

# DAY\_2


Carlos Costa #53

## Fase 2

Instance summary for i-004f415aa6f285034 (Day\_2\_Server) [Info](#)

Updated 1 minute ago

Instance ID

 i-004f415aa6f285034

IPv6 address

—


Hostname type

IP name: ip-172-31-82-201.ec2.internal

Answer private resource DNS name

IPv4 (A)

Auto-assigned IP address

 54.211.146.121 [Public IP]

IAM Role

—



IMDSv2

Required


Operator

—


Public IPv4 address

 54.211.146.121 | [open address](#) 

Instance state

 **Running**



Private IP DNS name (IPv4 only)

 ip-172-31-82-201.ec2.internal



Instance type

t2.micro


VPC ID

 vpc-02d093e8e76740d82 

Subnet ID

 subnet-04a46b908ef6e5407 

Instance ARN

 arn:aws:ec2:us-east-1:571267839083:instance/i-004f415aa6f285034

Instancia Ec2 criada com exito

datax-raw-storage-carlos [Info](#)

Objects


Metadata

Properties

Permissions

Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use

 Find objects by prefix


☐

Name

▲

Type

☐

 [clientes.csv](#)

csv

Bucket S3 criado e arquivo csv “clientes.csv” baixado

Fase 3

Crawler runs (1)

The list of crawler runs for this crawler.

Filter data

Filter by a date and time range

Start time (UTC)

End time (UTC)

Current/last duration

Status

January 28, 2025 at 14:02:32

-

58 s

Running

datax-crawler

Last updated (UTC)  
January 28, 2025 at 14:02:25

Crawler properties

Name

datax-crawler

Description

-

Maximum table threshold

-

IAM role

LabRole

Security configuration

-

Database

datax-db

Lake Formation configuration

-

State

READY

Table prefix

-

Crawler criado e rodando

ScriptDay\_2

ScriptJob detailsRunsData qualitySchedulesVersion ControlUpgrade analysis - preview

Script

Info

1import sys

2from pyspark.sql import SparkSession

3

4spark = SparkSession.builder.appName("DataX-ETL").getOrCreate()

5

6# Carregando dados CSV do S3

7df = spark.read.option("header", "true").csv("s3://datax-raw-storage-carlos/clientes.csv")

8

9# Filtrando clientes maiores de 30 anos

10df\_filtered = df.filter(df.Idade > 30)

11

12# Salvando o resultado processado no S3

13df\_filtered.write.csv("s3://datax-processed-storage-carlos", header=True)

14

15print("Processamento concluído.")

Script criado e pronto para o teste

## General purpose buckets (2) [Info](#) [All AWS Regions](#)

Buckets are containers for data stored in S3.

	Name	AWS Region
<input type="radio"/>	<a href="#">aws-glue-assets-571267839083-us-east-1</a>	US East (N. Virginia) us-east-1
<input type="radio"/>	<a href="#">datax-raw-storage-carlos</a>	US East (N. Virginia) us-east-1

Após o teste nota-se a aparição de um novo bucket s3

[Amazon S3](#) > [Buckets](#) > [aws-glue-assets-571267839083-us-east-1](#) > [scripts/](#)

## scripts/

### Objects

### Properties

#### Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use [A](#)

<input type="checkbox"/>	Name	Type
<input type="checkbox"/>	 <a href="#">ScriptDa.py</a>	py

O novo bucket adicionado possui uma pasta com o objeto desejado ja inserido.

### Conditions

Threshold type  
Anomaly detection

Whenever **CPUUtilization** is  
Greater (>)

Anomaly detection threshold  
50

#### ► Additional configuration

### tep 2: Configure actions

#### Actions

Notification  
When In alarm, send a notification to "Day\_2\_CloudWatch\_Alarms\_Topic"

CloudWatch configurado para mandar notificação caso CPU ultrapasse 50%

# Alarms (1)

<div><div>Q</div><div>Search</div></div>		
<input type="checkbox"/>	Name	▼
<input type="checkbox"/>	<a href="#">CpuAlarme</a>	

E-mail usado: laxare6755@andinews.com