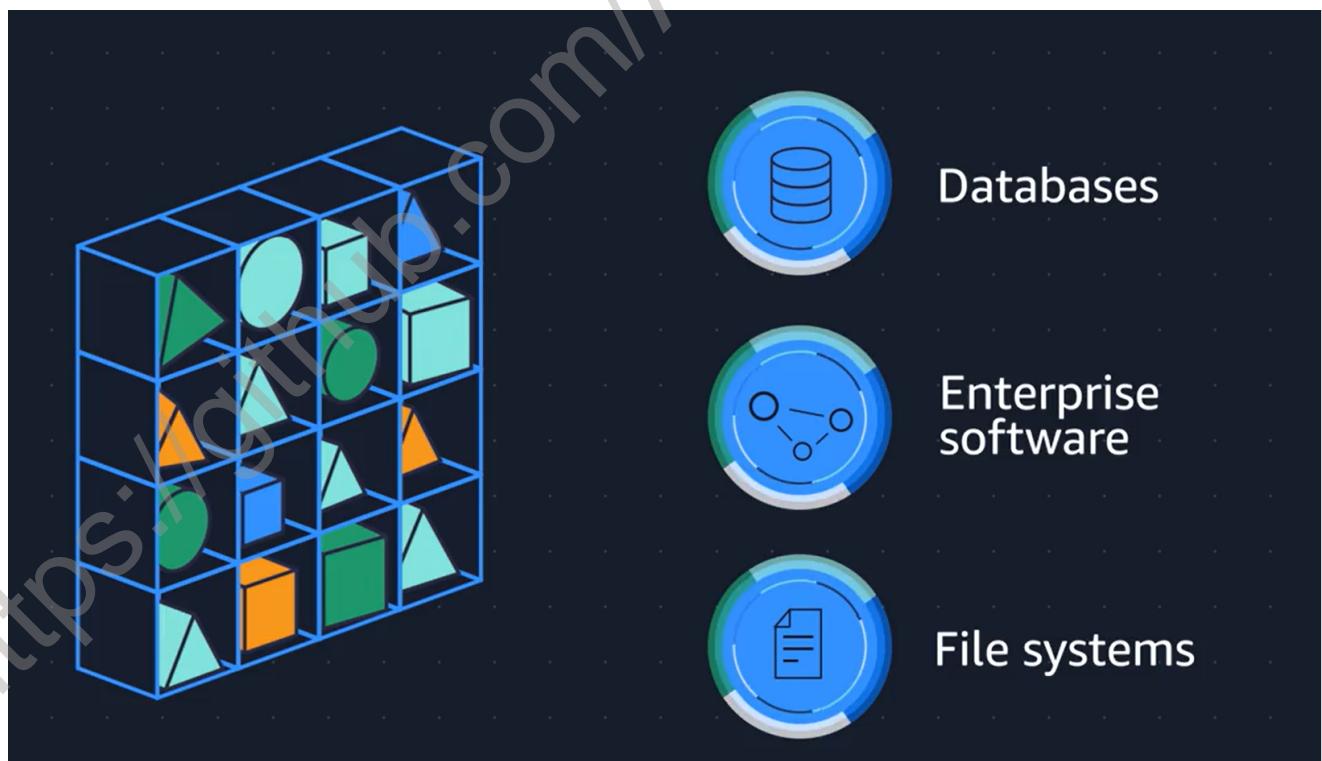
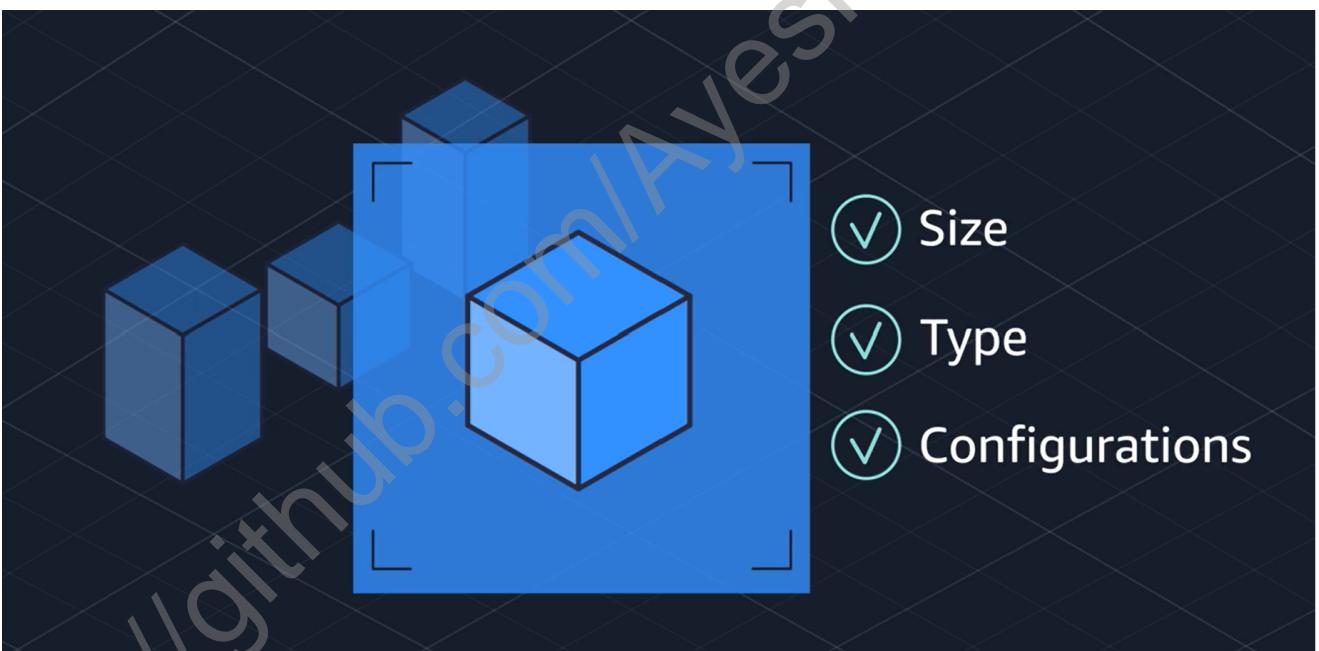
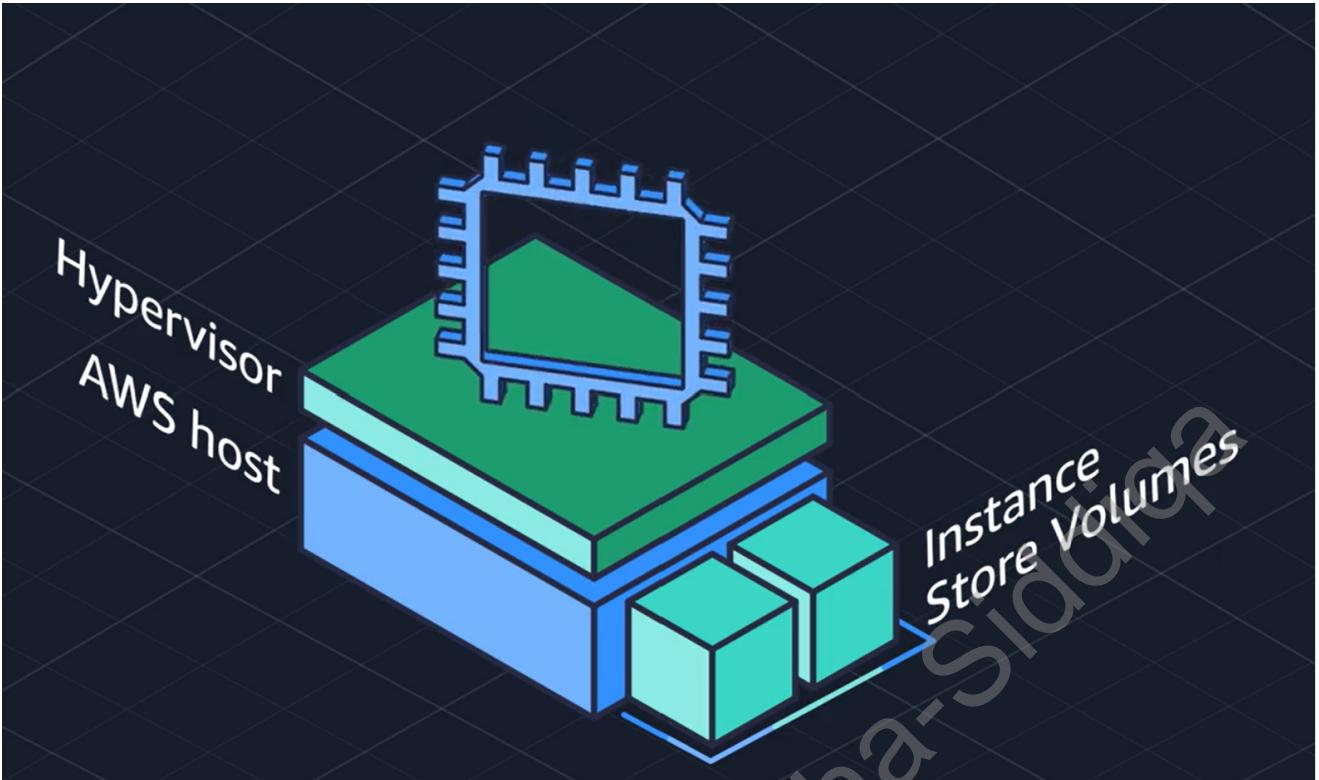


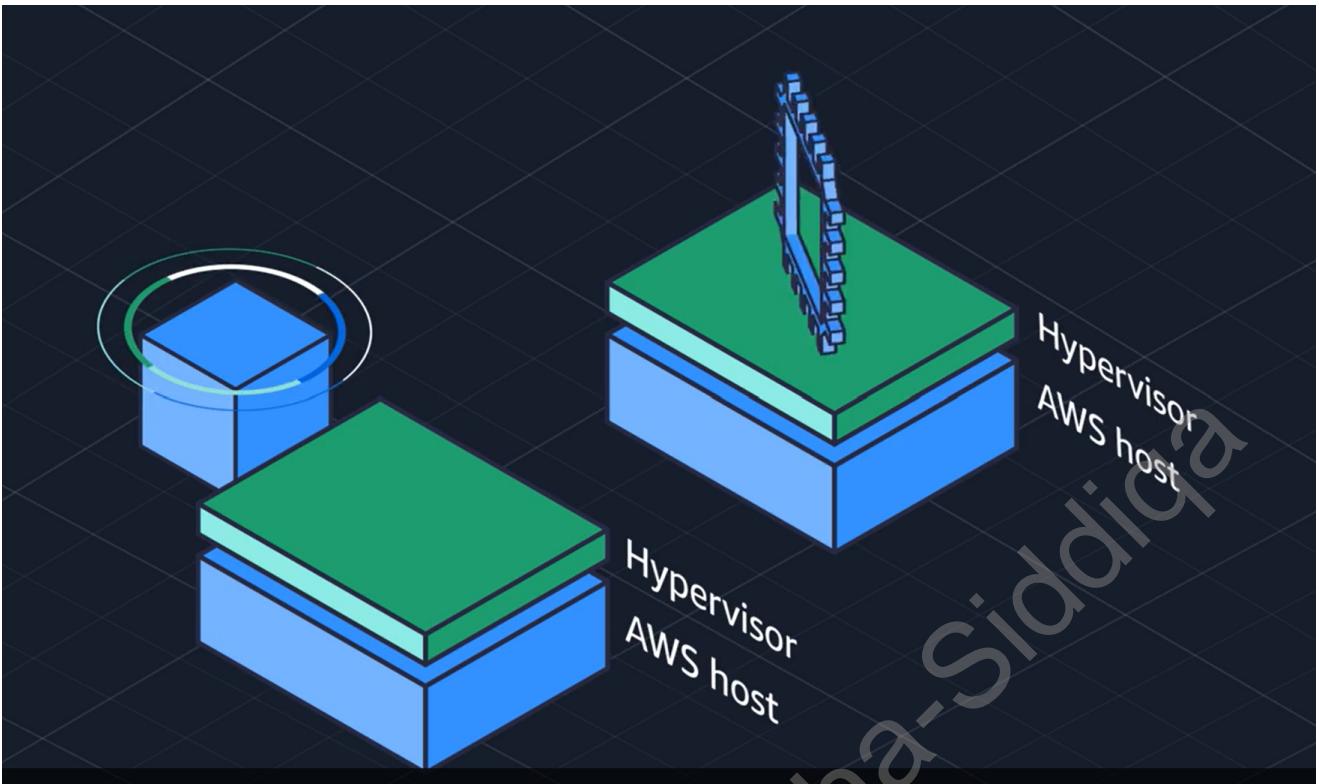
AWS Module 5 - Storage & Databases

Explain why instance store isn't preferred and how does Amazon EBS solve the problem.

- Block-level storage volumes behave like physical hard drives.
- 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.
- 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.
- Amazon EBS snapshots are incremental . 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.







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



Best for data that requires retention



Best for temporary data that is not kept long term



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

Explain Block 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.

Which service provides object-level storage?

- **Amazon S3** is a service that provides object-level storage. Amazon S3 stores data as objects in buckets.
- Amazon S3 offers unlimited storage space. 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.

Explain the different S3 storage classes

S3 Standard

- Designed for frequently accessed data
- Stores data in a minimum of three Availability Zones
- 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-IA

- Ideal for infrequently accessed data
- Similar to Amazon S3 Standard but has a lower storage price and higher retrieval price
- 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-IA

- Stores data in a single Availability Zone
- This makes it a good storage class to consider if:
- You want to save costs on storage.
- You can easily reproduce your data in the event of an Availability Zone failure.

S3 Intelligent-Tiering

- 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

S3 Glacier Flexible Retrieval

- Low-cost storage designed for data archiving
- Able to retrieve objects within a few minutes to hours

S3 Deep Archive

- Lowest-cost object storage class ideal for archiving
- 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

- Amazon S3 Outposts delivers object storage to your on-premises AWS Outposts environment.

Amazon S3

- Store data as objects.
- Store objects in buckets.
- Upload a maximum object size of 5 TB.
- Version objects.

You want to store data that is infrequently accessed but must be immediately available when needed. Which Amazon S3 storage class should you use?

- S3 Intelligent-Tiering
- S3 Glacier Deep Archive
- S3 Standard-IA
- S3 Glacier Flexible Retrieval

The correct response option is **S3 Standard-IA**.

The S3 Standard-IA storage class is ideal for data that is infrequently accessed but requires high availability when needed. Both S3 Standard and S3 Standard-IA store data in a minimum of three Availability Zones. S3 Standard-IA provides the same level of availability as S3 Standard but at a lower storage price.

The other response options are incorrect because:

- 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, S3 automatically moves it to the frequent access tier, S3 Standard.
- S3 Glacier Flexible Retrieval and S3 Glacier Deep Archive are low-cost storage classes that are ideal for data archiving. They would not be the best choice for this scenario, which requires high availability. You can retrieve objects stored in the S3 Glacier Flexible Retrieval storage class within a few minutes to a few hours. By comparison, you can retrieve objects stored in the S3 Glacier Deep Archive storage class within 12 hours.

The diagram is divided into two main sections. On the left, a dark blue vertical bar contains the letters "EBS" above a cylinder icon, with four white arrows pointing outwards from the cylinder. On the right, a light blue rectangular area contains the heading "Amazon Elastic Block Store" in bold. Below this are four horizontal boxes, each with a green checkmark icon and a small graphic: "Sizes up to 16 TiB" (with a bar chart icon), "Survive termination of their EC2 instance" (with a circuit board icon), "Solid state by default" (with a solid state drive icon), and "HDD options" (with a waveform icon).

EBS

Amazon Elastic Block Store

- Sizes up to 16 TiB
- Survive termination of their EC2 instance
- Solid state by default
- HDD options

Amazon Simple Storage Service

Unlimited storage



Individual objects
up to 5 TBs



Write once/read many



99.999999



S3



Amazon S3

- Web enabled
- Regionally distributed
- Offers cost savings
- Serverless

Amazon EBS

Volumes attach to EC2 instances

Availability Zone level resource

Need to be in the same Availability Zone to attach EC2 instances

Volumes do not automatically scale

What is Amazon EFS?

- It 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.

Comparing Amazon EBS and Amazon EFS

Select each of the cards below to review a comparison of Amazon EBS and Amazon EFS.

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.

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.

Amazon EFS

Multiple instances
reading and writing
simultaneously

Linux file system

Regional resource

Automatically scales

What is a relational Database?

- In a **relational database**, data is stored in a way that relates it to other pieces of data.

- Relational databases use **structured query language (SQL)** to store and query data.
- This approach allows data to be stored in an easily understandable, consistent, and scalable way.

ID	Product name	Size	Price
1	Medium roast ground coffee	12 oz.	\$5.30
2	Dark roast ground coffee	20 oz.	\$9.27

What is Amazon RDS?

- It is a service that enables you to run relational databases in the AWS Cloud.
- 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 fulfill 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).

What are the database engines available on Amazon RDS?

Amazon RDS offers six supported database engines, each optimized for specific aspects:

- **Amazon Aurora:** Optimized for high performance and reliability.
- **PostgreSQL:** Known for its open-source nature and advanced features.
- **MySQL:** A popular choice with a focus on compatibility and performance.
- **MariaDB:** Offers similar features to MySQL but with some enhancements.
- **Oracle Database:** Provides enterprise-grade features and support.
- **Microsoft SQL Server:** Ideal for Windows-based environments and applications.

What is Amazon Aurora, and what are its key features and advantages?

- Amazon Aurora is an enterprise-class relational database service provided by AWS.
- It is compatible with both MySQL and PostgreSQL relational databases.
- **Performance Advantages:**
 - Amazon Aurora is known for its exceptional performance, being up to five times faster than standard MySQL databases and up to three times faster than standard PostgreSQL databases.
- **Cost Reduction and Reliability:**
 - Amazon Aurora reduces database costs by minimizing unnecessary input/output (I/O) operations.
 - It ensures the reliability and availability of database resources.

- **High Availability:**
 - Amazon Aurora is an ideal choice for workloads that require high availability.
 - It replicates data across six copies in three Availability Zones and continuously backs up data to Amazon S3.

Amazon Aurora

MySQL

PostgreSQL

1/10th the cost of commercial databases

Data replication

Up to 15 read replicas

What are nonrelational databases?

- In a **nonrelational database**, you create tables.
- A table is a place where you can store and query data.

- Nonrelational databases are sometimes referred to as “NoSQL databases” because they use structures 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 (keys), and items have attributes (values). You can think of attributes as being different features of your data.
- In a key-value database, you can add or remove attributes from items in the table at any time. Additionally, not every item in the table has to have the same attributes.

Key	Value
	Name: John Doe
1	Address: 123 Any Street
	Favorite drink: Medium latte

Explain Amazon DynamoDB features.

- It is a key-value database service. It delivers single-digit millisecond performance at any scale.
- 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.
- As the size of your database shrinks or grows, 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.

Amazon DynamoDB

- Non-relational,
NoSQL database
- Purpose built
- Millisecond response
time
- Fully managed
- Highly scalable

Amazon RDS



Automatic high availability;
recovery provided



Customer ownership of data



Customer ownership of
schema



Customer control of network



Amazon DynamoDB

Key-value



Massive throughput
capabilities



PB size potential



Granular API access



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



Running a serverless database



Using SQL to organize data



Storing data in a key-value database



Scaling up to 10 trillion requests per day



Storing data in an Amazon Aurora database

What is **Amazon Redshift**?

- It 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.

What does **AWS Database Migration Service** enable you to do?

- It enables you to migrate relational databases, nonrelational databases, and other types of data stores.
- With AWS DMS, you move data between a source database and a target database.
- The source database or 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.
- Besides it also:
 - Enables in development and test database migrations by allowing developers to test applications against production data without affecting production users
 - In database consolidation by combining several databases into a single database
 - In continuous replication by sending ongoing copies of your data to other target sources instead of doing a one-time migration

- The service improves the performance of web applications by allowing you to retrieve information from fast, managed, in-memory caches, instead of relying entirely on slower disk-based databases.
- Amazon ElastiCache supports two open-source in-memory caching engines:
 - Redis - a fast, open-source, in-memory key-value data store for use as a database, cache, message broker, and queue. Amazon ElastiCache for Redis is a Redis-compatible in-memory service that delivers the ease-of-use and power of Redis along with the availability, reliability, and performance suitable for the most demanding applications.
 - Memcached - a widely adopted memory object caching system. ElastiCache for Memcached is protocol compliant with Memcached, so popular tools that you use today with existing Memcached environments will work seamlessly with the service.

Amazon Keyspaces (for Apache Cassandra)

- Amazon Keyspaces (for Apache Cassandra) is a scalable, highly available, and managed Apache Cassandra-compatible database service. With Amazon Keyspaces, you can run your Cassandra workloads on AWS using the same Cassandra application code and developer tools that you use today.

Amazon Timestream

- Amazon Timestream is a fast, scalable, fully managed time series database service for IoT and operational applications that makes it easy to store and analyze trillions of events per day at 1/10th the cost of relational databases.

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 Database (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.

Database services

Database type	Examples	AWS service
Relational	Traditional applications, enterprise resource planning (ERP), customer relationship management (CRM), ecommerce	 Amazon Aurora  Amazon RDS  Amazon Redshift
Key-value	High-traffic web applications, ecommerce systems, gaming applications	 Amazon DynamoDB
In-memory	Caching, session management, gaming leaderboards, geospatial applications	 Amazon ElastiCache  Amazon MemoryDB for Redis
Document	Content management, catalogs, user profiles	 Amazon DocumentDB (with MongoDB compatibility)
Wide column	High-scale industrial apps for equipment maintenance, fleet management, and route optimization	 Amazon Keyspaces
Graph	Fraud detection, social networking, recommendation engines	 Amazon Neptune
Time series	Internet of Things (IoT) applications, DevOps, industrial telemetry	 Amazon Timestream
Ledger	Systems of record, supply chain, registrations, banking transactions	 Amazon Ledger Database Services (QLDB)

AWS DMS

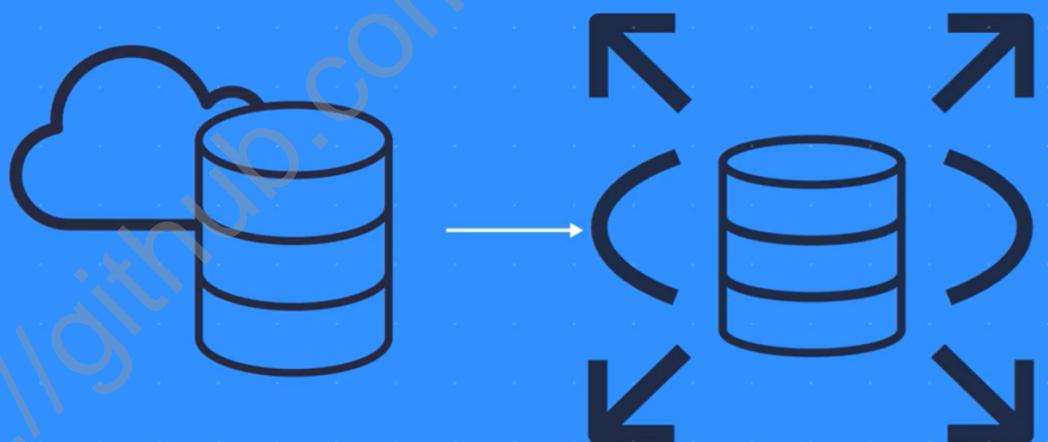
The source database remains fully operational during the migration.

Downtime is minimized for applications that rely on that database.

The source and target databases don't have to be of the same type.

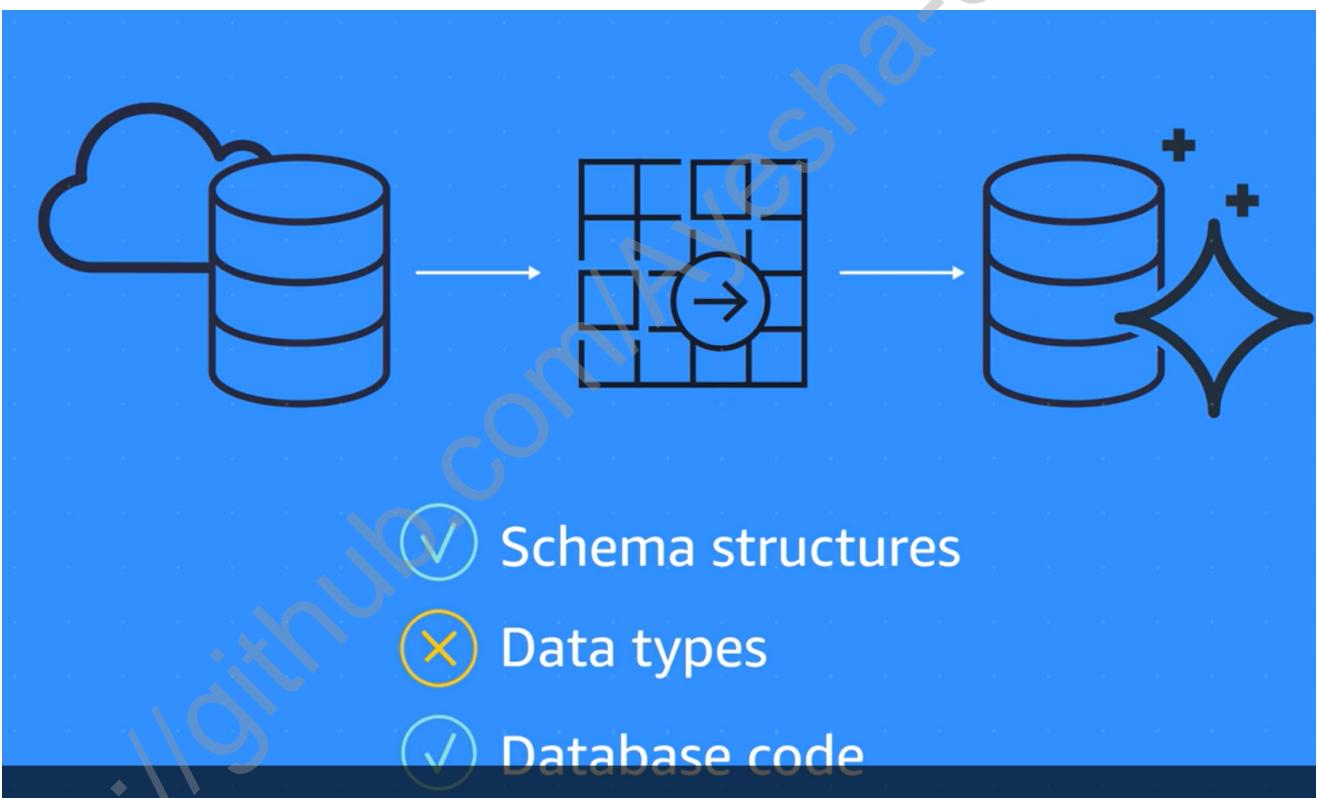
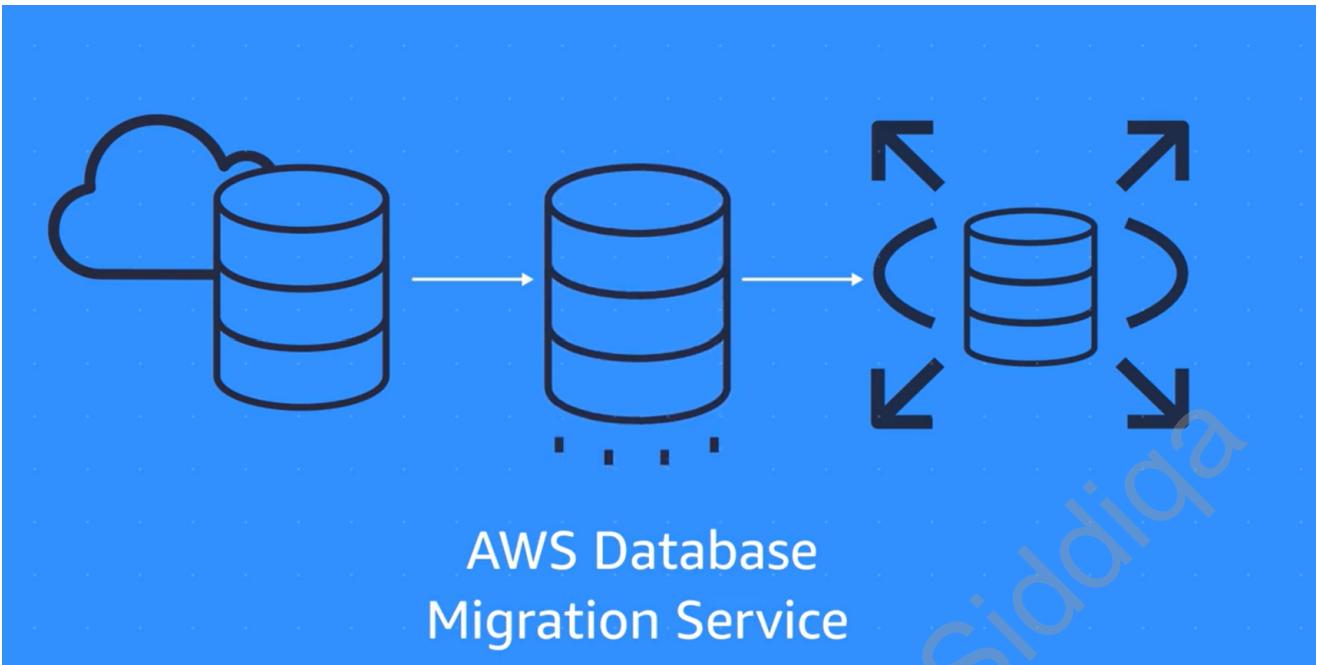


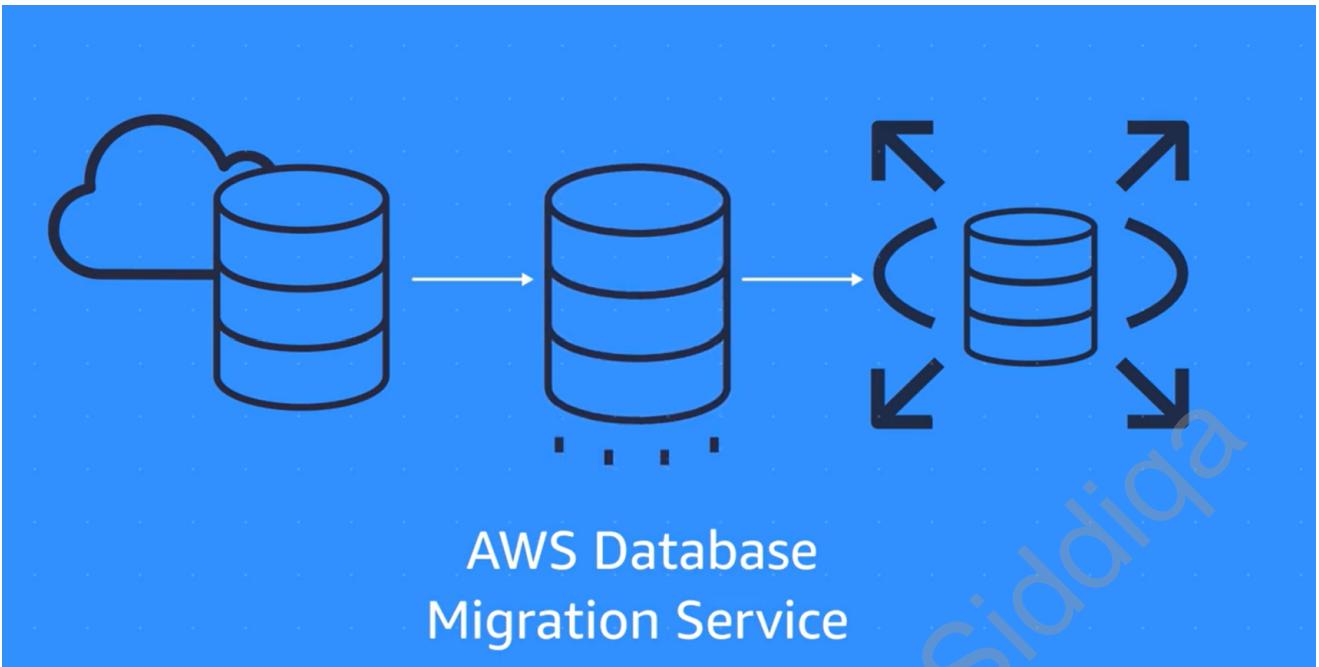
- ✓ Schema structures
- ✓ Data types
- ✓ Database code



On-premises
Amazon EC2
Amazon RDS

Amazon EC2
Amazon RDS





Development and test database migrations
Database consolidation
Continuous database replication





Amazon Neptune

<https://github.com/Ayesha-Siddiqah>



Amazon Quantum
Ledger Database
(Amazon QLDB)



Amazon ElastiCache



Amazon DynamoDB Accelerator (DAX)

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

Amazon S3 Standard

Amazon S3 Glacier Flexible Retrieval

Amazon S3 Intelligent-Tiering

Amazon S3 Standard-IA

Amazon S3 Glacier Deep Archive

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



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)



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