

# Data Modeling with Postgres

## Introduction

A startup called Sparkify wants to analyze the data they've been collecting on songs and user activity on their new music streaming app. The analytics team is particularly interested in understanding what songs users are listening to. Currently, they don't have an easy way to query their data, which resides in a directory of JSON logs on user activity on the app, as well as a directory with JSON metadata on the songs in their app.

They'd like a data engineer to create a Postgres database with tables designed to optimize queries on song play analysis, and bring you on the project. Your role is to create a database schema and ETL pipeline for this analysis. You'll be able to test your database and ETL pipeline by running queries given to you by the analytics team from Sparkify and compare your results with their expected results.

## Description of each file

Create\_tables.py: It is to create databases and tables to get prepared for further modeling;

sql\_queries: SQL Queries that are used in Postgres database and in this project, python is going to make use of these queries to create tables and insert values;

etl.py: A pipeline that reads Json logs and metadata and populate data into tables;

## Database Schema



