# Amazon S3

* Amazon S3 is one of the main building blocks of AWS
* It’s advertised as ”inﬁnitely scaling” storage
* Many websites use Amazon S3 as a backbone
* Many AWS services use Amazon S3 as an integration as well

# Use Cases:

* Backup and storage
* Disaster Recovery
* Archive
* Hybrid Cloud storage
* Application hosting
* Media hosting
* Data lakes G big data analytics
* Software delivery
* Static website

# Buckets:

* Amazon S3 allows people to store objects (ﬁles) in “buckets” (directories)
* Buckets must have a globally unique name (across all regions all accounts)
* Buckets are deﬁned at the region level
* S3 looks like a global service but buckets are created in a region

# Objects

* Objects (ﬁles) have a Key
* Just keys with very long names that contain slashes (“/”)
* Object values are the content of the body:
  + Max. Object Size is 5TB (5000GB)
  + If uploading more than 5GB, must use “multi-part upload”
* Metadata (list of text key / value pairs – system or user metadata)
* Tags (Unicode key / value pair – up to 10) – useful for security

/ lifecycle

* Version ID (if versioning is enabled

# Security of S3

User-Based:

* IAM Policies – which API calls should be allowed for a speciﬁc user from IAM

Resource-Based:

* Bucket Policies – bucket wide rules from the S3 console - allows cross account
* Object Access Control List (ACL) – ﬁner grain (can be disabled)
* Bucket Access Control List (ACL) – less common (can be disabled)

Note: an IAM principal can access an S3 object if

* The user IAM permissions ALLOW it OR the resource policy ALLOWS it
* AND there’s no explicit DENY

Encryption: encrypt objects in Amazon S3 using encryption keys

# Bucket Policies:

**JSON based policies:**

* Resources: buckets and objects
* Effect: Allow / Deny
* Actions: Set of API to Allow or Deny
* Principal: The account or user to apply the policy to

# Use S3 bucket for policy to:

* Grant public access to the bucket
* Force objects to be encrypted at upload
* Grant access to another account (Cross Account

|  |  |  |  |
| --- | --- | --- | --- |
| {  "Version": "2012-10-17", | | | |
| "Statement": [  {  "Sid": "publicRead",  "Effect": "Allow", | | | |
|  |  |  | "principal": "\*", |
|  |  |  | "Action": [ |
|  |  |  | "s3:GetObject" |
|  |  |  | ], |
|  |  |  | "resource": [ |
|  |  |  | "arn:aws:s3:::samplebucket/\*" |
|  |  |  | ] |
|  |  | } |  |
|  | ] |  |  |
| } |  |  |  |

# Static Website Hosting:

* S3 can host static websites and have them accessible on the Internet
* If you get a 403 Forbidden error, make sure the bucket policy allows public reads

# Task:

* Install Visual Studio code
* Create a ﬁle “index.html” and “!” and press tab
* Add some content and images in a folder
* Upload all the ﬁles in the folder to S3
* Enable static website hosting in Bucket Properties

# Versioning:

* You can version your ﬁles in Amazon S3
* It is enabled at the bucket level
* Same key overwrite will change the “version”: 1, 2, 3….
* It is best practice to version your buckets
* Protect against unintended deletes (ability to restore a version) • Easy roll back to previous version

Note:

* Any ﬁle that is not versioned prior to enabling versioning will have version “null”
* Suspending versioning does not delete the previous versions

# Replication:

* Must enable Versioning in source and destination buckets
* Cross-Region Replication (CRR)
* Same-Region Replication (SRR)
* Buckets can be in different AWS accounts
* Copying is asynchronous
* Must give proper IAM permissions to S3

Use cases:

* CRR – compliance, lower latency access, replication across accounts
* SRR – log aggregation, live replication between production and test account

# S3 Storage Classes

* Amazon S3 Standard - General Purpose
* Amazon S3 Standard-Infrequent Access (IA)
* Amazon S3 One Zone-Infrequent Access
* Amazon S3 Glacier Instant Retrieval
* Amazon S3 Glacier Flexible Retrieval
* Amazon S3 Glacier Deep Archive
* Amazon S3 Intelligent Tiering

Can move between classes manually or using S3 Lifecycle conﬁgurations

# S3 Durability & Availability Durability:

* High durability (99.999999999%, 11 9’s) of objects across

multiple AZ

* If you store 10,000,000 objects with Amazon S3, you can on average expect to incur a loss of a single object once every 10,000 years
* Same for all storage classes

# Availability:

* Measures how readily available a service is
* Varies depending on storage class
* Example: S3 standard has 99.99% availability = not available 53 minutes a year

# S3 Standard - General Purpose

* 99.99% Availability
* Used for frequently accessed data
* Low latency and high throughput
* Sustain 2 concurrent facility failures

Use Cases: Big Data analytics, mobile & gaming applications, content distribution…

# S3 Storage Classes - Infrequent Access

* For data that is less frequently accessed, but requires rapid access when needed
* Lower cost than S3 Standard

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

* 99.9% Availability
* Use cases: Disaster Recovery, backups

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

* High durability (99.999999999%) in a single AZ; data lost when AZ is destroyed
* 99.5% Availability

Use Cases: Storing secondary backup copies of on-premises data, or data you can recreate

# Amazon S3 Glacier Storage Classes

* Low-cost object storage meant for archiving / backup
* Pricing: price for storage + object retrieval cost

Amazon S3 Glacier Instant Retrieval

* Millisecond retrieval, great for data accessed once a quarter
* Minimum storage duration of 90 days

Amazon S3 Glacier Flexible Retrieval (formerly Amazon S3 Glacier):

* Expedited (1 to 5 minutes), Standard (3 to 5 hours), Bulk (5 to 12 hours) – free
* Minimum storage duration of 90 days
* Amazon S3 Glacier Deep Archive – for long term storage: • Standard (12 hours), Bulk (48 hours)
* Minimum storage duration of 180 days

# S3 Intelligent-Tiering

* Small monthly monitoring and auto-tiering fee
* Moves objects automatically between Access Tiers based on usage
* There are no retrieval charges in S3 Intelligent-Tiering
* Frequent Access tier (automatic): default tier
* Infrequent Access tier (automatic): objects not accessed for 30 days
* Archive Instant Access tier (automatic): objects not accessed for 90 days
* Archive Access tier (optional): conﬁgurable from 90 days to 700+ days
* Deep Archive Access tier (optional): conﬁg. from 180 days to 700+ days

# Shared Responsibility Model for S3:

|  |  |
| --- | --- |
| **AWS** | **Customer** |
| Infrastructure (global security, durability, availability, sustain concurrent loss of data in two facilities) | S3 Versioning |

|  |  |
| --- | --- |
| Conﬁguration and vulnerability analysis | S3 Bucket Policies |
| Compliance validation | S3 Replication Setup |
|  | Logging and Monitoring |
|  | S3 Storage Classes |
|  | Data encryption at rest and in transit |

**AWS Snow Family**

Highly-secure, portable devices to collect and process data at the edge, and migrate data into and out of AWS

# Data Migration:

* Snowcone
* Snowball Edge
* Snowmobile

# Edge Computing

* Snowcone
* Snowball Edge

# Challenges for Data Migrations:

* Limited Connectivity
* Limited Bandwidth
* High Network Cost
* Shared Bandwidth (Can’t maximize the line)
* Connection Stability

# Data Migration with Snow Family:

Ofﬂine devices to perform data migrations If it takes more than a week to transfer over the network, use Snowball devices!

# Snowball Edge (for data transfers)

* Physical data transport solution: move TBs or PBs of data in or out of AWS
* Alternative to moving data over the network (and paying network fees)
* Pay per data transfer job
* Provide block storage and Amazon S3 -compatible object storage

Snowball Edge Storage Optimized

* 80 TB of HDD capacity for block volume and S3 compatible object storage

Snowball Edge Compute Optimized

* 42 TB of HDD or 28TB NVMe capacity for block volume and S3 compatible object storage

Use cases: large data cloud migrations, DC decommission, disaster recovery

# AWS Snowcone & Snowcone SSD:

* Small, portable computing, anywhere, rugged & secure, withstands harsh environments
* Light (4.5 pounds, 2.1 kg)
* Device used for edge computing, storage, and data transfer
* Snowcone – 8 TB of HDD Storage
* Snowcone SSD – 14 TB of SSD Storage
* Use Snowcone where Snowball does not ﬁt (space - constrained environment)
* Must provide your own battery / cables
* Can be sent back to AWS ofﬂine, or connect it to internet and use AWS DataSync to send data

# AWS Snowmobile

* Transfer exabytes of data (1 EB = 1,000 PB = 1,000,000 TBs)
* Each Snowmobile has 100 PB of capacity (use multiple in parallel)
* High security: temperature controlled, GPS, 24/7 video surveillance
* Better than Snowball if you transfer more than 10 PB

# Snow Family Usage Process:

* Request Snowball devices from the AWS console for delivery
* Install the snowball client / AWS OpsHub on your servers
* Connect the snowball to your servers and copy ﬁles using the client
* Ship back the device when you’re done (goes to the right AWS facility)
* Data will be loaded into an S3 bucket
* Snowball is completely wiped

# What is Edge Computing:

* Process data while it’s being created on an edge location
* These locations may have
  + Limited / no internet access
  + Limited / no easy access to computing power
* We setup a Snowball Edge / Snowcone device to do edge computing
* Use cases of Edge Computing:
  + Preprocess data
  + Machine learning at the edge
  + Transcoding media streams
* Eventually (if need be) we can ship back the device to AWS (for transferring data for example)

# Snow Family - Edge Computing

* Snowcone & Snowcone SSD (smaller)
  + 2 CPUs, 4 GB of memory, wired or wireless access
  + USB-C power using a cord or the optional battery

Snowball Edge – Compute Optimized

* + 104 vCPUs, 416 GiB of RAM
  + Optional GPU (useful for video processing or machine learning)
  + 28TB NVMe or 42TB HDD usable storage
  + Storage Clustering available (up to 16 nodes)

Snowball Edge – Storage Optimized

* + Up to 40 vCPUs, 80 GiB of RAM, 80 TB storage

All: Can run EC2 Instances & AWS Lambda functions (using AWS IoT Greengrass)

Long-term deployment options: 1 and 3 years discounted pricing

# AWS OpsHub

* Historically, to use Snow Family devices, you needed a CLI (Command Line Interface tool)
* Today, you can use AWS OpsHub (a software you install on your computer / laptop) to manage your Snow Family Device
  + Unlocking and conﬁguring single or clustered devices
  + Transferring ﬁles
  + Launching and managing instances running on Snow Family Devices
  + Monitor device metrics (storage capacity, active instances on your device)
  + Launch compatible AWS services on your devices (ex: Amazon EC2 instances, AWS DataSync, Network File System (NFS)

# AWS Storage Cloud Native Options:

* Block → Amazon EBS & EC2 Instance Store
* File → Amazon EFS
* Object → Amazon S3 & Glacier

# Amazon S3 - Summary

* Buckets vs Objects:
  + global unique name, tied to a region
* S3 security:
  + IAM policy, S3 Bucket Policy (public access), S3 Encryption
* S3 Websites:
  + host a static website on Amazon S3
* S3 Versioning:
  + multiple versions for ﬁles, prevent accidental deletes
* S3 Replication:
  + same-region or cross-region, must enable versioning
* S3 Storage Classes:
  + Standard, IA, 1Z-IA, Intelligent, Glacier (Instant, Flexible, Deep)
* Snow Family:
  + Import data onto S3 through a physical device, edge computing
* OpsHub:
  + desktop application to manage Snow Family devices
* Storage Gateway:
  + hybrid solution to extend on-premises storage to S3