

Especialização Puc-Rio

SPRINT-03!

Detalhamento: A base de dados utilizada foi retirada do catálogo de dados abertos da cidade de Fortaleza-Ce. A base é da Autarquia Municipal de Trânsito e Cidadania (AMC).

- **Ano:** 2021
- **Tipo:** CSV
- **Tamanho:** 10418
- **Link:** <https://dados.fortaleza.ce.gov.br/dataset/sinistros-2021>

Nuvem utilizada: Amazon AWS - Ferramentas em nuvem utilizadas:

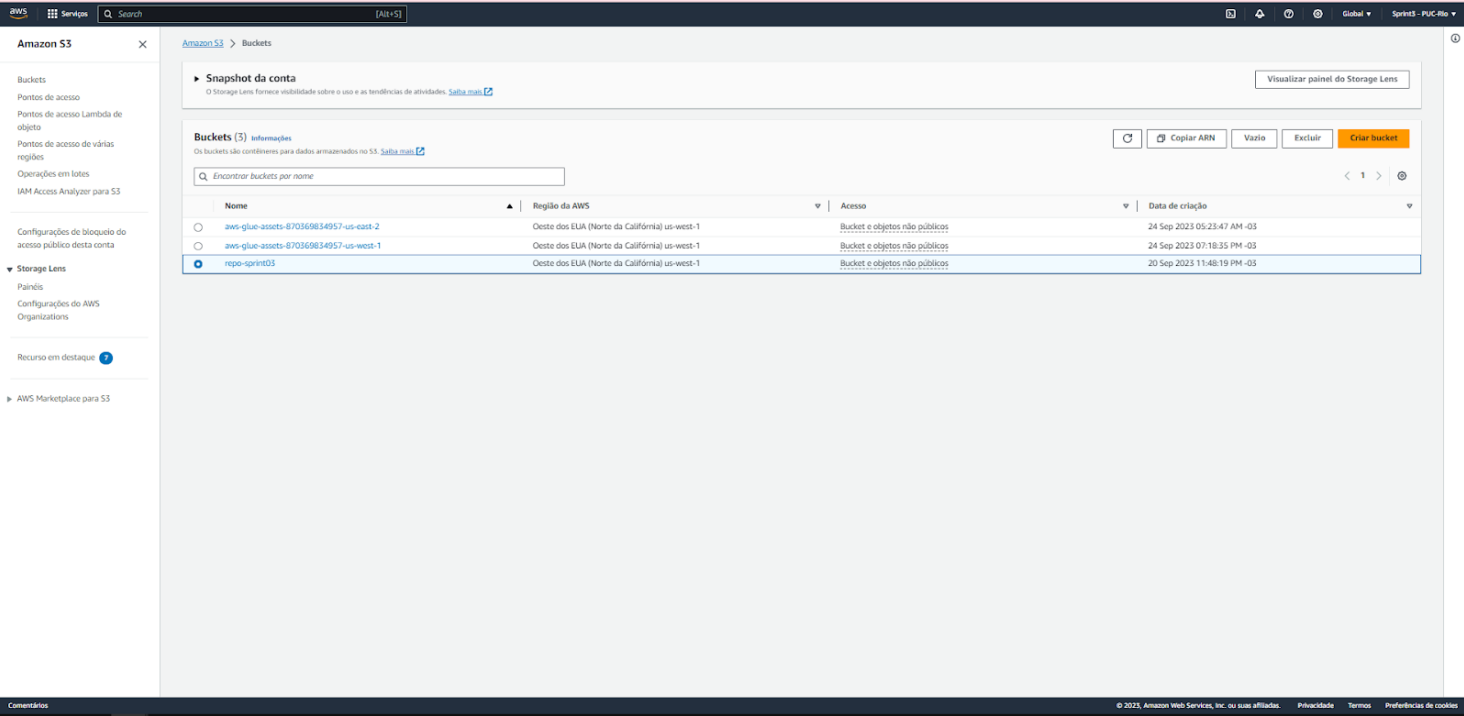
- S3 AWS - Armazenamento
- AWS Glue - Catálogo de dados e processo ETL
- Athena AWS - Análise via SQL

Objetivo: Responder às seguintes perguntas.

- Qual o transporte mais envolvido em acidentes?
- Dos transportes envolvidos em acidentes, quais houve mais mortes?
- Dos transportes envolvidos em acidentes, quais houve mais pessoas ilesas?
- Quais os 10 endereços responsáveis pela maior ocorrência de sinistros de trânsito?
- Qual mês do ano teve mais acidentes? Qual dia do mês? Qual o horário do dia?
- Qual mês do ano teve mais mortes? Qual dia do mês?
- Das ocorrências, em qual endereço houve mais mortes?
- Dos acidentes ocorridos, quais os 10 tipos mais frequentes?
- Dos acidentes ocorridos, quais os 10 tipos mais frequentes por endereço?
- A cada quantos acidentes temos ocorrência de morte?

Etapas Pré-análise

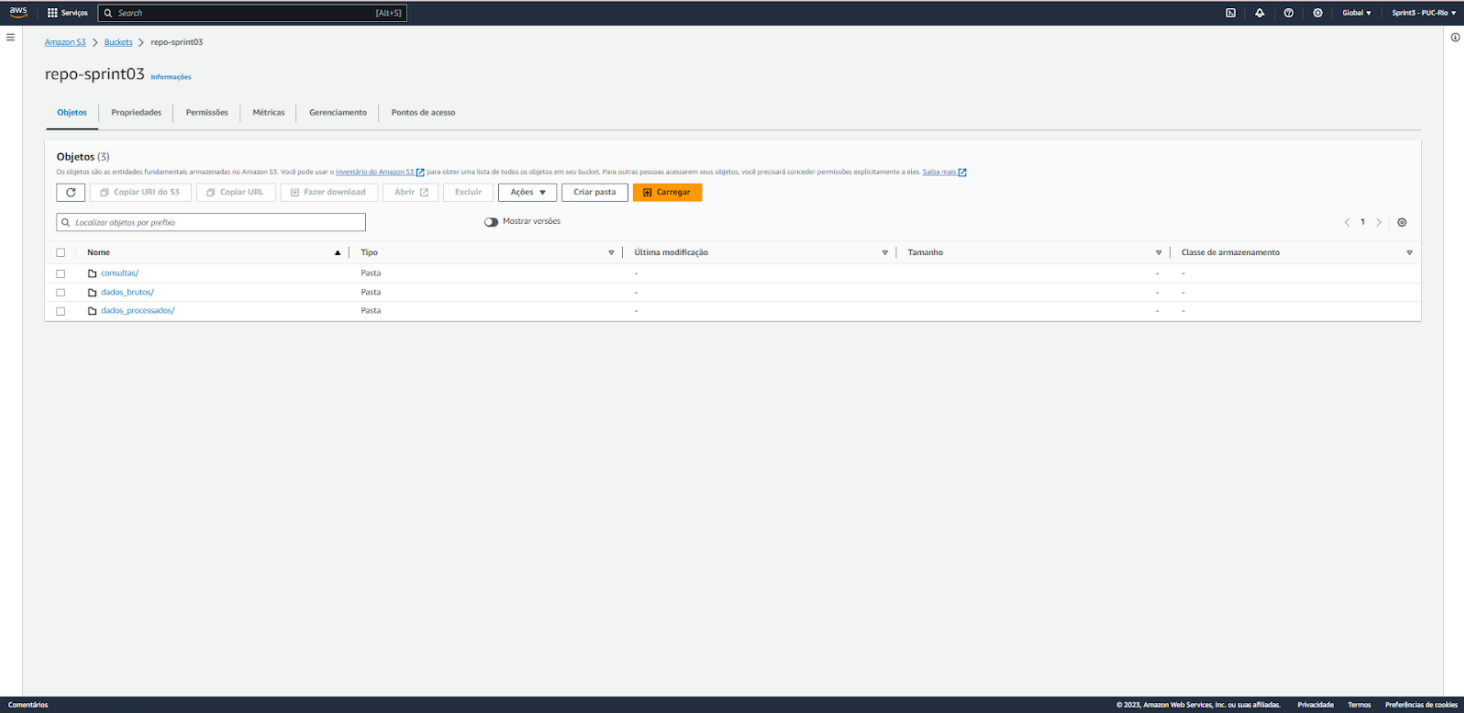
Buckets Utilizados.



O bucket **repo-sprint03** foi criado para armazenar o dataset e os demais foram criados automaticamente para armazenar os metadados do Athena e Glue AWS.

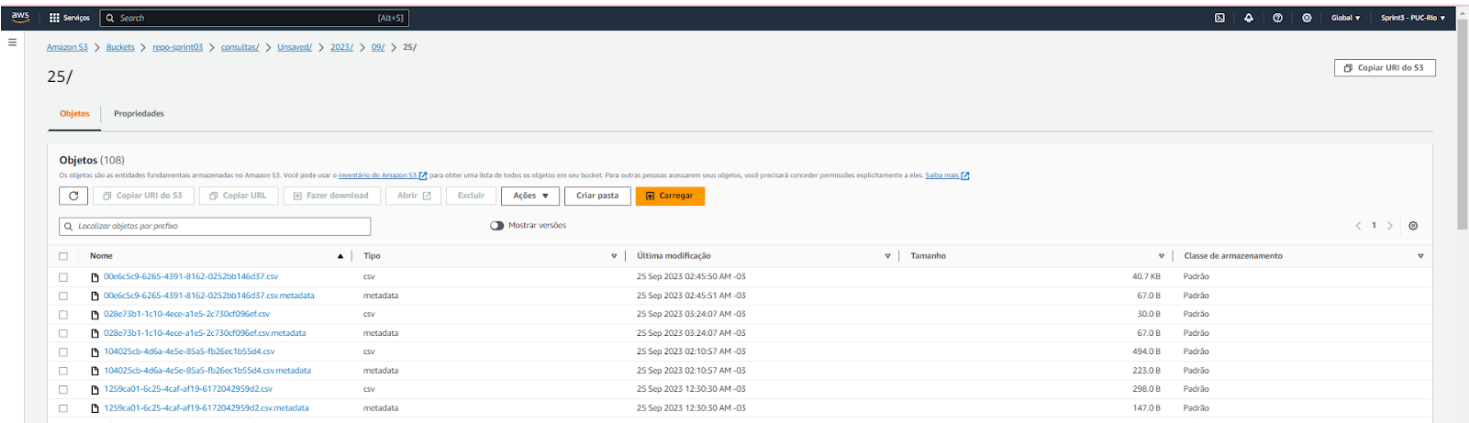
DataSets do bucket **repo-sprint03**

**

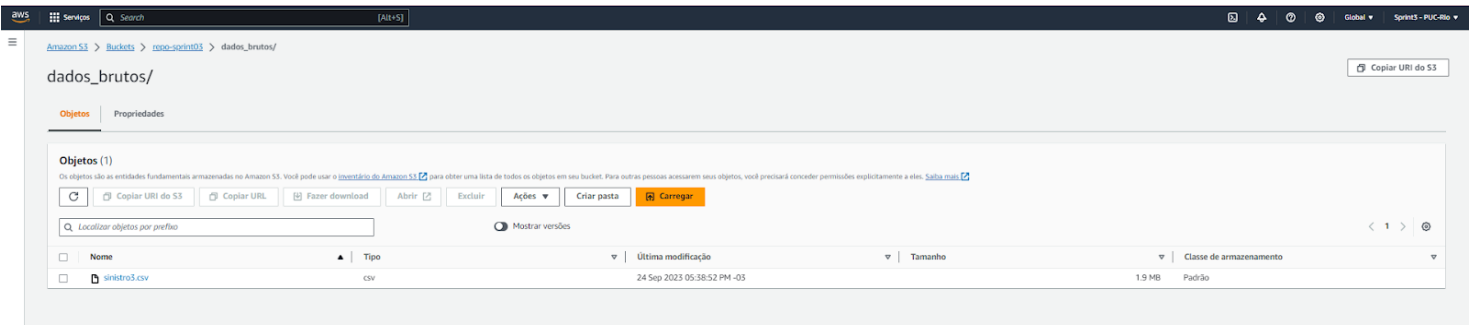


Temos então três pastas, uma para armazenar as consultas feitas no Athena (**a**), uma para armazenar os dados brutos (**b**), e outra para armazenar os dados processados (**c**).

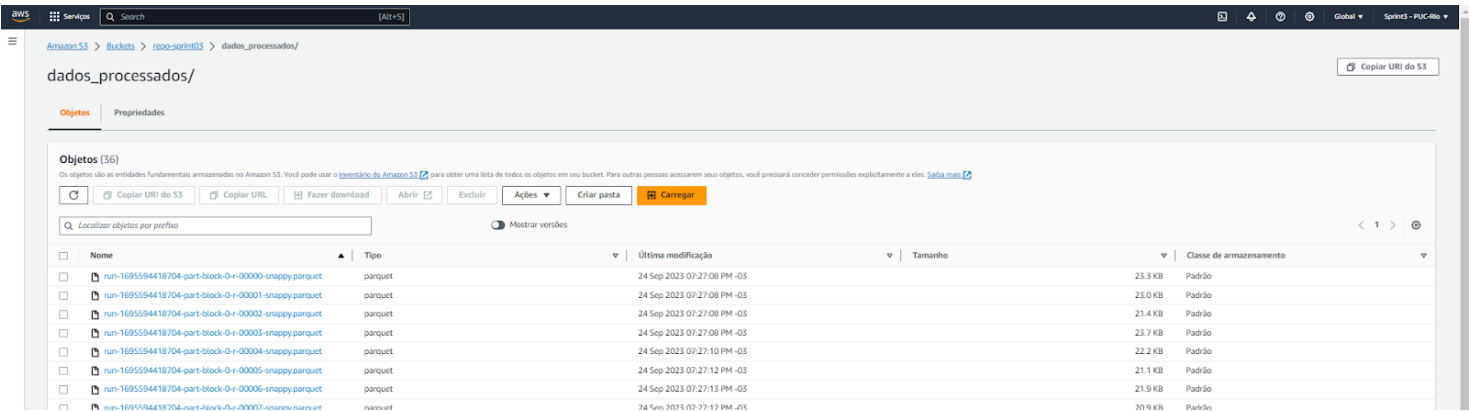
a)



b)



c)

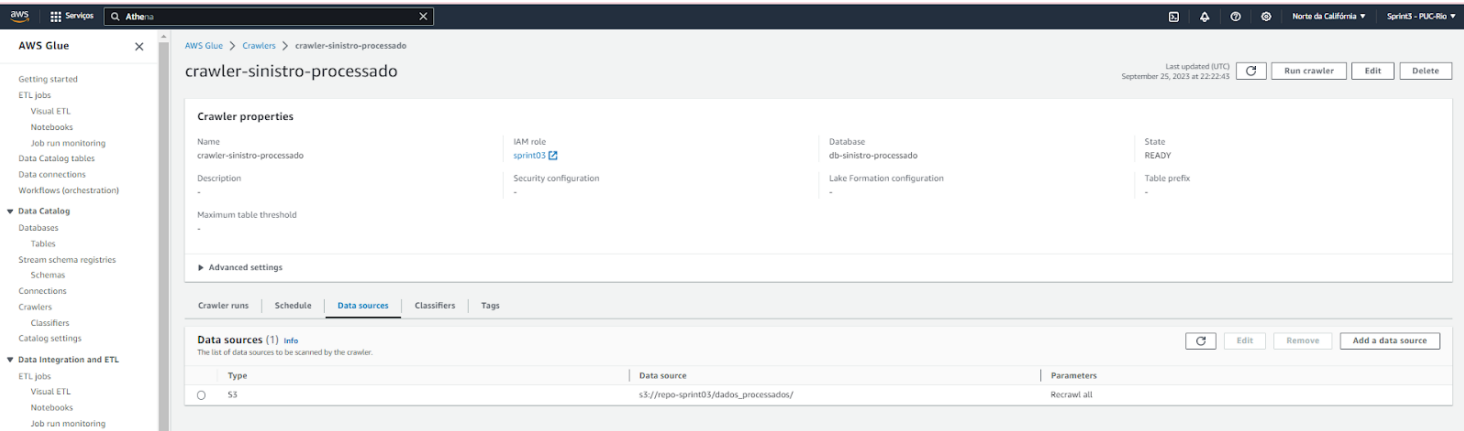
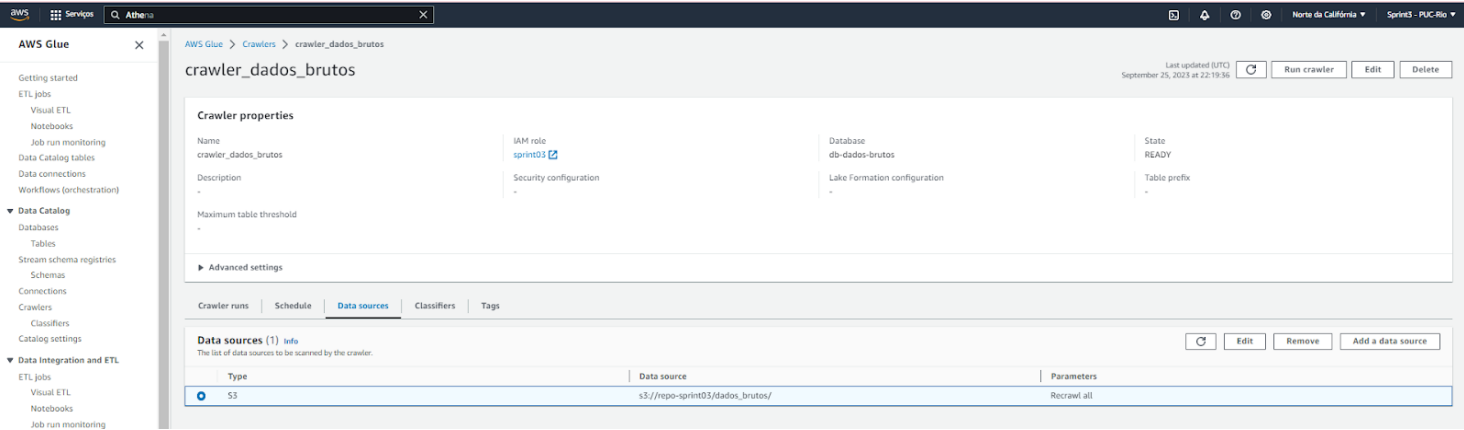
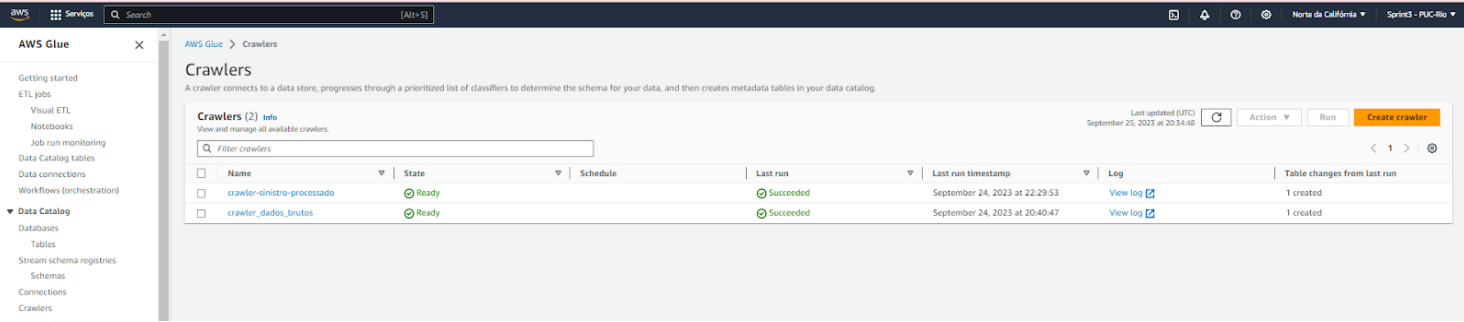


Uso do AWS Glue

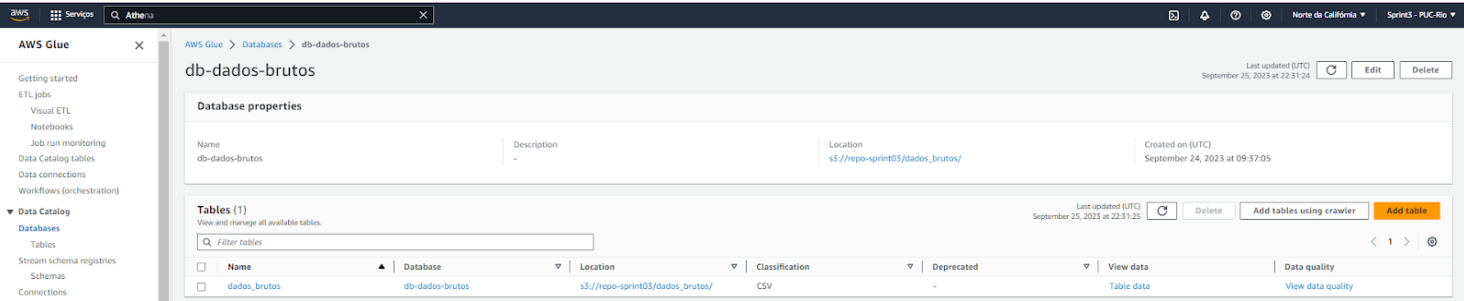
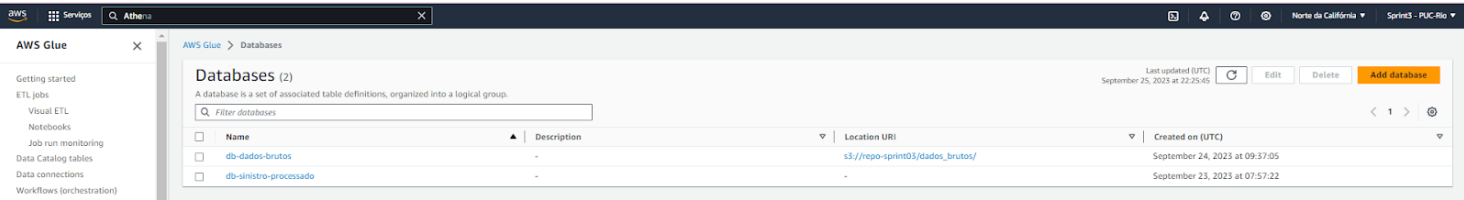
Utilizei o AWS Glue para criar dois crawlers, um deles para varrer os dados brutos e jogar em um uma base de dados disponível para outros serviços como o processo ETL. Trabalhar com os dados em um banco de dados no Glue é melhor do que diretamente no S3, algumas funcionalidades do processo de ETL como exclusão de colunas só ficou disponível quando na carga utilizei uma base de dados do Glue e não diretamente do S3. O processo de ETL resulta em uma nova base de dados no S3, mostrada na figura (c). Para fazer as consultas no Athena criei outro crawler para varrer os dados processados no S3

e carregar em um novo banco no Glue, esse banco será utilizado para as consultas em SQL. O registro dessas etapas segue nas figuras a seguir.

Crawlers



Bases de dados AWS Glue



AWS Glue **db-sinistro-processado**

Database properties

Name: db-sinistro-processado | Description: - | Location: - | Created on (UTC): September 23, 2023 at 07:57:22

Tables (1)

View and manage all available tables.

Filter tables

Name	Database	Location	Classification	Deprecated	View data	Data quality
tb_dados_processados	db-sinistro-processado	s3://repo-sprint03/dados_processados/	Parquet	-	Table data	View data quality

Processo ETL no AWS GLue

AWS Glue **AWS Glue Studio**

Create job info

☒ Visual with a source and target
Start with a source, ApplyMapping transform, and target.

☐ Visual with a blank canvas
Author using an interactive visual interface.

☐ Python Shell script editor
Write or upload your own Python shell script.

☐ Jupyter Notebook
Write your own code in a Jupyter Notebook for interactive development.

Source: Amazon S3 (JSON, CSV, or Parquet files stored in S3)

Target: Amazon S3 (S3 bucket by specifying a bucket path as the data target)

Your jobs (1)

Job name	Type	Last modified	AWS Glue version
processo_etl	Glue ETL	24/09/2023, 19:26:00	4.0

AWS Glue **processo_etl**

Visual | Script | Job details | Runs | Data quality | Schedules | Version Control

Data source properties - Data Catalog | Output schema | Data preview

Name: Carga de dados

Database: db-dados-brutos

Table: dados_brutos

Use runtime parameters

Transform - DropFields: Exclui campos

Transform - DropDuplica...: Exclui linhas duplicadas

Transform - Dynamic Tran...: Remove linhas Null

Transform - Dynamic Tran...: Insere ID

Data target - S3 bucket: Amazon S3 ALVO

AWS Glue

Getting started
ETL jobs

- Visual ETL
- Notebooks
- Job run monitoring
- Data Catalog tables
- Data connections
- Workflows (orchestration)

Data Catalog

- Databases
- Tables
- Stream schema registries
- Schemas
- Connections
- Crawlers
- Classifiers
- Catalog settings

Data Integration and ETL

- ETL jobs
- Visual ETL
- Notebooks
- Job run monitoring
- Interactive Sessions
- Data classification tools
- Sensitive data detection
- Record Matching
- Triggers
- Workflows (orchestration)
- Blueprints
- Security configurations

Legacy pages

- What's New
- Documentation
- AWS Marketplace

processo_etl

Last modified on 25/09/2023, 20:09:37 Try new UI Actions Save Run

Visual Script Job details Runs Data quality New Schedules Version Control

+

```

graph TD
    A[Data source - Data Catalog<br/>Carga de dados] --> B[Transform - DropFields<br/>Exclui campos]
    B --> C[Transform - DropDuplicates<br/>Exclui linhas duplicadas]
    C --> D[Transform - Dynamic Tran...<br/>Remove linhas Null]
    D --> E[Transform - Dynamic Tran...<br/>Insere ID]
    E --> F[Data target - S3 bucket<br/>Amazon S3 ALVO]
  
```

Transform Output schema Data preview

Name
Exclui campos

Node parents
Choose which nodes will provide inputs for this one.
Choose one or more parent node

Carga de dados
Catalog - DataSource

DropFields

Field	Data type
<input checked="" type="checkbox"/> cod_acidente	string
<input type="checkbox"/> dia	long
<input type="checkbox"/> mes	long
<input type="checkbox"/> ano	long
<input type="checkbox"/> hora	string
<input type="checkbox"/> log1	string
<input checked="" type="checkbox"/> log2	string
<input checked="" type="checkbox"/> numero	string
<input checked="" type="checkbox"/> latitude	double
<input checked="" type="checkbox"/> longitude	double
<input checked="" type="checkbox"/> tipocruzamento	string
<input checked="" type="checkbox"/> controletrafego	string
<input checked="" type="checkbox"/> usosoelo	string
<input type="checkbox"/> natureza	string
<input checked="" type="checkbox"/> severidade	string
<input type="checkbox"/> mortos	long
<input type="checkbox"/> feridos	long
<input type="checkbox"/> ilesos	long

AWS Glue

- Getting started
- ETL jobs
- Visual ETL
- Notebooks
- Job run monitoring**
- Data Catalog tables
- Data connections
- Workflows (orchestration)
- Data Catalog**
- Databases
- Tables
- Stream schema registries
- Schemas
- Connections
- Crawlers
- Classifiers
- Catalog settings
- Data Integration and ETL**
- ETL jobs
- Visual ETL
- Notebooks
- Job run monitoring**
- Interactive Sessions
- Data classification tools
- Sensitive data detection
- Record Matching
- Troopers

Monitoring

Date range: 7 Day

Job runs summary

Total runs	Running	Cancelled	Success	Failed	Success rate	DPU hours
21	0	0	3	18	14%	4

Job runs (21)

Filter job runs by property

	Job name	Type	Start time	End time	Run status	Run time	Capacity	Worker type	DPU hours
<input checked="" type="radio"/>	processo_etl	Glue ETL	09/24/2023 19:26:13	09/24/2023 19:27:26	✔ Succeeded	1 minute	10	G.1X	0.19
<input type="radio"/>	PROCESSOS-ETL	Glue ETL	09/24/2023 19:05:54	09/24/2023 19:06:22	✘ Failed	less than a minute	10	G.1X	0.17
<input type="radio"/>	PROCESSOS-ETL	Glue ETL	09/24/2023 18:57:57	09/24/2023 18:58:22	✘ Failed	less than a minute	10	G.1X	0.17
<input type="radio"/>	PROCESSOS-ETL	Glue ETL	09/24/2023 18:56:55	09/24/2023 18:57:27	✘ Failed	less than a minute	10	G.1X	0.17
<input type="radio"/>	PROCESSOS-ETL	Glue ETL	09/24/2023 18:20:58	09/24/2023 18:21:24	✘ Failed	less than a minute	10	G.1X	0.17
<input type="radio"/>	PROCESSOS-ETL	Glue ETL	09/24/2023 18:19:58	09/24/2023 18:20:28	✘ Failed	less than a minute	10	G.1X	0.17
<input type="radio"/>	PROCESSOS-ETL	Glue ETL	09/24/2023 18:18:07	09/24/2023 18:18:32	✘ Failed	less than a minute	10	G.1X	0.17
<input type="radio"/>	PROCESSOS-ETL	Glue ETL	09/24/2023 18:17:04	09/24/2023 18:17:37	✘ Failed	less than a minute	10	G.1X	0.17
<input type="radio"/>	PROCESSOS-ETL	Glue ETL	09/24/2023 18:07:05	09/24/2023 18:07:30	✘ Failed	less than a minute	10	G.1X	0.17
<input type="radio"/>	PROCESSOS-ETL	Glue ETL	09/24/2023 18:06:05	09/24/2023 18:06:35	✘ Failed	less than a minute	10	G.1X	0.17

Análise das informações via AWS Athena

Qual o transporte mais envolvido em acidentes?

Data source

AwsDataCatalog

Database

db-sinistro-processado

Tables and views

Create

Filter tables and views

Tables (1)

< 1 >

tb_dados_processados

Views (0)

< 1 >

Query 1

```
1 SELECT
2 (SELECT sum(automovel)
3 FROM "db-sinistro-processado"."tb_dados_processados") AS Carro,
4
5 (SELECT sum(motocicleta)
6 FROM "db-sinistro-processado"."tb_dados_processados") AS Motocicleta,
7
8 (SELECT sum(bicicleta)
9 FROM "db-sinistro-processado"."tb_dados_processados") AS Bicicleta,
10
11 (SELECT sum(microonibus)
12 FROM "db-sinistro-processado"."tb_dados_processados") AS Microonibus,
13
14 (SELECT sum(onibus)
15 FROM "db-sinistro-processado"."tb_dados_processados") AS Onibus,
16
17 (SELECT sum(caminhao)
18 FROM "db-sinistro-processado"."tb_dados_processados") AS Caminhao,
19
20 (SELECT sum(ciclomotor)
21 FROM "db-sinistro-processado"."tb_dados_processados") AS ciclomotor,
22
23 (SELECT sum(tracaoanimal)
24 FROM "db-sinistro-processado"."tb_dados_processados") AS Tracaoanimal,
25
26 (SELECT sum(trem)
27 FROM "db-sinistro-processado"."tb_dados_processados") AS Trem
28 FROM "db-sinistro-processado"."tb_dados_processados" DESC
29 LIMIT 1
```

SQL Ln 24, Col 75

Run again

Explain

Cancel

Clear

Create

Query results

Query stats

Completed

Time in queue: 117 ms

Run time: 609 ms

Data scanned: 41.38 KB

Results (1)

Search rows

Copy

Download results

#	Carro	Motocicleta	Bicicleta	Microonibus	Onibus	Caminhao	Ciclomotor	Tracaoanimal	Trem
1	6297	7949	683	12	215	257	25	2	4

CloudShell

Comentários

© 2023, Amazon Web Services, Inc. ou suas afiliadas. Privacidade Termos Preferências de cookies

Dos transportes envolvidos em acidentes, quais houve mais mortes?

Amazon Athena

Query editor

Editor

Recent queries

Saved queries

Settings

Workgroup

primary

Data source

AwsDataCatalog

Database

db-sinistro-processado

Tables and views

Create

Filter tables and views

Tables (1)

< 1 >

tb_dados_processados

Views (0)

< 1 >

Query 1

Query 2

Query 3

Query 4

Query 5

Query 6

Query 7

Query 8

Query 9

Query 10

Query 2

```
1 SELECT
2 (SELECT sum(mortos)
3 FROM "db-sinistro-processado"."tb_dados_processados" where automovel > 0) AS Carro,
4 (SELECT sum(mortos)
5 FROM "db-sinistro-processado"."tb_dados_processados" where motocicleta > 0) AS Motocicleta,
6 (SELECT sum(mortos)
7 FROM "db-sinistro-processado"."tb_dados_processados" where bicicleta > 0) AS Bicicleta,
8 (SELECT sum(mortos)
9 FROM "db-sinistro-processado"."tb_dados_processados" where microonibus > 0) AS Microonibus,
10 (SELECT sum(mortos)
11 FROM "db-sinistro-processado"."tb_dados_processados" where onibus > 0) AS Onibus,
12 (SELECT sum(mortos)
13 FROM "db-sinistro-processado"."tb_dados_processados" where caminhao > 0) AS Caminhao,
14 (SELECT sum(mortos)
15 FROM "db-sinistro-processado"."tb_dados_processados" where ciclomotor > 0) AS ciclomotor,
16 (SELECT sum(mortos)
17 FROM "db-sinistro-processado"."tb_dados_processados" where tracaoanimal > 0) AS Tracaoanimal,
18 (SELECT sum(mortos)
19 FROM "db-sinistro-processado"."tb_dados_processados" where trem > 0) AS Trem
20 FROM "db-sinistro-processado"."tb_dados_processados"
21 LIMIT 1
```

SQL Ln 21, Col 8

Run

Explain

Cancel

Clear

Create

Query results

Query stats

Completed

Time in queue: 214 ms

Run time: 1.152 sec

Data scanned: 45.65 KB

Results (1)

Search rows

Copy

Download results

#	Carro	Motocicleta	Bicicleta	Microonibus	Onibus	Caminhao	Ciclomotor	Tracaoanimal	Trem
1	74	107	14	1	13	16	1	0	0

CloudShell

Comentários

© 2023, Amazon Web Services, Inc. ou suas afiliadas. Privacidade Termos Preferências de cookies

Antes da consulta, já tinha julgado que esse seria o resultado, pelo menos para a motocicleta e carro. Em relação a motocicleta, se deve ao fato de não ser um transporte seguro por natureza, além de outros fatores que acometem as demais categorias como imprudência e falta de sinalização.

Dos transportes envolvidos em acidentes, quais houve mais pessoas ilesas?

Query 3

```
2 SELECT
3   (SELECT sum(ilesos)
4    FROM "db-sinistro-processado"."tb_dados_processados" where automovel > 0) AS Carro,
5   (SELECT sum(ilesos)
6    FROM "db-sinistro-processado"."tb_dados_processados" where motocicleta > 0) AS Motocicleta,
7   (SELECT sum(ilesos)
8    FROM "db-sinistro-processado"."tb_dados_processados" where bicicleta > 0) AS Bicicleta,
9   (SELECT sum(ilesos)
10   FROM "db-sinistro-processado"."tb_dados_processados" where microonibus > 0) AS Microonibus,
11   (SELECT sum(ilesos)
12   FROM "db-sinistro-processado"."tb_dados_processados" where onibus > 0) AS Onibus,
13   (SELECT sum(ilesos)
14   FROM "db-sinistro-processado"."tb_dados_processados" where caminhao > 0) AS Caminhao,
15   (SELECT sum(ilesos)
16   FROM "db-sinistro-processado"."tb_dados_processados" where ciclomotor > 0) AS Ciclomotor,
17   (SELECT sum(ilesos)
18   FROM "db-sinistro-processado"."tb_dados_processados" where tracaoanimal > 0) AS Tracaoanimal,
19   (SELECT sum(ilesos)
20   FROM "db-sinistro-processado"."tb_dados_processados" where trem > 0) AS Trem
21 FROM "db-sinistro-processado"."tb_dados_processados" DESC
22 LIMIT 1
```

Run again Explain Cancel Clear Create

Query results Query stats

Completed Time in queue: 240 ms Run time: 1.002 sec Data scanned: 47.44 KB

Results (1)

#	Carro	Motocicleta	Bicicleta	Microonibus	Onibus	Caminhao	Ciclomotor	Tracaoanimal	Trem
1	1268	645	35	7	109	158	8	0	0

Aqui vale uma explicação: transporte de tração animal e trem não teve pessoas ilesas porque estão dentro da categoria de feridos. Outra observação: é claro que o carro tem muitos aparatos de segurança como airbag e etc, no entanto, não é só por isso que tem mais ocorrências de pessoas ilesas, outro fator é que é muito utilizado e é um dos que têm mais ocorrências de sinistros, e obviamente vai ter um número maior de ilesos.

Quais os 10 endereços responsáveis pela maior ocorrência de sinistros de trânsito?

Query 4

```
1 SELECT log1 as "Endereço", count(log1) as "qtd de ocorrência de acidente"
2 FROM "db-sinistro-processado"."tb_dados_processados" group by log1 order by "qtd de ocorrência de acidente" desc limit 10
```

Run again Explain Cancel Clear Create

Query results Query stats

Completed Time in queue: 150 ms Run time: 491 ms Data scanned: 137.37 KB

Results (10)

#	Endereço	Qtd de ocorrência de acidente
1	Avenida General Osório de Paiva	205
2	Avenida Presidente Costa e Silva	169
3	Avenida Washington Soares	164
4	BR-116	140
5	Avenida Mister Hull	114
6	Avenida Godofredo Maciel	111
7	Rua Cônego de Castro	111
8	Avenida Bernardo Manuel	108
9	Avenida Senador Fernandes Távora	105
10	Avenida Presidente Castelo Branco	101

Qual mês do ano teve mais acidentes?

Services

Search

[Alt+S]

EditorRecent queriesSaved queriesSettings

Workgroupprimary

Data

Data source

AwsDataCatalog

Database

db-sinistro-processado

Tables and views

Create

Filter tables and views

Tables (1)

<1>

tb_dados_processados

Views (0)

<1>

Query 1Query 2Query 3Query 4Query 5

1 SELECT mes, count(mes) as "Qtd de ocorrencia no mês"

2 FROM "db-sinistro-processado"."tb_dados_processados" group by mes order by "Qtd de ocorrencia no mês" desc

SQLLn 2, Col 115

Run again

Explain

Cancel

Clear

Create

Query results

Query stats

Completed

Time in queue: 104 msRun time: 453 msData scanned: 5.98 KB

Copy

Download results

Results (12)

Search rows

#

mes

Qtd de ocorrencia no mês

1

7

1021

2

6

993

3

8

988

4

12

979

5

10

966

6

9

926

7

1

908

8

11

878

9

5

826

10

2

716

11

4

654

12

3

563

CloudShell

Comentários

© 2023, Amazon Web Services, Inc. ou suas afiliadas. PrivacidadeTermosPreferências de cookies

Qual dia do mês?

Services

Search

[Alt+S]

Amazon Athena > Query editor

EditorRecent queriesSaved queriesSettings

Workgroupprimary

Data

Data source

AwsDataCatalog

Database

db-sinistro-processado

Tables and views

Create

Filter tables and views

Tables (1)

<1>

tb_dados_processados

Views (0)

<1>

Query 1Query 2Query 3Query 4Query 5

1 SELECT dia, count(dia) as "Qtd de ocorrencia no dia do mês"

2 FROM "db-sinistro-processado"."tb_dados_processados" where mes = 7 group by dia order by "Qtd de ocorrencia no dia do mês" desc limit 1

SQLLn 2, Col 143

Run again

Explain

Cancel

Clear

Create

Query results

Query stats

Completed

Time in queue: 136 msRun time: 486 msData scanned: 26.44 KB

Copy

Download results

Results (1)

Search rows

#

dia

Qtd de ocorrencia no dia do mês

1

2

55

CloudShell

Comentários

© 2023, Amazon Web Services, Inc. ou suas afiliadas. PrivacidadeTermosPreferências de cookies

Qual o horário do dia?

Amazon Athena > Query editor

Editor Recent queries Saved queries Settings

Workgroup primary

Data

Data source: AwsDataCatalog Database: db-sinistro-processado

Tables and views: Filter tables and views

Tables (1): tb_dados_processados Views (0)

Query 1 : X Query 2 : X Query 3 : X Query 4 : X Query 5 : X

```
1 SELECT hora, count(hora) as "qtd de ocorrencia por hora do dia"
2 FROM "db-sinistro-processado"."tb_dados_processados" where mes = 7 and dia = 2 group by hora order by "qtd de ocorrencia por hora do dia" desc limit 1
```

SQL Ln 2, Col 158

Run again Explain Cancel Clear Create

Reuse query results up to 60 minutes ago

Query results Query stats

Completed Time in queue: 110 ms Run time: 512 ms Data scanned: 84.88 KB

Results (1)

Search rows

#	hora	Qtd de ocorrencia por hora do dia
1	06:47:00	2

CloudShell Comentários

© 2023, Amazon Web Services, Inc. ou suas afiliadas. Privacidade Termos Preferências de cookies

Qual mês do ano teve mais mortes?

Amazon Athena > Query editor

Editor Recent queries Saved queries Settings

Workgroup primary

Data

Data source: AwsDataCatalog Database: db-sinistro-processado

Tables and views: Filter tables and views

Tables (1): tb_dados_processados Views (0)

Query 1 : X Query 2 : X Query 3 : X Query 4 : X Query 5 : X Query 6 : X

```
1
2 SELECT mes, sum(mortos) as "qtd de mortes no mês"
3 FROM "db-sinistro-processado"."tb_dados_processados" group by mes order by "qtd de mortes no mês" desc;
```

SQL Ln 3, Col 111

Run Explain Cancel Clear Create

Reuse query results up to 60 minutes ago

Query results Query stats

Completed Time in queue: 97 ms Run time: 499 ms Data scanned: 10.25 KB

Results (12)

Search rows

#	mes	Qtd de mortes no mês
1	6	23
2	7	19
3	10	17
4	5	15
5	8	15
6	11	15
7	9	14
8	3	13
9	1	12
10	4	12
11	12	10
12	2	6

CloudShell Comentários

© 2023, Amazon Web Services, Inc. ou suas afiliadas. Privacidade Termos Preferências de cookies

Qual dia do mês?

Amazon Athena > Query editor

Editor Recent queries Saved queries Settings

Workgroup primary

Data

Data source: AwsDataCatalog Database: db-sinistro-processado

Tables and views: Filter tables and views

Tables (1): tb_dados_processados Views (0)

Query 1 Query 2 Query 3 Query 4 Query 5 Query 6

```
1 SELECT dia, sum(mortos) as "qtd de mortes no dia do mês"
2 FROM "db-sinistro-processado"."tb_dados_processados" where mes = 6 group by dia order by "qtd de mortes no dia do mês" desc limit 10;
```

SQL Ln 3, Col 140

Run again Explain Cancel Clear Create

Query results Query stats

Completed Time in queue: 102 ms Run time: 492 ms Data scanned: 30.71 KB

Results (10)

Search rows

#	dia	Qtd de mortes no dia do mês
1	29	3
2	11	2
3	10	2
4	2	2
5	27	2
6	4	2
7	16	1
8	25	1
9	1	1
10	14	1

Dos acidentes ocorridos, quais os 10 tipos mais frequentes?

Amazon Athena > Query editor

Editor Recent queries Saved queries Settings

Workgroup primary

Data

Data source: AwsDataCatalog Database: db-sinistro-processado

Tables and views: Filter tables and views

Tables (1): tb_dados_processados Views (0)

Query 1 Query 2 Query 3 Query 4 Query 5 Query 6 Query 7 Query 8

```
1 SELECT natureza as Tipo,count(natureza) as "qtd de acidentes por tipo"
2 FROM "db-sinistro-processado"."tb_dados_processados" group by natureza order by "qtd de acidentes por tipo" desc limit 10;
```

SQL Ln 3, Col 114

Run again Explain Cancel Clear Create

Query results Query stats

Completed Time in queue: 95 ms Run time: 479 ms Data scanned: 14.06 KB

Results (10)

Search rows

#	Tipo	Qtd de acidentes por tipo
1	Colisão	5588
2	Queda	1937
3	Atropelamento	1169
4	Colisão Transversal	456
5	Colisão Lateral	420
6	Colisão Traseira	315
7	Choque c/ Obstáculo Fixo	222
8	Colisão Frontal	60
9	Capotamento	60
10	Atropelamento de Animal	42

Dos acidentes ocorridos, quais os 10 tipos mais frequentes por endereço?

The screenshot shows the Amazon Athena Query Editor interface. The query being executed is:

```
SELECT natureza as tipo, log1 as "Endereço", count(natureza) as "Qtd de acidentes por tipo e endereço"
FROM "db-sinistro-processado"."tb_dados_processados" group by natureza, log1 order by "Qtd de acidentes por tipo e endereço" desc limit 10;
```

The query results are displayed in a table with 10 rows and 4 columns: #, Tipo, Endereço, and Qtd de acidentes por tipo e endereço.

#	Tipo	Endereço	Qtd de acidentes por tipo e endereço
1	Colisão	Avenida Washington Soares	115
2	Colisão	Avenida General Osório de Paiva	111
3	Colisão	Avenida Presidente Costa e Silva	89
4	Colisão	BR-116	69
5	Colisão	Avenida Bernardo Manuel	65
6	Colisão	Rua Cônego de Castro	64
7	Colisão	Avenida Mister Hull	63
8	Colisão	Avenida Godofredo Maciel	56
9	Colisão	Avenida Senador Fernandes Távora	55
10	Colisão	Avenida Doutor Silas Manguba	50

Qual a relação de mortes por acidentes? Ou seja, a cada quantos acidentes temos pelo menos uma morte?

The screenshot shows the Amazon Athena Query Editor interface. The query being executed is:

```
SELECT (((temp2.total_mortos*1.0) / tab_temp.total) * 100) as "Proporção de mortes por acidente"
FROM (
  SELECT count(*) AS total
  FROM "db-sinistro-processado"."tb_dados_processados" AS tab_temp,
  (
    SELECT count(mortos) AS total_mortos
    FROM "db-sinistro-processado"."tb_dados_processados"
    WHERE mortos > 0
  ) AS temp2
) AS temp1
```

The query results are displayed in a table with 1 row and 2 columns: # and Proporção de mortes por acidente.

#	Proporção de mortes por acidente
1	1.6221925593779997

Podemos entender essa informação da seguinte forma, a cada **10418 mil** sinistros de trânsito ocorrem pelo menos **1,62 mortes** na cidade de Fortaleza. Levando em consideração o tamanho da base de **10418** registros de sinistros referente ao ano de **2021**.

Considerações finais

Esse trabalho fez uma análise de uma base de dados de sinistros de trânsito da cidade de Fortaleza, Ceará. Para isso, utilizou a nuvem Amazon AWS para armazenar, catalogar, fazer todo processo de ETL e análise. As perguntas feitas sobre os dados são pertinentes levando em consideração minha visão de autor, obviamente existem muitas outras perguntas pertinentes que esse trabalho não respondeu. Durante o processo foram encontradas algumas dificuldades em relação a plataforma, especialmente sobre a região escolhida, que em outros serviços da mesma plataforma não eram visíveis, pois outra região era escolhida de forma automática, tendo que refazer alguns passos. Sobre os dados, trata-se de uma base com muitos campos, logo alguns foram removidos no processo ETL. Pude perceber que alguns campos foram preenchidos a mão no sistema que gerou os dados, isso porque continha erros de digitação, principalmente no endereço das ocorrências. Por conta disso, os dados ainda podem apresentar ruídos. Levando em consideração o todo posso dizer que objetivo do trabalho foi atingido.