**문제해결 빅데이터 활용 프로젝트 기획안**

기획안 작성일자 : 2022.09.23

|  |  |
| --- | --- |
| 조 명 | 3조 : 주말에 영화 보시조 |
| 조 원 | 조장 : 이서정 조원 : 손지수, 유승종, 이상엽, 정현진 |
| 프로젝트 주제 및 개요 | 영화 산업 흥행 요인 분석을 위한 데이터 DB 구축 |
| 프로젝트 수행 방향 | 1. 영화 산업 흥행 요인 분석 필요 데이터 수집 - HDFS  - 영화진흥위원회 API 활용 영화 정보(영화명, 상영관, 관객수 등) 수집  - 네이버 검색 트렌드/ 영화 API 활용 영화 검색 수 수집  - 인스타그램 영화명 해시태그 게시글 수 크롤링  - 네이버 영화 대중 평점/ 전문가 평점 크롤링  2. DataWarehouse, DataMart 설계 및 구축 - Oracle cloud  - ERD Cloud 활용 DB 테이블 설계  - Oracle cloud 계정에 DW/DM 구축  3. 데이터 파이프라인 구축  - 각 ETL과정을 클래스화 하여 app으로 구현  4. 영화 산업 흥행요인 분석 및 프로덕트 개발  - 각 요인들과 흥행정도의 상관관계 분석 시각화 -BI tool  - Rest API 서버 개발 - Django Rest Framework  ● 수행도구, 데이터소개  - Python, PySpark, Django  - Docker, Spark, Hadoop, Oracle  - 영화진흥위원회, 네이버 검색 트렌드/영화 API |
| 프로젝트 조직  (구성원 및 역할) | ● 역할분담  공통 : 주제 선정 및 포트폴리오 작성, Rest API 서버 구축, 시각화  조 장: 이서정 - 환경 구축 및 네이버 영화 검색수 ETL 구현  조원1: 손지수 - 네이버 영화 평점 ETL 구현  조원2: 유승종 - 인스타그램 게시글 수 ETL 구현  조원3: 이상엽 - 일별 박스오피스 정보 ETL 구현  조원4: 정현진 - 영화 상세 정보 ETL 구현 |
| 프로젝트 추진 일정 | ● 일정  9/22 ~ 9/23: 주제 선정 및 일정 수립  9/23 ~ 9/26: 기본 환경 구축  9/27 ~ 9/28: 데이터 수집 및 HDFS 설계  9/29 ~ 10/3: 데이터 가공 및 DW/DM 설계  10/4 ~ 10/5: Rest API 구현 및 API 사이트 구현  10/6 ~ 10/7: 시각화 및 포트폴리오 작성 |