**프로젝트 계획서**

**신한은행 기업 연계 프로젝트**

**2021. 10. 13**

**김예찬, 김윤서, 방지환, 이규정, 정희진**

**목차**

1. 프로젝트 개요... 3

1-1. 주제... 3

1-2. 개요... 4

1-3.. 사용대상

1-4. 팀원별 담당 역할... 5

2. 개발 계획... 6

2-1. 일정 계획... 6

2-2. 개발 언어 및 활용 기술... 6

2-3. 오픈소스 활용 및 개발…6

2-4. 예산... 7

3. 분석 및 설계... 8

3-1. 요구사항 정의... 8

3-2. 애플리케이션 아키텍쳐... 9

# **1. 프로젝트 개요**

## **1-1. 주제**

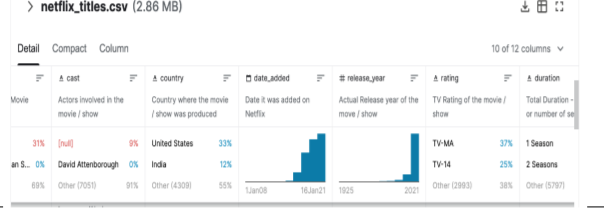
* 빅데이터 분석 및 시각화 기능 개발 (데이터셋 feature 시각화)
* 代 은행용 `kaggle` 제작

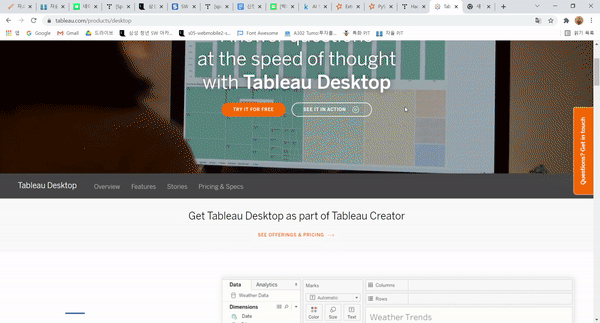
## **1-2. 개요**

* **빅데이터**와 **AI**의 발전에 따라 당행 내부에서도 이를 위한 전행적 데이터 활용역량 향상이 요구되고 있고, 관련 업무들도 다양해짐
* **데이터분석**이나 **ML**(Machine Learning) / **DL**(Deep Learning)모델링시 첫번째 작업은 데이터셋의 **특성을 파악**하는 것이고, 데이터셋의 특성을 쉽게 파악할 수 있다면 데이터 **분석과 활용**의 로드맵을 세우는 작업이 용이해질 것임
* 데이터셋의 대략적 특성을 파악할 수 있는 시각화와 통계측정치를 제공하는 웹페이지 개발을 통해 사용자의 데이터 접근 용이성 향상 도모

*\* 벤치마킹 또는 유사 서비스 사례 소개*

[ `kaggle.com`의 dataset, ‘tablaue’]





## **1-3. 사용대상**

* 행내 **데이터 분석가** 및 **ML / DL 모델러**

## **1-4. 목표**

1) 데이터 분석 및 ML/DL 모델링 시 참고 용도로 활용

2) 향후 전행 데이터셋 대상으로 확대

3) 학습한 기술을 접목하여 서비스로서의 완성도와 안정성을 모두 갖춘 프로젝트 완성 및 팀원들간 백엔드, 프론트엔드 등의 구분 없이 서비스 기획, 설계 및 각 기술스택에 대한 전반적인 이해를 모두 갖추며 개발자로서의 역량 향상

## **1-5. 팀원별 담당 역할**

|  |  |  |
| --- | --- | --- |
| **이름** | **역할** | **담당 업무** |
| 김예찬 | 팀장 | Backend |
| 김윤서 | 서기 | Frontend / Data |
| 방지환 | 서버 담당 | Server / Data |
| 이규정 | 부팀장 | Frontend / Data |
| 정희진 | 데이터 팀장 | Frontend / Data |

\*서기는 돌아가면서 역할 분배

\*front 컴포넌트 구성 후 Spring 학습하며 Backend 참여

# **2. 개발 계획**

## **2-1. 일정 계획**

|  |  |  |  |
| --- | --- | --- | --- |
| **시작일** | **종료일** | **내용** | **담당자** |
| 10/13 | 10/15 | 기획서 작성 |  |
| 10/18 | ~ | 화면 기획(화면 정의서 작성) |  |
| 10/25 | ~ | 개발 환경 구성  개발: 백엔드 / DB 스키마  개발: 사용자 화면 개발  개발: 어드민 화면 개발  완성 기능 리뷰  통합 테스트  개선 사항 추가 개발 |  |
| 11/15 | ~ | 발표자료 준비  사이트 런칭 |  |

## **2-2. 개발 언어 및 활용 기술**

|  |  |  |
| --- | --- | --- |
| **항목** | **적용 대상** | **비고** |
| Spring Boot | Backend |  |
| Vue.js | Frontend |  |
| Database | DB | 미정 |
| Hadoop | 분산 처리 |  |
| Python | 데이터분석 | pyspark |
| Spark | 데이터 분석, 처리 | scala |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## **2-3. 예산**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **항목** | **상세** | **수량** | **단가** | **비용(원)** |
| **서버** | **S3 서버 (AWS)** | **30** | **8970** | **8970** |
| **서버** | **분산 클러스터** | **-** | **-** | **-** |
| **강의** | **따라하며 배우는 도커와 CI환경** | **1** | **24200** | **24200** |
| **강의** | **Hadoop 3.2.1 and Big data** | **1** | **16500** | **16500** |
| **도서(이북)** | **스파크를 다루는 기술 Spark in Action** | **1** | **28000** | **28000** |
| **도서** | **빅데이터 시각화** | **1** | **25200** | **25200** |
| **도서(이북)** | **Do it! 자바 프로그래밍 입문** | **1** | **16500** | **16500** |
| **도서** | **Programming in Scala 4/e** | **1** | **45000** | **45000** |
| **도서(이북)** | **스파크 완벽 가이드** | **1** | **38400** | **38400** |
| **도서(이북)** | **9가지 사례로 익히는 고급 스파크** | **1** | **20800** | **20800** |
| **합계** |  |  |  | **228,900** |

**3. 분석 및 설계**

## **3-1. 요구사항 정의**

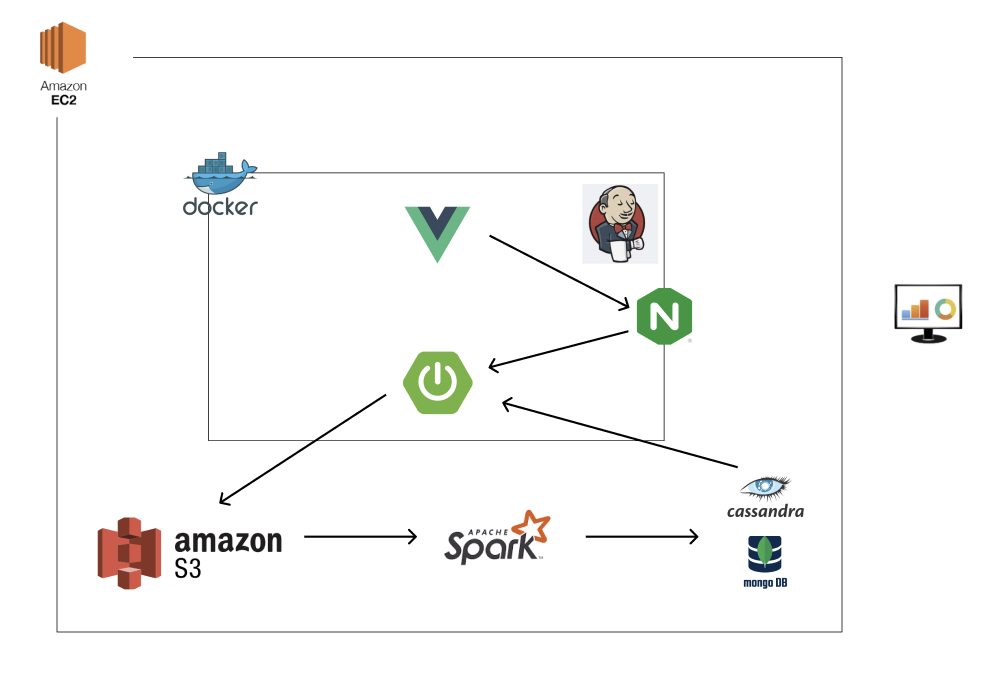
* 필수 구현 기능

1. 데이터셋 feature 추출 (예상 스킬 : Spark, Scala)
2. 분산 처리를 통해 빅데이터셋 처리 성능 확보

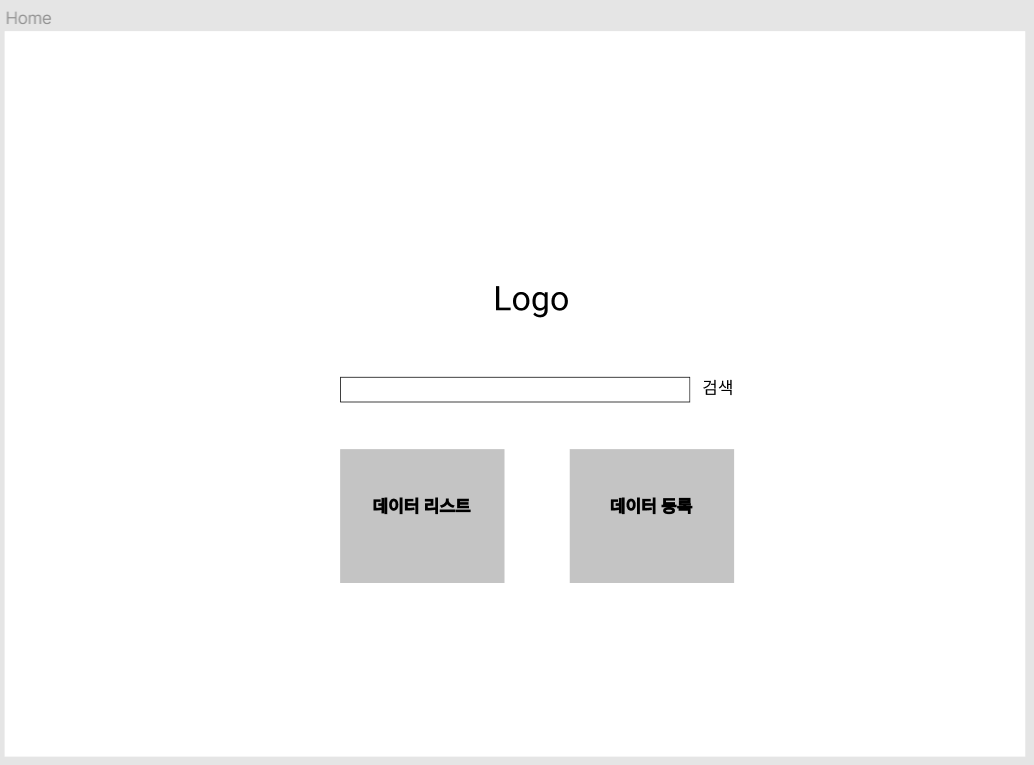
|  |  |  |
| --- | --- | --- |
| **ID** | **요구사항명** | **설명** |
| Req. 1. | 파일 업로드 기능 (데이터셋 등록) | 추출하고자 하는 데이터를 등록(업로드)한다. |
| Req. 2. | 데이터셋 분석 | 업로드 한 데이터 분석 (확장자, 데이터 타입, etc) |
| Req. 3. | 히스토그램 출력 | column 별 value의 도수 분포 |
| Req. 4. | unique value 개수 출력 | 각 column별 unique value 개수 출력 |
| Req. 5. | 데이터 리스트 페이지 | 등록된 데이터들을 볼 수 있는 페이지 |
| Req. 6. | 커스터마이징 | kaggle의 범위 설정 기능  필터 설정 |
| Req. 7. | 기초 통계량 출력 | 평균, 중위값, 최대, 최소, 표준편차 |
| Req. 8. | 데이터 검색 | 찾고자 하는 데이터, 키워드 검색 |

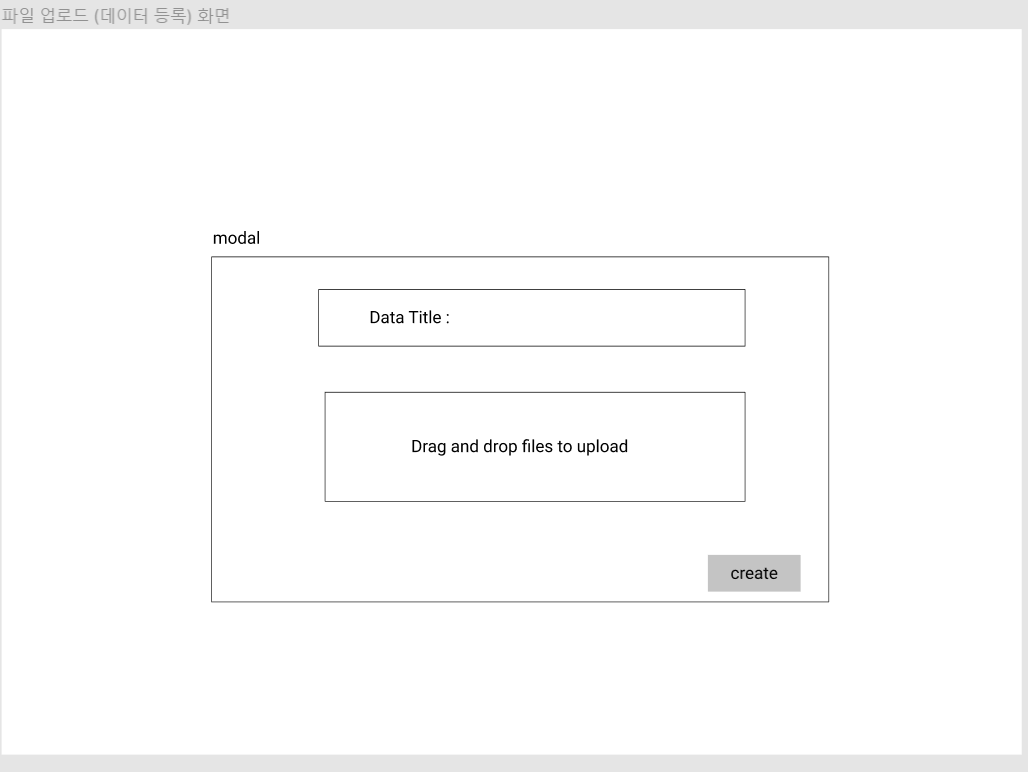
**3-2. 애플리케이션 아키텍쳐**

1) 다이어그램

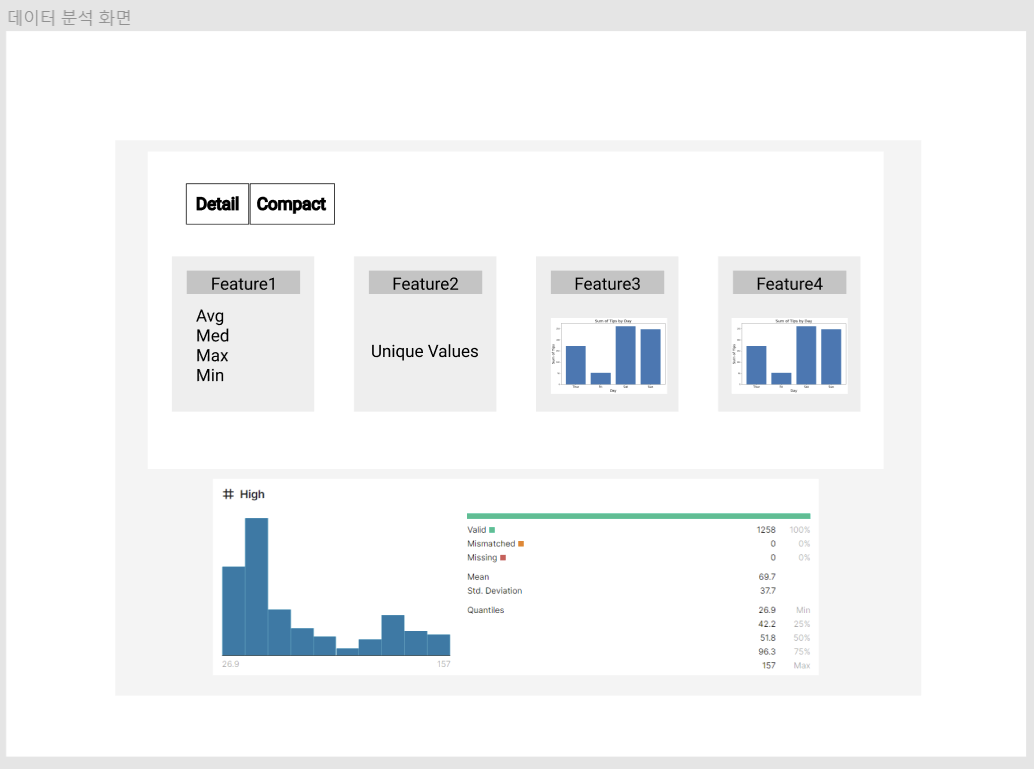


2) 화면 예시 (wire frame)





1번 메인페이지 : 데이터 등록, 2번 파일 업로드 모달 데이터 리스트 , 검색



3번 데이터 시각화 결과 페이지 4번 데이터 시각화 리스트