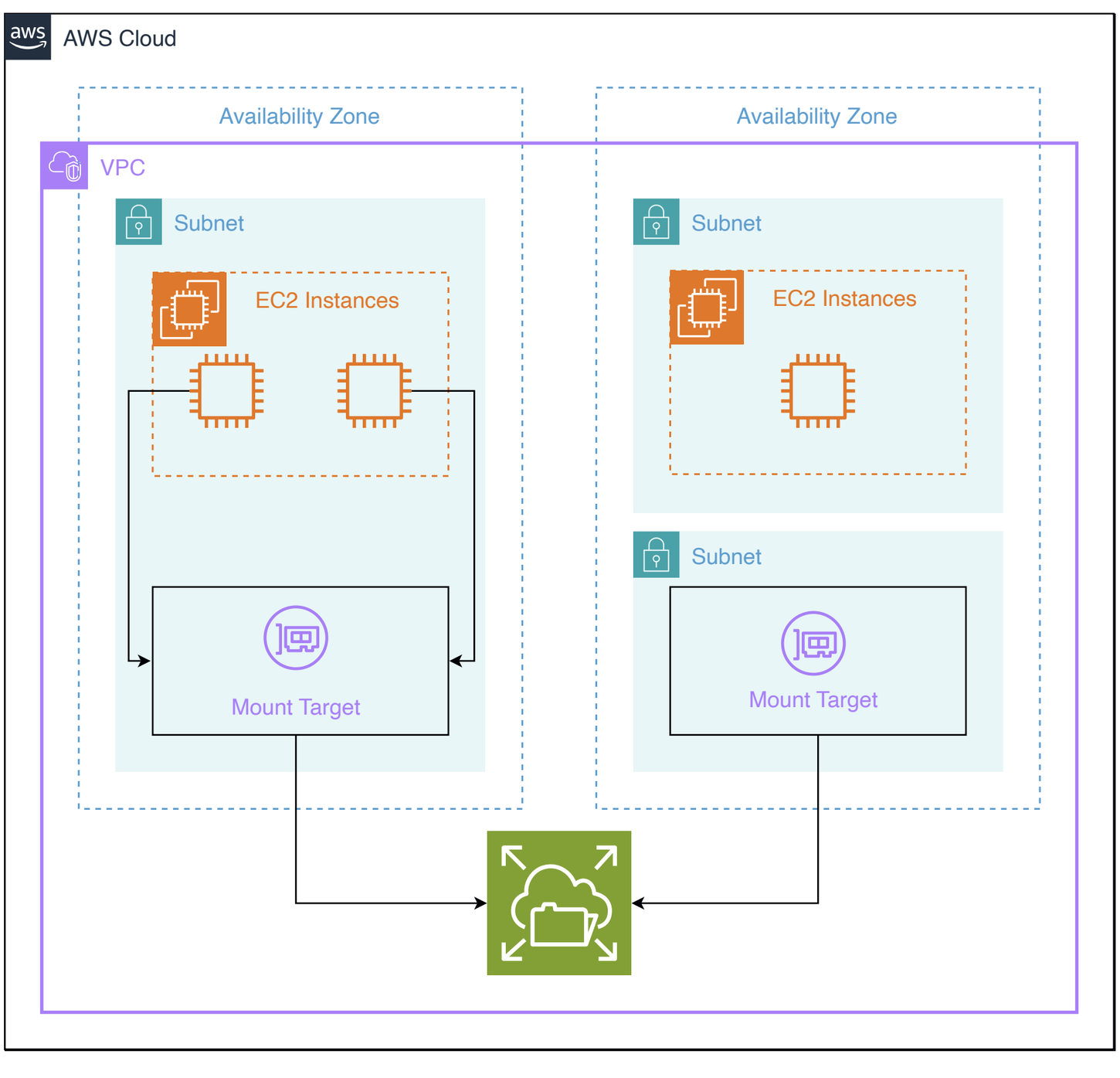
- Allows us to create network file systems (NFS).
- We can simultaneously mount the NFS system to multiple instances across a region.
- Connects to EC2 instances through minute targets and communicates using the NFS protocol.
- Can scale up to store petabytes of data.
- Supports two types of modes:

Mode	Purpose
Bursting	Adjusts throughput based on the system's size.
Provisioned	Fixed amount of throughput.

- EFS offers three types of storage classes. It also allows us to define life cycle policies to automate the data transfer based on access patterns.



- Amazon EFS supports two file system encryption forms: data encryption in transit and rest.
- Data in transit is encrypted by using transport layer security (TLS) when mounting a file system using the Amazon EFS mount helper.
- AWS manages encryption at rest and does not require the user to change the application.



AWS Storage Services

educative

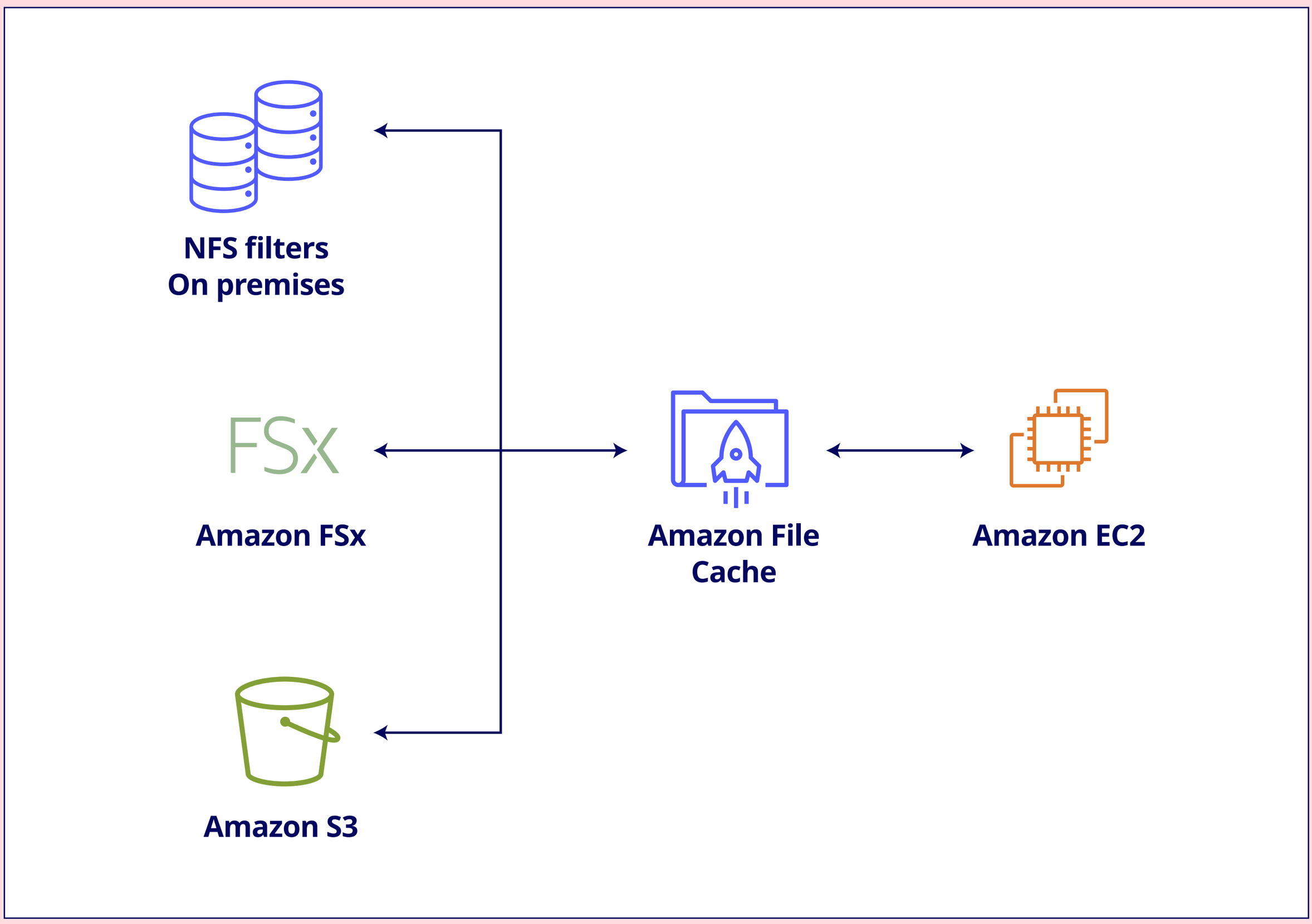
File System (FSx)

- Allows us to run high-performance file systems on the cloud.
- Can connect to an on-premises system using AWS Direct Connect or VPN.
- Offers four different types of file systems.

FSx for NetApp ONTAP	FSx for OpenZFS	FSx for Windows File Server	FSx for Lustre
Supports multiprotocol access to data using NFS, SMB, and iSCSI.	Accessible by EC2 instance over NFS protocol.	Supports SMB, Windows NTFS and integration of Active Directory AD.	A file system optimized for processing large workloads in parallel.

FileCache

- Offers a high-speed cache that can store data and link to an Amazon Simple Storage Service (Amazon S3) or network file system (NFS) data repository
- Used to provide a unified view of dispersed data to reduce access latency for compute-intensive workloads.
- We need the Lustre client on the machine before mounting the cache



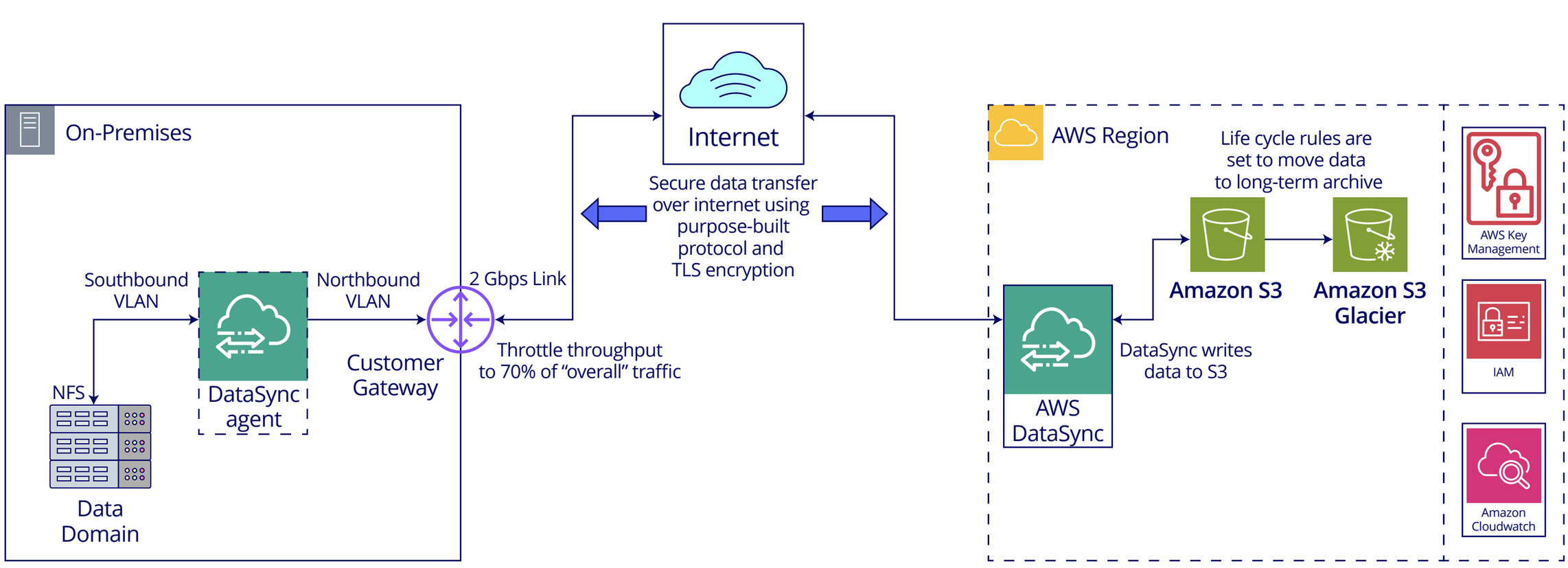
Key Benefits of AWS Storage

DataSync

- Allows us to migrate on-premises data to AWS S3, EFS, and FSx with end-to-end security over AWS Direct Connect.
- Requires DataSync Agent installed on the source system.
- Created a data transfer task that specifies the destination and source of data.
- A single DataSync task can fully utilize 10 Gbps over a network link between the on-premises environment and AWS.
- Charges for the amount of data copied.

Migrating hundreds of TB of data to Amazon S3 with AWS DataSync:

The following is a high-level architecture diagram of the implementation of Autodesk's DataSync solution:



Workflow:

1. Deploy DataSync Agent (OVA file) on-premises for local storage access.
2. Secure data transfer over the internet using purpose-built protocol and TLS encryption.
3. DataSync services write data to S3.
4. Life cycle rules are set to move data to long-term archival (Glacier).

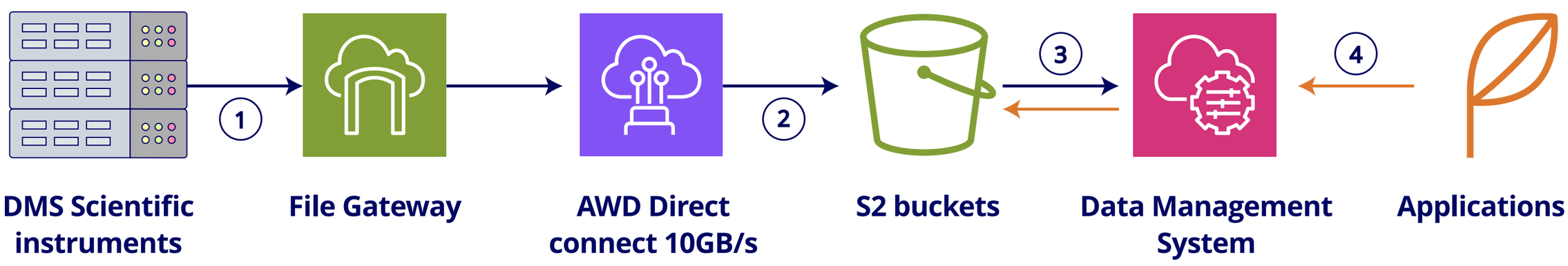
Storage Gateways

- Enables hybrid storage between on-premises and on-cloud storage.

File Gateway for S3	File Gateway for FSx	Volume Gateway	Tape Gateway
Used to retrieve objects in S3 using SMB and NFS protocol.	Seamlessly integrates on-premise and AWS FSx storage.	Allows using disk volumes through the iSCSI block protocol.	Allows us to archive backup data on Amazon Glacier.

How Bristol Myers Squibb Uses Amazon S3 and AWS Storage Gateway to Manage Scientific Data:

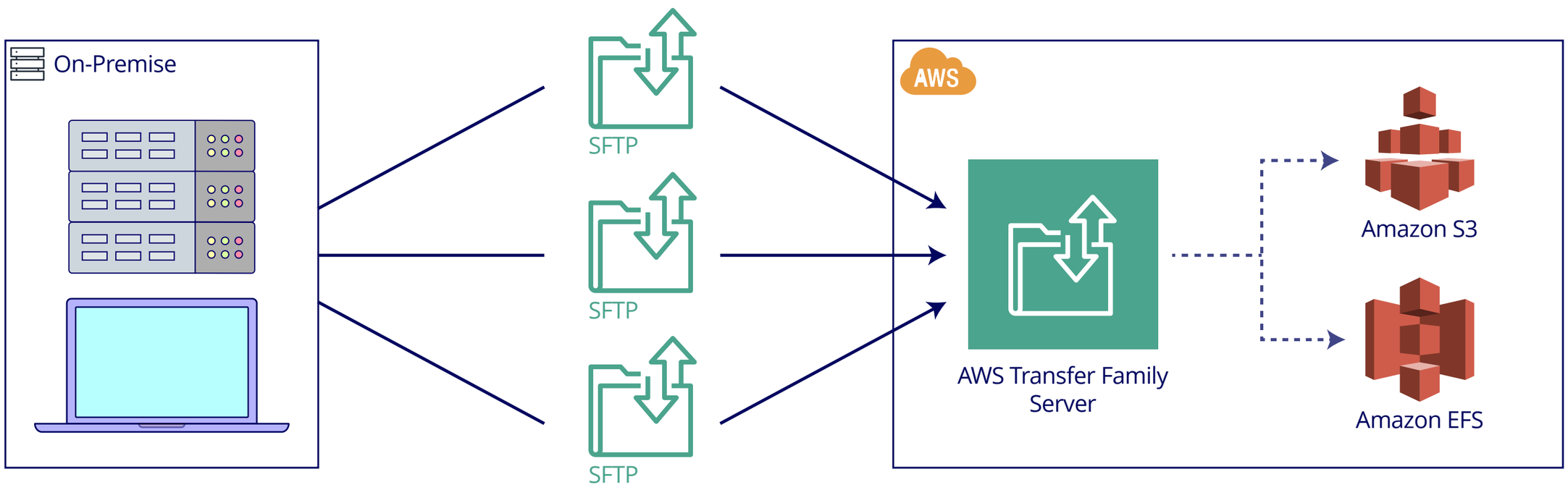
- Petabytes of raw data from scientific instruments, transferred to a Storage Gateway file share.
- S3 standard is used for most data for easy and quick access. Idle data that needs to be archived for compliance purposes is stored in S3 Glacier.
- Uses S3 Intelligent tiering to automatically move data among different S3 classes based on access frequency.



1. Instruments write raw data into AWS Storage Gateway File Share.
2. AWS Storage Gateway transfers files to Amazon S3 buckets.
3. Data Management System scans S3 buckets regularly.
4. Applications request data via Data Management System's meta catalog.

Transfer Family

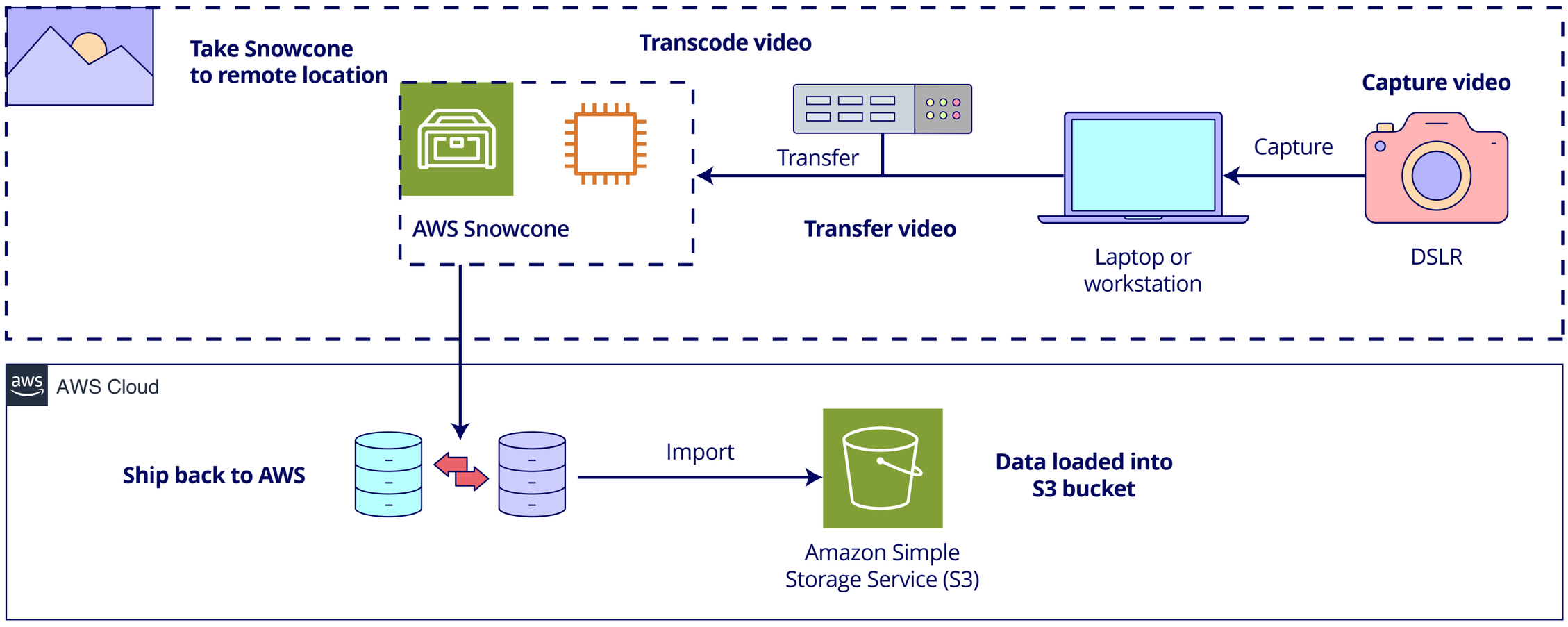
- Allows secure data transfer between multiple storage services without the need to manage the file transfer protocol infrastructure.
- Supports multiple file transfer protocols, including FTP, FTPS, and SFTP.
- Easily integrates with S3 and EFS.
- Charges on an hourly basis after the configuration of the server endpoint. Offers different pricing structures for different protocols.



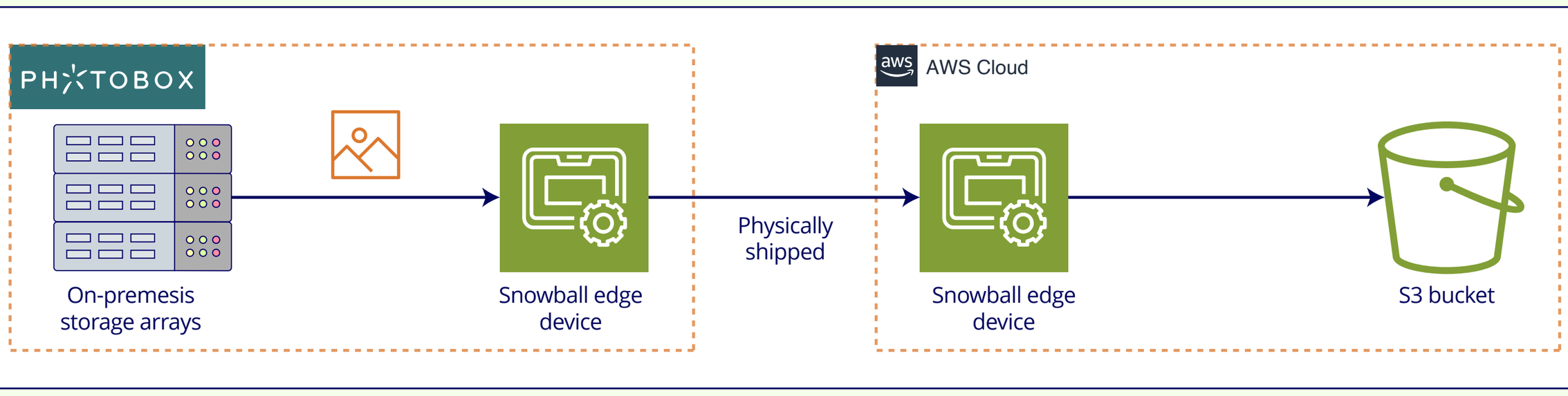
Snow Family

- Allows secure transfer of data into and out of AWS.
  - **Snowcone**
    - It is a portable, secure device with edge computing and data transfer services.
    - Transfers data by shipping the device to AWS offline or through DataSync online.
    - Can be used to collect data at edge devices such as sensors or machines and transfer it to AWS storage services.

AWS Storage Services



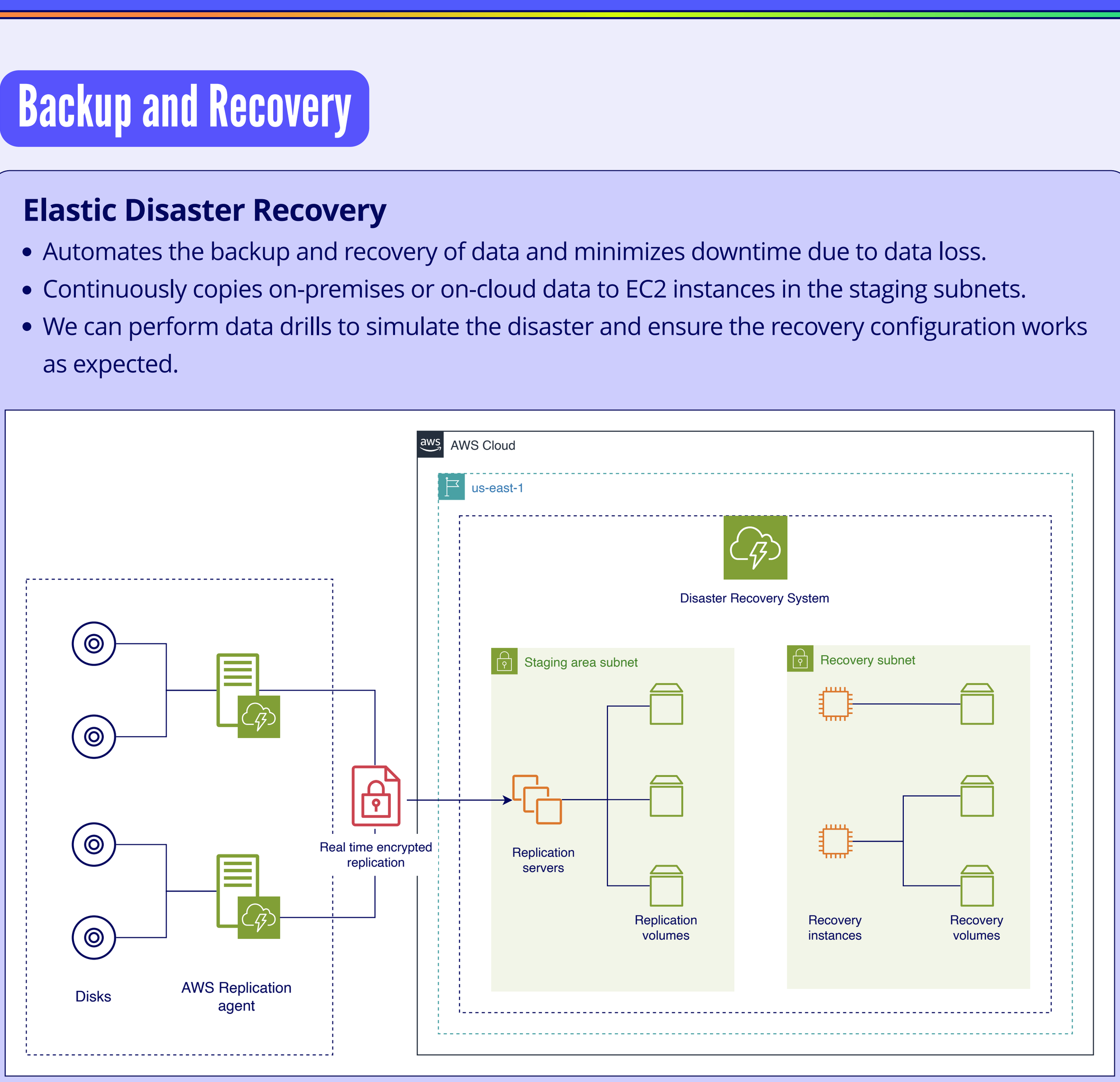
- **Snowball**
  - A compute-optimized device that can migrate petabytes of data in limited network conditions.
  - Offers two types of snowball devices: compute-optimized and storage-optimized.
  - Can be used to collect data at extremely remote locations and shipped back.
  - To save on costs, Photobox migrated 10PB of its data from its EMC Isilon and IBM
  - Cleversafe on-premises storage arrays to Amazon S3 using Snowball edge devices.



Backup and Recovery

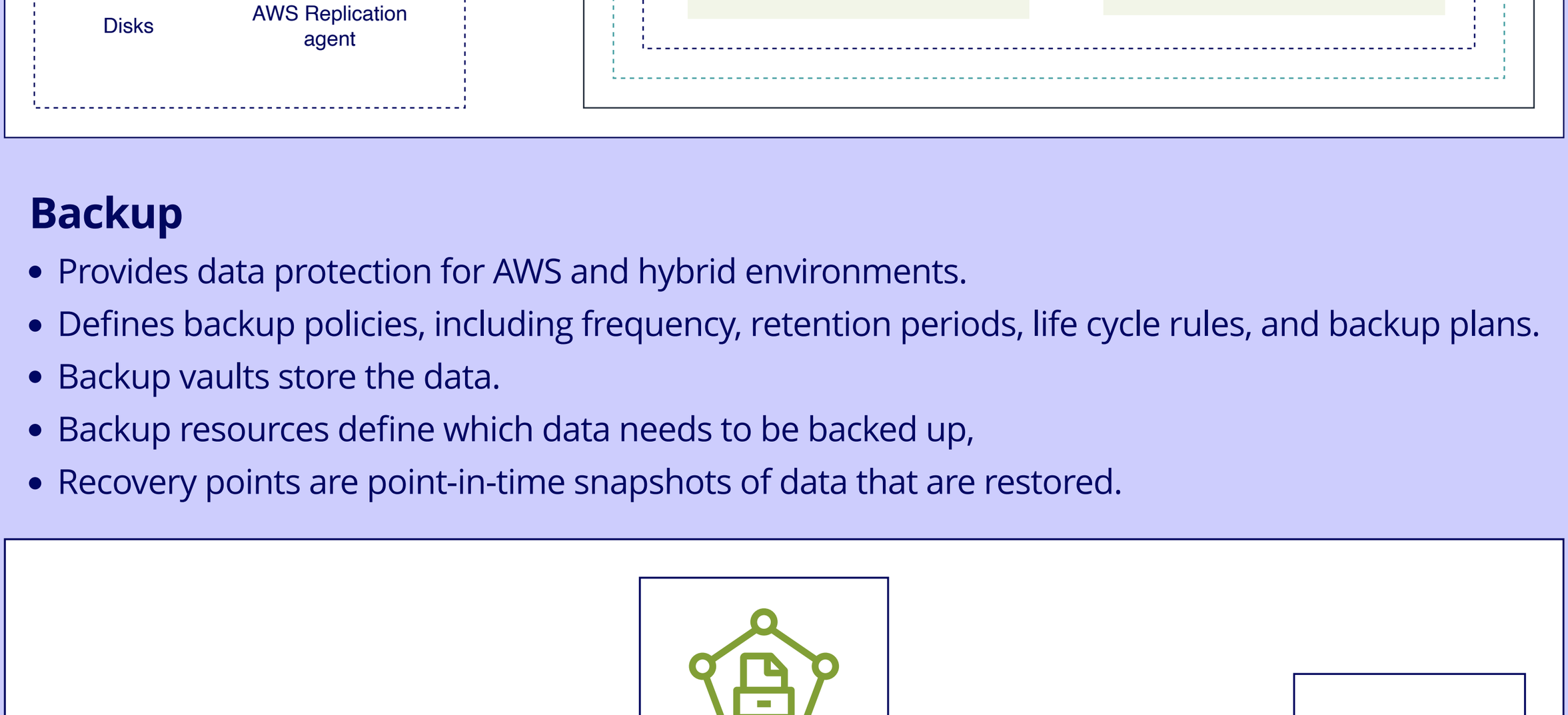
Elastic Disaster Recovery

- Automates the backup and recovery of data and minimizes downtime due to data loss.
- Continuously copies on-premises or on-cloud data to EC2 instances in the staging subnets.
- We can perform data drills to simulate the disaster and ensure the recovery configuration works as expected.



Backup

- Provides data protection for AWS and hybrid environments.
- Defines backup policies, including frequency, retention periods, life cycle rules, and backup plans.
- Backup vaults store the data.
- Backup resources define which data needs to be backed up,
- Recovery points are point-in-time snapshots of data that are restored.



Cost Optimization

Identify Your Needs

For backup/long-term retention	• EBS offers cold HDD (sc1) to retain data long-term in block format. • S3 offers S3 Glacier, Flexible Retrieval, S3 Glacier Instant Retrieval, and S3 Glacier Deep Archive for long-term data retention in object form.
For data lake	• S3 stores a large amount of data economically. Additionally, it can easily integrate with AWS Data services.
Access frequency/request rate	• S3 stores a large amount of data economically. Additionally, it can easily integrate with AWS Data services. • S3 and EFS volumes offer multiple storage classes.

Performance Optimization

- **Parallelization:** Upload and download objects from an S3 bucket in parts while establishing multiple concurrent connections.
- **Prefixes:** Adjust prefixes in the bucket such that no prefix throttles during peak hours and traffic is divided among multiple prefixes.
- **S3 Storage lens:** Continuously monitor the performance of an S3 bucket and ways to optimize it.