Storage

Đà Nẵng, năm COVID thứ 2

Instance Storage types for EC2:

- Instance Store: ephemeral
- Elastic Block Store (EBS): GP-SSD (default), P-IOPS, HDD, …
- Elastic File System (EFS): NAS.

Elastic Block Storage (EBS):

- EBS volumes can be attached to your EC2 instances and seen as disks by the OS. EBS Snapshots:
- You can create point-in-time snapshots of EBS volumes, which are persisted to Amazon S3 (but you don’t see them in your S3 bucket).
- Subsequent snapshots are incremental: only the blocks on the device that have changed after your most recent snapshot are saved.

EBS Volume Types:

- Provisioned IOPS (io1, io2, io2BE):
  - SSD
  - IOPS: 256K for io2BE, 64K for io1/io2. 64K is only achievable on Nitro instances. Other instances are capped to 32K IOPS.
  - 4,000 MB/s for io2BE, 1,000 MB/s for io1/io2
  - You can provision from 100 IOPS up to 64,000 IOPS per volume on Nitro-based Instances and up to 32,000 on other instances.
  - maximum ratio of provisioned IOPS to volume size (in GiB) is 50:1 for io1 volumes, and 500:1 for io2 volumes.

- General Purpose (gp2, gp3):
  - SSD
  - 16K IOPS.
  - 1,000 MB/s for gp3 and 250 MB/s for gp2.
  - The performance of gp2 volumes is tied to volume size, which determines the baseline performance level of the volume and how quickly it accumulates I/O credits; larger volumes have higher baseline performance levels and accumulate I/O credits faster.

- Throughput Optimized HDD (st1):
  - HDD
  - 500 IOPS
  - 500 MB/s
  - Cannot be used as boot drive
- Cold HDD (sc1):
  - HDD
  - 250 IOPS
  - 250 MB/s
  - Cannot be used as boot drive
- Magnetic:
  - Previous generation HDD.
  - 40-200 IOPS
  - 90 MB/s

Elastic File System (EFS):

- Simple, petabyte scalable NFS file storage for EC2 instances.
- Protocol: NFS v4.
- Elastic: automatically grows and strings as you add or remove files.
- Stored redundantly across multiple AZs,
- Can be accessed concurrently by 1000s of EC2 instances from different AZs.
- On-premise supported via direct connect.

Amazon S3 (Simple Storage Service):

- Consistency:
  - new objects: strong read-after-write consistency.
  - replaced objects: eventual consistency.
- HA: S3 data replicated across 3 AZs at least (except the One Zone-IA and RRS classes).
- Block Public Access option: to ensure that public access to all your S3 buckets and objects is blocked. This setting applies account-wide for all current and future buckets.

S3 Storage Classes (per object):

- S3 Standard: low latency, high performance, high cost. Default class. No retrieval fees.
- S3 Standard-IA: Infrequent Access, same performance as S3-Standard, medium cost, min storage 30 days, retrieval fee.
- S3 One Zone-IA: Infrequent Access, stored only in one AZ
- S3 Glacier: Archival, Retrieval fee, min storage 90 days, 3 retrieval speeds (from minutes to 12 hours).
- S3 Glacier Deep Archive: Archival, Retrieval fee, min storage 180 days, 2 retrieval speeds (12 or 24 hours). Internally uses S3 Glacier.
- S3 Intelligent-Tiering: automatically moving data to the most cost-effective access tier, without operational overhead. Min 30 days. No retrieval fees.

S3 Presigned URLs:

- A presigned URL gives you access to the object identified in the URL for specified duration.
- When you create a presigned URL, you associate it with a specific action.
- Anyone using the URL will act on the object as if he were the original signing user.

Prefixes and S3 Performance:

- A prefix is a logical grouping of the objects in a bucket.
- The prefix value is similar to a directory name that enables you to store similar data under the same directory in a bucket.
- Can create a hierarchy.
- You can chose your delimiter for the bucket, such as slash (/).
- S3 Request rates: 3,500 writes and 5,500 reads per second PER PREFIX!
- To parallelize reads/writes, use multiple prefixes.

S3 Transfer Acceleration:

- A bucket-level feature that enables fast, easy, and secure transfers of files over long distances between your client and an S3 bucket.
- Takes advantage of the globally distributed edge locations in Amazon CloudFront.
- As the data arrives at an edge location, the data is routed to Amazon S3 over an optimized network path.
- Popular use cases:
  - Your customers upload to a centralized bucket from all over the world.
  - You transfer gigabytes to terabytes of data on a regular basis across continents.
  - You can’t use all of your available bandwidth over the internet when uploading to Amazon S3.
- To access the bucket that is enabled for Transfer Acceleration, you must use the endpoint {bucket-name}.s3-accelerate.amazonaws.com.
- You can use the Amazon S3 Transfer Acceleration Speed Comparison tool to compare accelerated and non-accelerated upload speeds across Amazon S3 Regions.

S3 Encryption - Server-Side Encryption (SSE): You have three mutually exclusive options, depending on how you choose to manage the encryption keys:

- Server-Side Encryption with S3-Managed Keys (SSE-S3).
- Server-Side Encryption with CMKs stored in KMS (SSE-KMS).
- Server-Side Encryption with Customer-Provided Keys (SSE-C).

Query in Place options in S3:

- S3 Select:
  - you can use simple SQL statements to filter contents of S3 objects and retrieve just the subset of data that you need.
  - Works on objects stored in CSV, JSON, or Apache Parquet format.
  - Supports compressed objects in GZIP and BZIP2 for CSV and JSON.
  - supports server-side encrypted objects.
- Amazon Athena.
- Amazon Redshift Spectrum.

S3 Object Lock:

- Blocks object version deletion during a customer-defined retention period so that you can enforce retention policies as an added layer of data protection or for regulatory compliance.
- If a user attempts to delete an object before its "Retain Until Date" has passed, the operation will be denied.
- Two modes for locking in S3:
  - Governance Mode: specific IAM permissions are able to remove the lock.
  - Compliance Mode: lock protection cannot be removed by any user, including root.
- Two methods for locking in S3 Glacier:
  - Vault Access policy: can be modified.
  - Vault Lock policy: can be locked for compliance.
- S3 Object Legal Hold:
  - Prevents an object version from being overwritten or deleted.
  - Doesn’t have an associated retention period and remains in effect until removed.
  - Legal holds can be freely placed and removed by any user who has the s3:PutObjectLegalHold permission.