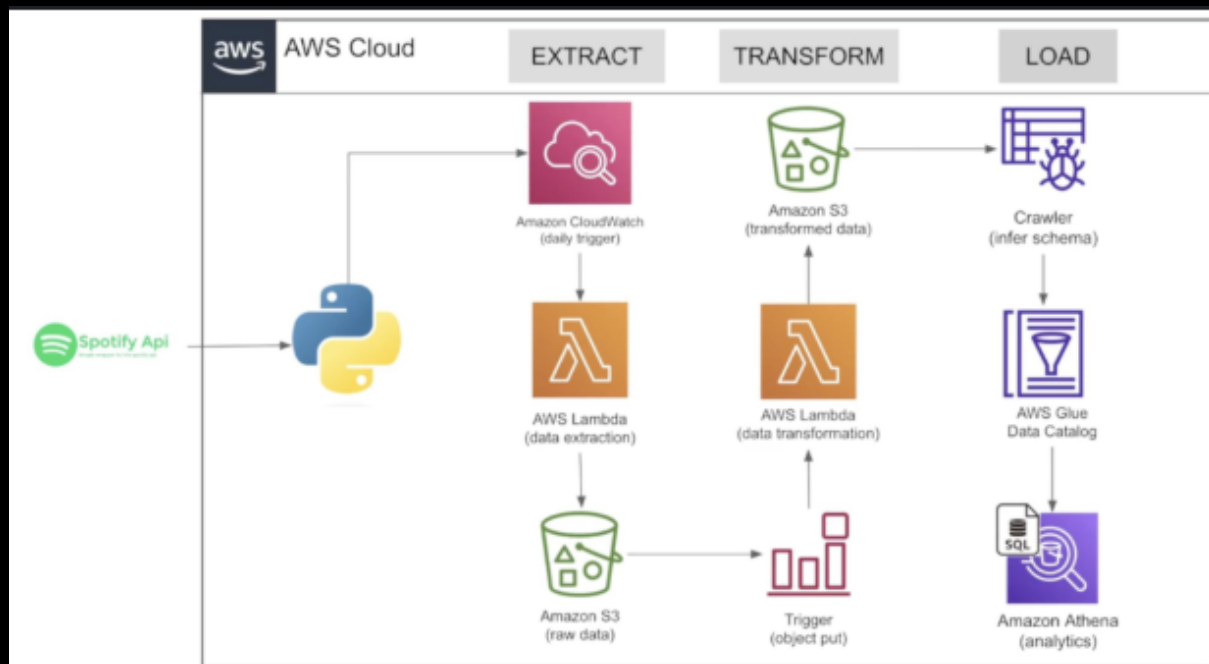# 1.SPOTIFY DATA EXTRACTION AND LOADING

Business probelm: User wants to explore weekly trending songs and improve the app based on it.

-collect the data on a weekly basis so that after one or two years he can find different insights ;so we need to build an ETL pipeline for it.

Initially we are gonna build this pipeline:

EXTRACTION:
Spotify api- data extraction.
AWS LAMBDA: Deploy code in aws lambda
AWS CLOUDWATCH: we can schedule ,to run lambda on particular interval we provide
First we collect data from spotify API and store it in s3(raw data)
TRANSFORM:
   +   Add trigger
AWS LAMBDA(data transformation)
= s3(transformed data)
Whole process is automated
Then
LOAD:
GLUE CRAWLER: Go through each column it understand the data ,diffrent column names,data type and it will build the GLUE CATALOG(all details)
ATHENA : Directly run sql queries.

SPOTIFY API : Music artists,albums, and tracks

AMAZON S3:
-simple storage service
-storage service that can store and retrieve any amount of data from anywhere on the web.
-used to store and distribute large media files,data backups,and static website files.

COMPUTE:(AWS LAMBDA):
-Serverless computing service.
- To run code response to events like changes in s2,dynamodb,or other aws services.

LOGS/TRIGGER:
-Collect and monitor log files,alarms

PROJECT ::

1.Open https://developer.spotify.com/dashboard
Create app
App name:
Client id : b6e073250b1044eca162cee8c1e53ec9
Client secret :8c2e08d41bed40678503fd92a7bd10af

```
client_credentials_manager = SpotifyClientCredentials(client_id =
"b6e073250b1044eca162cee8c1e53ec9",client_secret="8c2e08d41bed40678503f
d92a7bd10af")
```

This is the authentication

```
sp = spotipy.Spotify(client_credentials_manager =
client_credentials_manager)
```

This is authorization.

Copy pasted the URL link

From URL link extracted the id of it

From the data we are extracting the different lists from that we are converting it into
dictionary for json format

Next we are creating a list
List with pandas we create a dataframe
Changed date time and maintained correct data types

CREATE S3 BUCKET- WE WILL STORE THE DATA
-create raw and transformed folders

CREATE LAMBDA TRANSFORMATIONS
-CREATE EXTRACT FUCNTION
 Lambda - add the code
1.you will get one problem like pandas not installed so for aws lambda you should do following process ;
Lambda→ layers→ add the xip file and create a layer and then add it in the code choose custom layer

Import boto3

Entyire json file to s3

Mow we are gonna automate this by event bridge