

# Tech Experts

Semana 4

# Relational Database Services (RDS)

- Banco de dados Relacional - nAuto gerenciado
- Engines: Mysql, Microsoft SQL Server, MariaDB, Amazon Aurora, Oracle, PostgreSQL
- Acessível por um CNAME
- Failover:
  - Single-AZ
    - EBS alocado na mesma zona de disponibilidade que as instancias EC2
    - Vulnerável à zona de disponibilidade
  - Multi-AZ
    - Instancias alocadas em diferentes zonas de disponibilidades
    - Standby replica sempre é mantida síncrona com a instancia primaria
    - A replica Standby não pode ser acessada diretamente, você não pode acessa-la a não ser que ocorra um failover.
    - Standby não trás benefícios de performance (apenas para backup)
    - Sempre deverá ser alocada na mesma região da primaria, na mesma VPC, mas em subnet diferente
    - Não prove tolerância a falha (60-120sec de indisponibilidade em eventos de failover)

# Relational Database Services

- Backup
  - **RPO (Recovery Point Objective)**
    - Tempo entre o backup mais recente e um incidente
    - Para reduzir RPO necessário reduzir a recorrência dos backups automáticos
  - **RTO (Recovery Time Objective)**
    - Tempo entre o incidente e o horário que o sistema é totalmente reestabelecido
    - RTO pode ser reduzido com boa documentação e uso de hardware sobressalente
  - Automatic Back-ups e Manual Snapshots
    - AWS-managed S3 – Não pode ser visto por usuários através do console.
    - Region resiliente (multi AZ)
    - Single AZ, backup é feito a partir da instancia primaria
    - Multi AZ, backup é feito a partir da standby replica
    - Ocorrem de forma incremental (primeira versão tem todos os dados, as próximas apenas a diferença)

# Relational Database Services

- Backup
  - **RPO (Recovery Point Objective)**
    - Tempo entre o backup mais recente e um incidente
    - Para reduzir RPO necessário reduzir a recorrência dos backups automáticos
  - **RTO (Recovery Time Objective)**
    - Tempo entre o incidente e o horário que o sistema é totalmente reestabelecido
    - RTO pode ser reduzido com boa documentação e uso de hardware sobressalente
  - Automatic Back-ups e Manual Snapshots
    - AWS-managed S3 – Não pode ser visto por usuários através do console.
    - Region resiliente (multi AZ)
    - Single AZ, backup é feito a partir da instancia primaria
    - Multi AZ, backup é feito a partir da standby replica
    - Ocorrem de forma incremental (primeira versão tem todos os dados, as próximas apenas a diferença)
    - Quando ocorre uma restauração de backup, uma nova instancia RDS é criada.
    - Recomendo a leitura:  
[https://docs.aws.amazon.com/pt\\_br/AmazonRDS/latest/UserGuide/USER\\_RestoreFromSnapshot.html](https://docs.aws.amazon.com/pt_br/AmazonRDS/latest/UserGuide/USER_RestoreFromSnapshot.html)

# Relational Database Services

- Read replica(Amazon RDS Read Replicas)
  - Um novo endpoint, apenas para leitura
  - Atualizado de forma assíncrona
  - É diferente de standby replica
  - Pode ser promovida para read-write caso a primaria falhe
  - Podem existir até 5 réplicas de leitura para oracle e até 15 para as demais engines.
- Cross-Region Read Replicas

# Relational Database Services

- Aurora
  - Shared cluster volume(SSD) – Maximo 64TB
  - Maior disponibilidade e performance
  - Cluster Endpoint e Reader Endpoint
    - Se não possuir replicas, os dois serão o mesmo
  - Multi-Master
    - Permite mais instancias com permissão de leitura/escrita

# Redshift

- Datawarehouse -> Não é usado para transações real time (não é um OLTP e sim OLAP)
- Petabyte scale
- Redshift Spectrum: Query no S3 sem necessidade de carregar no Redshift
- QuickSight pode ser integrado ao redshift
- Single AZ.
- Leader Node e Compute Nodes

# Dynamodb

- NoSQL
- Auto gerenciado
- Chave valor

# Neptune

- NoSQL
- Auto gerenciado
- Grafos



# MemCache

- NoSQL
- Auto gerenciado
- Chave valor
- Armazenamento em memória

# Q1

A Machine Learning Specialist has completed a proof of concept for a company using a small data sample, and now the Specialist is ready to implement an end- to-end solution in AWS using Amazon SageMaker. The historical training data is stored in Amazon RDS.

Which approach should the Specialist use for training a model using that data?

- A. Write a direct connection to the SQL database within the notebook and pull data in
- B. Push the data from Microsoft SQL Server to Amazon S3 using an AWS Data Pipeline and provide the S3 location within the notebook.
- C. Move the data to Amazon DynamoDB and set up a connection to DynamoDB within the notebook to pull data in.
- D. Move the data to Amazon ElastiCache using AWS DMS and set up a connection within the notebook to pull data in for fast access.

# Q1

A Machine Learning Specialist has completed a proof of concept for a company using a small data sample, and now the Specialist is ready to implement an end- to-end solution in AWS using Amazon SageMaker. The historical training data is stored in Amazon RDS.

Which approach should the Specialist use for training a model using that data?

- A. Write a direct connection to the SQL database within the notebook and pull data in (Sagemaker Training Jobs, Batch transform, processing job etc, só leem dados do S3)
- B. Push the data from Microsoft SQL Server to Amazon S3 using an AWS Data Pipeline and provide the S3 location within the notebook.
- C. Move the data to Amazon DynamoDB and set up a connection to DynamoDB within the notebook to pull data in.
- D. Move the data to Amazon ElastiCache using AWS DMS and set up a connection within the notebook to pull data in for fast access.

# Q2

A Data Science team is designing a dataset repository where it will store a large amount of training data commonly used in its machine learning models. As Data Scientists may create an arbitrary number of new datasets every day, the solution has to scale automatically and be cost-effective. Also, it must be possible to explore the data using SQL. Which storage scheme is MOST adapted to this scenario?

- A. Store datasets as files in Amazon S3.
- B. Store datasets as files in an Amazon EBS volume attached to an Amazon EC2 instance.
- C. Store datasets as tables in a multi-node Amazon Redshift cluster.
- D. Store datasets as global tables in Amazon DynamoDB.

# Q2

A Data Science team is designing a dataset repository where it will store a large amount of training data commonly used in its machine learning models. As Data Scientists may create an arbitrary number of new datasets every day, the solution has to scale automatically and be cost-effective. Also, it must be possible to explore the data using SQL. Which storage scheme is MOST adapted to this scenario?

- A. Store datasets as files in Amazon S3.
- B. Store datasets as files in an Amazon EBS volume attached to an Amazon EC2 instance.
- C. Store datasets as tables in a multi-node Amazon Redshift cluster.
- D. Store datasets as global tables in Amazon DynamoDB.

# Q3

- A machine learning (ML) specialist is setting up an ML environment that multiple data scientists will access. The ML specialist is deploying one Amazon SageMaker notebook instance for each data scientist. The ML specialist must ensure that each data scientist has access to only their personal notebook instance.
- What should the ML specialist do to meet this requirement?
  - A. Attach an IAM policy to the IAM users of the data scientists to grant access to only their personal notebook instance
  - B. Use port forwarding to prevent all internet traffic from being forwarded to the notebook instances
  - C. Use Amazon CloudWatch to invoke an AWS Lambda function that restricts unauthorized access
  - D. Attach an Amazon S3 bucket policy to restrict access to the S3 buckets that contain notebook instances of other users.

# Q3

- A machine learning (ML) specialist is setting up an ML environment that multiple data scientists will access. The ML specialist is deploying one Amazon SageMaker notebook instance for each data scientist. The ML specialist must ensure that each data scientist has access to only their personal notebook instance.
- What should the ML specialist do to meet this requirement?
  - A. Attach an IAM policy to the IAM users of the data scientists to grant access to only their personal notebook instance
  - B. Use port forwarding to prevent all internet traffic from being forwarded to the notebook instances
  - C. Use Amazon CloudWatch to invoke an AWS Lambda function that restricts unauthorized access
  - D. Attach an Amazon S3 bucket policy to restrict access to the S3 buckets that contain notebook instances of other users.

# Q4

- Which two of the following are correct?
  - A. Multi-AZ:Same Region::Read Replica:Multiple Region
  - B. Multi-AZ:Multiple Region::Read Replica:Same Region
  - C. Multi-AZ:Synchronous Replication::Read Replica:Asynchronous Replication
  - D. Multi-AZ:ASynchronous Replication::Read Replica:Synchronous Replication



# Q4

- Which two of the following are correct?
  - A. Multi-AZ:Same Region::Read Replica:Multiple Region
  - B. Multi-AZ:Multiple Region::Read Replica:Same Region
  - C. Multi-AZ:Synchronous Replication::Read Replica:Asynchronous Replication
  - D. Multi-AZ:ASynchronous Replication::Read Replica:Synchronous Replication