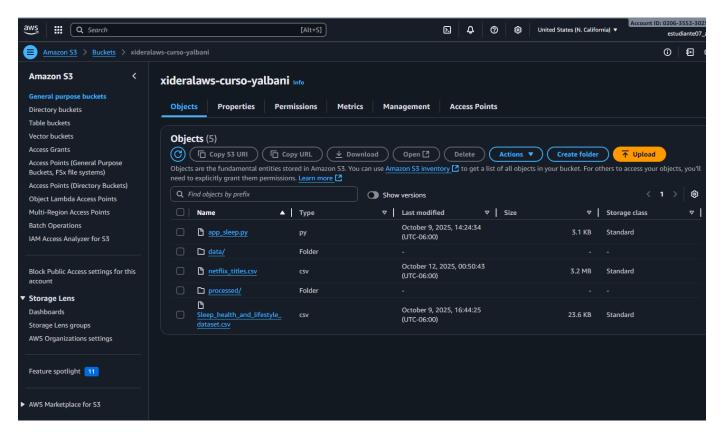
LAMBDA A STRAMLIT

Contenido S3



Funcion LAMBDA. Guarda procesamiento en carpeta processed

```
=
        EXPLORER
                                 lambda_function.py ×
                                                                                                                                                                                                                         □ ...
       V XIDERALAWS-YALBANI
                                   lambda_function.py
Ф
       lambda_function.py
                                          import boto3
Q
                                          import pandas as pd
                                          import io
                                          import datetime
₽
                                          def lambda_handler(event, context):
8
                                              # --- Configuracion de S3 ---
bucket = "xideralaws-curso-yalbani"
key = "netflix_titles.csv"
OUTPUT_PREFIX = "processed/"
\mathbb{A}
                                    11
                                    13
                                    14
15
                                               timestamp = datetime.datetime.now().strftime("%Y %m %d_%H %M %S")
                                              OUTPUT_KEY = f"{OUTPUT_PREFIX}netflix_directores_count_{timestamp}.csv"
s3 = boto3.client("s3")
                                    16
17
                                    18
19
                                                   # LECTURA
                                                    obj = s3.get_object(Bucket=bucket, Key=key)
                                    22
23
24
                                                   body = obj["Body"].read()
df = pd.read_csv(io.BytesIO(body))
      ∨ DEPLOY
                                    25
26
27
                                                   # PROCESAMIENTO EJEMPLO
                                                    df processed = df.dropna(subset=['director'])
                                                   df_final = df_processed.groupby('director').size().reset_index(name='total_titulos')
df_final = df_final.sort_values(by='total_titulos', ascending=False)
                                                   # Convertir el DataFrame procesado a CSV en memoria (StringIO)
                                                   csv_buffer = io.StringIO()
       V TEST EVENTS [SELECTED: T...
                                                                                                                                                                                                         ∨ <u>≡</u> 6 ... ^ ×
                                  PROBLEMS OUTPUT CODE REFERENCE LOG TERMINAL
                                                                                                                                                                                   Execution Results

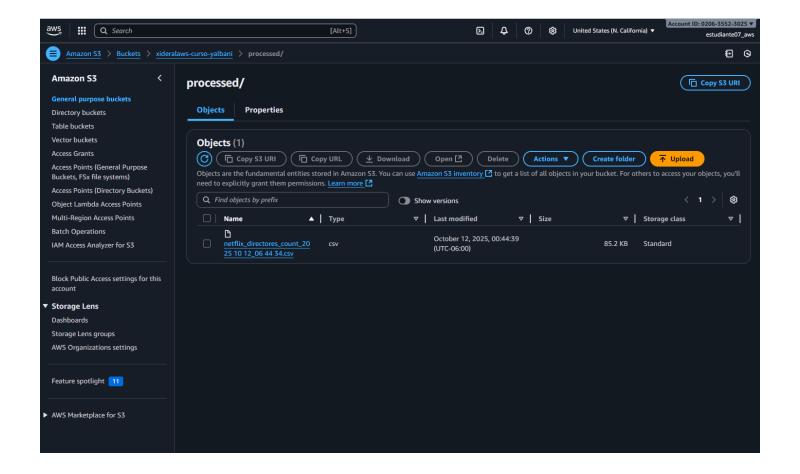
✓ △ Private saved ev...

                                  Status: Succeeded
                                   Test Event Name: test
           test
                                                {\"status\": \"success\", \"message\": \"Datos de Netflix procesados y guardados en S3.\", \"output_file\": \"s3://xideralaws-curso-yalbani/
                                     processed/netflix_directores_count_2025 10 12_06 44 34.csv\"}"
                                                                                                                                                            1 No changes to deploy
```

• Con el csv guardado, se activa el streamlit

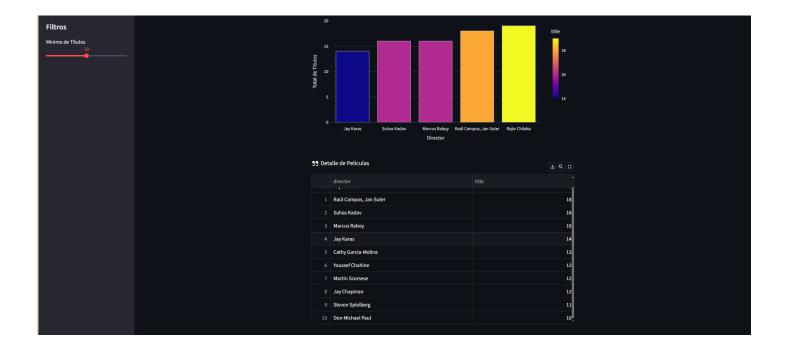
```
Jupyter app_lambda.py Last Checkpoint: 6 minutes ago
   File Edit View Settings Help
   1 import pandas as pd
   2 import streamlit as st
3 import boto3
   4 import io
5 import requests
  6 import json
7 import plotly.express as px
  9 # --- Carga de Datos Procesados desde S3 ---
 10 @st.cache_data
11 def cargar_datos_procesados():
12 s3 = boto3.client("s3")
13 bucket = "xideralaws-curso-yalbani"
         OUTPUT_PREFIX = "processed/"
 14
15
16
17
18
19
          try:
# Listar todos los archivos en la carpeta procesada
- Nota objects v2(Bucket=bucket, Prefix
                response = s3.list_objects_v2(Bucket=bucket, Prefix=OUTPUT_PREFIX)
 20
21
22
23
24
25
26
27
28
30
31
32
33
34
35
36
37
38
39
40
41
42
                     st.warning("Lambda no ha ejecutado el procesamiento. No se encontraron archivos procesados.")
return pd.DataFrame()
                # Encontrar el archivo mas reciente (timestamp mas alto)
              all_files = response["Contents"]
# Filtrar solo archivos CSV si fuera necesario, y ordenar por la ultima modificacion
               all_files = sorted(all_files, key=lambda x: x['LastModified'], reverse=True)
             # El archivo mas reciente es el primero
latest_key = all_files[0]['Key']
              st.info(f"Cargando el ultimo archivo procesado: {latest_key.split('/')[-1]}")
obj = s3.get_object(Bucket=bucket, Key=latest_key)
              body = obj["Body"].read()
              df_final = pd.read_csv(io.BytesIO(body))
return df_final
         except Exception as e:
             st.error(f"Error al cargar el CSV procesado desde S3: {e}")
                return pd.DataFrame()
43
44 # --- Streamlit ---
45
```

Aquí se guarda el csv procesado

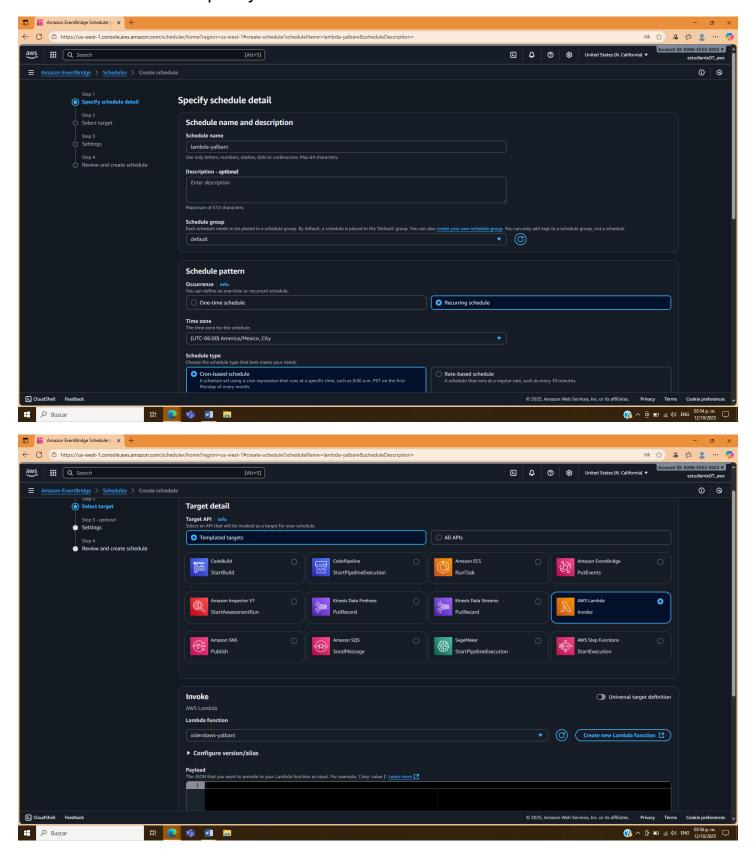


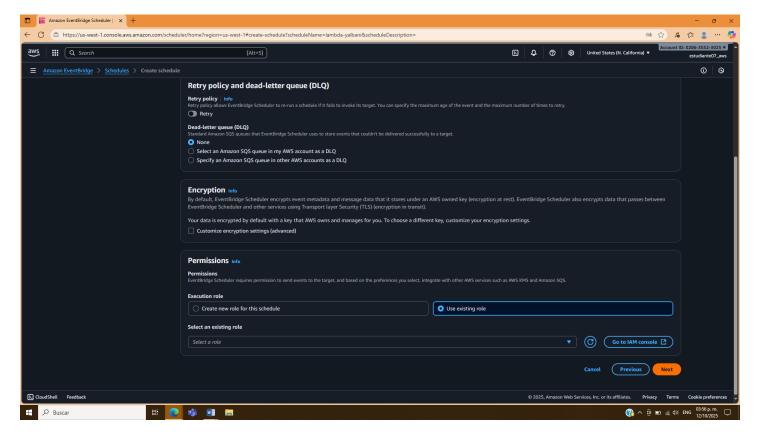
• Resultado Streamlit



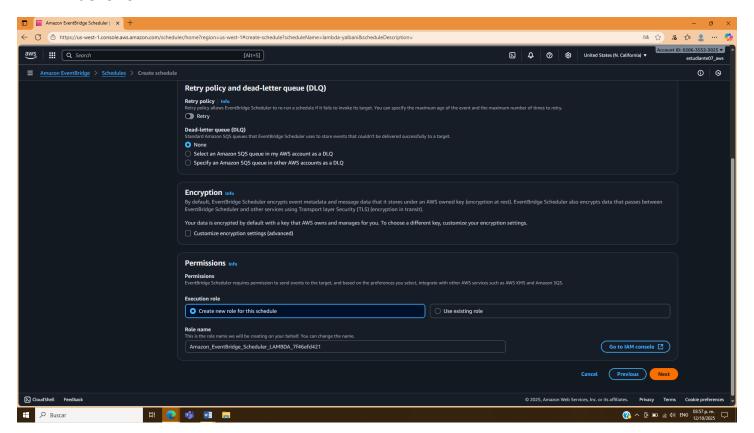


• Crear un CRON que ejecute diario a las 2 am





Nuevo rol



• No hay permisos para crear nuevas reglas

