SOEN363

Data Systems for Software Engineers

October 7, Fall 2024

Project Proposal

Christa Abou-Arraje 40226631 (team lead)

Despina Koulisakis 40190212

Nour Hassoun 40233077

Sara Abellard 40184667

The name: Global Music Insights Database
The topic: Building a database to track global music trends, including popular songs, artists, and albums from different countries.
The scope: The database will gather data on top-ranking songs and albums from different countries and music platforms, like Spotify and Last.fm. It will include information on artists and the genre they belong to, as well as listeners and their favorite songs. This will help analyze trends in music popularity across different regions and over time.

Database description (focus on entities):

- Songs: Title, artist, album, genre, release date, duration, and popularity score.
- Artists: Name, genre, country, debut year, and associated albums/songs.
- Albums: Title, artist, release date, genre, tracklist.
- Music Charts: Chart name, country, date, top songs (ranking, song, artist).
- Listeners: User demographics (age, country, gender), genre preferences.

Datasources:

- *Spotify API*: Provides data on songs, albums, artists, playlists, and streaming statistics
  -API type: RESTFul API
  -Link: https://developer.spotify.com/documentation/web-api/
- *Last.fm API*: Provides music charting and user behavior data, including track plays, top artists, and user demographics.
  -API type: RESTful API
  -Link: https://www.last.fm/api

Implementation platform:

- Relational Database: PostgreSQL for the relational database implementation.
- NoSQL Database: MongoDB for Phase II (migration to NoSQL).

Programming platform:
-Language: Python for API consumption and data collection, utilizing libraries such as requests for making API calls and psycopg2 for PostgreSQL interactions.
-Scripting Tools: Pandas for data transformation and cleaning before loading into the relational database.
Estimated plan for complete data collection:

- **Week 1-2**: Finalize API selection and gather documentation for the Spotify and Last.fm APIs.
- **Week 3-4**: Develop Python scripts to extract data on songs, artists, and albums
- **Week 5-6**: Collect data on user playlists, listening habits, and chart rankings.
- **Week 7**: Begin organizing and cleaning data before loading into the PostgreSQL database.
- **Week 8**: Populate the relational database
- **Week 9-10**: Phase II: begin the process of migrating data into MongoDB.