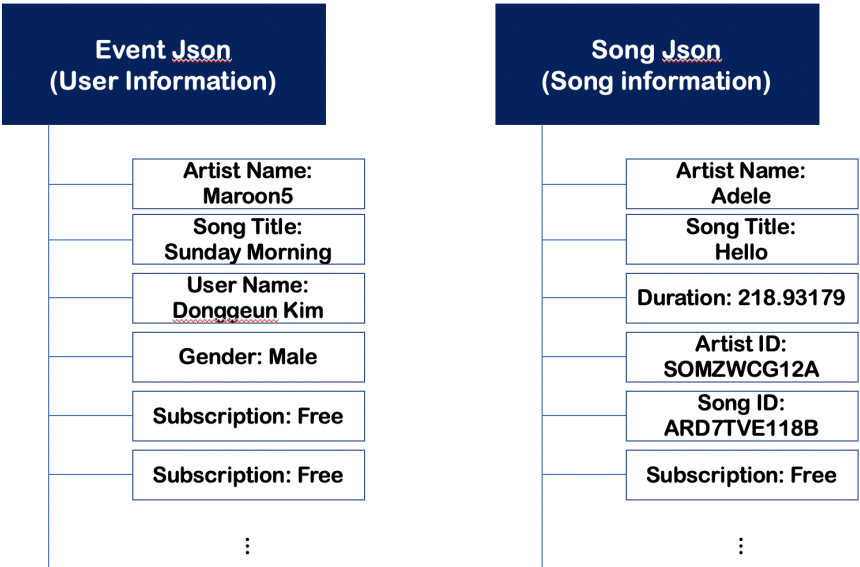


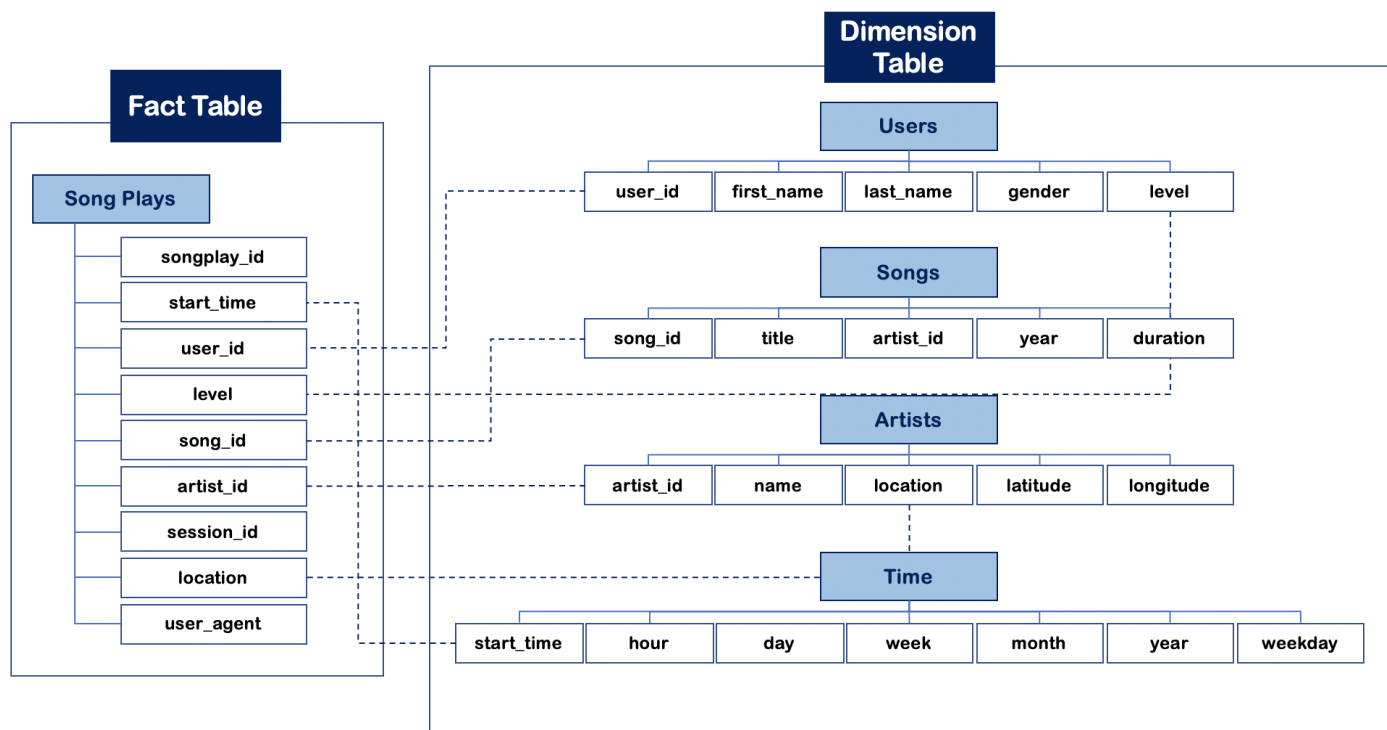
Schema for Song Play Analysis

This project builds on top of the SQL ETL project. Run `create_tables.py` and then `etl.py`. The primary difference is that this project loads data from Amazon S3 and Fact/Dimension table in Amazon Redshift Cluster. Using the song and user datasets, this project attempts to create a star schema optimized for queries on song play analysis. The data feeds are as following.



Schema Illustration

The schema organizes data above into following set of tables. The fact table supports general analysis of data feed whereas dimension tables highlight distinctive features worthy of attentions on their own.



Core files or, in other words, final output consists of following. ¶

1. **create_tables.py**

This file initializes database for this project, drops table if there is any, and creates tables discussed above.

2. **sql_queries.py**

It is a collection of sql queries necessary for dropping and creating tables. It also contains queries for inserting data into each table, which is utilized in `etl.py`

3. **etl.py**

It grabs relevant files, processes files and feeds relevant data to tables.

4. **dwh.cfg**