

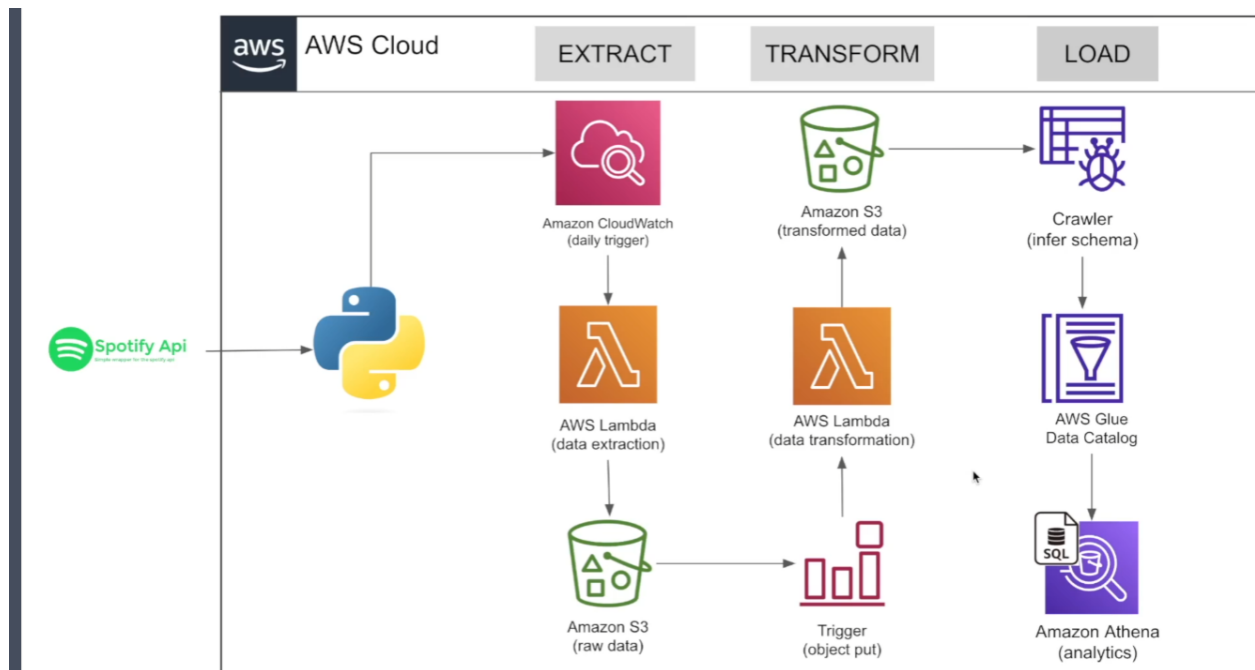
Spotify Pipeline

spotify api (src) → extract data and deploy code to lambda → and store raw data in S3

from S3 we do transformation using lambda and store in S3

use crawler to read metadata and infer schema and store in Data Catalog

use Athena to analyze data stored in Data Catalog



In S3 we created 2 folders

1. raw data →
 - a. already processed

- b. to be processed
- 2. transformed Data →
 - a. album data
 - b. artist data

to be processed will have extract data from spotify playlist

which will be processed by lambda to create files in album and artist data

after src files in to be processed should be moved already processed folder and then delete those srcfiles from to be processed folder

Later use GLue crawler to infer schema and metadata and load to GLue catalog

Connect to Athena to analyze from Glue catalog