Music Recommendation Architecture

Bogdan Petre
Yuying Chen
Zhuoyun Kan

Objective

Recommend music for users based on their interests and playing history by using existing and streaming data

Data

Two major data sources

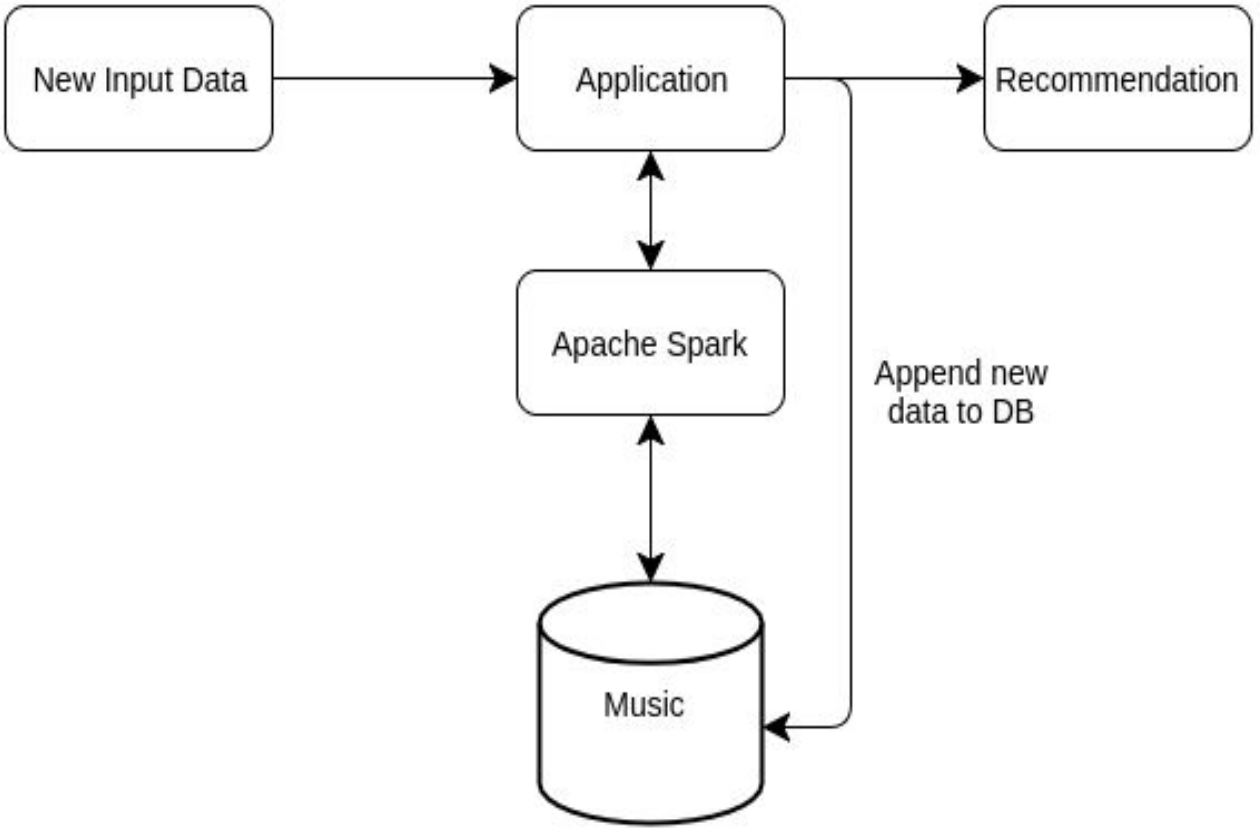
- Music data collected from the Internet as pre-stored data for training
- Live streaming data from the Internet or generated synthetically

Possible sources: Last.fm, Spotify, or some generated user data

Architecture

Technologies used:

- Python
- Spark
- Cassandra



Plan of Implementation

- Explore possible sources of data collection
- Build scalable databases (Cassandra) to store music data
- Implement Spark to do Map/Reduce (e.g., feature extraction), and do machine learning (online or) for music recommendation models
- Store the trained results to database
- Retrieve the model data for User and generate recommendation list (Content Matching)