**Data Modeling with Postgres**

**Purpose of Project**

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. They'd like a data engineer to create a Postgres database with tables designed to optimize queries on song play analysis.

## Project Description

In this Project, fact and dimension tables will be designed for a star schema for this particular purpose. An ETL pipeline will also be designed to transfer data from files in two local directories into these tables in Postgres using Python and SQL.

## Database schema

This database uses the star schema: One Fact Table and Four Dimension Tables.

## 

## File Structure

ETL.py will work as following for processing the data:

* Connect to the sparkify datase. And it will drop and create all the tables.
* Parse out each json file and load all of the files into dataframe.
* Song\_data and Log\_data will be loaded into the fact and dimension tables.

Files will be excuted in the following orders:

* 1.create\_tables.py
* 2.etl.py
* 3.test.ipynb

## Summary

This project will provide Sparkify to answer some business question like 'Which are the top popular songs users are listening to".