**STORAGES AND DATABASES**

**Instance stores and Amazon Elastic Block Store(Amazon EBS)**

An **instance store** provides temporary block-level storage for an Amazon EC2 instance. An instance store is disk storage that is physically attached to the host computer for an EC2 instance, and therefore has the same lifespan as the instance. When the instance is terminated, you lose any data in the instance store.

**Amazon Elastic Block Store (Amazon EBS)** is a service that provides block-level storage volumes that you can use with Amazon EC2 instances. If you stop or terminate an Amazon EC2 instance, all the data on the attached EBS volume remains available.

To create an EBS volume, you define the configuration (such as volume size and type) and provision it. After you create an EBS volume, it can attach to an Amazon EC2 instance.Because EBS volumes are for data that needs to persist, it’s important to back up the data. You can take incremental backups of EBS volumes by creating Amazon EBS snapshots.



An **EBS snapshot** is an incremental backup. This means that the first backup taken of a volume copies all the data. For subsequent backups, only the blocks of data that have changed since the most recent snapshot are saved.

Incremental backups are different from full backups, in which all the data in a storage volume copies each time a backup occurs. The full backup includes data that has not changed since the most recent backup.

**Amazon Simple Storage Service(Amazon s3)**

Store and retrieve unlimited amount of data.Store data as **objects**. Objects stored in **buckets**.Maximum upload object size is **5TB**. Can create multiple buckets.

In **object storage** each object contains data metadata and a key.The data might be an image, video, text document, or any other type of file. Metadata contains information about what the data is, how it is used, the object size, and so on. An object’s key is its unique identifier.

**Amazon Simple Storage Service (Amazon S3)** is a service that provides object-level storage. Amazon S3 stores data as objects in buckets.The maximum file size for an object in Amazon S3 is 5 TB. When you upload a file to Amazon S3, you can set permissions to control visibility and access to it. You can also use the Amazon S3 versioning feature to track changes to your objects over time.

**Amazon S3 storage classes**

* **S3 standard** : Designed for **frequently accessed data**, Stores data in a minimum of three Availability Zones. Amazon S3 Standard provides high availability for objects. This makes it a good choice for a wide range of use cases, such as websites, content distribution, and data analytics. Amazon S3 Standard has a higher cost than other storage classes intended for infrequently accessed data and archival storage
* **S3 standard infrequent Access(s3 standard IA):** Ideal for i**nfrequently accessed data** ,Similar to Amazon S3 Standard but has a lower storage price and higher retrieval price. Amazon S3 Standard-IA is ideal for data infrequently accessed but requires high availability when needed. Both Amazon S3 Standard and Amazon S3 Standard-IA store data in a minimum of three Availability Zones. Amazon S3 Standard-IA provides the same level of availability as Amazon S3 Standard but with a lower storage price and a higher retrieval price.
* **S3 one zone-infrequent access(s3 onezone IA):** Stores data in a single Availability Zone, Has a lower storage price than Amazon S3 Standard-IA. One Zone-IA stores data in a single Availability Zone. This makes it a good storage class to consider if the following conditions apply: You want to save costs on storage.You can easily reproduce your data in the event of an Availability Zone failure.
* **S3 intelligent tiering:** Ideal for data with unknown or c**hanging access patterns** Requires a small monthly monitoring and automation fee per object. In the S3 Intelligent-Tiering storage class, Amazon S3 monitors objects’ access patterns. If you haven’t accessed an object for 30 consecutive days, Amazon S3 automatically moves it to the infrequent access tier, S3 Standard-IA. If you access an object in the infrequent access tier, Amazon S3 automatically moves it to the frequent access tier, S3 Standard.
* **S3 glacier instant retrieval:** Works well for **archived data** that requires immediate access. Can retrieve objects within a few milliseconds. When you decide between the options for archival storage, consider how quickly you must retrieve the archived objects. You can retrieve objects stored in the S3 Glacier Instant Retrieval storage class within milliseconds, with the same performance as S3 Standard.
* **S3 glacier flexible retrieval:** Low-cost storage designed for **data archiving**. Able to retrieve objects within a few minutes to hours. S3 Glacier Flexible Retrieval is a low-cost storage class that is ideal for data archiving. For example, you might use this storage class to store archived customer records or older photos and video files. You can retrieve your data from S3 Glacier Flexible Retrieval from 1 minute to 12 hours.
* **S3 glacier deep archive:** Lowest-cost object storage class ideal for **archiving**. Able to retrieve objects within 12 hours. S3 Deep Archive supports long-term retention and digital preservation for data that might be accessed once or twice in a year. This storage class is the lowest-cost storage in the AWS Cloud, with data retrieval from 12 to 48 hours. All objects from this storage class are replicated and stored across at least three geographically dispersed Availability Zones.
* **S3 Outposts:** Creates S3 buckets on Amazon S3 Outposts. Makes it easier to retrieve, store, and access data on AWS Outposts. Amazon S3 Outposts delivers object storage to your on-premises AWS Outposts environment. Amazon S3 Outposts is designed to store data durably and redundantly across multiple devices and servers on your Outposts. It works well for workloads with local data residency requirements that must satisfy demanding performance needs by keeping data close to on-premises applications

**Amazon EBS vs Amazon S3**

Amazon EBS:

* Size up to 16tb
* Survive termination of their ec2 instance.
* Solid state by default.
* HDD options.

Amazon s3:

* Unlimited storage.
* Individual objects up to 5tb
* Write once read many
* 99.99999999% durable.

This means, if you are using complete objects or only occasional changes, S3 is victorious. If you are doing complex read, write, change functions, then, absolutely, EBS is your knockout winner.

**Amazon Elastic File System(Amazon EFS)**

Amazon Elastic File System (Amazon EFS)(opens in a new tab) is a scalable file system used with AWS Cloud services and on-premises resources. As you add and remove files, Amazon EFS grows and shrinks automatically. It can scale on demand to petabytes without disrupting applications.

**Amazon EBS vs Amazon EFS**

**EBS:**

An Amazon EBS volume stores data in a single Availability Zone.

To attach an Amazon EC2 instance to an EBS volume, both the Amazon EC2 instance and the EBS volume must reside within the same Availability Zone.

Volumes attached to EC2 instance

Availability zone level resource

Need to be in the same availability zone to attach ec2 instance.

Volumes do not automatically scale.

**EFS:**

Amazon EFS is a regional service. It stores data in and across multiple Availability Zones.

The duplicate storage enables you to access data concurrently from all the Availability Zones in the Region where a file system is located. Additionally, on-premises servers can access Amazon EFS using AWS Direct Connect.

Multiples instances reading and writing simultaneously.

Regional resource.

Automatically scales.

**Amazon Relational Database Services(Amazon RDS)**

Amazon RDS is a managed service that automates tasks such as hardware provisioning, database setup, patching, and backups. With these capabilities, you can spend less time completing administrative tasks and more time using data to innovate your applications. You can integrate Amazon RDS with other services to fulfil your business and operational needs, such as using AWS Lambda to query your database from a serverless application

Amazon RDS provides a number of different security options. Many Amazon RDS database engines offer encryption at rest (protecting data while it is stored) and encryption in transit (protecting data while it is being sent and received).

**Amazon RDS Engines**

* Amazon Aurora
* PostgreSQL
* MySQL
* MariaDB
* Oracle Database
* Microsoft SQL Server

**Amazon Aurora**

Amazon Aurora is an **enterprise-class relational database**. It is compatible with MySQL and PostgreSQL relational databases. It is up to five times faster than standard MySQL databases and up to three times faster than standard PostgreSQL databases.

Amazon Aurora helps to reduce your database costs by reducing unnecessary input/output (I/O) operations, while ensuring that your database resources remain reliable and available.

Consider Amazon Aurora if your workloads require high availability. It replicates six copies of your data across three Availability Zones and continuously backs up your data to Amazon S3.

**Amazon DynamoDB**

A serverless database.A non relational database. Millisecond response time.

**Amazon DynamoDB** is a key-value database service.DynamoDB is serverless, which means that you do not have to provision, patch, or manage servers. You also do not have to install, maintain, or operate software.DynamoDB automatically scales to adjust for changes in capacity while maintaining consistent performance. This makes it a suitable choice for use cases that require high performance while scaling.

**RDS:**

* Automatic high availability. Recovery provided.
* Customer ownership of data.
* Customer ownership of schema.
* Customer control of the network.

**DynamoDB:**

* Key value
* Massive throughput capabilities
* PB(peta Byte) size potential
* Granular API access.

**Amazon RedShift**

**Amazon Redshift** is a **data warehousing** service that you can use for big data analytics. It offers the ability to collect data from many sources and helps you to understand relationships and trends across your data. Mainly business intelligence.

**AWS Data Migration Services**

**AWS Database Migration Service (AWS DMS)** enables you to migrate relational databases, non relational databases, and other types of data stores.

With AWS DMS, you move data between a source database and a target database. The source and target databases can be of the same type or different types. During the migration, your source database remains operational, reducing downtime for any applications that rely on the database.

For example, suppose that you have a MySQL database that is stored on premises in an Amazon EC2 instance or in Amazon RDS. Consider the MySQL database to be your source database. Using AWS DMS, you could migrate your data to a target database, such as an Amazon Aurora database.

**Homogenous** : mysql to amazon RDS for mysql

**Heterogeneous** : 2 steps, we first need to convert them using the AWS Schema Conversion Tool. This will convert the source schema and code to match that of the target database. The next step is then to use DMS to migrate data from the source database to the target database.

But these are not the only use cases for DMS. Others include d**evelopment and test database migrations, database consolidation, and even continuous database replication.**

**Development and test migration** is when you want to develop this to test against production data, but without affecting production users. In this case, you use DMS to migrate a copy of your production database to your dev or test environments, either once-off or continuously.

**Database consolidation** is when you have several databases and want to consolidate them into one central database.

Finally, **continuous replication** is when you use DMS to perform continuous data replication. This could be for disaster recovery or because of geographic separation.

**Additional Database Services**

* **Amazon DocumentDB** : Amazon DocumentDB is a document database service that supports MongoDB workloads. (MongoDB is a document database program.)
* **Amazon Neptune** : Amazon Neptune is a graph database service. You can use Amazon Neptune to build and run applications that work with highly connected datasets, such as recommendation engines, fraud detection, and knowledge graphs.
* **Amazon Quantum Ledger Databases(Amazon QLDB)**: Amazon Quantum Ledger Database (Amazon QLDB) is a ledger database service. You can use Amazon QLDB to review a complete history of all the changes that have been made to your application data.
* **Amazon Managed Blockchain** : Amazon Managed Blockchain is a service that you can use to create and manage blockchain networks with open-source frameworks. Blockchain is a distributed ledger system that lets multiple parties run transactions and share data without a central authority.
* **Amazon ElastiCache** : Amazon ElastiCache is a service that adds caching layers on top of your databases to help improve the read times of common requests. It supports two types of data stores: Redis and Memcached.
* **Amazon DynamoDB Accelerator**: Amazon DynamoDB Accelerator (DAX) is an in-memory cache for DynamoDB. It helps improve response times from single-digit milliseconds to microseconds.