

빅데이터 프로그래밍 기말프로젝트 계획서

팀원: 김아영 60201561, 김보겸 60211648, 남재희 60210521, 박상재 60201674, 이성우 60201079

깃허브: <https://github.com/24-bdp-team8>

1. 주제 및 목표

주제	사용자 맞춤 음악 추천 시스템
목표	사용자 선호도와 음악 청취 기록을 기반으로 맞춤형 음악을 추천하는 시스템을 개발하고자 한다.
기대 효과	다양한 장르, 아티스트, 플레이리스트 분석 및 데이터를 활용하여 개인화된 음악 경험을 제공함으로써 사용자 만족도를 증가시키고자 한다.

2. 데이터 수집

- Spotify API: 음악 트랙, 앨범, 아티스트, 플레이리스트 데이터를 수집
<https://developer.spotify.com/documentation/web-api>
- Million Song Dataset: 대규모 음악 데이터셋으로 청취 기록과 메타데이터 포함
사용자 행동 데이터 (공개 데이터 셋, 캐글)
<https://www.kaggle.com/c/msdchallenge/data>
- 크롤링: 음악 관련 블로그, 리뷰 사이트 크롤링으로 감성 분석 데이터 확보

3. 데이터 전처리 및 저장

HDFS: 음악 메타데이터, 사용자 청취 기록 등을 대용량 저장소에 통합
MapReduce/Spark: 데이터 정제, 중복 제거 및 포맷 정리
Hive: 정리된 데이터를 구조화하여 쿼리 및 분석

4. 추천 시스템 설계

1) 콘텐츠 기반 추천

Spotify API로 가져온 메타데이터(장르, 템포, 키, 에너지 등)를 기반으로 유사한 음악 추천

2) 협업 필터링

Spark MLlib로 사용자가 청취한 음악과 다른 사용자의 데이터를 비교하여 추천