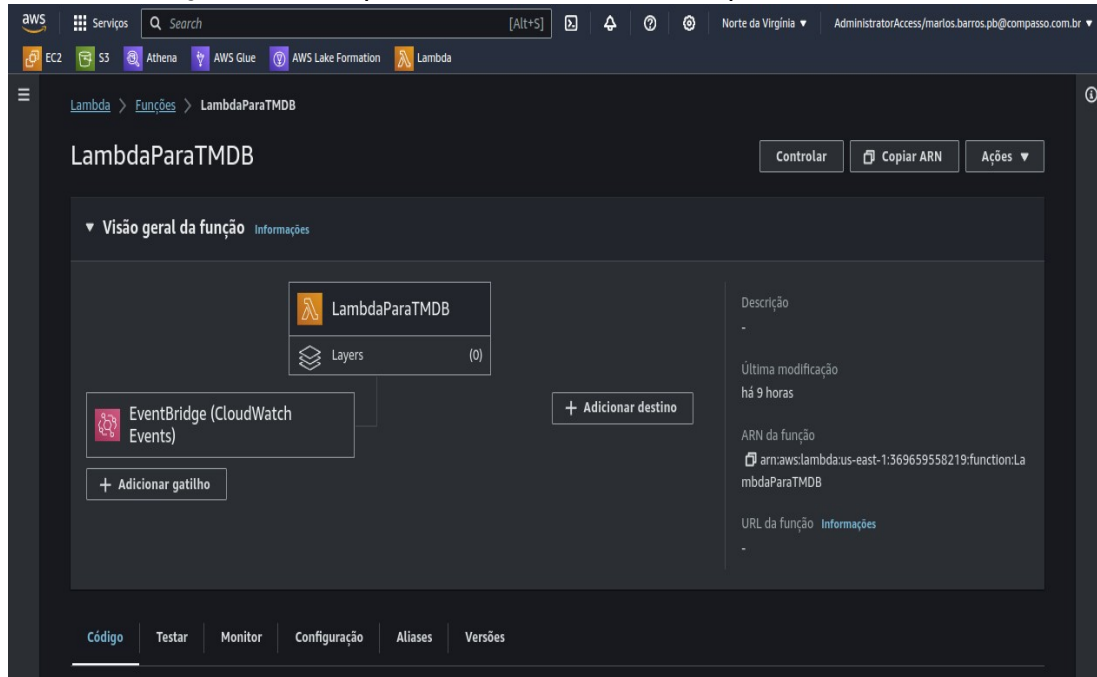
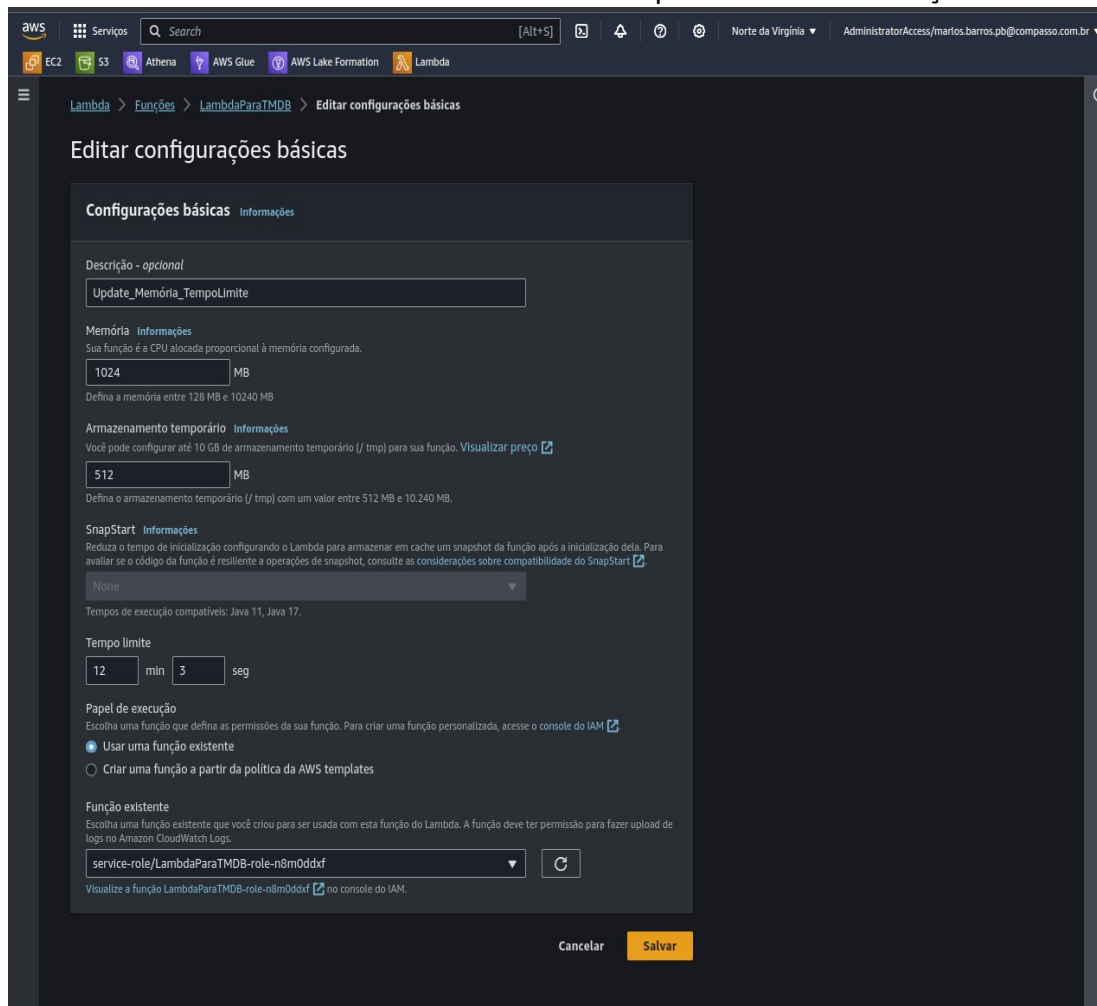


Criei uma função lambda para fazer chamadas a api TMDB:



Aumentei a memória da CPU alocada e o tempo limite da execução:



Capturei os dados do TMDb utilizando uma função AWS Lambda para fazer chamadas à API. Os dados que coletei foram armazenados no Amazon S3, na camada RAW Zone, mantendo o formato original (JSON). Além disso, consegui agrupá-los em arquivos contendo no máximo 100 registros cada:

```
aws
Serviços
[Search] [Alt+S] [Norte da Virgínia] AdministratorAccess/marios.barros.pb@compasso.com.br

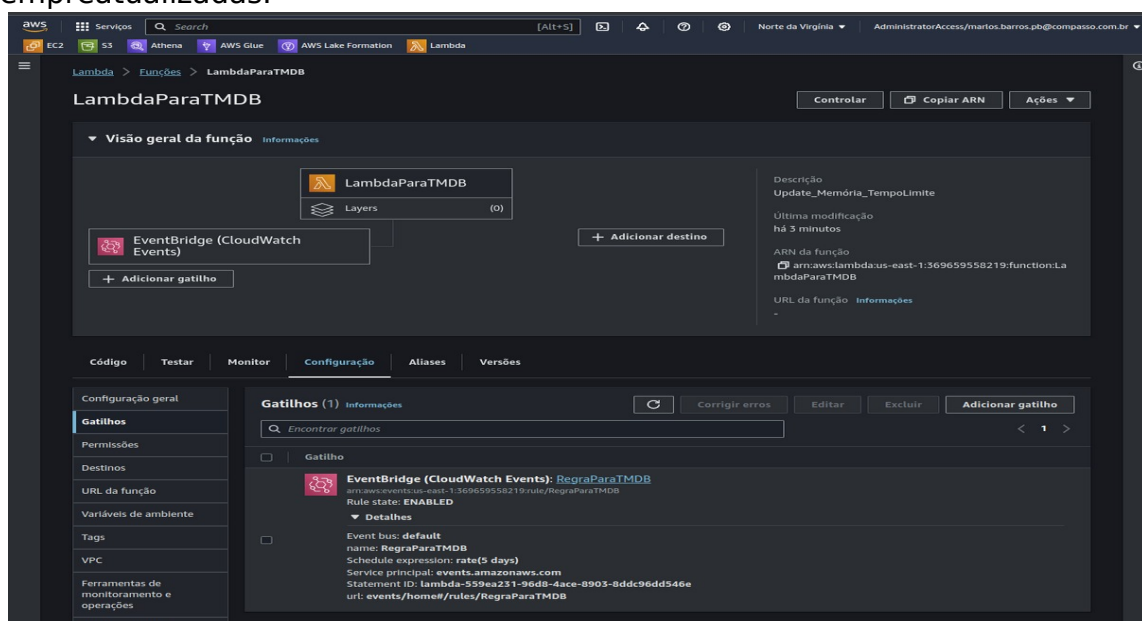
EC2 S3 Athena AWS Glue AWS Lake Formation Lambda

File Edit Find View Go Tools Window Test Deploy

Go to Anything
Environment
  - Environment
  - Lambda
  - Environment Var
  - Execution results

1 import json
2 import boto3
3 import requests
4 from datetime import datetime
5
6 REGIAO_AWS = 'us-east-1'
7 ID_ACESSO_IAM_AWS = 'AKIAWMELNFF7PM222ZS'
8 CHAVE_SECRET_IAM_AWS = 'SmYmBG1q3o4j60LZoYVyz3VnOU8XMMQc-lP1NVE'
9
10 TIPO_MIDIA = {'tv': 'Series',
11              'movie': 'Movies'}
12
13 GENEROS = {'28,12': 'Ação-Aventura',
14            '28': 'Ação',
15            '12': 'Aventura',
16            '18759': 'Ação-Aventura'}
17
18 def obter_dados_tmdb(pagina, tipo_midia, genero):
19     url = (f'https://api.themoviedb.org/3/discover/{tipo_midia}?api_key=cdfef3627e7ecf677ae84cd7fee138e28'
20           f'&language=pt-BR&sort_by=popularity.desc&include_adult=false&'
21           f'&include_video=false&page={pagina}&'
22           f'&with_genres={genero}')
23
24     resposta = requests.get(url)
25     return resposta.json()
26
27 def enviar_para_s3(dados, indice, tipo_midia, genero):
28     s3 = boto3.client('s3', aws_access_key_id=ID_ACESSO_IAM_AWS,
29                       aws_secret_access_key=CHAVE_SECRET_IAM_AWS,
30                       region_name=REGIAO_AWS)
31
32     nome_bucket = 'data-lake-narlos-igor'
33     nome_arquivo = (f'raw/TMDB/JSON/{TIPO_MIDIA[tipo_midia]}/{GENEROS[genero]}/'
34                    f'{datetime.now().strftime("%Y/%m/%d")}.json')
35     s3.put_object(Body=json.dumps(dados), Bucket=nome_bucket, Key=nome_arquivo)
36
37 def processar_midia(tipo_midia, generos):
38     for genero in generos:
39         dados = obter_dados_tmdb(1, tipo_midia, genero)
40         todos_resultados = dados.get('results', [])
41         total_paginas = dados.get('total_pages', 1)
42
43         for pagina in range(2, total_paginas + 1):
44             dados = obter_dados_tmdb(pagina, tipo_midia, genero)
45             todos_resultados.extend(dados.get('results', []))
46
47             for i in range(0, len(todos_resultados), 100):
48                 pedaco = todos_resultados[i:i + 100]
49                 enviar_para_s3(pedaco, i // 100, tipo_midia, genero)
50
51 def lambda_handler(event, context):
52     try:
53         processar_midia('movie', ['28,12', '28', '12'])
54         processar_midia('tv', ['18759'])
55         return {
56             'statusCode': 200,
57             'body': json.dumps('Upload para o S3 realizado com sucesso!')
58         }
59     except Exception as e:
60         return {
61             'statusCode': 500,
62             'body': json.dumps(f'Erro: {str(e)}')
63         }
```

Configurei um sistema de automação utilizando o Amazon EventBridge, que aciona a coleta de dados a cada 5 dias para manter nossas informações sempre atualizadas:



Capturei os filmes e séries e os separei por pasta, cada arquivo JSON na pasta de filmes contém 100 registros do gênero Ação/Aventura:

Exemplo: Filmes com ID: 28. 12 do gênero Ação/Aventura
s3://data-lake-marlos-igor/Raw/TMDB/JSON/Movies/Ação-Aventura/2023/10/10/

EC2S3AthenaAWS GlueAWS Lake FormationLambda

Tamanho total calculado com êxito
Visualize os detalhes abaixo.

Calcular o tamanho totalInformações

As informações abaixo não estarão mais disponíveis depois que você sair desta página.

Resumo

Origem

s3://data-lake-marlos-igor/Raw/TMDB/JSON/Movies/Ação-Aventura/2023/10/10/

Número total de objetos

68

Tamanho total

3.3 MB

Objetos especificados

Encontrar objetos por nome

| Nome | Tipo | Última modificação | Tamanho | Número total de objetos | Erro |
|--------------|------|-----------------------------|---------|-------------------------|------|
| data_0.json | json | 10 Oct 2023 06:36:20 AM -03 | 72.4 KB | 1 | - |
| data_1.json | json | 10 Oct 2023 06:36:20 AM -03 | 76.6 KB | 1 | - |
| data_10.json | json | 10 Oct 2023 06:36:21 AM -03 | 80.1 KB | 1 | - |
| data_11.json | json | 10 Oct 2023 06:36:21 AM -03 | 74.9 KB | 1 | - |
| data_12.json | json | 10 Oct 2023 06:36:22 AM -03 | 70.6 KB | 1 | - |
| data_13.json | json | 10 Oct 2023 06:36:22 AM -03 | 64.9 KB | 1 | - |
| data_14.json | json | 10 Oct 2023 06:36:22 AM -03 | 66.9 KB | 1 | - |
| data_15.json | json | 10 Oct 2023 06:36:22 AM -03 | 62.8 KB | 1 | - |
| data_16.json | json | 10 Oct 2023 06:36:22 AM -03 | 62.6 KB | 1 | - |
| data_17.json | json | 10 Oct 2023 06:36:22 AM -03 | 56.8 KB | 1 | - |
| data_18.json | json | 10 Oct 2023 06:36:22 AM -03 | 54.5 KB | 1 | - |
| data_19.json | json | 10 Oct 2023 06:36:22 AM -03 | 54.3 KB | 1 | - |

Exemplo: Filmes com ID: 28 do gênero Ação
s3://data-lake-marlos-igor/Raw/TMDB/JSON/Movies/Ação/2023/10/10/

EC2S3AthenaAWS GlueAWS Lake FormationLambda

Tamanho total calculado com êxito
Visualize os detalhes abaixo.

Calcular o tamanho totalInformações

As informações abaixo não estarão mais disponíveis depois que você sair desta página.

Resumo

Origem

s3://data-lake-marlos-igor/Raw/TMDB/JSON/Movies/Ação/2023/10/10/

Número total de objetos

100

Tamanho total

6.1 MB

Objetos especificados

Encontrar objetos por nome

| Nome | Tipo | Última modificação | Tamanho | Número total de objetos | Erro |
|--------------|------|-----------------------------|---------|-------------------------|------|
| data_0.json | json | 10 Oct 2023 06:37:45 AM -03 | 67.7 KB | 1 | - |
| data_1.json | json | 10 Oct 2023 06:37:45 AM -03 | 73.5 KB | 1 | - |
| data_10.json | json | 10 Oct 2023 06:37:46 AM -03 | 78.9 KB | 1 | - |
| data_11.json | json | 10 Oct 2023 06:37:46 AM -03 | 76.6 KB | 1 | - |
| data_12.json | json | 10 Oct 2023 06:37:46 AM -03 | 79.4 KB | 1 | - |
| data_13.json | json | 10 Oct 2023 06:37:46 AM -03 | 77.0 KB | 1 | - |
| data_14.json | json | 10 Oct 2023 06:37:46 AM -03 | 75.8 KB | 1 | - |
| data_15.json | json | 10 Oct 2023 06:37:46 AM -03 | 78.0 KB | 1 | - |
| data_16.json | json | 10 Oct 2023 06:37:46 AM -03 | 76.0 KB | 1 | - |
| data_17.json | json | 10 Oct 2023 06:37:47 AM -03 | 78.8 KB | 1 | - |
| data_18.json | json | 10 Oct 2023 06:37:47 AM -03 | 76.4 KB | 1 | - |
| data_19.json | json | 10 Oct 2023 06:37:47 AM -03 | 77.6 KB | 1 | - |
| data_2.json | json | 10 Oct 2023 06:37:45 AM -03 | 73.0 KB | 1 | - |

Exemplo: Filmes com ID: 12 do gênero Aventura
s3://data-lake-marlos-igor/Raw/TMDB/JSON/Movies/Aventura/2023/10/10/

Serviços

Search

[Alt+S]

EC2

S3

Athena

AWS Glue

AWS Lake Formation

Lambda

Global

AdministratorAccess/marlos.barros.pb@compasso.com.br

Tamanho total calculado com êxito

Visualize os detalhes abaixo.

Calcular o tamanho total

Informações

Fechar

As informações abaixo não estarão mais disponíveis depois que você sair desta página.

Resumo

Origem

s3://data-lake-marlos-igor/Raw/TMDB/JSON/Movies/Aventura/2023/10/10/

Número total de objetos

100

Tamanho total

5.5 MB

Objetos especificados

Encontrar objetos por nome

< 1 >

| Nome | Tipo | Última modificação | Tamanho | Número total de objetos | Erro |
|--------------|------|-----------------------------|---------|-------------------------|------|
| data_0.json | json | 10 Oct 2023 06:38:39 AM -03 | 74.9 KB | 1 | - |
| data_1.json | json | 10 Oct 2023 06:38:40 AM -03 | 75.4 KB | 1 | - |
| data_10.json | json | 10 Oct 2023 06:38:41 AM -03 | 78.4 KB | 1 | - |
| data_11.json | json | 10 Oct 2023 06:38:41 AM -03 | 82.6 KB | 1 | - |
| data_12.json | json | 10 Oct 2023 06:38:41 AM -03 | 79.9 KB | 1 | - |
| data_13.json | json | 10 Oct 2023 06:38:41 AM -03 | 83.9 KB | 1 | - |
| data_14.json | json | 10 Oct 2023 06:38:41 AM -03 | 78.2 KB | 1 | - |
| data_15.json | json | 10 Oct 2023 06:38:41 AM -03 | 82.9 KB | 1 | - |
| data_16.json | json | 10 Oct 2023 06:38:42 AM -03 | 77.8 KB | 1 | - |
| data_17.json | json | 10 Oct 2023 06:38:42 AM -03 | 78.5 KB | 1 | - |
| data_18.json | json | 10 Oct 2023 06:38:42 AM -03 | 76.6 KB | 1 | - |
| data_19.json | json | 10 Oct 2023 06:38:42 AM -03 | 79.3 KB | 1 | - |
| data_2.json | json | 10 Oct 2023 06:38:40 AM -03 | 75.2 KB | 1 | - |

CloudShell

Comentários

© 2023, Amazon Web Services, Inc. ou suas afiliadas.

Privacidade

Termos

Preferências de cookies

Exemplo: Series com ID: 10759 do gênero Ação/Aventura
s3://data-lake-marlos-igor/Raw/TMDB/JSON/Series/Ação-Aventura/2023/10/10/

Serviços

Search

[Alt+S]

EC2

S3

Athena

AWS Glue

AWS Lake Formation

Lambda

Global

AdministratorAccess/marlos.barros.pb@compasso.com.br

Tamanho total calculado com êxito

Visualize os detalhes abaixo.

Calcular o tamanho total

Informações

Fechar

As informações abaixo não estarão mais disponíveis depois que você sair desta página.

Resumo

Origem

s3://data-lake-marlos-igor/Raw/TMDB/JSON/Series/Ação-Aventura/2023/10/10/

Número total de objetos

71

Tamanho total

3.9 MB

Objetos especificados

Encontrar objetos por nome

< 1 >

| Nome | Tipo | Última modificação | Tamanho | Número total de objetos | Erro |
|--------------|------|-----------------------------|---------|-------------------------|------|
| data_0.json | json | 10 Oct 2023 06:39:07 AM -03 | 71.8 KB | 1 | - |
| data_1.json | json | 10 Oct 2023 06:39:07 AM -03 | 79.7 KB | 1 | - |
| data_10.json | json | 10 Oct 2023 06:39:09 AM -03 | 77.4 KB | 1 | - |
| data_11.json | json | 10 Oct 2023 06:39:09 AM -03 | 72.0 KB | 1 | - |
| data_12.json | json | 10 Oct 2023 06:39:09 AM -03 | 75.8 KB | 1 | - |
| data_13.json | json | 10 Oct 2023 06:39:09 AM -03 | 78.9 KB | 1 | - |
| data_14.json | json | 10 Oct 2023 06:39:09 AM -03 | 69.3 KB | 1 | - |
| data_15.json | json | 10 Oct 2023 06:39:09 AM -03 | 76.0 KB | 1 | - |
| data_16.json | json | 10 Oct 2023 06:39:10 AM -03 | 69.6 KB | 1 | - |
| data_17.json | json | 10 Oct 2023 06:39:10 AM -03 | 69.6 KB | 1 | - |
| data_18.json | json | 10 Oct 2023 06:39:10 AM -03 | 73.3 KB | 1 | - |
| data_19.json | json | 10 Oct 2023 06:39:10 AM -03 | 70.5 KB | 1 | - |

CloudShell

Comentários

© 2023, Amazon Web Services, Inc. ou suas afiliadas.

Privacidade

Termos

Preferências de cookies