

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`

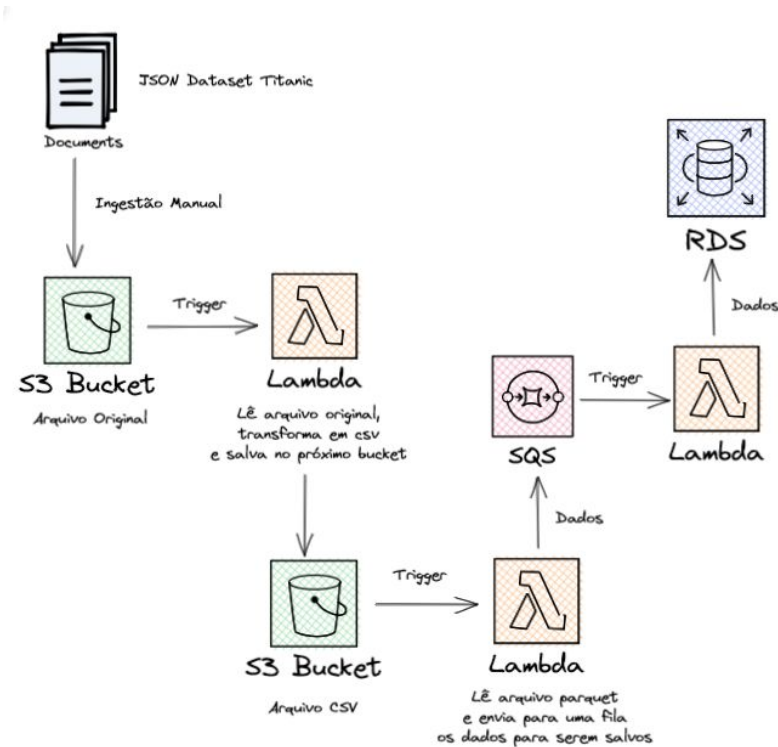


Links

- AWS: https://signin.aws.amazon.com/signup?request_type=register
- 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 dois buckets no S3
- Criar fila SQS
- Criar Políticas
- Criar instância no RDS
 - Instalar DBeaver ou similar
 - Configurar segurança da instância para receber conexões públicas.
- 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 Conta AWS



Explore Free Tier products with a new AWS account.

To learn more, visit aws.amazon.com/free.



Sign up for AWS

Root user email address

Used for account recovery and as described in the [AWS Privacy Notice](#)

AWS account name

Choose a name for your account. You can change this name in your account settings after you sign up.

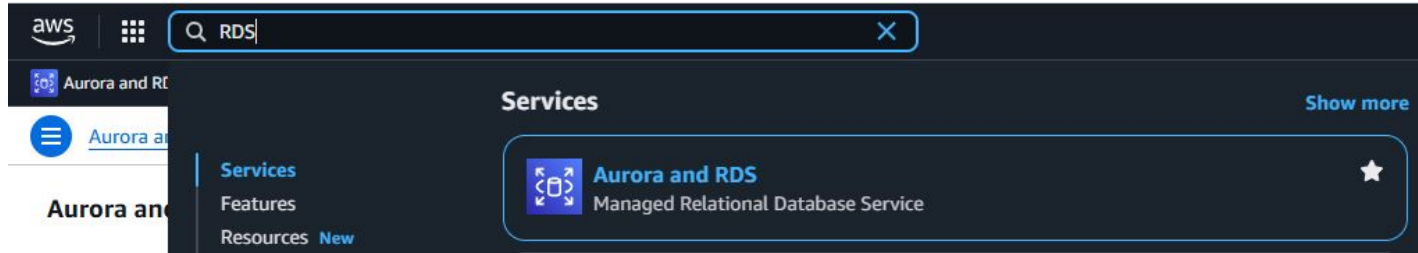
Verify email address

OR

Sign in to an existing AWS account



Criação RDS



Recursos

Você está usando os recursos do Amazon RDS a seguir na região US East (N. Virginia) (usado/cota)

Instâncias de banco de dados (1/40)

Armazenamento alocado (0.005 TB/100 TB)

As instâncias e o armazenamento incluem Neptune e DocumentDB. Aumentar o limite de instâncias de banco de dados

Clusters de banco de dados (0/40)

Instâncias reservadas (0/40)

Snapshots (1)

Manual

Cluster de banco de dados (0/100)

Instância de banco de dados (0/100)

Grupos de parâmetros (3)

Padrão (2)

Personalizado (1/100)

Grupos de opções (2)

Padrão (2)

Personalizado (0/20)

Grupos de sub-redes (1/50)

Plataformas compatíveis VPC

Rede padrão vpc-beb2a9c4



Criação RDS

Criar banco de dados



Escolher um método de criação de banco de dados

☒ Criação padrão

Defina todas as opções de configuração, incluindo as de disponibilidade, segurança, backups e manutenção.

☐ Criação fácil

Use as configurações recomendadas de melhores práticas. Algumas opções de configuração podem ser alteradas após a criação do banco de dados.

☒ MariaDB



☒ Nível gratuito

Use o nível gratuito do RDS para desenvolver novas aplicações, testá-las ou obter uma experiência prática com o Amazon RDS. [Informações](#)



Criação RDS

Configurações

Identificador da instância de banco de dados [Informações](#)

Digite um nome para a instância de banco de dados. O nome deve ser exclusivo entre todas as instâncias de banco de dados de propriedade de sua conta da AWS na região atual da AWS.

database-1

O identificador da instância de banco de dados não diferencia maiúsculas de minúsculas, mas é armazenado com todas as letras minúsculas (como em "mydbinstance"). Restrições: 1 a 63 caracteres alfanuméricos ou hífens. O primeiro caractere deve ser uma letra. Não pode conter dois hífens consecutivos. Não pode terminar com um hífen.

▼ Configurações de credenciais

Nome do usuário principal [Informações](#)

Digite um ID de login para o usuário principal de sua instância de banco de dados.

admin

De um a 16 caracteres alfanuméricos. O primeiro caractere deve ser uma letra.

Gerenciamento de credenciais

Você pode usar o AWS Secrets Manager ou gerenciar suas credenciais de usuário principal.



Gerenciado no AWS Secrets Manager - *mais seguro*

O RDS gera uma senha para você e a gerencia durante todo o ciclo de vida usando o AWS Secrets Manager.



Autogerenciada

Crie sua própria senha ou faça com que o RDS crie uma senha para você gerenciar.



Gerar senha automaticamente

O Amazon RDS pode gerar uma senha para você, ou você pode especificar sua própria senha.

Senha principal [Informações](#)

Password strength

Restrições mínimas: pelo menos 8 caracteres ASCII imprimíveis. Não pode conter nenhum dos seguintes símbolos: / ' " @

Confirmar senha principal [Informações](#)



Criação RDS

Configuração da instância

As opções de configuração da instância de banco de dados abaixo são limitadas àquelas compatíveis com o mecanismo selecionado acima.

Classe da instância de banco de dados [Informações](#)

▼ Ocultar filtros

☐ Mostrar classes de instância compatíveis com gravações otimizadas do Amazon RDS

Informações

As gravações otimizadas do Amazon RDS melhoram o throughput de gravação em até duas vezes sem custo adicional.

☐ Incluir as classes de geração anteriores

☐ Classes padrão (inclui classes m)

☐ Classes otimizadas para memória (inclui classes r e x)

☒ Classes com capacidade de intermitência (inclui classes t)

db.t4g.micro

2 vCPUs 1 GiB RAM Rede: até 2085 Mbps

Armazenamento

Tipo de armazenamento [Informações](#)

Volumes de armazenamento SSD de IOPS provisionadas (io2) agora estão disponíveis.

SSD de uso geral (gp2)

Performance de linha de base determinada pelo tamanho do volume

Armazenamento alocado [Informações](#)

20

GiB

O valor de armazenamento alocado deve ser de 20 GiB a 6144 GiB

► Configuração adicional de armazenamento



Criação RDS

Conectividade [Informações](#)



Recurso de computação

Escolha se deseja configurar uma conexão com um recurso de computação para esse banco de dados. A configuração de uma conexão altera automaticamente as configurações de conectividade, para que o recurso de computação possa se conectar a esse banco de dados.

☒ Não se conectar a um recurso de computação do EC2

Não configure uma conexão com um recurso de computação para esse banco de dados. Você poderá configurar uma conexão com um recurso de computação manualmente mais tarde.

☐ Conectar-se a um recurso de computação do EC2

Configure uma conexão com um recurso de computação do EC2 para esse banco de dados.

Tipo de rede [Informações](#)

Para usar o modo de pilha dupla, certifique-se de associar um bloco CIDR do IPv6 a uma sub-rede na VPC especificada.

☒ IPv4

Os seus recursos só podem comunicar-se por meio do protocolo de endereçamento IPv4.

☐ Modo de pilha dupla

Os seus recursos podem comunicar-se por meio do IPv4, IPv6 ou ambos.

Nuvem privada virtual (VPC) [Informações](#)

Escolha a VPC. A VPC define o ambiente de rede virtual dessa instância de banco de dados.

Default VPC (vpc-beb2a9c4)

6 Sub-redes, 6 Zonas de disponibilidade

Somente as VPCs com um grupo de sub-redes de banco de dados correspondente são listadas.

① Depois de criar o banco de dados, não é possível alterar a VPC.

Grupo de sub-redes de banco de dados [Informações](#)

Escolha o grupo de sub-redes de banco de dados. O grupo de sub-redes de banco de dados define quais sub-redes e intervalos de IP a instância de banco de dados pode usar na VPC selecionada.

default

6 Sub-redes, 6 Zonas de disponibilidade

Acesso público [Informações](#)

☒ Sim

O RDS atribui um endereço IP público ao banco de dados. As instâncias do Amazon EC2 e outros recursos fora da VPC podem se conectar ao seu banco de dados. Os recursos dentro da VPC também podem se conectar ao banco de dados. Escolha um ou mais grupos de segurança de VPC que especificam quais recursos podem se conectar ao banco de dados.

☐ Não

O RDS não atribui um endereço IP público ao banco de dados. Somente instâncias do Amazon EC2 e outros recursos dentro da VPC podem se conectar ao seu banco de dados. Escolha um ou mais grupos de segurança de VPC que especificam quais recursos podem se conectar ao banco de dados.



Criação RDS

Custos mensais estimados

O nível gratuito do Amazon RDS ficará disponível para você por 12 meses. A cada mês, o nível gratuito permite o uso gratuito dos recursos do Amazon RDS listados abaixo:

- 750 horas do Amazon RDS em uma instância db.t2.micro, db.t3.micro ou db.t4g.micro Single-AZ.
- 20 GB de armazenamento de uso geral (SSD).
- 20 GB de armazenamento para backup automatizado e qualquer snapshot de banco de dados iniciado pelo usuário.

[Saiba mais sobre o nível gratuito da AWS.](#)

Quando seu uso gratuito expirar ou se seu uso do aplicativo ultrapassar os níveis de uso gratuito, você simplesmente pagará taxas de serviço padrão de acordo com o uso, conforme descrito na [página de preços do Amazon RDS.](#)

Bancos de dados (1) Recursos do grupo Modificar Ações Criar banco de dados

	Identificador de banco de dados	Status	Função	Mecani...	Região ...	Tamanh...
<input type="radio"/>	<u>database-1</u>	✓ Disponível	Instância	MariaDB	us-east-1f	db.t4g.m



Criação RDS

Regras de grupos de segurança (3)

Q Filtrar por Regras de grupos de segurança

Grupo de segurança

	Tipo
default (sg-56ea3873)	EC2 Security Group - Inbound
default (sg-56ea3873)	CIDR/IP - Inbound
default (sg-56ea3873)	CIDR/IP - Outbound

Editar regras de entrada

sgr-03e1fe30f5877cb7a

MYSQL/Aurora

TCP

3306

Persona...

Q

0.0.0.0/0 X

Adicionar regra

Grupos de segurança (1) [Informações](#)

Q Encontre grupos de segurança por atributo ou etiqueta

default X

Limpar filtros

<input type="checkbox"/>	Name	ID do grupo de segurança	Nome do grupo de segurança
<input type="checkbox"/>	-	sg-56ea3873	default



Dúvidas?



Entregável

O aluno deverá processar o um arquivo [link](#), salvar cada informação retornada dos links e extrair as informações abaixo:

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

[Documentação da API](#)

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 mensal** 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 após 7 dias, porém é necessário armazenar por 1 mês.

