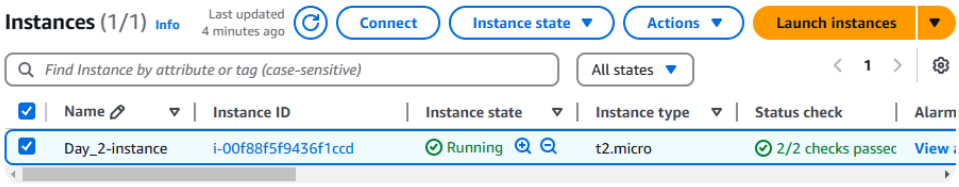
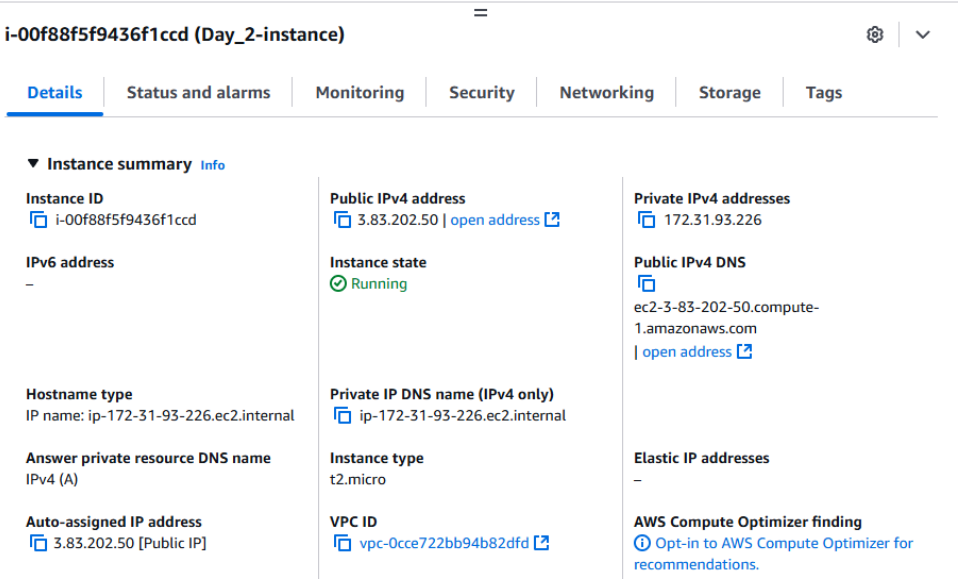
Relatório Day_2 – Fase 2 (Ana Letícia #53)

Link do repositório do GitHub: https://github.com/anaoliveira07/Desempate_Estadual

1- Ec2

Criei a instância Ec2 como foi solicitado.





2- Conectar a instância via terminal

Usei o comando: sudo yum install python3 git aws-cli -y

```
   ,     #_
   ~\_  ####_        Amazon Linux 2
  ~~  \_#####\
  ~~     \###|       AL2 End of Life is 2026-06-30.
  ~~       \#/ ___
   ~~       V~' '->
    ~~~         /    A newer version of Amazon Linux is available!
      ~~._.   _/
         _/ _/       Amazon Linux 2023, GA and supported until 2028-03-15.
       _/m/'           https://aws.amazon.com/linux/amazon-linux-2023/

[ec2-user@ip-172-31-93-226 ~]$ sudo yum update -y
Loaded plugins: extras_suggestions, langpacks, priorities, update-motd
amzn2-core
                                                                                                    | 3.6 kB  00:00:00

No packages marked for update
[ec2-user@ip-172-31-93-226 ~]$ sudo yum install python3 git aws-cli -y
Loaded plugins: extras_suggestions, langpacks, priorities, update-motd
Package python3-3.7.16-1.amzn2.0.8.x86_64 already installed and latest version
Package awscli-1.18.147-1.amzn2.0.2.noarch already installed and latest version
Resolving Dependencies
--> Running transaction check
---> Package git.x86_64 0:2.47.1-1.amzn2.0.2 will be installed
--> Processing Dependency: git-core = 2.47.1-1.amzn2.0.2 for package: git-2.47.1-1.amzn2.0.2.x86_64
--> Processing Dependency: git-core-doc = 2.47.1-1.amzn2.0.2 for package: git-2.47.1-1.amzn2.0.2.x86_64
--> Processing Dependency: perl-Git = 2.47.1-1.amzn2.0.2 for package: git-2.47.1-1.amzn2.0.2.x86_64
--> Processing Dependency: perl(Git) for package: git-2.47.1-1.amzn2.0.2.x86_64
--> Processing Dependency: perl(Term::ReadKey) for package: git-2.47.1-1.amzn2.0.2.x86_64
--> Running transaction check
---> Package git-core.x86_64 0:2.47.1-1.amzn2.0.2 will be installed
---> Package git-core-doc.noarch 0:2.47.1-1.amzn2.0.2 will be installed
---> Package perl-Git.noarch 0:2.47.1-1.amzn2.0.2 will be installed
--> Processing Dependency: perl(Error) for package: perl-Git-2.47.1-1.amzn2.0.2.noarch
---> Package perl-TermReadKey.x86_64 0:2.30-20.amzn2.0.2 will be installed
--> Running transaction check
---> Package perl-Error.noarch 1:0.17020-2.amzn2 will be installed
--> Finished Dependency Resolution

Dependencies Resolved

================================================================================================
================================================================================================
```

i-00f88f5f9436f1ccd (Day_2-instance)

PublicIPs: 3.83.202.50 PrivateIPs: 172.31.93.226

3- S3

Criei o Bucket no S3

datax-raw-storage-ana Info

Objects | Metadata | Properties | Permissions | Metrics | Management | Access Points

Objects (0)

Copy S3 URI | Copy URL | Download | Open | Delete | Actions

Create folder | Upload

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

Find objects by prefix | Show versions | 1

Name	Type	Last modified

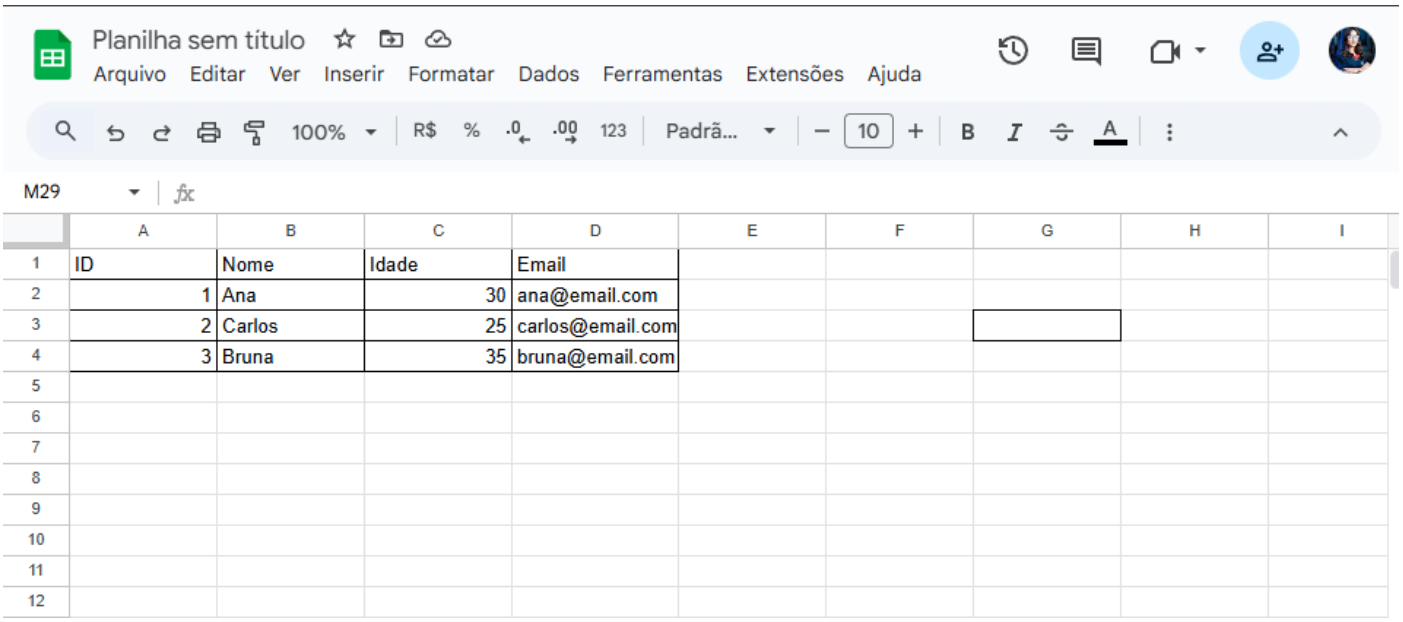
No objects

You don't have any objects in this bucket.

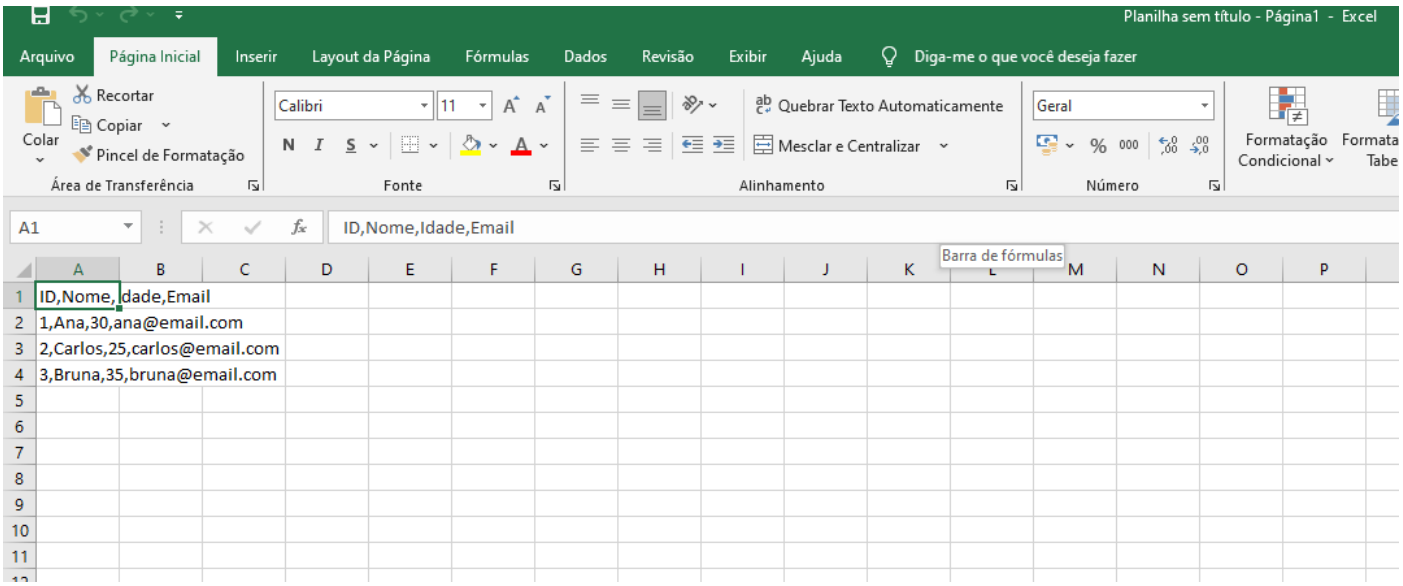
Upload

4- **Criação do arquivo .csv**

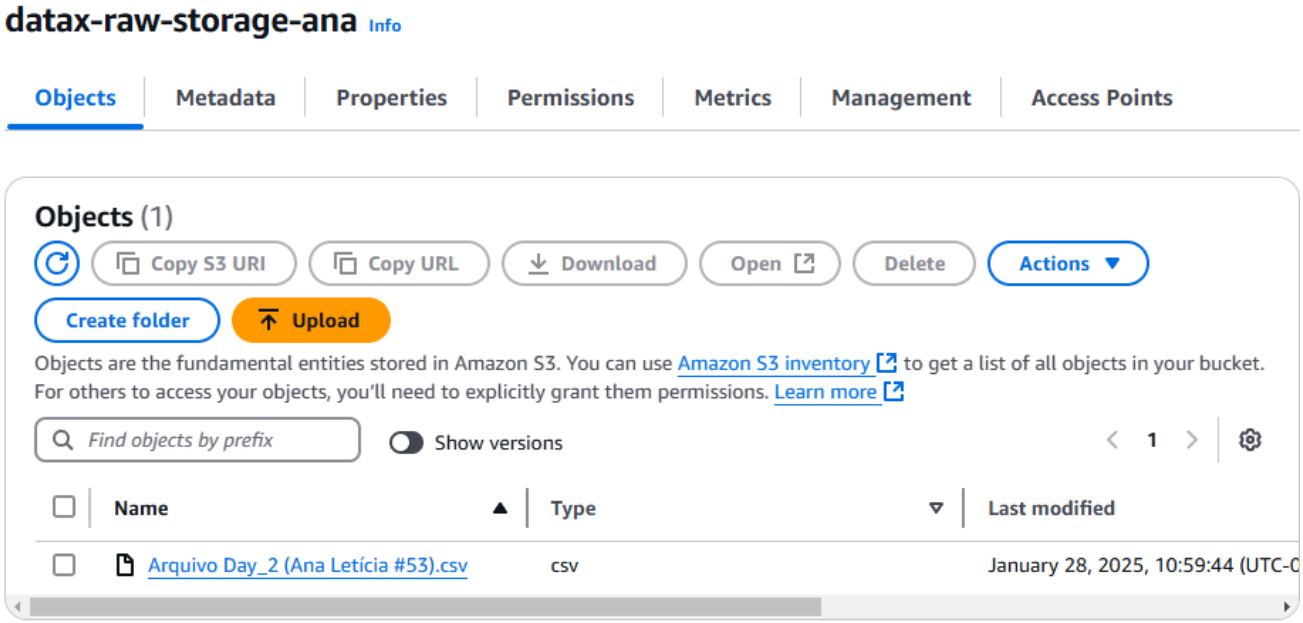
Criei no Sheets no Google uma tabela como foi solicitada no PDF e salvei como arquivo .csv



Planilha depois de salva:



5- Subir o arquivo no Bucket do S3



*OBS: Eu não soube o arquivo do S3 pela instância usando o AWS CLI

FASE 3

1- AWS Glue

Criação do Crawler:

Definir propriedades do rastreador

Detalhes do rastreador Informações

Nome

datax-crawler

O nome pode ter até 255 caracteres. Alguns conjuntos de caracteres, incluindo caracteres de controle, são proibidos.

Escolha fontes de dados e classificadores

Configuração da fonte de dados

Seus dados já estão mapeados para tabelas do Glue?

- Ainda não — Selecione uma ou mais fontes de dados a serem rastreadas.
- Sim — Selecione tabelas existentes no seu Catálogo de Dados do Glue.

Fontes de dados (1) Informações

Editar | Remover | Adicionar uma fonte de dados

A lista de fontes de dados a serem escaneadas pelo rastreador.

Tipo	Fonte de dados	Parâmetros
S3	s3://datax-armazenamento-bruto-ana	Rastrear tudo novamente

Set output and scheduling

Output configuration Info

Target database

datax-db

Clear selection | Add database

Crawler criado:

Crawlers

A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.

Crawlers (1) Info

Last updated (UTC) January 28, 2025 at 14:13:55

Action | Run | Create crawler

View and manage all available crawlers.

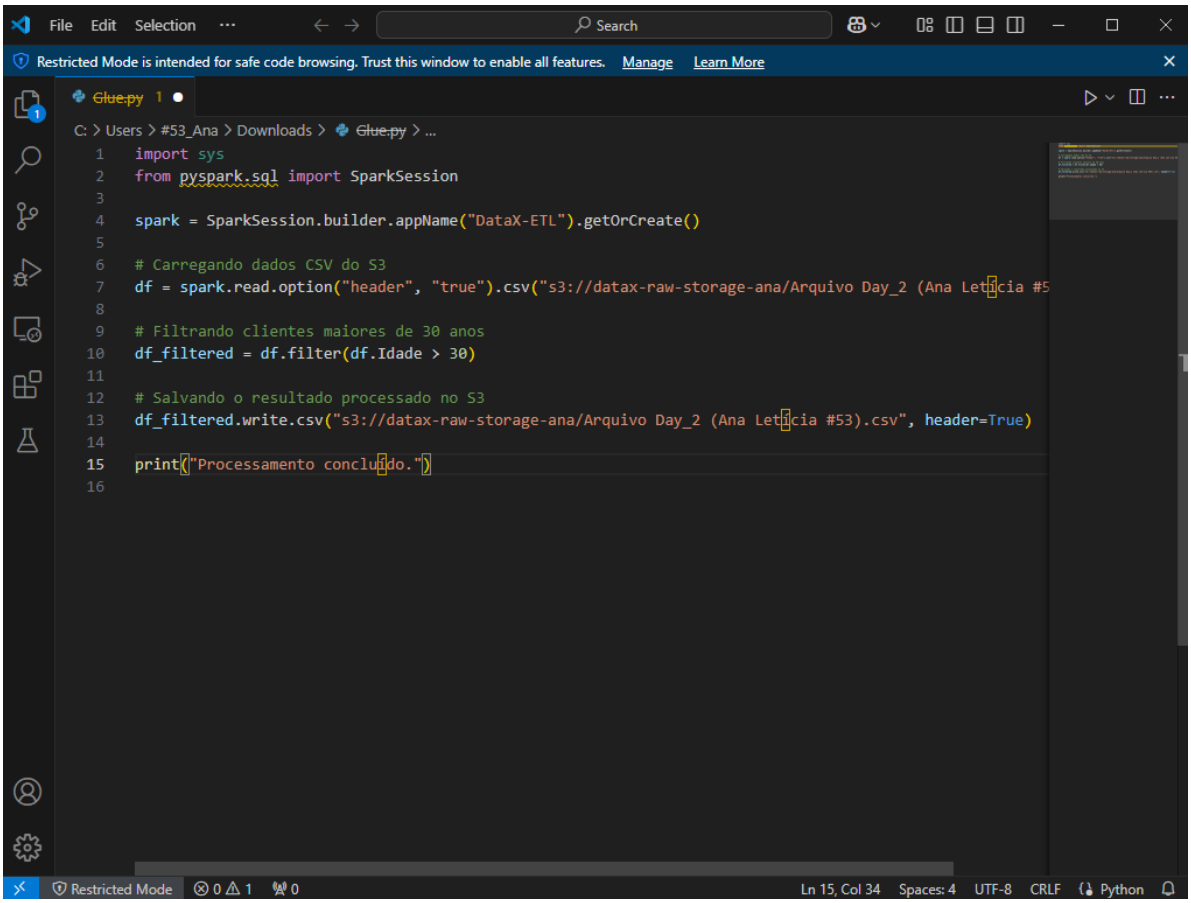
Filter crawlers

1

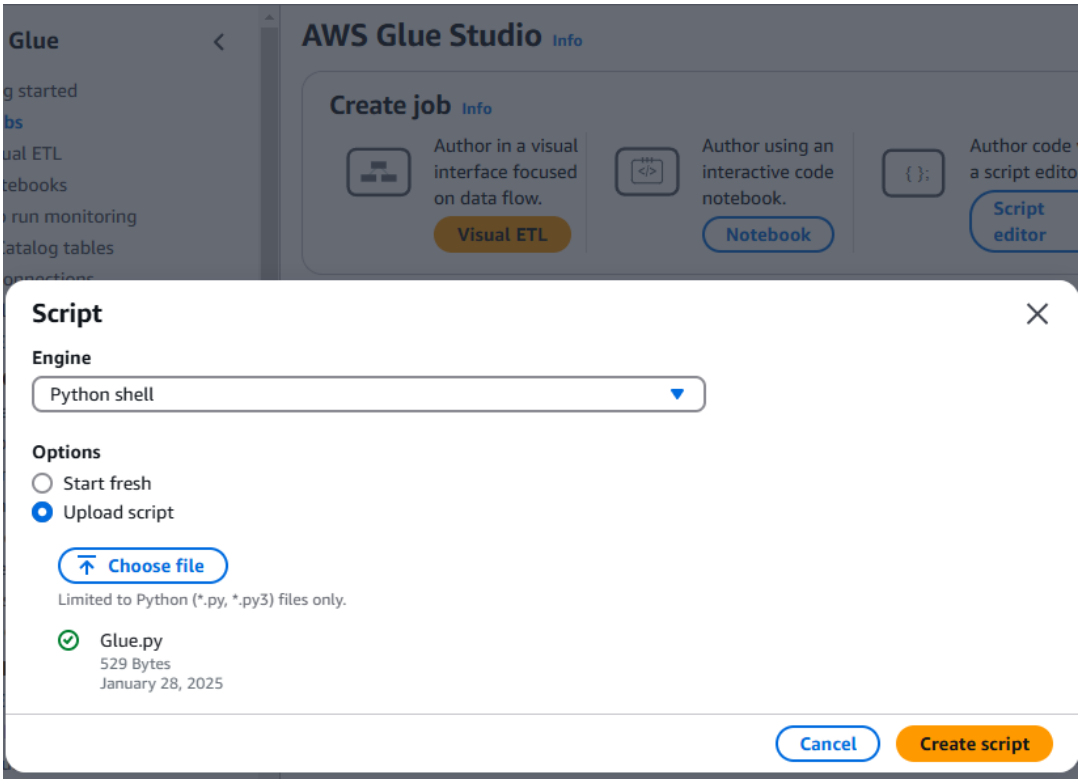
Name	State	Schedule
datax-crawler	Ready	

2- Código Python:

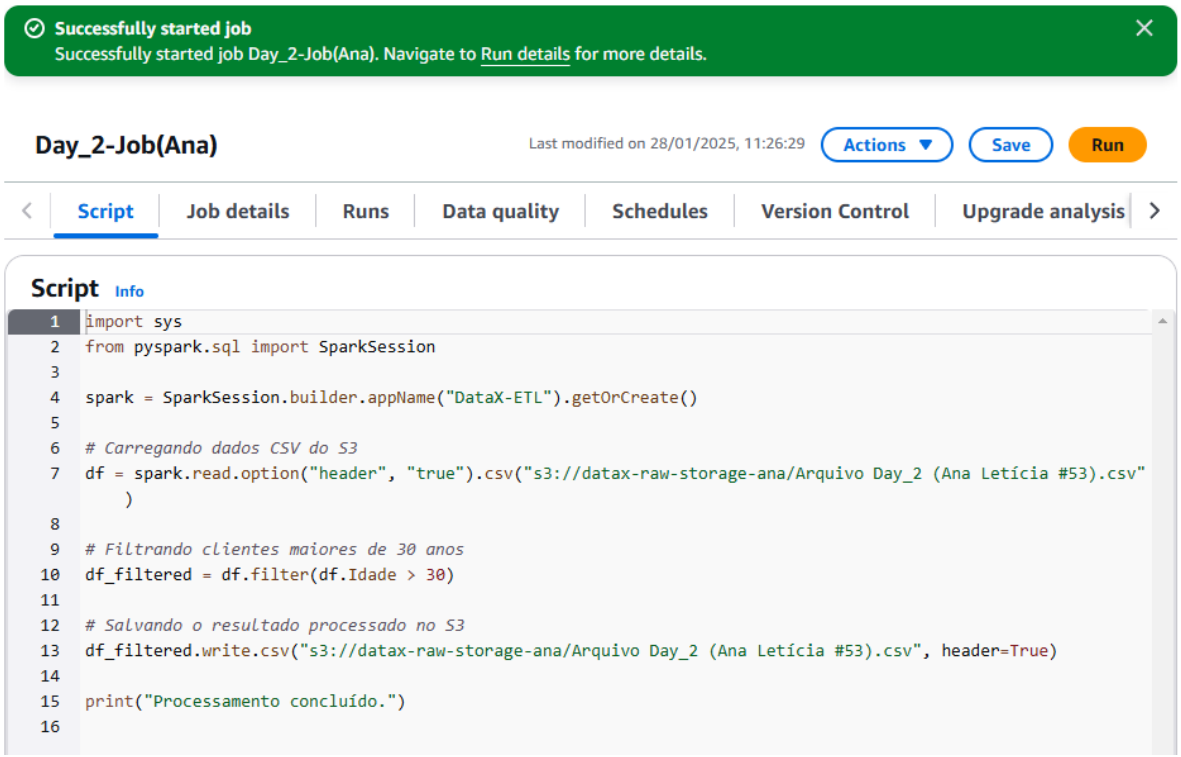
Baixei o arquivo .py do drive e modifiquei o que precisava



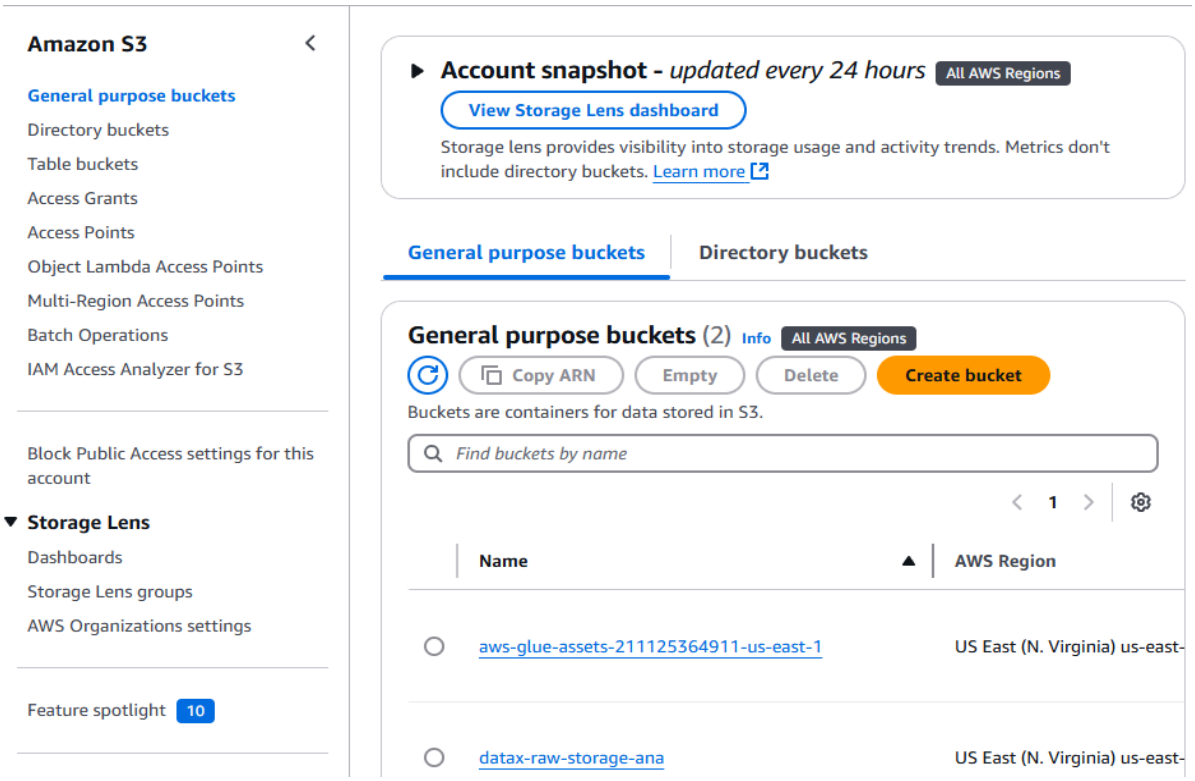
3- Criação do Job



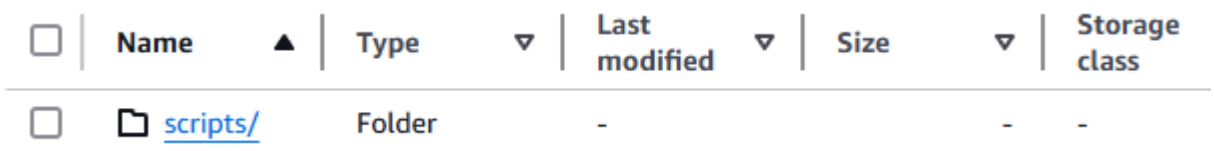
4- Executar Job



5- Depois de executar o job



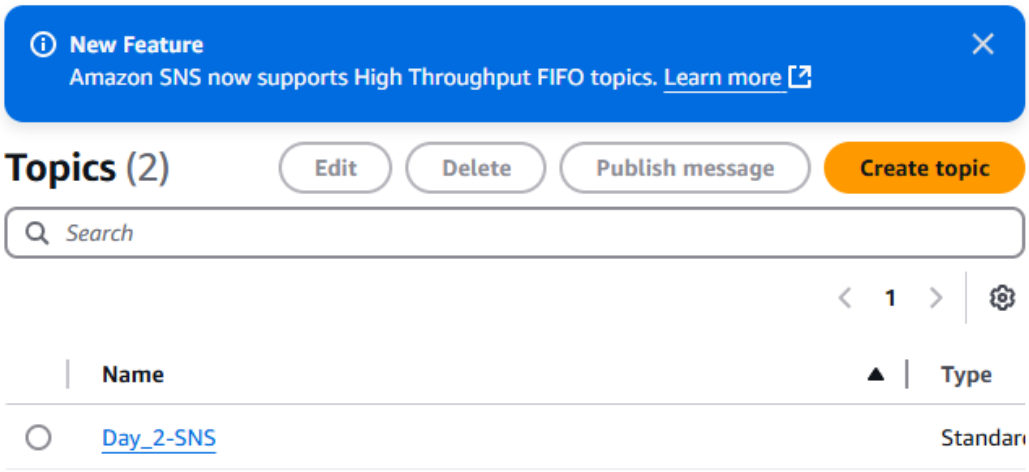
6- Dentro do bucket criado pelo Job do AWS Glue:



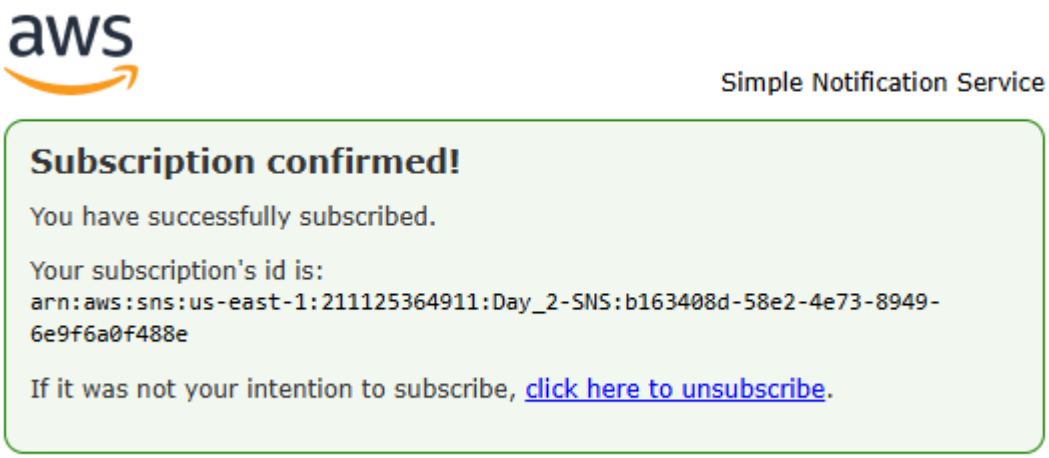
Ana Letícia #53

7- SNS

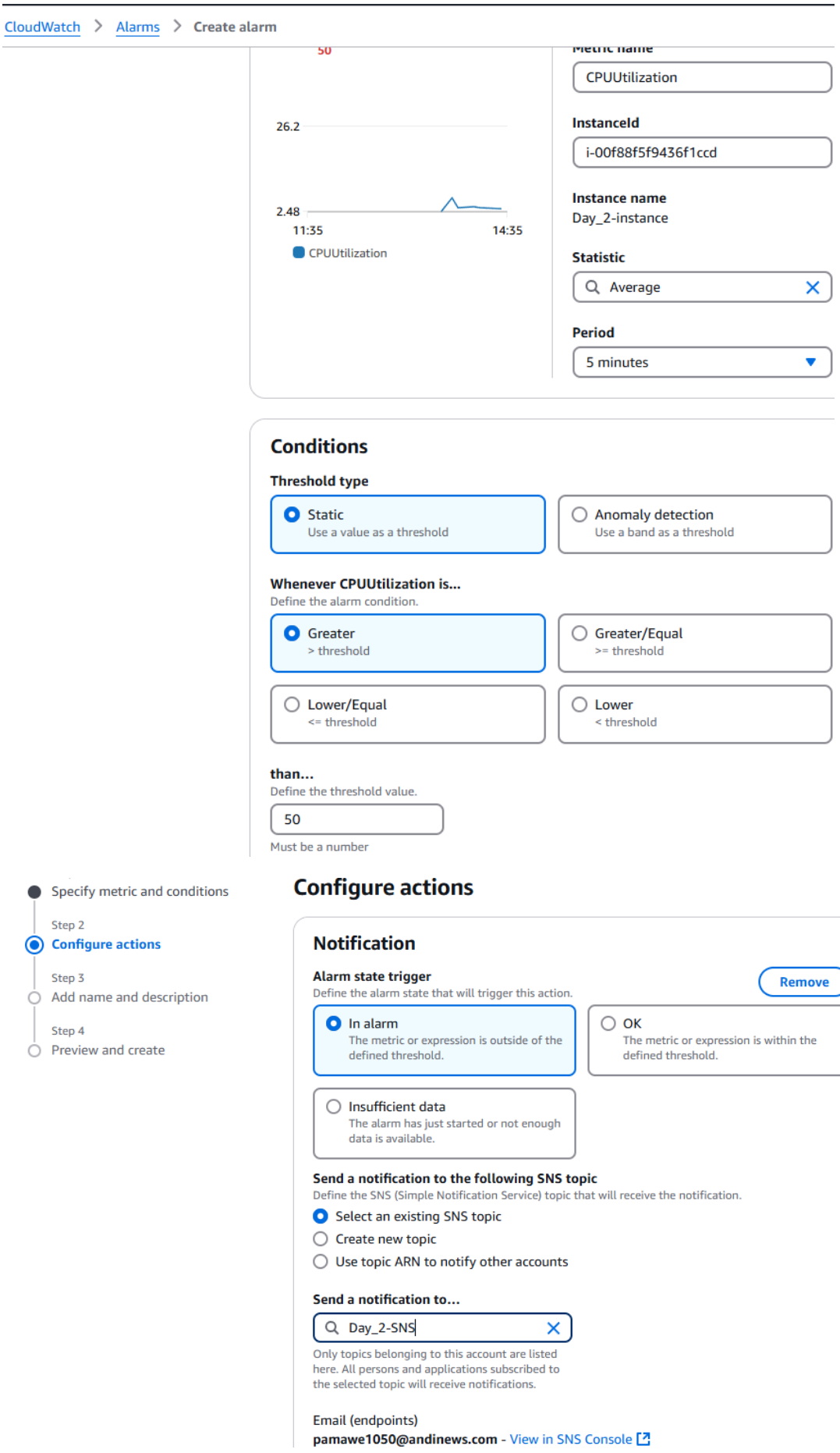
Tive que criar um SNS:



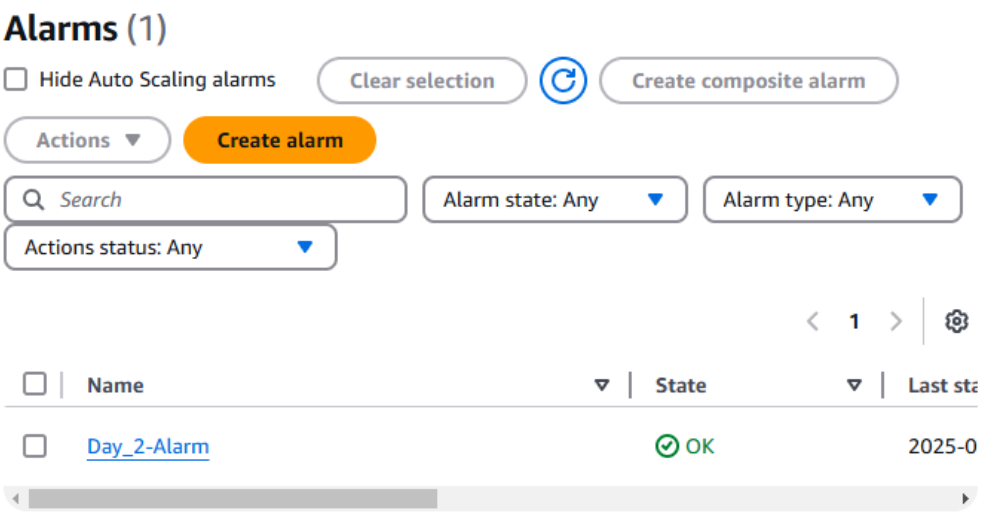
Usei um email criado no site que foi fornecido no Teams: pamawe1050@andinews.com



8- Criação do CloudWatch



9- Alarme criado



Serviços usados:

- S3
- AWS Glue
- EC2
- SNS
- CloudWatch