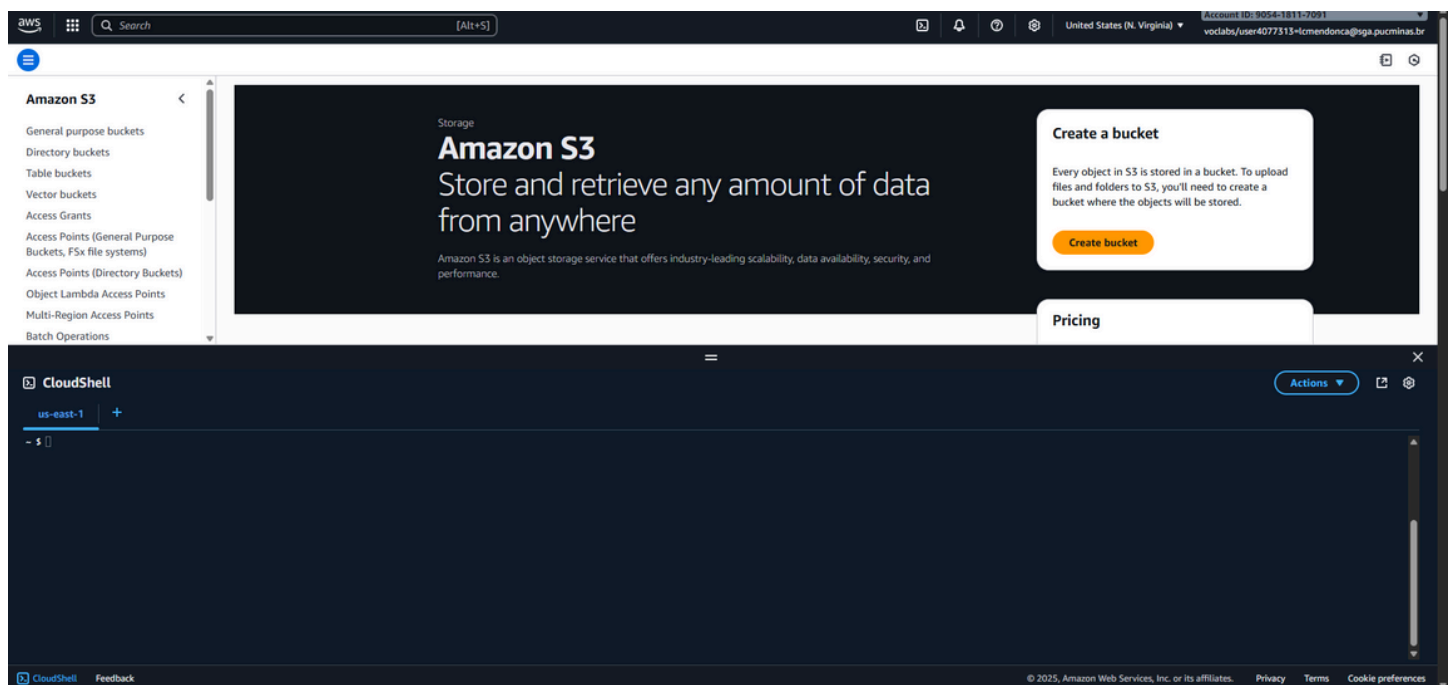
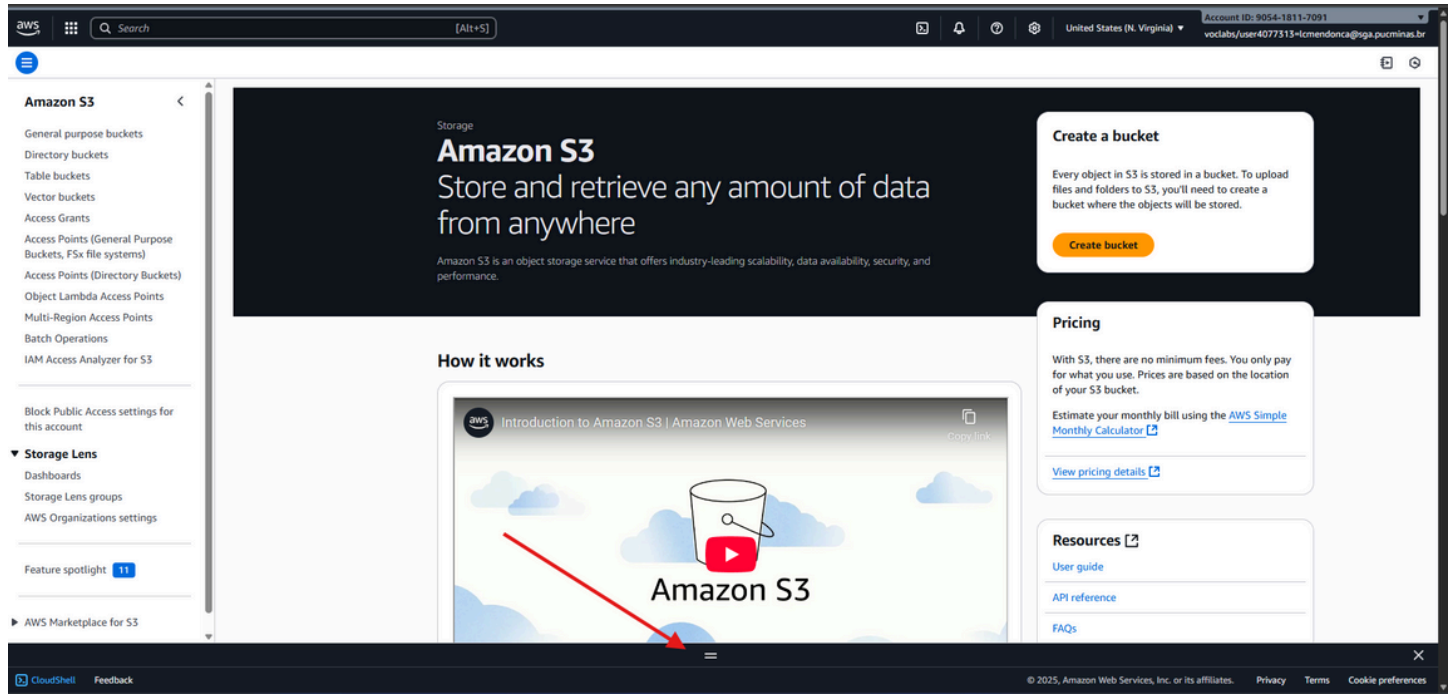


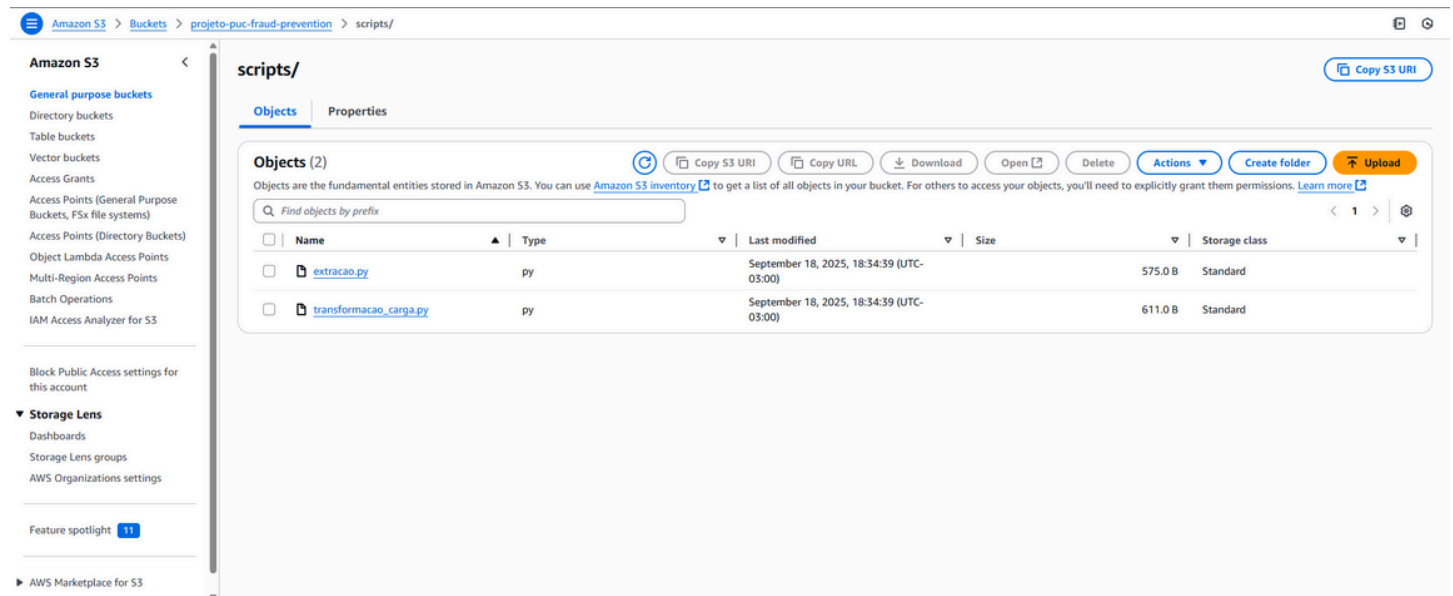
Como executar os códigos

Passo 1: Acessar o CloudShell na parte inferior da tela:

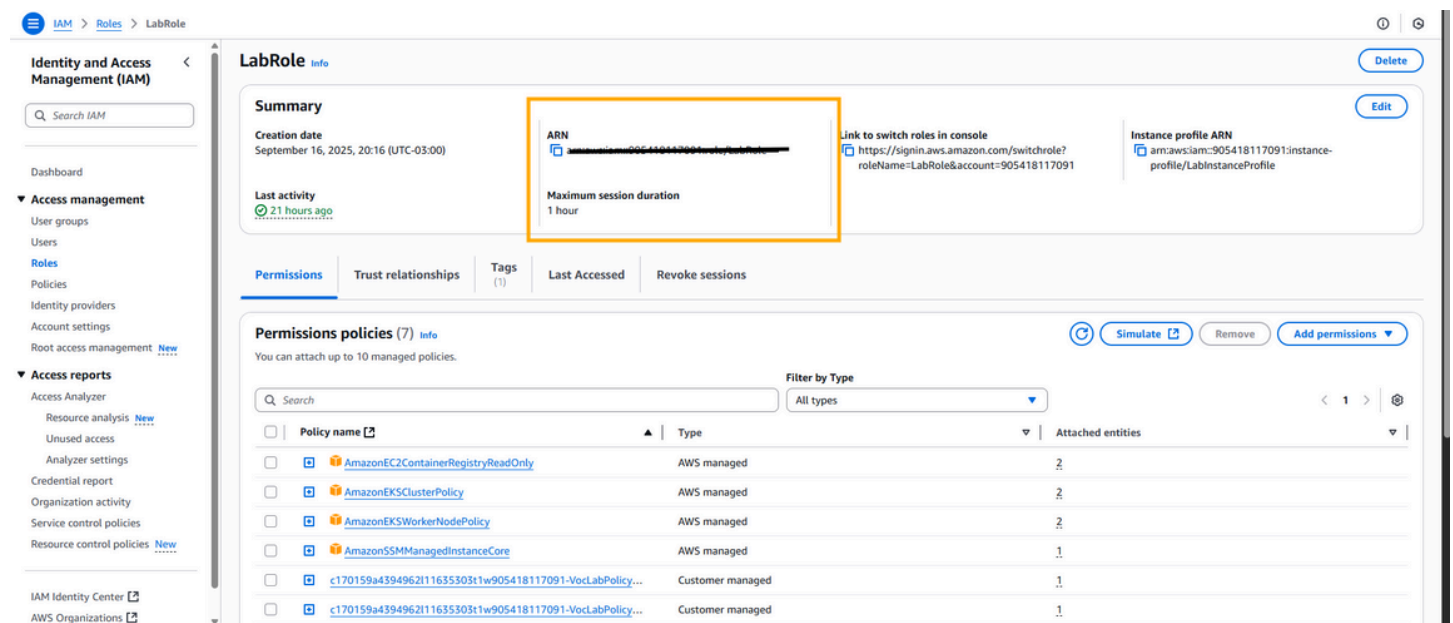


Passo 2: Colocar os códigos que estão no GIT de nome “criar_bucket_pastas” e executar no cliente.

Passo 3: Subir os códigos ETL python “extracao.py” & “transformacao_carga.py” para o bucket no diretório scripts/

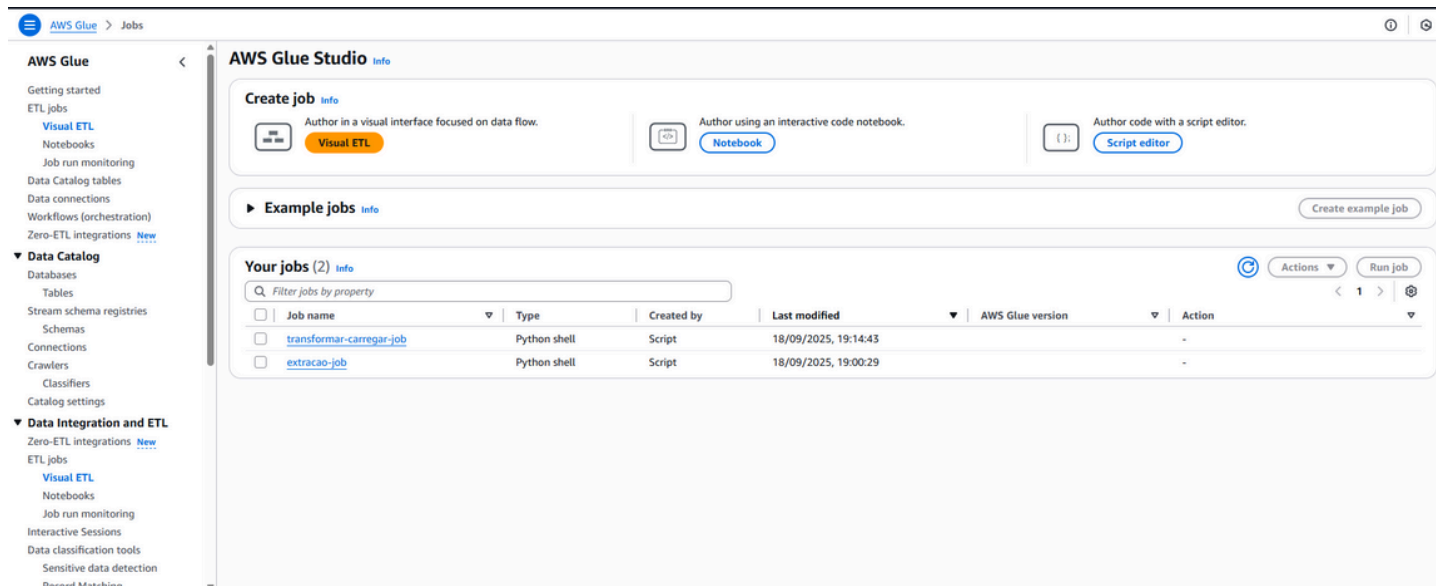


Passo 4: Pegar o seu IAM LabRole indo no serviço IAM na aba Role, e pegar o ARN:



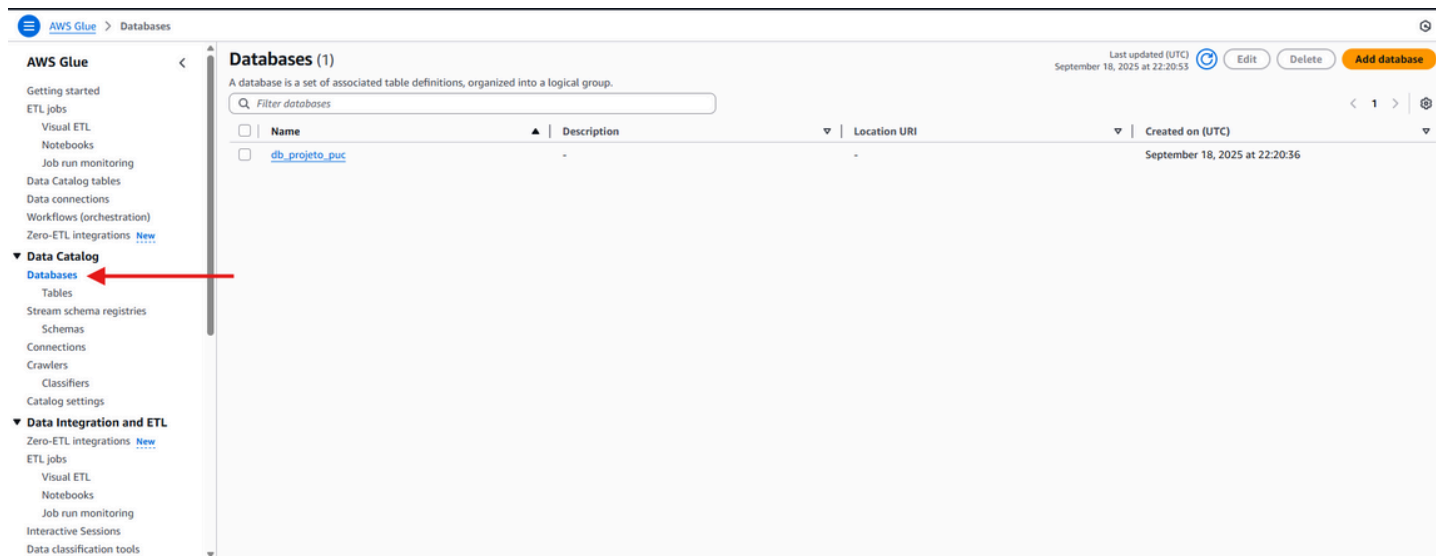
Passo 5: Com o ARN em mãos edite o código no GIT “job_cli_extracao” e o “job_cli_transformar_carregar” colocando o seu ARN no comando --role.

Passo 6: Após executar os scripts no CloudShell, o resultado esperado é encontrado no Visual ETL no serviço AWS Glue:



Passo 7: Criar um DataBase, um database AWS não é um provimento DaaS. É só uma forma confusa da aws para colocar um parquet em formato colunar. O use o comando do arquivo “cli_criar_db”

É possível saber se o database foi criando, dentro do próprio serviço AWS Glue, na aba DataCatalog/Databases:



Passo 8: Criar os crawlers usando os arquivos “cli_criar_raw_crawler” e “cli_criar_processed_crawler”, após isso é só rodar o crawler, mas lembre que o raw precisa ser primeiro.

É possível verificar a “saúde” dos crawlers na aba crawler ainda no AWS Glue:

AWS Glue

Crawlers

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Zero-ETL integrations

Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

Data Integration and ETL

Zero-ETL integrations

ETL jobs

Visual ETL

Crawlers (2)

info

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.

View and manage all available crawlers.

Filter crawlers

Name

State

Schedule

Last run

Last run timestamp

Log

Table changes from last r...

processed-crawler

Ready

Succeeded

September 18, 2025 at 22:...

View log

-

raw-crawler

Ready

Succeeded

September 18, 2025 at 22:...

View log

-

Last updated (UTC)

September 18, 2025 at 22:27:45

Action

Run

Create crawler

<

1

>

Passo 9: Ao final é só ir ao Athena, apontar para o S3, escolher o data base e realizar as consultas SQL que quiser.