EML 4 -

Introdução à computação em nuvens

INFORMAÇÃO,

TECNOLOGIA

& INOVAÇÃO

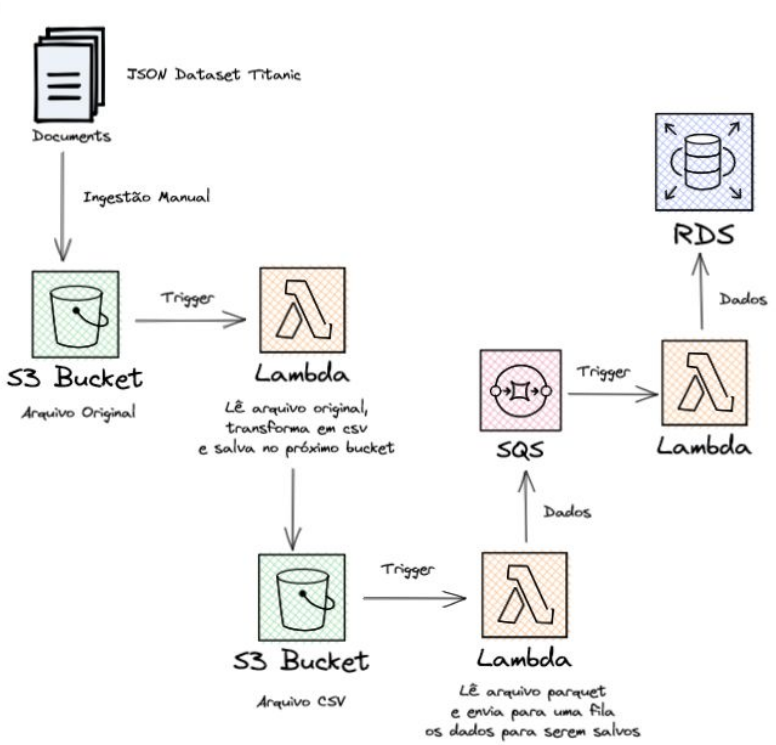
Atividade

- Desenvolver na sua conta AWS um pipeline de Extração, Transformação e Carregamento de Dados
- Depois de configurar os serviços, a ingestão do arquivo deverá ser feita manualmente para o serviço de Storage, a partir desse ponto todos os eventos deverão ser iniciados automaticamente
- O banco de dados final deverá conter os dados originais do arquivo
- Arquivo original encontra-se no repositório do Github: titanic_data.json
- Arquivo com script sql para gerar tabela para salvar os dados: titanic_database.sql

Links

- Github: https://github.com/gpgomes/ml-mba-activity
  - Instruções
  - Scripts SQL
  - Templates para os lambdas
  - Dataset para ingestão
- DBeaver (DB Client): https://dbeaver.io/download/
- Kaggle: https://www.kaggle.com/ (Referência do dataset)

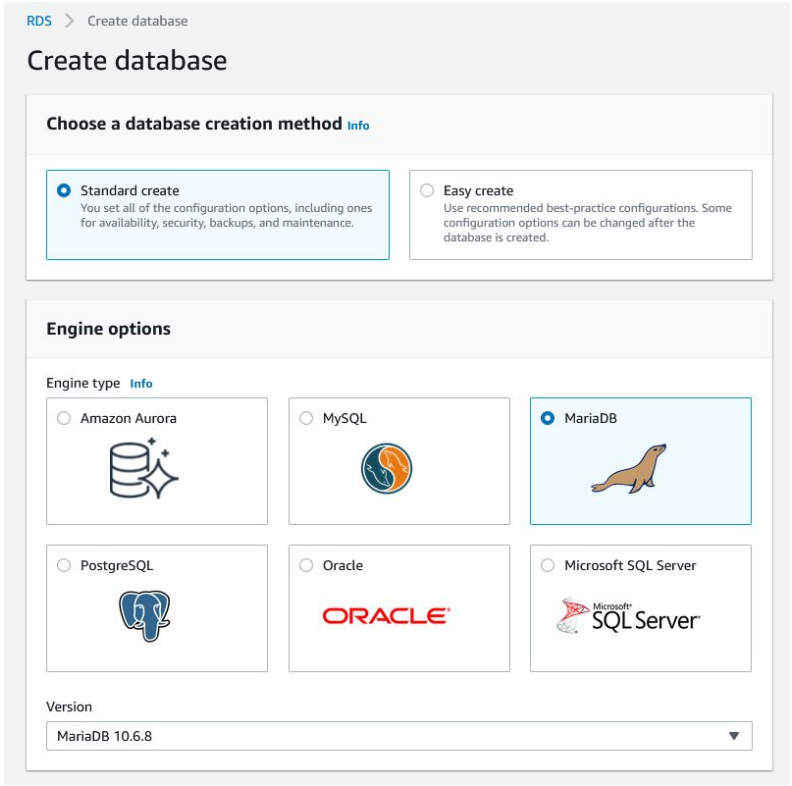
Arquitetura sugerida



Passo-a-passo

- Criar instância no RDS
  - Instalar DBeaver ou similar
  - Configurar segurança da instância para receber conexões públicas.
- Criar dois buckets no S3
- Desenvolver lambda com trigger de S3 para leitura do arquivo original e criação do arquivo csv
- Desenvolver lambda com trigger de S3 para leitura do arquivo csv e envio de mensagem para o SQS
- Desenvolver lambda com trigger de SQS para salvar dados no RDS

Criação RDS


RDS > Create database
Create database
Choose a database creation method Info
Standard create
You set all of the configuration options, including ones for availability, security, backups, and maintenance.
Easy create
Use recommended best-practice configurations. Some configuration options can be changed after the database is created.
Engine options
Engine type Info
Amazon Aurora
MySQL
MariaDB
PostgreSQL
Oracle
Microsoft SQL Server
Version
MariaDB 10.6.8


Templates
Choose a sample template to meet your use case.
Production
Use defaults for high availability and fast, consistent performance.
Dev/Test
This instance is intended for development use outside of a production environment.
Free tier
Use RDS Free Tier to develop new applications, test existing applications, or gain hands-on experience with Amazon RDS.
Info

Criação RDS

Settings

DB instance identifier Info

Type a name for your DB instance. The name must be unique across all DB instances owned by your AWS account in the current AWS Region.

ml-mba-test-1

The DB instance identifier is case-insensitive, but is stored as all lowercase (as in "mydbinstance"). Constraints: 1 to 60 alphanumeric characters or hyphens. First character must be a letter. Can't contain two consecutive hyphens. Can't end with a hyphen.

▼ Credentials Settings

Master username Info

Type a login ID for the master user of your DB instance.

admin

1 to 16 alphanumeric characters. First character must be a letter.

☐ Auto generate a password

Amazon RDS can generate a password for you, or you can specify your own password.

Master password Info

••••••••••

Constraints: At least 8 printable ASCII characters. Can't contain any of the following: / (slash), '(single quote), "(double quote) and @ (at sign).

Confirm password Info

••••••••••

Instance configuration

The DB instance configuration options below are limited to those supported by the engine that you selected above.

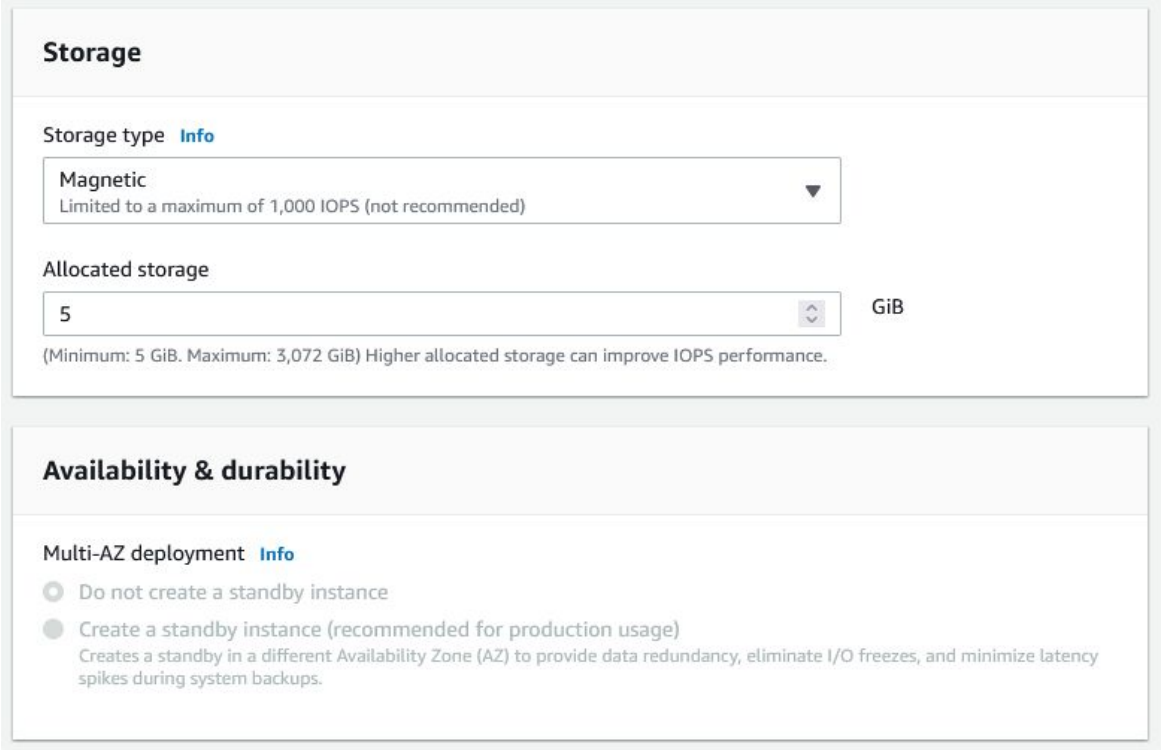
DB instance class Info

- ○ Standard classes (includes m classes)
- ○ Memory optimized classes (includes r and x classes)
- ◉ Burstable classes (includes t classes)

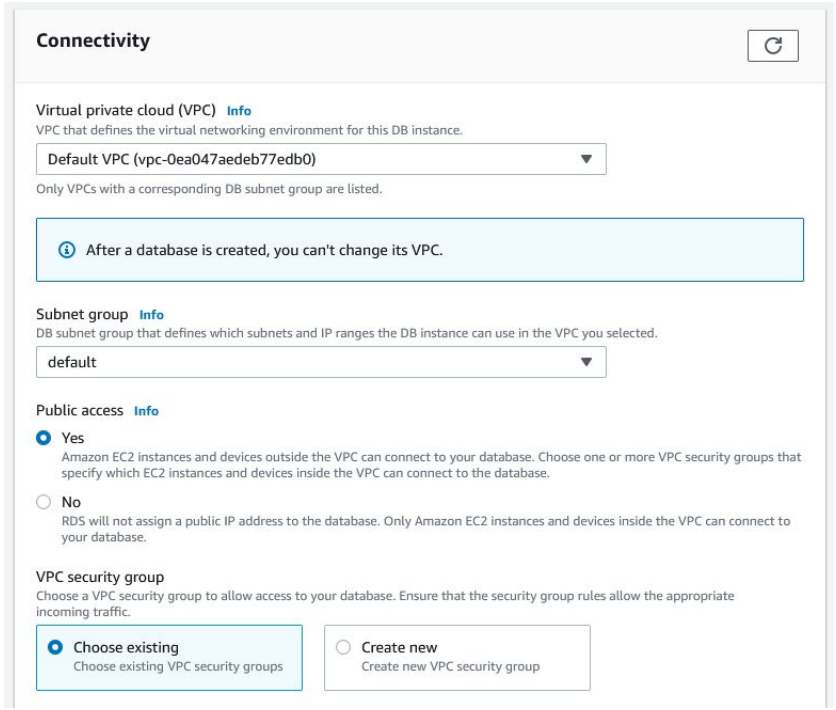
db.t3.micro
2 vCPUs 1 GiB RAM Network: 2,085 Mbps ▼

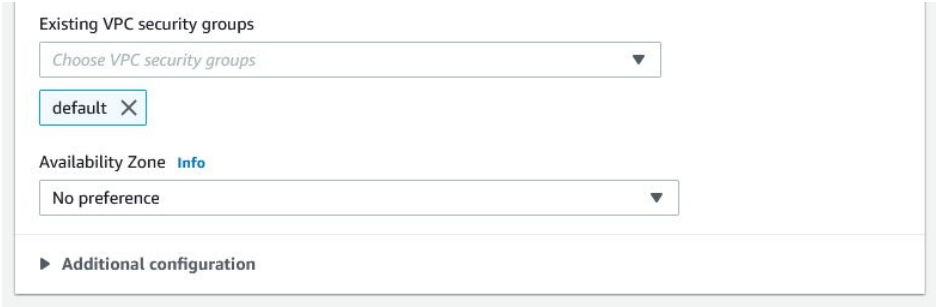
⬭ Include previous generation classes

Criação RDS

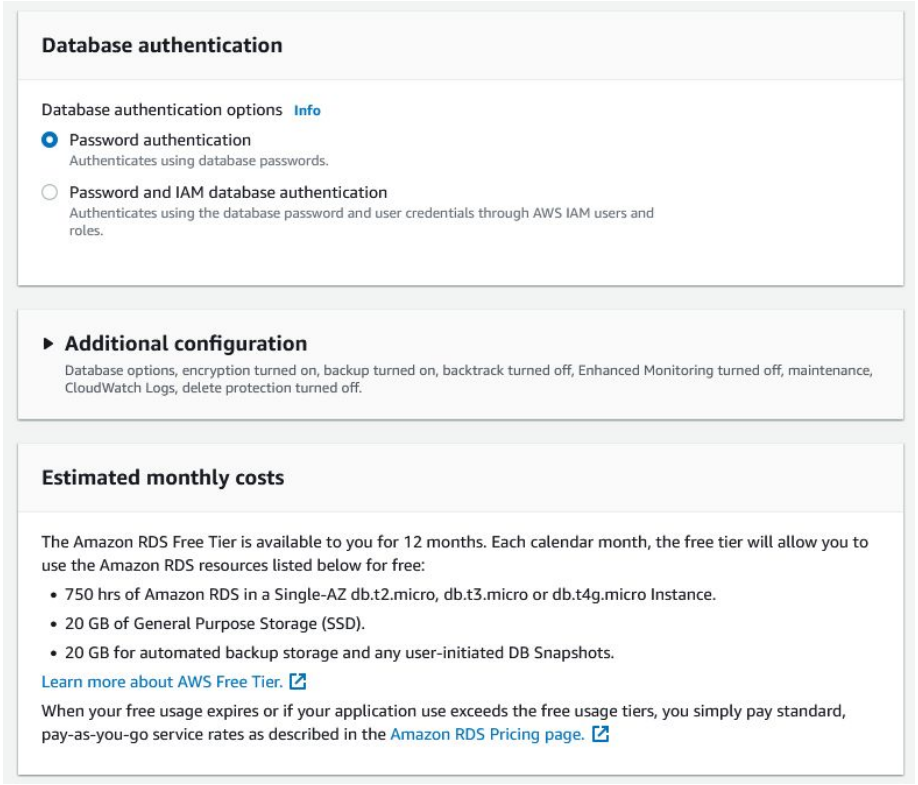

Storage
Storage type Info
Magnetic
Limited to a maximum of 1,000 IOPS (not recommended)
Allocated storage
5
GiB
(Minimum: 5 GiB. Maximum: 3,072 GiB) Higher allocated storage can improve IOPS performance.
Availability & durability
Multi-AZ deployment Info
Do not create a standby instance
Create a standby instance (recommended for production usage)
Creates a standby in a different Availability Zone (AZ) to provide data redundancy, eliminate I/O freezes, and minimize latency spikes during system backups.

Criação RDS

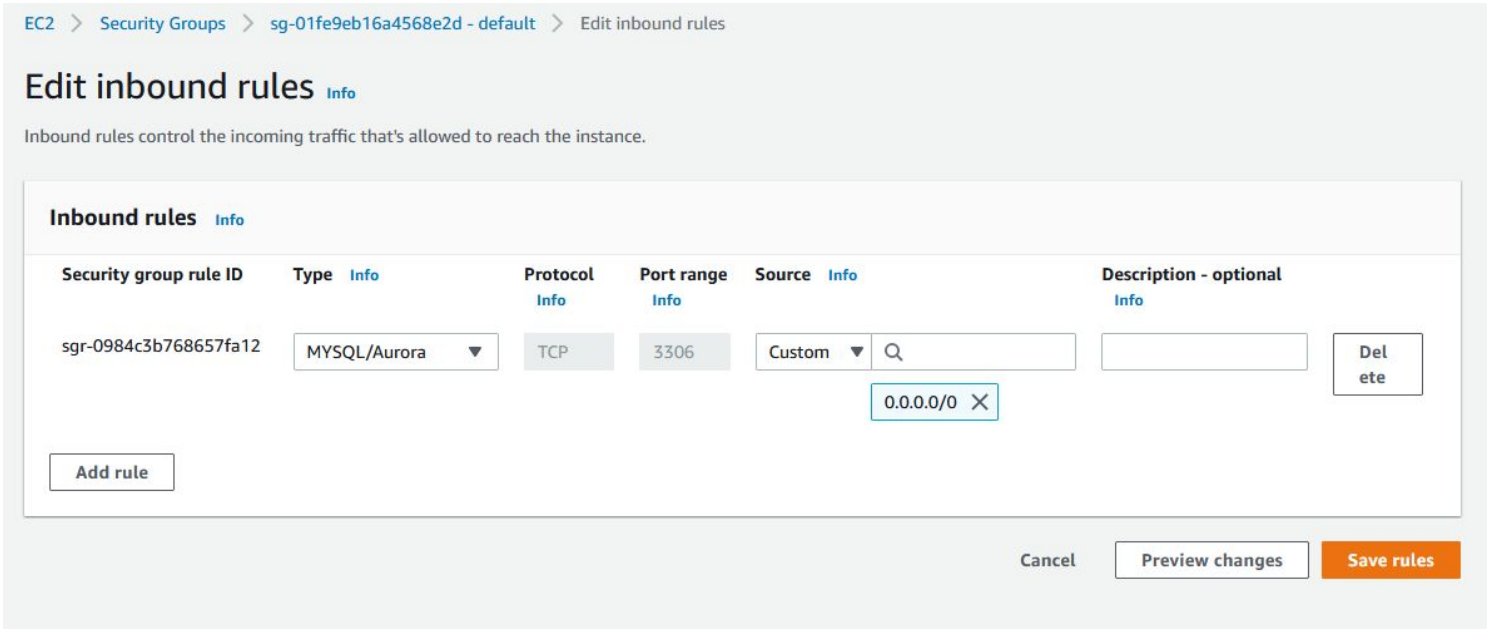

Connectivity
Virtual private cloud (VPC) Info
VPC that defines the virtual networking environment for this DB instance.
Default VPC (vpc-0ea047aedeb77edb0)
Only VPCs with a corresponding DB subnet group are listed.
After a database is created, you can't change its VPC.
Subnet group Info
DB subnet group that defines which subnets and IP ranges the DB instance can use in the VPC you selected.
default
Public access Info
Yes
Amazon EC2 instances and devices outside the VPC can connect to your database. Choose one or more VPC security groups that specify which EC2 instances and devices inside the VPC can connect to the database.
No
RDS will not assign a public IP address to the database. Only Amazon EC2 instances and devices inside the VPC can connect to your database.
VPC security group
Choose a VPC security group to allow access to your database. Ensure that the security group rules allow the appropriate incoming traffic.
Choose existing
Choose existing VPC security groups
Create new
Create new VPC security group


Existing VPC security groups
Choose VPC security groups
default
Availability Zone Info
No preference
Additional configuration

Criação RDS


Database authentication
Database authentication options Info
Password authentication
Authenticates using database passwords.
Password and IAM database authentication
Authenticates using the database password and user credentials through AWS IAM users and roles.
Additional configuration
Database options, encryption turned on, backup turned on, backtrack turned off, Enhanced Monitoring turned off, maintenance, CloudWatch Logs, delete protection turned off.
Estimated monthly costs
The Amazon RDS Free Tier is available to you for 12 months. Each calendar month, the free tier will allow you to use the Amazon RDS resources listed below for free:
750 hrs of Amazon RDS in a Single-AZ db.t2.micro, db.t3.micro or db.t4g.micro Instance.
20 GB of General Purpose Storage (SSD).
20 GB for automated backup storage and any user-initiated DB Snapshots.
Learn more about AWS Free Tier.
When your free usage expires or if your application use exceeds the free usage tiers, you simply pay standard, pay-as-you-go service rates as described in the Amazon RDS Pricing page.

Criação RDS


EC2 > Security Groups > sg-01fe9eb16a4568e2d - default > Edit inbound rules
Edit inbound rules Info
Inbound rules control the incoming traffic that's allowed to reach the instance.
Inbound rules Info
Security group rule ID
Type Info
Protocol Info
Port range Info
Source Info
Description - optional Info
sgr-0984c3b768657fa12
MYSQL/Aurora
TCP
3306
Custom
0.0.0.0/0
Delete
Add rule
Cancel
Preview changes
Save rules

Entregável

O aluno deverá processar esse arquivo, salvar cada informação retornada dos links e extrair as informações abaixo:

- ID
- Nome
- Tipo(s)
- Altura
- Peso
- Quantidade de movimentos que pode aprender

O aluno deve se organizar a fim de conseguir encontrar as informações necessárias e salvar em um Banco de Dados.

Os alunos têm total liberdade para escolher quais serviços devem ser utilizados e como o projeto deve ser organizado.

É importante deixar claro as ferramentas e a lógica que utilizou para solucionar o problema.

O cálculo de custo deve **ser anua**l e deve seguir os seguinte pontos:

- Aumentar a volumetria em 100x.
- O processo criado é executado 1 vez por dia.
- O dado para de ser acessado em 6 meses, porém é necessário armazenar por 1 ano.