**Module 5**

**Storage and Database**

Instance stores and Amazon Elastic Block Store.

What is block storage?

* A place to store files e.g., your hard drive
* Block level storage behaves like hard drives
* A file being a series of bytes that are stored in blocks on disc
* When file is updated, the whole series of blocks aren’t all overwritten
* It updates just the pieces change

What is 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.
* Provide temporary block level storage for an Amazon EC2 instance
* When the instance is terminated you lose any data in the instance store.

**Example of how instances store work:**

**What is 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 terminate the Amazon EC2 instance all the data on the attached EBS volume remains available.
* To create EBS volume you need to define the configuration (such as volume size and type) and provision it. After created it can attach to Amazon EC2 instance
* You can take incremental backups of EBS volumes by creating Amazon EBS snapshots

**EBS Snapshot**

* Is an incremental backup.
* This means that the first backup taken of a volume copy all the data.
* For subsequent backups only the data that have changed since the most recent snapshot are saved

**Incremental vs Full Backup**

* 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

**Different types of storages**

1. **Amazon Simple Storage Service (Amazon S3)**

* It is a data store that allows you to store and retrieve an unlimited amount of data at any scale
* Data is stored as object in a bucket
* Might be used to store images, videos, text files.
* The maximum object size that you can upload is five terabytes
* S3 can be used to host static website

**Different tiers of Data**

1. **S3 Standard**

* Comes with 11 nines of durability
* Means an object is stored in S3 standard has a 99.999999999 percentage
* It will remain intact after a period of one year

1. **S3 Frequent access or S3-A1**

* Used for data that is accessed less frequently but requires rapid access when needed
* Perfect place to store backups, disaster recovery files or any object that requires a long-term storage

**Object Storage**

* Each consist of data, metadata and key
* Data might be images, video, text documents, or any type of file
* Metadata contains information about what the data is, how it used, the object size
* Object key is its unique identifier

**Amazon S3 Storage classes**

* You pay only what you use
* Are purpose-built to provide the lowest cost storage for different access pattern

**Types of S3 Storage classes**

1. **S3 Intelligent**

* For automatic cost savings for data with unknown or changing access patterns

1. **S3 Standard**

* For frequently accessed data

1. **S3 Standard-Infrequent Access (S3 Standard-IA)/ S3 One Zone-Infrequent Access**

* For less frequently accessed data

1. **S3 Glacier Instant Retrieval**

* For achieve data that needs immediate access

1. **S3 Glacier Flexible Retrieval (Formerly S3 Glacier)**

* For rarely accessed long term data that does not require immediate access

1. **Amazon S3 Glacier Deep Archive (S3 Glacier Deep Archive)**

* For long-term achieve and digital preservation with retrieval in hours at the lowest level cost storage in the cloud.

**Factors to consider when choosing Amazon S3 storage class**

* How often you plan to retrieve data
* How available you need your data to be

**Differences**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Amazon S3 Standard** | **Amazon S3 Intelligent-Tiering (S3 Intelligent)** | **Amazon S3 Standard-Infrequent Access (S3 Standard-IA)** | **Amazon S3 Glacier Instant Retrieval** | **Amazon S3 Glacier Flexibility (Formerly S3 Glacier)** | **Amazon S3 Glacier Deep Achieve (S3 Glacier Deep Archive)** | **S3 on Outputs** |
| Designed for frequently accessed data | For data with unknown or changing access patterns | For infrequently accessed data | Low-cost storage designed for data archiving | Delivers low cost storage up to 10% lower for archive data that accessed 1-2 times per year and retrieved asynchronously | Low-cost object storage class ideals for archiving | Delivers objects to your on-premises AWS Outposts environment |
| Stores data in a minimum of three Availability Zones | Requires a small monthly monitoring and automation fee per object | Similar to S3 standard but has a lower storage price and higher retrieval price | Able to retrieve objects within a few minutes to hours |  | Able to retrieve objects within 12 hours | Designed to durably and redundantly store on your Outposts |
| Provides high availability for objects | Amazon S3 Objects access patterns |  |  |  |  | Encryption using SSE-S3 and SSE-C |
| Good choice for wide range of use cases such as website, content distribution and data analytics | If you haven’t accessed an object for 30 consecutive days, Amazon S3 automatically moves it to infrequent access tier, S3 Standard-IA |  |  |  |  | Authentication and authorization using IAM, and S3 Access Points |
| Has a high cost than other storage intended for in frequently accessed data and archival storage |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

**Comparing Amazon EBS and Amazon EFS**

**File Storage**

* Multiple clients can access data that is stored in shared files folders
* A storage server uses block storage with a local file system to organize files
* Is ideal for use cases in which number of resources need to access the same data the same time

**What is Amazon Elastic File System (Amazon EFS)?**

* 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

|  |  |
| --- | --- |
| **Amazon EBS** | **Amazon EFS** |
| Stores data in a single availability zone | Stores data in a cross multiple availability zone |
| To attach an Amazon EC2 instance to an EBS volume, both the Amazon EC2 instance and EBS volume must reside with the same availability zone | The duplicate storage enables access data concurrently from all the availability zones in the region where file system can access Amazon EFS using AWS Direct Connect |
|  |  |

**Amazon Relational Database Service (Amazon RDS)**

**How’s data stored in relational database?**

* Data is stored in a way that it relates it to other pieces of data
* Uses structured query language to store and query data

**What is Amazon Relational Database Service?**

* Is a service that enables you to run relational database in the AWS Cloud
* Is managed service that automates task such as hardware provisioning, database setup, patching, and backups.
* Provides a number of securities options, offers encryption rest (protecting data while is stored) and encryption in transits (protecting data while it is being sent and received)

Amazon RDS is available on 6 database Engine

Which optimise for memory, performance, or input/output(I/O). Supported database engines include:

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

1. **Amazon Aurora**

* Is an enterprise-class relational database.
* It is compatible with MySQL and PostgreSQL relational database
* It is up to 5 times faster than standard MySQL and up to three times faster than standard PostgreSQL database
* Helps to reduce database costs by reducing unnecessary input/output(I/O) operations, while ensuring that your database resources remain reliable and available.
* Use if your workload requires high availability
* Continuously backup data to Amazon S3

1. **Amazon DynamoDB**

* Is a key-value database service. It delivers single-digit millisecond performance at any scale
* It’s a serverless database, meaning you don’t need to manage the underlying instances or infrastructure powering it.
* You do not have to install, maintain or operate software
* Consist of tables, is a place where you can store and query data
* Data is organized into items and items have attributes
* You don’t have to worry about scaling up or down
* Is a non-relational database
* It isn’t the best fit for every workload out there
* Automatically scales to adjust for changes in capacity while maintaining consistent performance.

**Nonrelational Databases**

* Sometimes referred as NOSQL databases because they use structured other than rows and columns to organize data
* One type of structural approach for nonrelational databases is key-value pairs, with key-value pairs, data is organized into items and items have attributes.
* You can add or remove attributes from items in the table at anytime

1. **Amazon Redshift**

* Is a data warehousing service that you can use for big data analytics
* Offers ability to collect data from many sources and help you to understand relationships and trends across your data.
* This is data warehousing as a service.
* Its massively scalable

**AWS Database Migration Service (AWS DMS)**

* Enables you to migrate relational databases, nonrelation databases and other types of data stores
* With AWS DMS, you move data between a source database and target database
* During the migration, your source database remains operational, reducing downtime for applications that rely on the database

**Other use cases of AWS DMS**

* Enables developers to test application against production data without affecting production users
* Combining several databases into a single database
* Sending ongoing copies of your data to other targe sources instead of doing one-time migration

|  |  |
| --- | --- |
| **Amazon DocumentDB** | Is a document database service that support MongoDB workloads. (MongoDB is a document database program) |
| **Amazon Neptune** | * Is a graph database service * You can use Amazon Neptune to build and run application that work with highly connected datasets, such as recommendation engines, fraud detection and knowledge graphs |
| **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 |
| **Amazon Managed Blockchain** | * Is a service that you can use to create and managed blockchain networks with open-source frameworks * Blockchain is a distributed ledger system that lets multiple parties run transaction and share data without a central authority |
| **Amazon ElasticCache** | * Is a service that adds catching layers on top of your databases to help improve the read times of common request * It supports two types of data stores: Redis and Memcatched |
| **Amazon DynameDB Accelerator** | * **Is an in-memory cache for DynamoDB** * It helps improve response time from single-digit milliseconds to microseconds |