module 5: storage and databases

learning objectives -

- summarize the basic concept of storage and databases.
- describe the benefits of Amazon EBS
- describe the benefits of Amazon S3
- describe the benefits Amazon EFS
- summarize various storage solutions
- describe the benefits of Amazon RDS
- describe the benefits of Amazon DynamoDB
- summarize various database services

▼ ***How can you explain block-level storage?***

A place to store files - a series of bytes that are stored in blocks on disc.

▼ ***When a file is updated, how does the data change?***

Only the pieces (blocks) that change are overwritten, not the entire thing.

▼ ***Where does block level storage exist?***

Your hard drive - EC2 instances have hard drives as well

▼ ***What is local storage in EC2 instances?***

Instance store volumes

▼ ***How are these volumes connected?***

They are physically attached to the host, and your EC2 instances are running on top

▼ ***What is the catch?***

If you stop or terminate the EC2 instance, all data written to the instance store volume will be deleted.

If you stop or terminate the EC2 - when you start it up again from the stop state, it’s likely that the EC2 instance will start up on another host

▼ ***What is a more stable version of this storage?***

Elastic Block Store

▼ ***What are EBS storage units called?***

EBS volumes

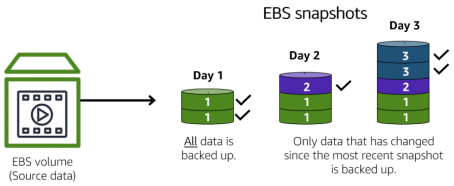
▼ ***What other functionality does EBS provide?***

Incremental backups

▼ ***How are incremental backups different from full backups?***

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.

Amazon EBS snapshots



Incremental backups of EBS volumes with Amazon EBS snapshots. On Day 1, two volumes are backed up. Day 2 adds one new volume and the new volume is backed up. Day 3 adds two more volumes for a total of five volumes. Only the two new volumes are backed up.

Which of the following are characteristics of the Amazon EBS service? (Select TWO.)

- [x] Best for data that requires retention
- [ ] Best for temporary data that is not kept long term
- [x] Separate drives from the host computer of an EC2 instance
- [ ] Physically attached to the host computer of an EC2 instance
- [ ] Data is deleted when an EC2 instance is stopped

▼ ***What is the main advantage of S3?***

It's a data store that allows you to store and retrieve a virtually unlimited amount of data at any scale.

▼ ***How do you store data using S3?***

In buckets

▼ ***How do you protect from accidental deletion?***

By versioning the objects

▼ ***What other features are there of buckets?***

You can create multiple and store them across different classes or tiers of data. You can also set permissions.

▼ ***What is the first storage class?***

S3 standard - 11 nines of durability

▼ ***In how many facilities is this data stored in?***

3

▼ ***How else can you use S3?***

As a static website, a collection of HTML files and each file is akin to a physical page of the actual site.

▼ ***How to create a static website?***

Upload all your HTML, static web assets, and so forth

▼ ***What is S3 Standard Infrequent Access and when would it be used?***

It's used for data that is accessed less frequently but requires rapid access. Perfect for backups, disaster recovery, or long term storage.

▼ ***What service can you use for auditing purposes?***

S3 Glacier Flexible Retrieval to Archive that data

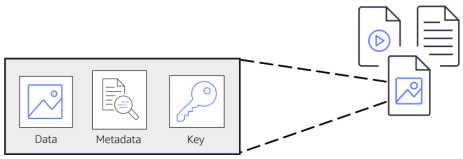
▼ ***How would you use S3 Glacier Flexible Retrieval?***

First you move data; Then you create vaults, populate. Use lock policies. Specify controls.

▼ ***What are lifecycle policies?***

You can move shit around???

Object storage



In **object storage**, each object consists of 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.

▼ ***What is the maximum file size in S3?***

5TB

▼ ***What are the two factors in consider when thinking of storage classes?***

How often you plan to retrieve your data

How available you need your data to be

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 infrequently 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 One Zone-IA)

- Stores data in a single Availability Zone
- Has a lower storage price than Amazon S3 Standard-IA

Compared to S3 Standard and S3 Standard-IA, which store data in a minimum of three Availability Zones, S3 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 changing 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

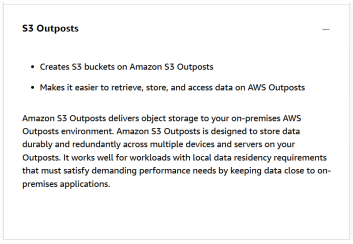
- Works well for archived data that requires immediate access
- Can retrieve objects within a few milliseconds

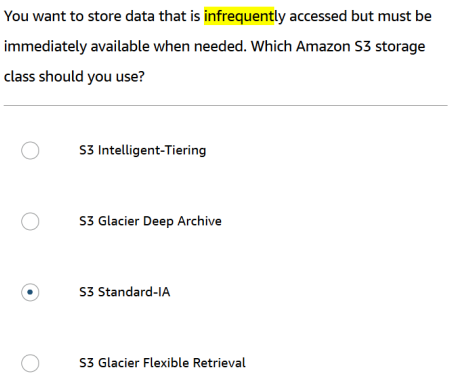
When you decide between the options for archival storage, consider how quickly you must retrieve the archived objects. You can retrieve objects stored in the S3 Glacier Instant Retrieval storage class within milliseconds, with the same performance as S3 Standard.

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.





▼ ***How much memory does block storage offer?***

16 tebibytes each

▼ ***What advantage does block storage offer?***

The ability to survive the termination of EC2 instances

▼ ***How much memory does regional object storage offer?***

Unlimited storage with 99.999 999 999% durable.

▼ ***Between working with an 80 GB video file and a photo analysis website, which storage fits with each?***

Photo analysis - S3 (regional object storage) & Video file - EBS (block storage)

▼ ***What is EFS?***

Elastic File System - managed file system

▼ ***What is the use case for having a shared file system?***

You might have multiple servers running analytics on large amounts of data being stored in a shared file system

▼ ***What does EFS allow you to do -***

Have multiple instances accessing the data in EFS at the same time.

Scale up and down as needed without you needing to do anything to make that scaling happen.

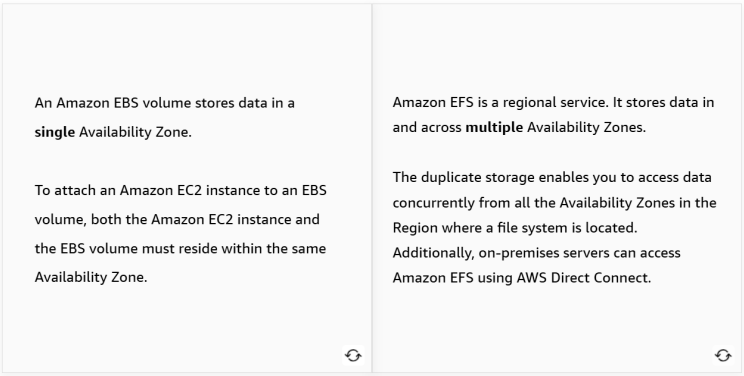
▼ ***Difference between EFS and EBS?***

EBS - you need to be in the same AZ. You can save files on it. Run databases on top of it. Or store applications on it. It's a hard drive and it automatically scale

EFS - can have multiple instances reading and writing from it at the same time

▼ ***If EBS is a hard drive, then what is EFS?***

It's a true file system for Linux, a regional resource which means any EC2 instance in the region can write to the EFS file system.



▼ ***What is lift and shift?***

A way to migrate your databases to run on Amazon EC2.

▼ ***What things do you still have control over?***

1. On prem
2. OS
3. Memory
4. CPU

5. Storage Capacity

▼ ***What are the added benefits for RDS?***

1. Automated patching
2. Backups
3. Redundancy
4. Failover
5. Disaster Recovery

▼ ***What is Amazon Aurora?***

The most managed relational database option. Priced at 1/10th cost

▼ ***What are the added benefits of Aurora?***

Data is replicated across facilities (6 copies at a given time).

▼ ***What is DynamoDB?***

Nonrelational DB option (serverless)

▼ ***How does DynamoDB store this data?***

Redundantly across AZ and mirrors the data across multiple drives under the hood for you

▼ ***What is a use case for DynamoDB?***

When u have lots of users

▼ ***What can be some issues with relational databases?***

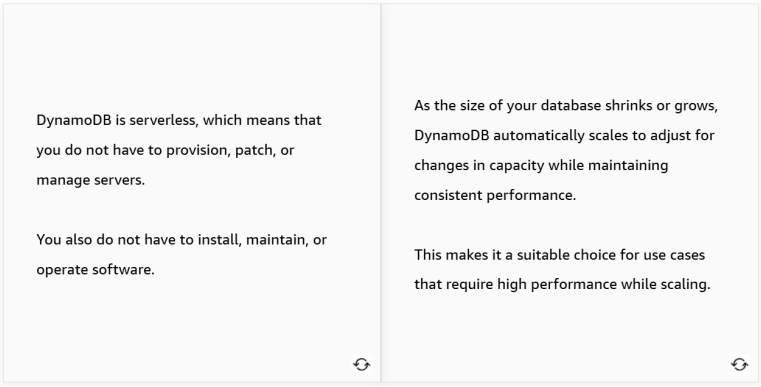
They might be rigid and have performance and scaling issues. Not good for a dataset that is being accessed at a very high rate.

▼ ***Can you remove and add attributes from tables?***

Yes, it provides great flexibility

▼ ***What is a way to describe NoSQL Databases?***

Key value pairs



What are the scenarios in which you should use Amazon Relational Database Service (Amazon RDS)? (Select TWO.)

- [ ] Running a serverless database
- [x] Using SQL to organize data
- [ ] Storing data in a key-value database
- [ ] Scaling up to 10 trillion requests per day
- [x] Storing data in an Amazon Aurora database

▼ ***When do we need data warehouses?***

If our biz questions are to look backwards.

▼ ***What do we use for ADW?***

Redshift

▼ ***What does AWS Database Migration Service do?***

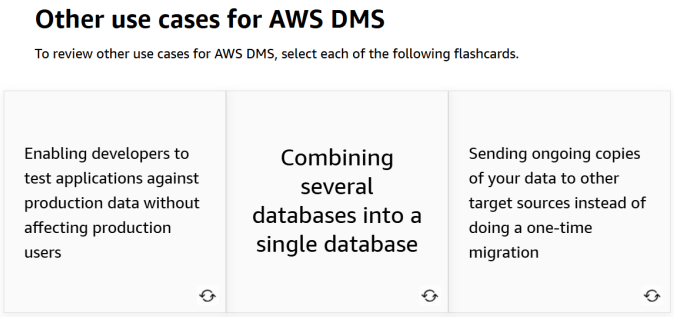
It enables you to migrate relational, nonrelational databases and other types of data stores.

▼ ***Can the source and target databases be the same type?***

They can and they can be different types too.

▼ ***During the migration, is there downtime?***

No



▼ ***What is Amazon DocumentDB?***

A doc DB service that supports MongoDB workloads

▼ ***What is Amazon Neptune?***

Graph DB service - for highly connected datasets such as rec engines

▼ ***What is Amazon Quantum Ledger Database?***

Ledger DB - review a complete history of all the changes

▼ ***What is Amazon Managed Blockchain?***

Creates blockchain networks with open source frameworks - a distributed ledger system that lets multiple parties run transactions and share data without a central authority

▼ ***What is Elasticache?***

A service that adds caching layers on top of your databases to help improve the read times of common requests

▼ ***What is DynamoDB Accelerator?***

In memory cache for DynamoDB → improves response times

Which Amazon S3 storage classes are optimized for archival data? (Select TWO.)

- [ ] Amazon S3 Standard
- [x] Amazon S3 Glacier Flexible Retrieval
- [ ] Amazon S3 Intelligent-Tiering
- [ ] Amazon S3 Standard-IA
- [x] Amazon S3 Glacier Deep Archive

Which statement or statements are TRUE about Amazon EBS volumes and Amazon EFS file systems?

- [x] EBS volumes store data within a single Availability Zone. Amazon EFS file systems store data across multiple Availability Zones.
- [ ] EBS volumes store data across multiple Availability Zones. Amazon EFS file systems store data within a single Availability Zone.
- [ ] EBS volumes and Amazon EFS file systems both store data within a single Availability Zone.
- [ ] EBS volumes and Amazon EFS file systems both store data across multiple Availability Zones.

You want to store data in an object storage service. Which AWS service is best for this type of storage?

- [ ] Amazon Managed Blockchain
- [ ] Amazon Elastic File System (Amazon EFS)
- [ ] Amazon Elastic Block Store (Amazon EBS)
- [x] Amazon Simple Storage Service (Amazon S3)

Which statement best describes Amazon DynamoDB?

- ✗ A service that enables you to run relational databases in the AWS Cloud
- ✓ A serverless key-value database service
- ✗ A service that you can use to migrate relational databases, nonrelational databases, and other types of data stores
- ✗ An enterprise-class relational database

Which service is used to query and analyze data across a data warehouse?

- ✓ Amazon Redshift
- ✗ Amazon Neptune
- ✗ Amazon DocumentDB
- ✗ Amazon ElastiCache