

# Spotify ETL Pipeline



This project collects data from the Spotify API and sends it to Snowflake for easy analysis, all done automatically using AWS services.

- Extract: A Lambda function connects to Spotify, pulls playlist data, and saves it to an Ś3 bucket as a raw JSON file.
- Transform: When a new file is uploaded, another Lambda function is triggered. It cleans and organizes the data into tables (artist, album, and songs), and saves them as CSV files in S3.
- Load: Snowpipe in Snowflake automatically detects these new CSV files and loads them into structured database tables.

This pipeline runs without any manual steps. It's fully automated, cloud-based, and perfect for handling real-time music data at scale.



## AMAZON CLOUDWATCH

Trigger the data extraction process automatically on a schedule



#### **AWS LAMBA**

Runs a Python function to extract data from the Spotify API which will be our raw data



## S3 BUCKET

The landing zone for the raw data that's been pulled from the Spotify API by the Lambda function.





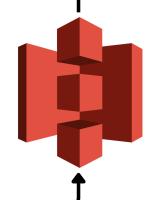
### S3 BUCKET

The landing zone for the transformed, clean data ready for loading into Snowflake.



#### **AWS LAMBA**

Processes raw data by cleaning, transforming, and saving structured outputs to S3



# **TRIGGER** (OBJECT PUT)

S3 upload event auto triggers the transformation Lambda function.

**TRANSFORM** 



Automatically loads new CSV files from S3 into Snowflake tables as soon as they arrive.



Stores the final structured data in relational tables for analysis and querying.

**LOAD**