* MusicBrainz Database

The MusicBrainz Database is built on the PostgreSQL relational database engine and contains all of MusicBrainz' music metadata. This data includes information about [artists](https://musicbrainz.org/doc/Artist), [release groups](https://musicbrainz.org/doc/Release_Group), [releases](https://musicbrainz.org/doc/Release), [recordings](https://musicbrainz.org/doc/Recording), [works](https://musicbrainz.org/doc/Work), and [labels](https://musicbrainz.org/doc/Label), as well as the many [relationships](https://musicbrainz.org/doc/Relationships) between them. The database also contains a full history of all the changes that the MusicBrainz community has made to the data.

* ETL Process

Basic entities below are extracted, cleansed, and transformed

Dimensions.

Dim.date added

AreaTypes

1. id, name, parent, child\_order columns extracted
2. Unknown records added with the value of 0
3. Data loaded as Dim.Areas

ArtistsTypes

1. id, name, parent, child\_order columns extracted
2. Unknown records added with the value of 0
3. Data loaded as Dim.ArtistTypes

EventTypes

1. id, name, parent, child\_order columns extracted
2. Unknown records added with the value of 0
3. Data loaded as Dim.EventTypes

PlaceTypes

1. id, name, parent, child\_order columns extracted
2. Unknown records added with the value of 0
3. Data loaded as Dim.PlaceTypes

Genders

1. id, name, parent, child\_order columns extracted
2. Unknown records added with the value of 0
3. Data loaded as Dim.Genders

Events

1. id, **gid**, begin\_date\_year, begin\_date\_month, begin\_date\_day, end\_date\_year, end\_date\_month, end\_date\_day, time, type, cancelled, endedcolumns extracted
2. events without begin and end dates cleansed
3. events without gid cleansed
4. date columns and merged based on dim.date primary keys created
5. duration column created
6. data loaded as Dim.Events

Artists

1. id, **gid**, name, sort\_name, type, area, gender, ended columns extracted
2. type, area, genders columns with null values filled with 0 zero which indicates unknown records
3. ended column renamed to is\_active column to improve the readablitly
4. data loaded as Dim.Artists

Areas

1. id, **gid**, name, type, ended columns extracted
2. type column with null values filled with 0 zero which indicates unknown records
3. ended column renamed to is\_active column to improve the readablitly
4. data loaded as Dim.Areas

Places

1. id, **gid**, name, type, area, ended columns extracted
2. type, area columns with null values filled with 0 zero which indicates unknown records
3. ended column renamed to is\_active column to improve the readablitly
4. data loaded as Dim.Places

Appendix 1

MusicBrainz Schema

