

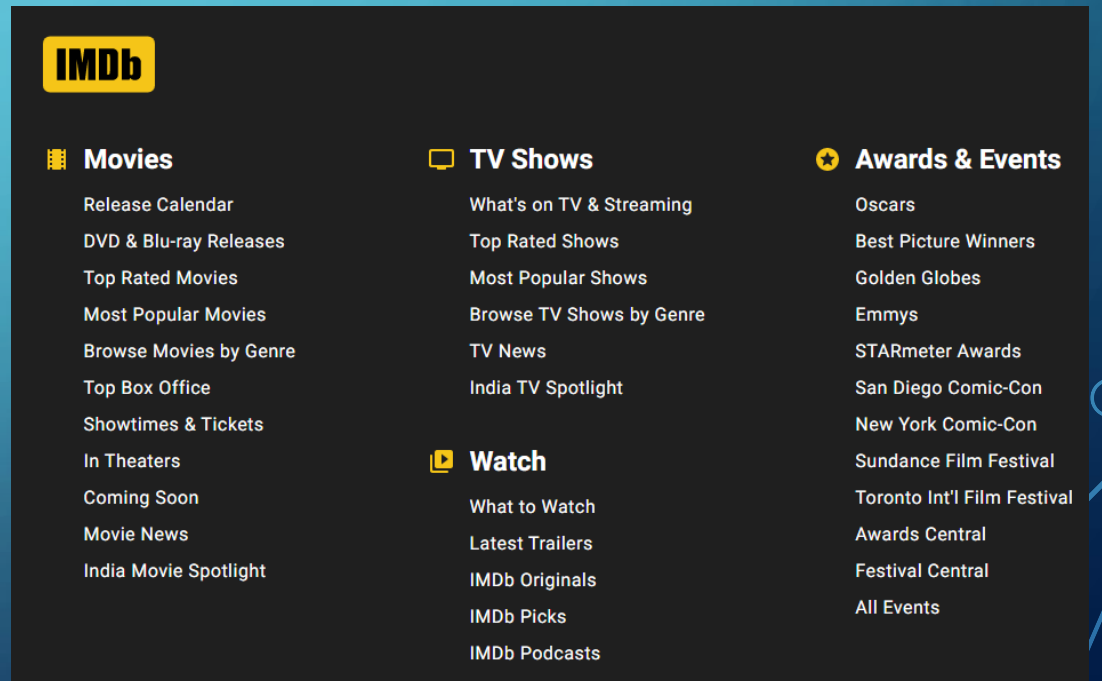
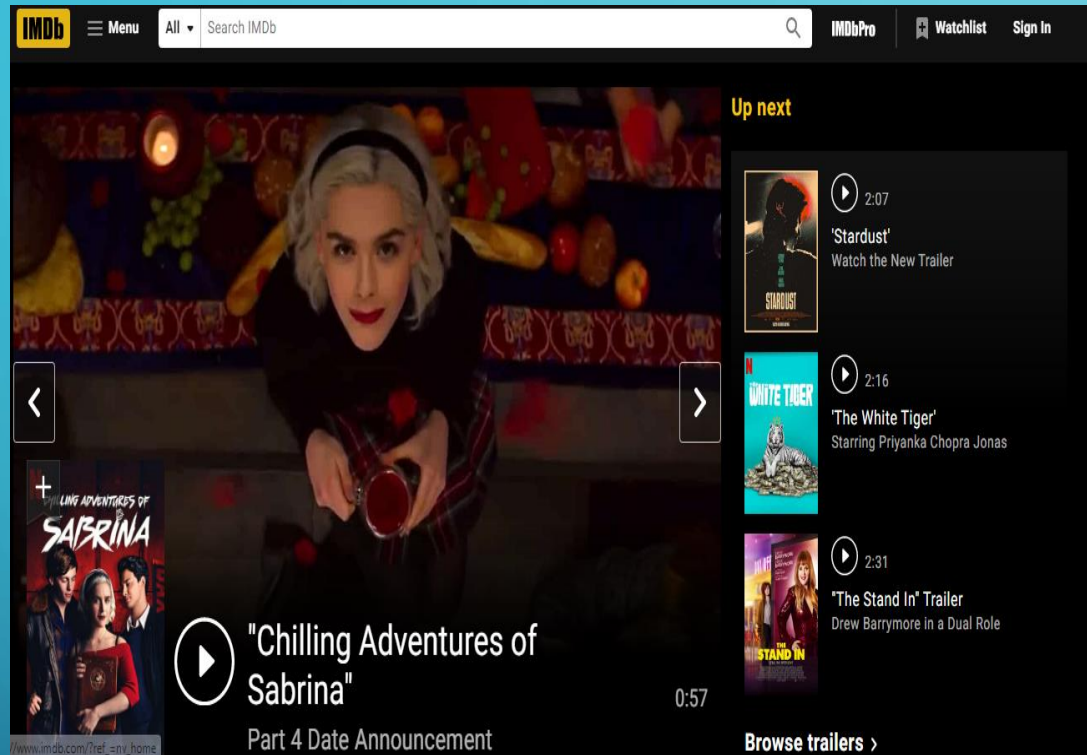
Big Data Analytics Specialist

Proyecto de AWS

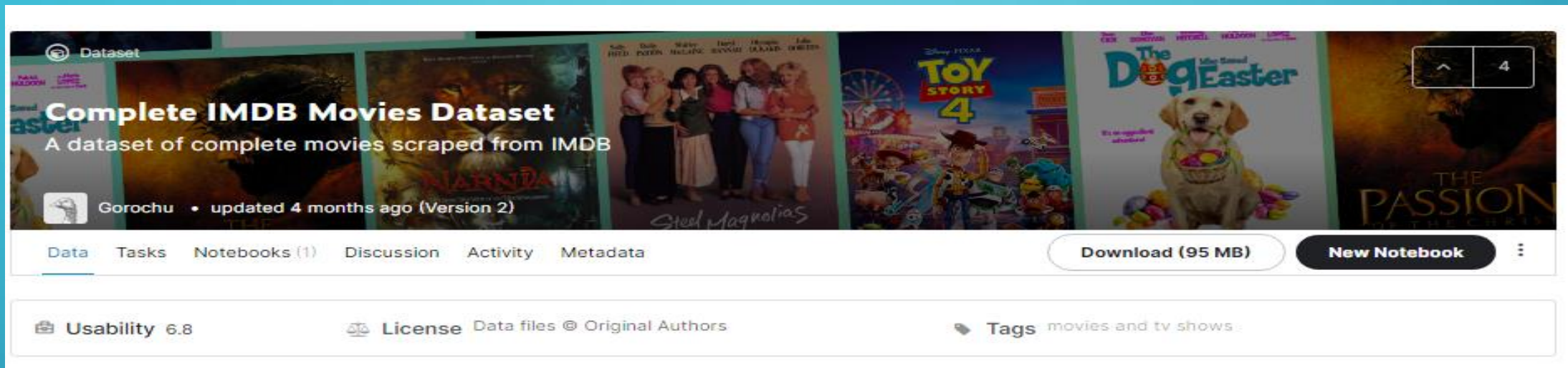


Julia Dávila
2020

CASO : INTERNET MOVIE DATABASE (IMDb)



DATASET : IMDB MOVIES



Complete IMDB Movies Dataset
A dataset of complete movies scraped from IMDB

Gorochu • updated 4 months ago (Version 2)

Data Tasks Notebooks (1) Discussion Activity Metadata

Download (95 MB) New Notebook

Usability 6.8 License Data files © Original Authors Tags movies and tv shows

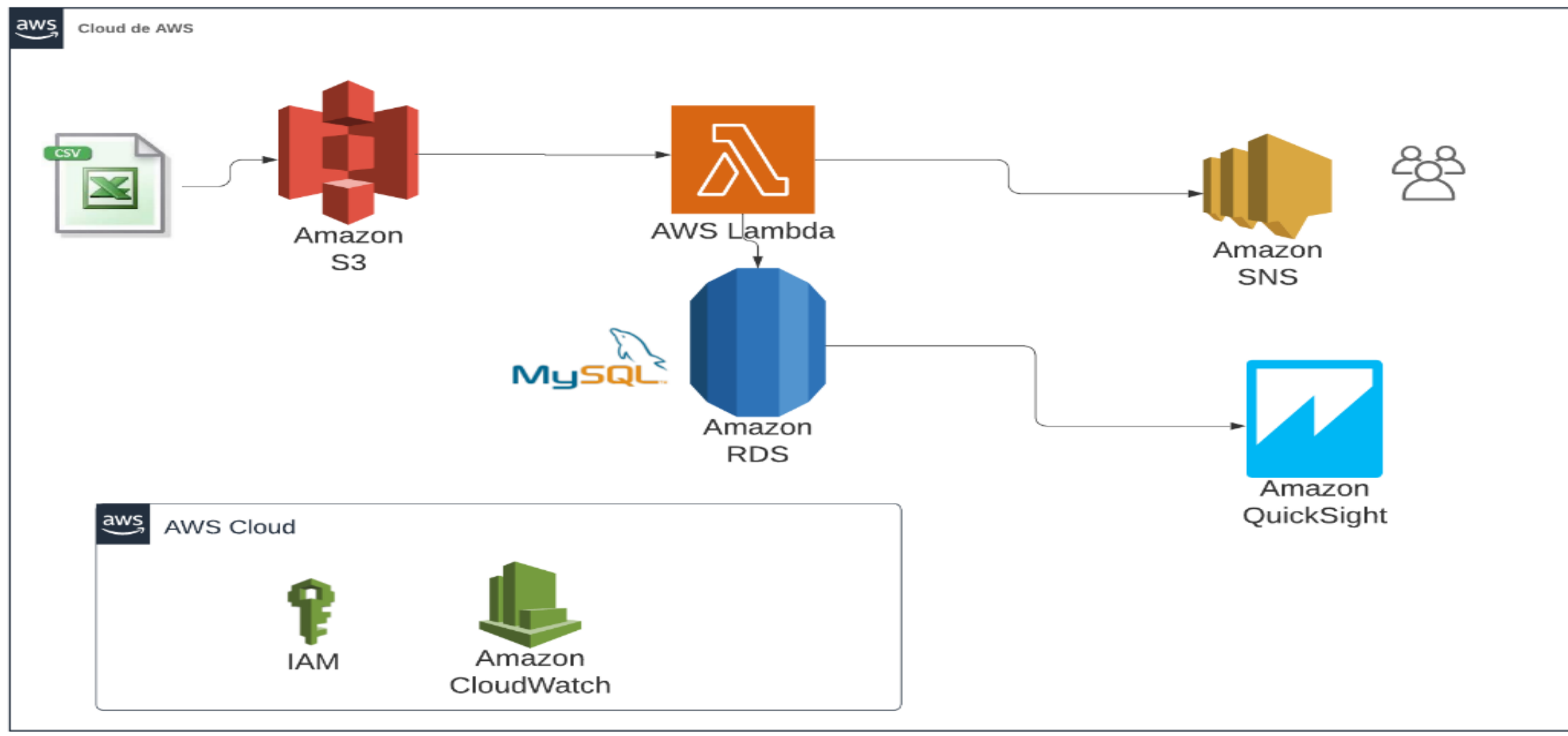
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
title	rating	year	users_rating	votes	metascore	img_url	countries	languages	actors	genre	tagline	description	directors	runtime	imdb_url
The Avengers	PG-13	2012	8	1233746	69	https://m.m	['USA']	['English', 'Ru	['Robert Dow	['Action', 'Ad	Avengers As	Earth's might	['Joss Whedc	143 min	https://www.i
Good Will Hunti	R	1997	8.3	826970	70	https://m.m	['USA']	['English']	['Matt Damo	['Drama', 'Ro	Wildly charis	Will Hunting	['Gus Van Sai	126 min	https://www.i
2001: A Space O	G	1968	8.3	580491	84	https://m.m	['UK', 'USA']	['English', 'Ru	['Keir Dullea	['Adventure', '50	Years Ago	After discove	['Stanley Kut	149 min	https://www.i
Scarface	R	1983	8.3	712283	65	https://m.m	['USA']	['English', 'Sp	['Al Pacino',	['Crime', 'Dra	The world is	In 1980 Mian	['Brian De Pa	170 min	https://www.i
Gran Torino	R	2008	8.1	701410	72	https://m.m	['Germany', 'U	['English', 'Hi	['Clint Eastw	['Drama']		Disgruntled	['Clint Eastw	116 min	https://www.i
Three Billboards	R	2017	8.2	401685	88	https://m.m	['UK', 'USA']	['English']	['Frances Mc	['Comedy', 'Crime', 'Dram	A mother pe	['Martin McD	115 min	https://www.i	
A Beautiful Mini	PG-13	2001	8.2	819184	72	https://m.m	['USA']	['English']	['Russell Cro	['Biography', 'I	need to bel	After	['Ron Howar	135 min	https://www.i
Fargo	R	1996	8.1	594745	85	https://m.m	['USA', 'UK']	['English']	['William H. I	['Crime', 'Dra	An ordinary	Jerry Lundeg	['Joel Coen',	98 min	https://www.i
Vertigo	PG	1958	8.3	346708	100	https://m.m	['USA']	['English']	['James Stew	['Mystery', 'R	A tall story a	A former pol	['Alfred Hitc	128 min	https://www.i
Blade Runner	R	1982	8.1	670715	84	https://m.m	['USA']	['English', 'Ge	['Harrison Fo	['Action', 'Sci	The star of "I	A blade runn	['Ridley Scot	117 min	https://www.i
Before Sunrise	R	1995	8.1	257987	77	https://m.m	['USA', 'Austri	['English', 'Ge	['Ethan Hawk	['Drama', 'Ro	... When lov	A young mar	['Richard Lin	101 min	https://www.i
Fiddler on the R	G	1971	8	37637	67	https://m.m	['USA']	['English', 'He	['Topol', 'Nor	['Drama', 'Fal	"Fiddler on t	In prerevolu	['Norman Jev	181 min	https://www.i
The Night of the	Not Rated	1955	8	79158	99	https://m.m	['USA']	['English']	['Robert Mitc	['Crime', 'Dra	The wedding	A religious fa	['Charles Lau	92 min	https://www.i

<https://www.kaggle.com/gorochu/complete-imdb-movies-dataset/metadata>

ARQUITECTURA PROPUESTA

PROYECTO BDA AWS (2020)

Sofia Davila | October 29, 2020





ROLES

Roles > Role_vpc_s3_sns

Summary

Delete role

Role ARN	arn:aws:iam::762011018360:role/Role_vpc_s3_sns 
Role description	Allows Lambda functions to call AWS services on your behalf. Edit
Instance Profile ARNs	
Path	/
Creation time	2020-10-26 22:28 EST
Last activity	2020-10-29 20:34 EST (Today)
Maximum session duration	1 hour Edit

Permissions

Trust relationships

Tags (1)






Access Advisor

Revoke sessions

▼ Permissions policies (5 policies applied)

Attach policies

[+ Add inline policy](#)

Policy name ▼	Policy type ▼	
▶  AmazonRDSFullAccess	AWS managed policy	×
▶  AWSLambdaFullAccess	AWS managed policy	×
▶  AmazonS3FullAccess	AWS managed policy	×
▶  AmazonVPCFullAccess	AWS managed policy	×
▶  AmazonSNSFullAccess	AWS managed policy	×

▶ Permissions boundary (not set)

BUCKET S3

aws

Services ▼

admin_aws @ 7620-1101-8360 ▼

Global ▼

Su

proyecto-bda-jds

Bucket overview

Region	Amazon resource name (ARN)	Creation date	Access
US East (Ohio) us-east-2	arn:aws:s3:::proyecto-bda-jds	October 25, 2020, 20:58 (UTC-05:00)	<u>Objects can be public</u>

Objects

Properties

Permissions

Metrics

Management

Access points

Drag and drop files and folders you want to upload here, or choose **Upload**.

Objects (1)

Objects are the fundamental entities stored in Amazon S3. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Delete

Actions ▼

Create folder

Upload

BASE DE DATOS RDS MYSQL



db-rds-mysql

Modify

Actions ▼

Summary

DB identifier
db-rds-mysql

Role
Instance

CPU

1.90%

Current activity

3 Connections

Info

Available

Engine

MySQL Community

Class

db.t2.micro

Region & AZ

us-east-2c

Connectivity & security

Monitoring

Logs & events

Configuration

Maintenance & backups

Tags

Connectivity & security

Endpoint & port

Endpoint

db-rds-mysql.cjxytg1tnk.us-east-2.rds.amazonaws.com

Port

3306

Networking

Availability zone

us-east-2c

VPC

vpc-a774c0cc

Subnet group

default-vpc-a774c0cc

Subnets

Security

VPC security groups

rds_vpc (sg-04b8ed92816e707f6)
(active)

Public accessibility

Yes

Certificate authority

rds-ca-2019

CONEXIÓN A MYSQL RDS DESDE WORKBENCH

Setup New Connection

Connection Name: Type a name for the connection

Connection Method: Method to use to connect to the RDBMS

Parameters

Hostname: Port: Name or IP address of the server host - and TCP/IP port.


Username: Name of the user to connect with.

Password: The user's password. Will be requested later if it's not set.

Default Schema: The schema to use as default schema. Leave blank to select it later.

Store Password For Connection

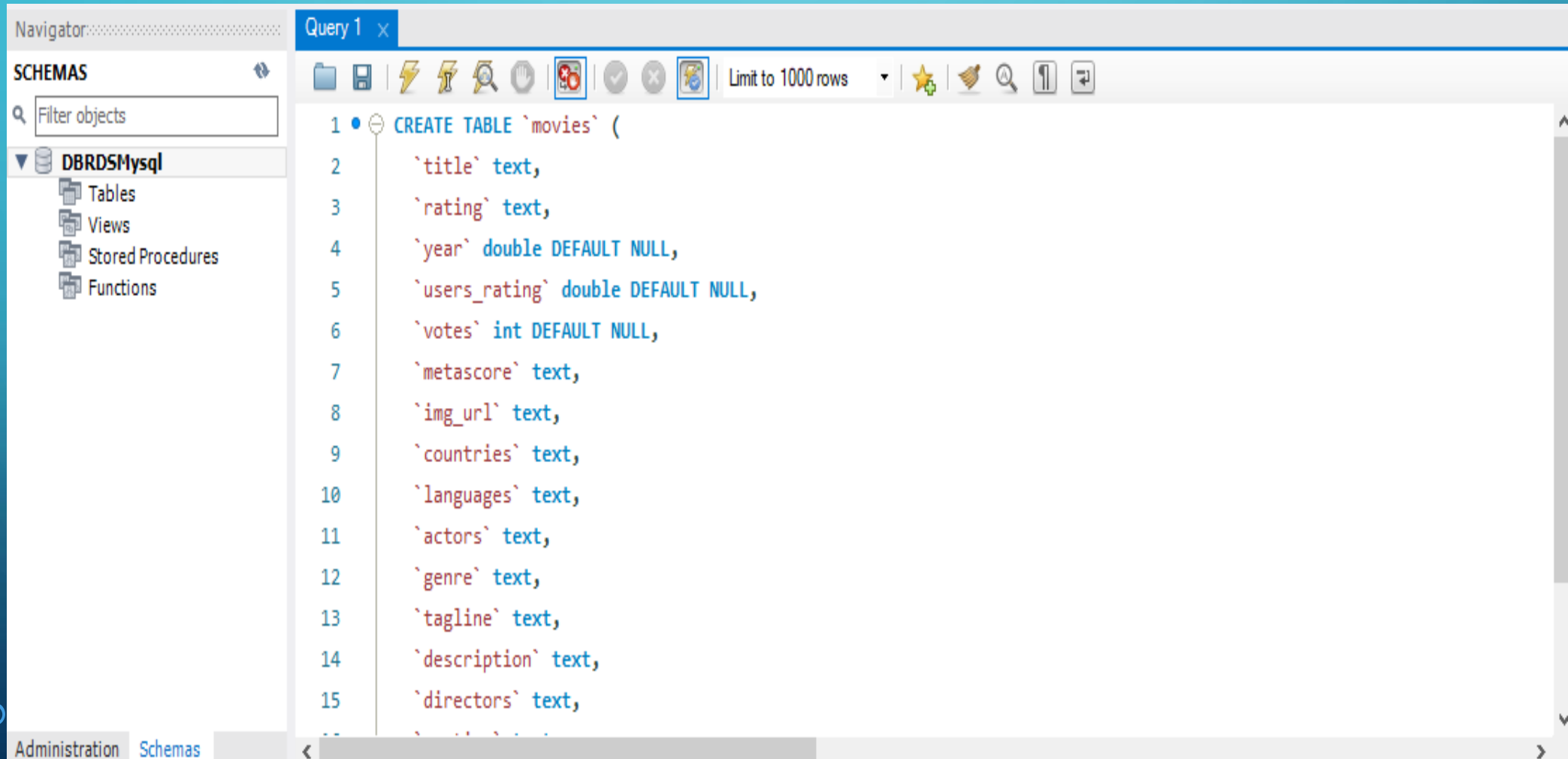
Please enter password for the following service:

 **Service:** Mysql@db-rds-mysql.cjyxytjg1tnk.us-east-2.rds.amazonaws.com:3306



User: admin_rds

Password:

CONEXIÓN A MYSQL RDS DESDE WORKBENCH



FUNCIÓN LAMBDA

 Your changes have been saved. 

lambda_trigger_s3_to_RDS

ThrottleQualifiers ▼Actions ▼s3prueba ▼Test

▼ Designer

lambda_trigger_s3_to_RDS

Layers (0)

S3

+ Add trigger

Amazon SNS

+ Add destination

FUNCIÓN LAMBDA

lambda_trigger_s3_to_RDS

Throttle

Qualifiers ▾

Actions ▾

s3prueba ▾

Test

Function code [Info](#)

Deploy

Actions ▾

File Edit Find View Go Tools Window Test ▾ Deploy

Environment

- lambda_trigger_s3_
- └─ pymysql
- └─ lambda_function.py
- └─ rds_config.py

```
45 bucket = event['Records'][0]['s3']['bucket']['name']
46 key = event['Records'][0]['s3']['object']['key']
47 download_path = '/tmp/' + key
48 print(download_path, bucket, key)
49 s3_client.download_file(bucket, key, download_path)
50 with open(download_path) as csvfile:
51     reader = csv.DictReader(csvfile, delimiter=';')
52     rowCount = 0
53     for row in reader:
54         rowCount += 1
55         print(row['title'], row['year'], row['votes'])
56         row_insert=(row['title'],row['rating'], row['year'], row['users_rating'],row['votes'],row['metascore'],row['img_url'],row['countries'],
57                     row['languages'],row['actors'],row['genre'],row['tagline'], row['description'],row['directors'],row['runtime'],row['imdb_url'])
58         inserta_new_movies(row_insert)
59
60 return 'Se insertaron'+str(rowCount)+" registros en la tabla movies"
```

58:29 Python Spaces: 4

Execution Result

Status: Succeeded Max Memory Used: 82 MB Time: 1851.11 ms ▾ Execution results

Response:
"Se insertaron15 registros en la tabla movies"

Request ID:
"d4ce8548-d0ce-4706-a4f7-9c29729ee318"

Function logs:
START RequestId: d4ce8548-d0ce-4706-a4f7-9c29729ee318 Version: \$LATEST
[INFO] 2020-10-31T03:45:37.654Z d4ce8548-d0ce-4706-a4f7-9c29729ee318 Found credentials in environment variables.

/tmp/new_movies.csv proyecto-bda-jds new_movies.csv
The Avengers 2012 9.1332746

FUNCIÓN LAMBDA

```
import json
import boto3
import boto3
import logging
import os
import sys
import uuid
import pymysql
import csv
import rds_config

#credenciales de bd
db_host = rds_config.db_host
db_username = rds_config.db_username#
db_password = rds_config.db_password
db_name = rds_config.db_name

#conectando a la base de datos
logger = logging.getLogger()
logger.setLevel(logging.INFO)

try:
    conn = pymysql.connect(db_host, user=db_username, passwd=db_password, db=db_name, connect_timeout=5)
    logger.info("SUCCESS: Connection to RDS MySQL instance succeeded")
except pymysql.MySQLError as e:
    logger.error("ERROR: Unexpected error: Could not connect to MySQL instance.")
    logger.error(e)
    sys.exit()
```

FUNCIÓN LAMBDA

```
def inserta_new_movies(row_insert):
    try:
        with conn.cursor() as cur:
            cur.execute('INSERT INTO movies (title,rating,year,users_rating,votes,metascore,img_url,countries,languages,\'
            \'actors,genre,tagline,description,directors,runtime,imdb_url)values (%s, %s, %s,%s, %s, %s,%s, %s, %s,%s, %s,%s,%s, %s,%s,%s)',row_insert)
            conn.commit()
    except Exception as e:
        logger.error(e)
    return "ok new movies"

#funcion que obtiene la pelicula top
def top_movie():
    with conn.cursor() as cur:
        cur.execute("select title, actors,description,genre,runtime,year, imdb_url ,rating from movies order by users_rating desc limit 1")
        for row in cur:
            titulo=row[0]
            actores=row[1]
            descripcion=row[2]
            genero=row[3]
            duracion=row[4]
            anio=row[5]
            link=row[6]
            calificacion=row[7]
            correo="Le recomendamos ver la pelicula "+titulo+"("+str(anio)+")|"+calificacion+"|"+str(duracion)+"|"+genero+"|"+actores+"."+descripcion+"
            Veala en:"+link
            logger.info(correo)

    return correo
```

FUNCIÓN LAMBDA

```
def envia_email(mensaje,topico,asunto):
    sns = boto3.client('sns')
    sns.publish(
        TopicArn=topico,
        Message=mensaje,
        Subject=asunto,
        MessageStructure='raw'
    )
    return "Mensaje enviado"

def lambda_handler(event, context):
    s3_client = boto3.client('s3')
    bucket = event['Records'][0]['s3']['bucket']['name']
    key = event['Records'][0]['s3']['object']['key']
    download_path = '/tmp/'+key
    print(download_path,bucket,key)
    s3_client.download_file(bucket, key,download_path)
    with open(download_path) as csvfile:
        reader = csv.DictReader(csvfile, delimiter=',')
        rowCount = 0
        for row in reader:
            rowCount += 1
            print(row['title'], row['year'], row['votes'])
            row_insert=(row['title'],row['rating'], row['year'], row['users_rating'],row['votes'],row['metascore'],row['img_url'],row['countries'],
                        row['languages'],row['actors'],row['genre'],row['tagline'], row['description'],row['directors'],row['runtime'],row['imdb_url'])
            inserta_new_movies(row_insert)
    topico="arn:aws:sns:us-east-2:762011018360:top_movies"
    asunto='Pelicula Top del dia en IMDB'
    mensaje=top_movie()
    print(mensaje)
    envia_email(mensaje,topico,asunto)
    return 'Se insertaron'+str(rowCount)+" registros en la tabla movies"
```


ENVIO DE NOTIFICACIONES SNS

aws

Services ▾

admin_aws @ 7620-1101-8360 ▾

Ohio ▾

Sa

Amazon SNS

×

Dashboard

Topics

Subscriptions

▼ Mobile

Text messaging (SMS)

top_movies

Edit

Delete

Publish message

Details

Name

top_movies

ARN

arn:aws:sns:us-east-2:762011018360:top_movies

Type

Standard

Display name

-

Topic owner

762011018360

Subscriptions

Access policy

Delivery retry policy (HTTP/S)

Delivery status logging

Encryption

Tags

Subscriptions (1)

Edit

Delete

Request confirmation

Confirm subscription

Create subscription

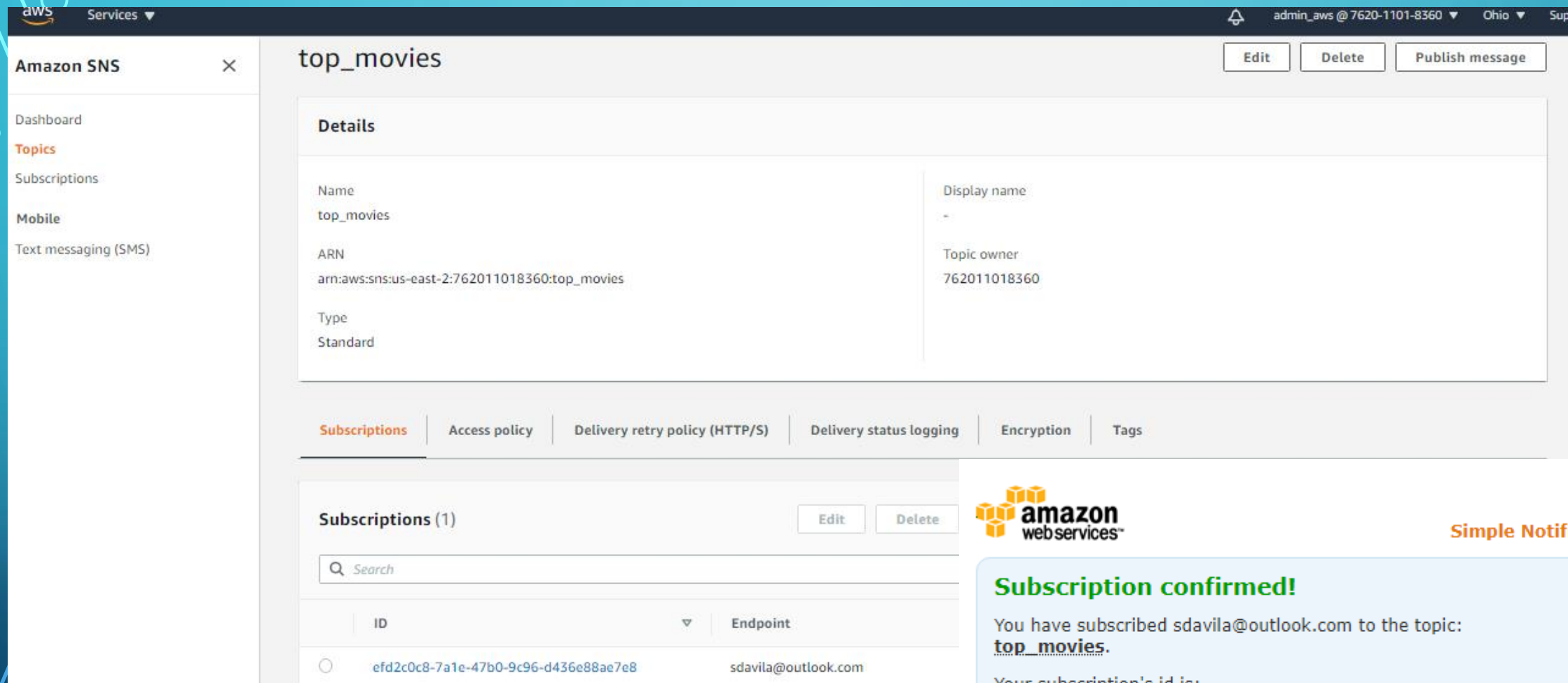
Q Search

< 1 >

⚙

ID	Endpoint	Status	Protocol
<input type="radio"/> efd2c0c8-7a1e-47b0-9c96-d436e88ae7e8	sdavila@outlook.com	✔ Confirmed	EMAIL

SUSCRIPCIÓN SNS



The screenshot shows the AWS Management Console for the Amazon SNS service. The left sidebar contains navigation links: Dashboard, Topics, Subscriptions, Mobile, and Text messaging (SMS). The main content area is titled 'top_movies' and includes buttons for Edit, Delete, and Publish message. Below the title is a 'Details' section with the following information:

Field	Value
Name	top_movies
ARN	arn:aws:sns:us-east-2:762011018360:top_movies
Type	Standard
Display name	-
Topic owner	762011018360

Below the details section are tabs for Subscriptions, Access policy, Delivery retry policy (HTTP/S), Delivery status logging, Encryption, and Tags. The 'Subscriptions' tab is active, showing a table with one subscription:

ID	Endpoint
efd2c0c8-7a1e-47b0-9c96-d436e88ae7e8	sdavila@outlook.com

Overlaid on the bottom right is a confirmation message from Amazon Web Services:

Subscription confirmed!

You have subscribed sdavila@outlook.com to the topic: **top_movies**.

Your subscription's id is:
arn:aws:sns:us-east-2:762011018360:top_movies:efd2c0c8-7a1e-47b0-9c96-d436e88ae7e8

If it was not your intention to subscribe, [click here to unsubscribe](#).

CORREO DE CONFIRMACIÓN

Pelicula Top del dia en IMDB

① ¿Tiene demasiado correo? [Cancelar suscripción](#)

 Traducir mensaje a: [Español](#) | [No traducir nunca de: Inglés](#)



AWS Notifications <no-reply@sns.amazonaws.com>

Vie 30/10/2020 23:13

Para: Usted



Le recomendamos ver la película The Avengers(2012.0)[PG-13|143 min][['Action', 'Adventure', 'Sci-Fi']][['Robert Downey Jr.', 'Chris Evans', 'Mark Ruffalo', 'Chris Hemsworth', 'Scarlett Johansson', 'Jeremy Renner', 'Tom Hiddleston', 'Clark Gregg', 'Cobie Smulders', 'Stellan Skarsgård', 'Samuel L. Jackson', 'Gwyneth Paltrow', 'Paul Bettany', 'Alexis Denisof', 'Tina Benko']:Earth's mightiest heroes must come together and learn to fight as a team if they are going to stop the mischievous Loki and his alien army from enslaving humanity. Veala en:<https://www.imdb.com/title/tt0848228/>


--

If you wish to stop receiving notifications from this topic, please click or visit the link below to unsubscribe:

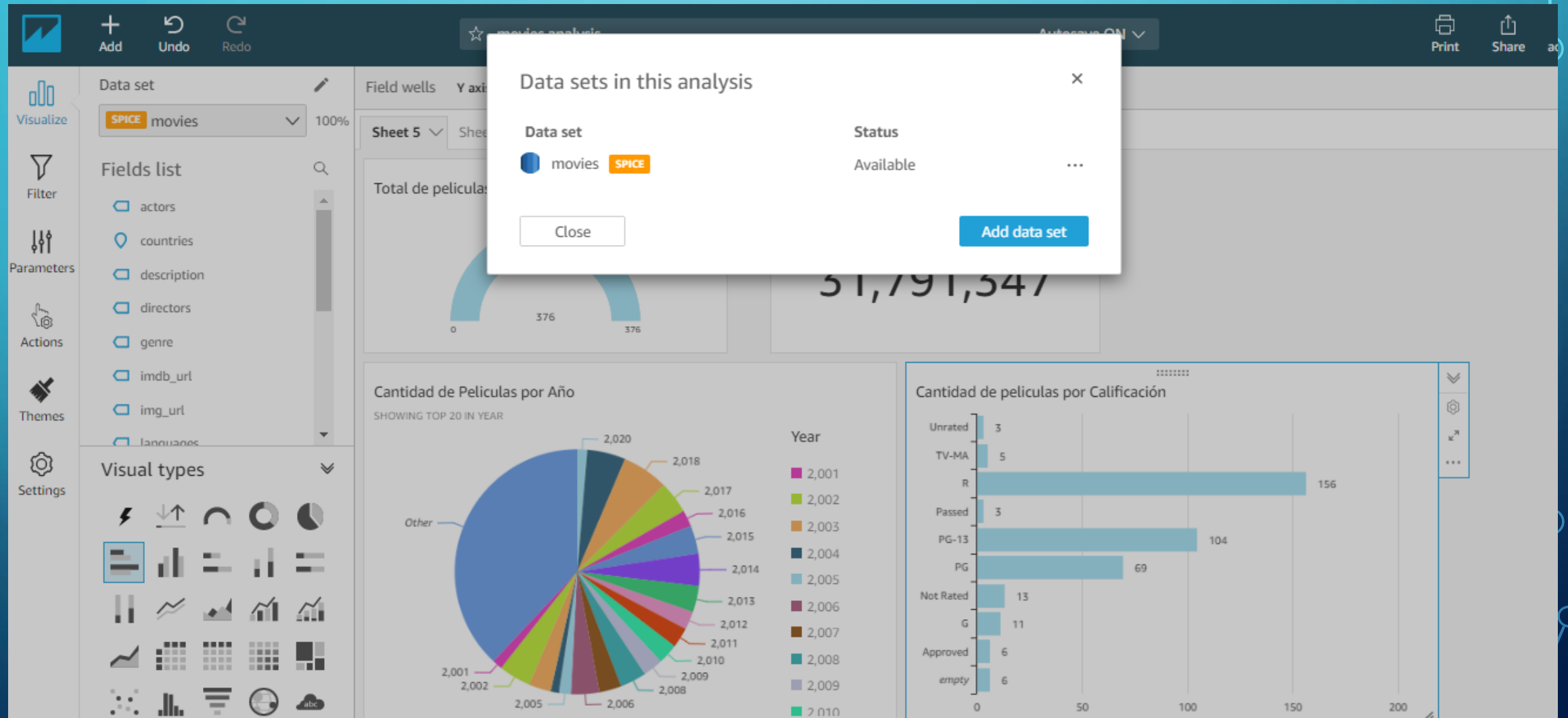
https://sns.us-east-2.amazonaws.com/unsubscribe.html?SubscriptionArn=arn:aws:sns:us-east-2:762011018360:top_movies:efd2c0c8-7a1e-47b0-9c96-d436e88ae7e8&Endpoint=sdavila@outlook.com

Please do not reply directly to this email. If you have any questions or comments regarding this email, please contact us at

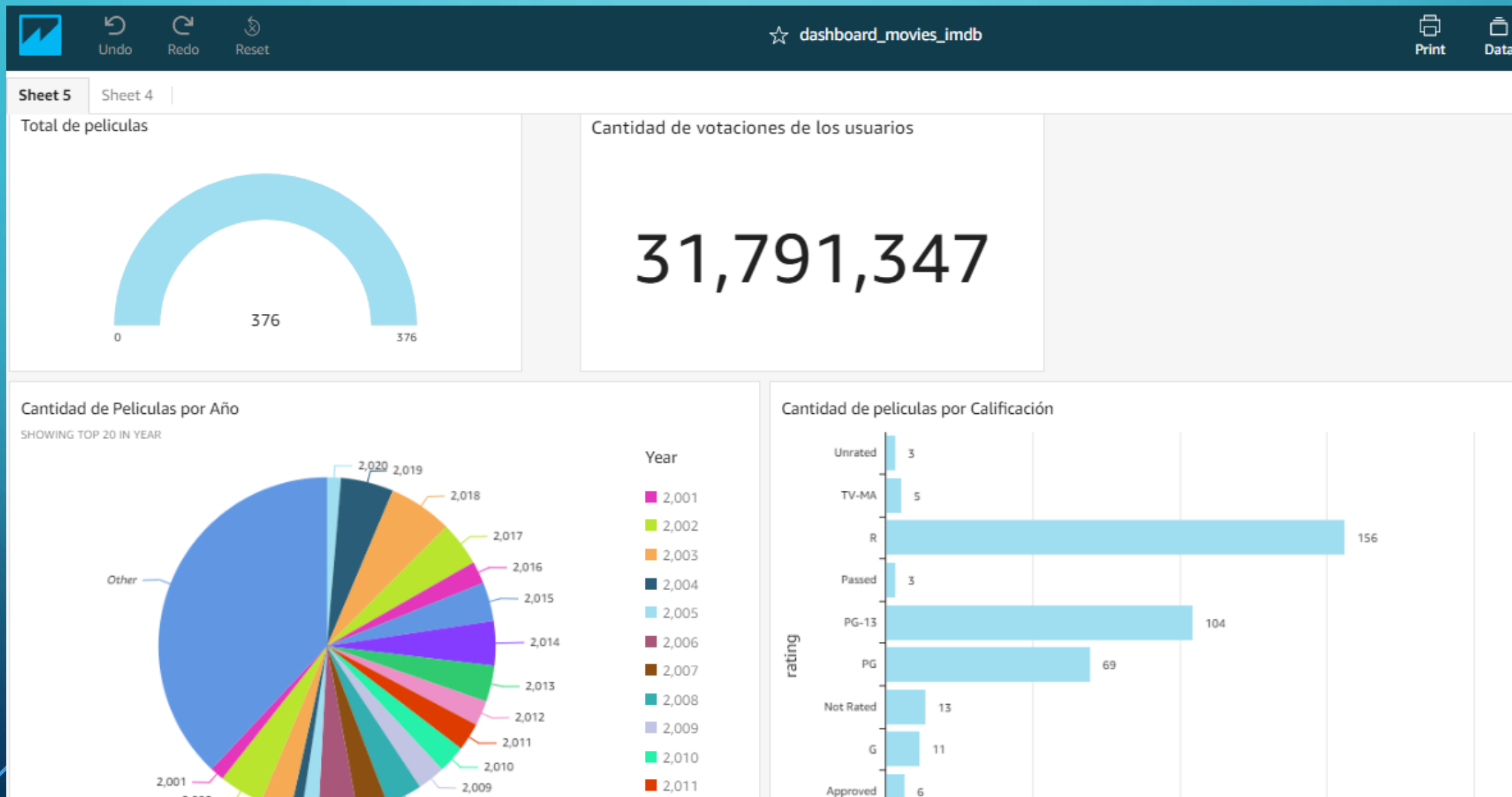
<https://aws.amazon.com/support>

 Programar una reunión

DASHBOARD EN QUICKSIGHT



DASHBOARD EN QUICKSIGHT





<https://www.youtube.com/watch?v=9ZqhWUCui-o>

