

# Passo a passo AWS Athena

## Etapa 1: Configurar Athena:

- Crie uma pasta dentro do *bucket* chamada **queries**. O **AWS Athena** usará esta pasta para armazenar as consultas executadas.

Amazon S3 > Buckets > testeum.com > Criar pasta

### Criar pasta Informações

Use pastas para agrupar objetos em buckets. Quando você cria uma pasta, o S3 cria um objeto usando o nome especificado seguido por uma barra (/). Esse objeto aparece como pasta no console. [Saiba mais](#)

**Sua política de bucket pode bloquear a criação de pastas**

Se a política de bucket impedir o upload de objetos sem tags, metadados ou usuários favorecidos da lista de controle de acesso (ACL) específicos, você não poderá criar uma pasta usando essa configuração. Em vez disso, use a [configuração de upload](#) para fazer upload de uma pasta vazia e especifique as configurações apropriadas.

**Pasta**

Nome de pasta

 /

Os nomes de pastas não podem conter "/". [Consulte as regras para nomenclatura](#)

🟢 Pasta "queries" criada com êxito.

Objetos | Propriedades | Permissões | Métricas | Gerenciamento | Pontos de acesso

**Objetos (4)**

Os objetos são as entidades fundamentais armazenadas no Amazon S3. Você pode usar o [inventário do Amazon S3](#) para obter uma lista de todos os objetos em seu bucket. Para outras pessoas acessarem seus objetos, você precisará conceder permissões explicitamente a eles. [Saiba mais](#)

Copiar URI do S3 Copiar URL Fazer download Abrir Excluir Ações Criar pasta Carregar

<input type="checkbox"/>	Nome	Tipo	Última modificação	Tamanho	Classe de armazenamento
<input type="checkbox"/>	404.html	html	25 Oct 2023 08:16:47 PM -03	0 B	Padrão
<input type="checkbox"/>	dados/	Pasta	-	-	-
<input type="checkbox"/>	index.html	html	25 Oct 2023 08:16:48 PM -03	261.0 B	Padrão
<input type="checkbox"/>	queries/	Pasta	-	-	-

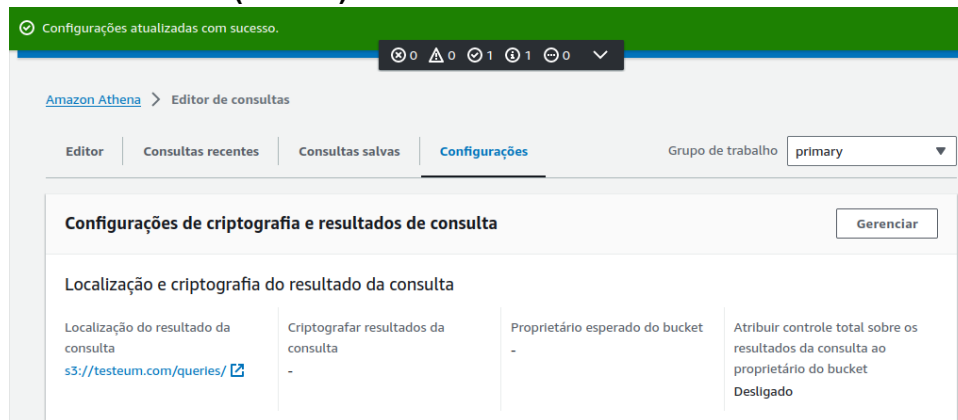
- Acesse o AWS Athena e clique em **Explore the Query Editor (Explorar o editor de consultas)**.

The screenshot shows the Amazon Athena console. At the top, there's a search bar with 'Athena' entered. Below it, a sidebar on the left lists various services and features. The main content area displays the 'Amazon Athena' service card, which includes a 'Comece a usar' (Get started) section with options to use SQL Trino or PySpark. Below this, there's a blue banner about Apache Spark. The bottom part of the screenshot shows the 'Editor de consultas' (Query editor) interface, which includes tabs for 'Editor', 'Consultas recentes', 'Consultas salvas', and 'Configurações'. The 'Configurações' tab is active, showing settings for 'Localização e criptografia do resultado da consulta' (Query result location and encryption).

- No Athena, escolha **View Settings (Visualizar configurações)** para configurar um local para os resultados de consultas no Amazon S3.
- Na guia **Settings (Configurações)**, escolha **Manage (Gerenciar)**.
- Em **Manage settings (Gerenciar configurações)**, faça um dos seguintes procedimentos:
  - Na caixa de texto **Query result location (Localização dos resultados da consulta)**, insira o caminho para o bucket criado no Amazon S3 para resultados de consultas. Adicione o prefixo `s3://` ao caminho. E aponte para a pasta **queries criadas**.

The screenshot shows the 'Gerenciar configurações' (Manage settings) page in the Amazon Athena console. The page has a breadcrumb trail: 'Amazon Athena > Editor de consultas > Gerenciar configurações'. The main heading is 'Gerenciar configurações'. Below it, there's a section titled 'Localização e criptografia do resultado da consulta' (Query result location and encryption) with a sub-heading 'Localização do resultado da consulta - opcional' (Query result location - optional). The text says: 'Insira um prefixo S3 na região atual em que o resultado da consulta será salvo como um objeto.' (Enter an S3 prefix in the current region where the query result will be saved as an object). There's a search bar with the text 's3://testeum.com/queries' and a 'Visualizar' (View) button. To the right, there's a 'Navegue pelo S3' (Browse S3) button. Below this, there's a blue box with an information icon and text: 'Você pode criar e gerenciar regras de ciclo de vida para esse bucket' (You can create and manage lifecycle rules for this bucket). It also says: 'Use as regras de ciclo de vida do Amazon S3 para armazenar seus resultados de consulta e metadados de forma econômica ou para excluí-los após um período.' (Use the lifecycle rules of Amazon S3 to store your query results and metadata in a cost-effective way or to delete them after a period). There's a 'Configuração do ciclo de vida' (Lifecycle configuration) link.

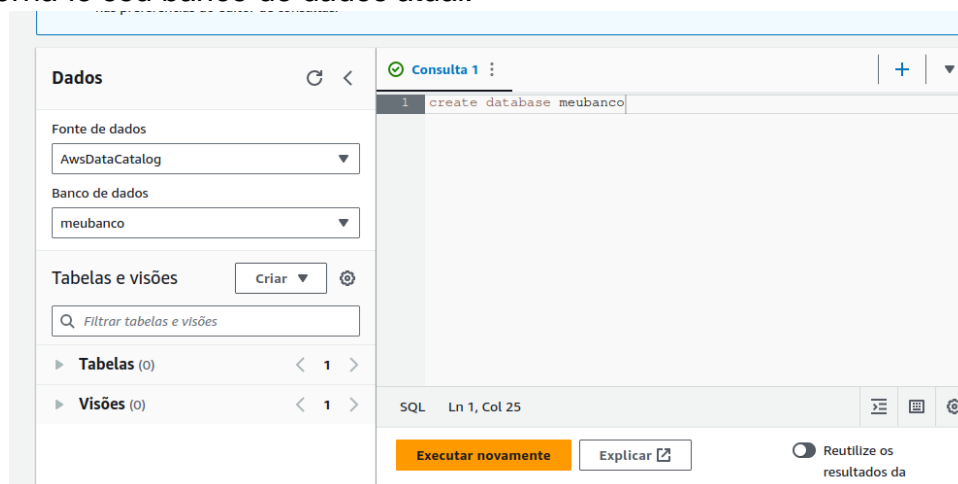
- Escolha **Browse S3 (Navegar no S3)**, escolha o *bucket* do Amazon S3 que você criou na região atual e escolha **Choose (Escolher)**.
- Escolha **Save (Salvar)**.



- Selecione **Editor** para alternar para o editor de consultas.

## Etapa 2: Criar um banco de dados:

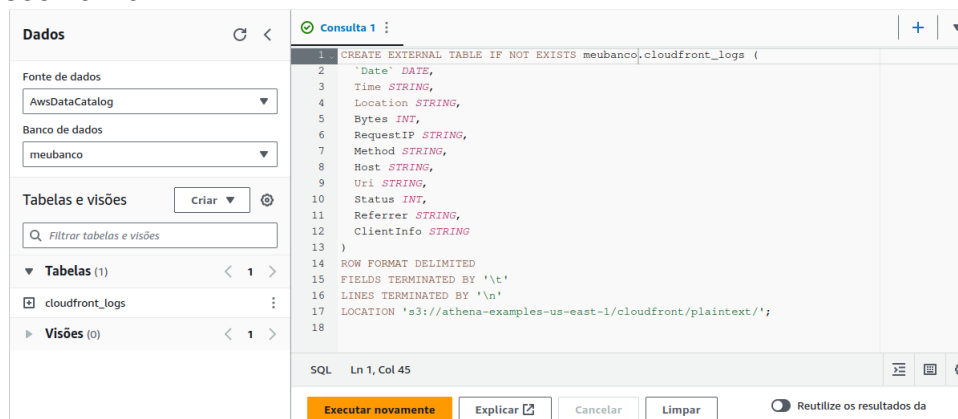
- À direita do painel de navegação, você pode usar o editor de consultas do Athena para inserir e executar as consultas e instruções.
- Para criar um banco de dados denominado meubanco, insira a instrução **CREATE DATABASE**
- Selecione **Run (Executar)** ou pressione **Ctrl+ENTER**.
- Na lista **Database (Banco de dados)** à esquerda, escolha meubanco para torná-lo seu banco de dados atual.



### Etapa 3: Criar uma tabela:

Agora que você tem um banco de dados, pode criar uma tabela do Athena para ele. A tabela criada será baseada nos dados de log de exemplo do Amazon CloudFront, no local `s3://athena-examples-myregion/cloudfront/plaintext/`, em que *myregion* é a sua Região da AWS atual.

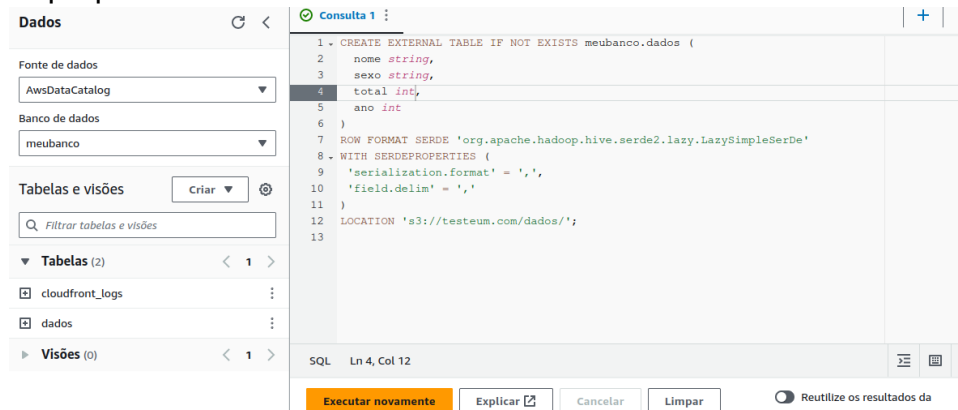
- Escolha **Run**.



- Se a importação for bem-sucedida, você verá uma mensagem verde **Completed (Concluído)**

### Alguns itens a observar para a atividade:

- Definimos todos os campos no conjunto de dados e demos a eles um tipo apropriado.



- Informamos ao Athena para usar o analisador *LazySimpleSerDe* CSV. Usamos esse analisador porque ele permite valores nulos para números. Ele não suporta valores entre aspas.
- Informamos ao analisador que os campos são delimitados por vírgulas e que a primeira linha contém nomes de campos que podem ser ignorados.
- Especificamos o local do arquivo CSV. Precisamos apenas fornecer a pasta, não o arquivo em si.

- Teste os dados com a seguinte consulta, substituindo o nome dos campos, banco de dados e tabela pelos nomes que você criou anteriormente:

SELECT NOME FROM NOME DO BANCO.NOME DA TABELA WHERE ANO = 1999 ORDER BY TOTAL LIMIT 15;

**Dados**

Fonte de dados: AwsDataCatalog

Banco de dados: meubanco

Tabelas e visões: cloudfront\_logs, dados

**Consulta 1**

```
select nome from meubanco.dados where ano = 1999 order by total limit 15;
```

**Resultados (15)**

#	nome
1	Adalah
2	Aanya
3	Abinaya
4	Abriona
5	Achsah
6	Aalijah
7	Aania

- Crie uma consulta que lista os 3 nomes mais usados em cada década desde o 1950 até hoje.

**Consulta 1**

```
1 WITH nomesUsados AS (
2   SELECT
3     nome,
4     FLOOR(ano / 10) * 10 AS decada,
5     sum(total) AS quantidade,
6     ROW_NUMBER() OVER (PARTITION BY FLOOR(ano / 10) * 10 ORDER BY sum(total) DESC) AS rank
7   FROM dados
8   WHERE ano >= 1950
9   GROUP BY nome, FLOOR(ano / 10) * 10
10 )
11 SELECT
12   nome,
13   decada,
14   quantidade
15 FROM nomesUsados
16 WHERE rank <= 3
17 order by decada, quantidade;
```

## Resultado Da Consulta

nome	decada	quantidade
Robert	1950	832110
Michael	1950	839312
James	1950	845821
John	1960	716364
David	1960	736709
Michael	1960	837068
Christopher	1970	478445
Jennifer	1970	583688
Michael	1970	712793
Jessica	1980	471446
Christopher	1980	558407
Michael	1980	668724
Matthew	1990	352315
Christopher	1990	361223
Michael	1990	464210
Joshua	2000	232246
Michael	2000	251189
Jacob	2000	274165
Isabella	2010	96445
Emma	2010	98776
Sophia	2010	104441