

DMZ

국내 영화 및 드라마 아카이브 / 지도

A602 김예진 김경륜 박성준 유승윤 정예지

1. 기획

기획 배경 – 관계자 인터뷰
기존 서비스 조사
기획 의도

2. 서비스

컨셉 및 기능
시연

3. 개발

아키텍처
데이터 수집

아카이브 – 데이터 처리
지도 – 관계 계산 및 렌더링
프론트엔드
기술스택

4. 마무리

일정 및 산출물 관리
팀원 및 역할 소개

DMZ

01 기획

대상

5년차 이상 드라마 기획 프로듀서

불편사항

장르와 키워드별로 작품을 검색하기 어려움
작품을 한눈에 보는 기능이 제공되지 않음
연도별 정렬이 제공되지 않아 체크하기 불편함

요구사항

모든 드라마 정보를 한 페이지에서 보이는 기능이 있으면 좋겠음
작품 검색 시 다양한 필터링을 제공하고 연도별 정렬 기능이 제공되면 좋겠음
성별/연령대를 설정해 배우를 검색하는 기능이 제공되면 좋겠음
제작진을 검색하는 기능이 제공되면 좋겠음

WATCHA PEDIA

국내 드라마 검색 시 사용

줄거리, OTT, 리뷰를 보기 위해 주로 이용
찜한 리스트 정보를 보기에 가장 편함
전체 검색 리스트 뷰가 불편하고, 장르 검색이 안 됨

NAVER

대략적인 정보를 검색하기에 용이
한 번에 한 작품만 검색 가능
필터링 기능이 없어 불편함



왓챠피디아에서 서비스되지 않는 작품 검색

보고싶어요

▼ 담은 순



광해, 왕이 된 남자
예상 ★ 2.8



미드나잇 선
예상 ★ 3.2



행복까지 30일
예상 ★ 3.7



멜로렌드 드라이브
예상 ★ 2.5



오직 사랑하는 이...
예상 ★ 2.8



스포트라이트
예상 ★ 3.6



위 호스
예상 ★ 3.3



마이 걸
예상 ★ 3.4



다양한 검색 필터링 기능 제공
한국 작품을 찾는 것에 한계가 있음
한국어로 정보를 찾기 힘들

Advanced Search

Movies, TV & Video Games

Want comedies from 1970 with at least 1000 votes and average rating of 7.5?

[Advanced Title Search](#) ▶

People

Want women born between 1950 and 1990 who have been nominated for Oscars?

[Advanced Name Search](#) ▶

Collaborations

Want a list of titles in which both Brad Pitt and George Clooney appeared?

[Search Collaborations](#) ▶

Search Within a Topic

Movies, TV and Video Games

Plot ▼

E.g. horses

SEARCH

People

Biographies ▼

E.g. arrested

SEARCH

Browse titles by: [Genre](#) | [Keyword](#)

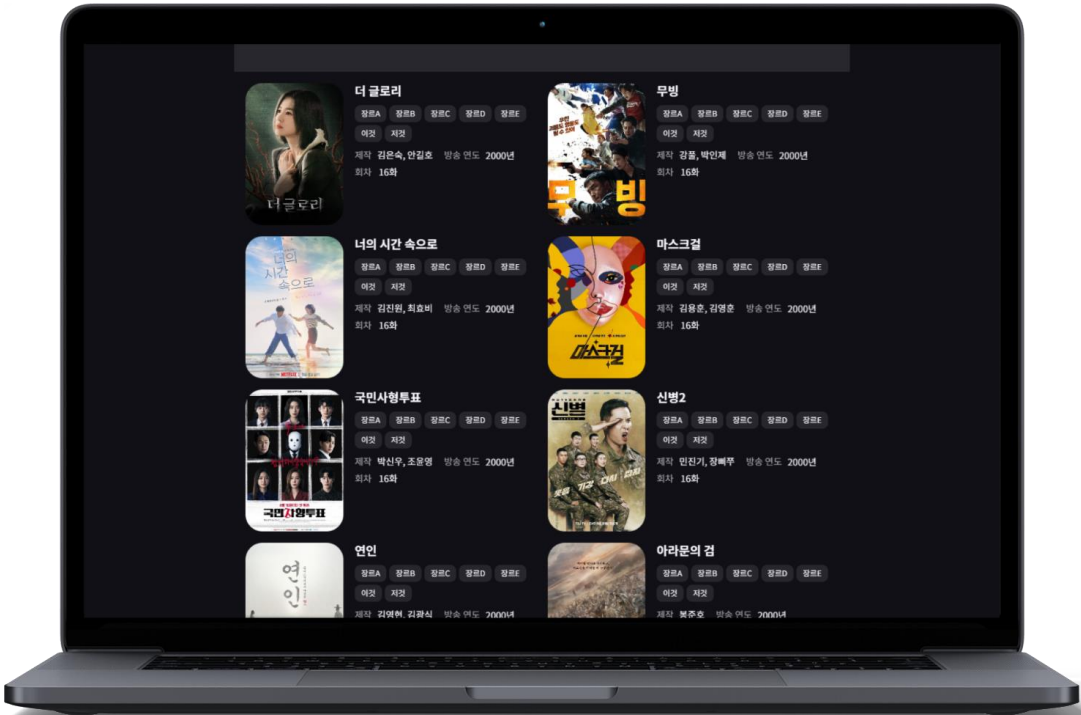
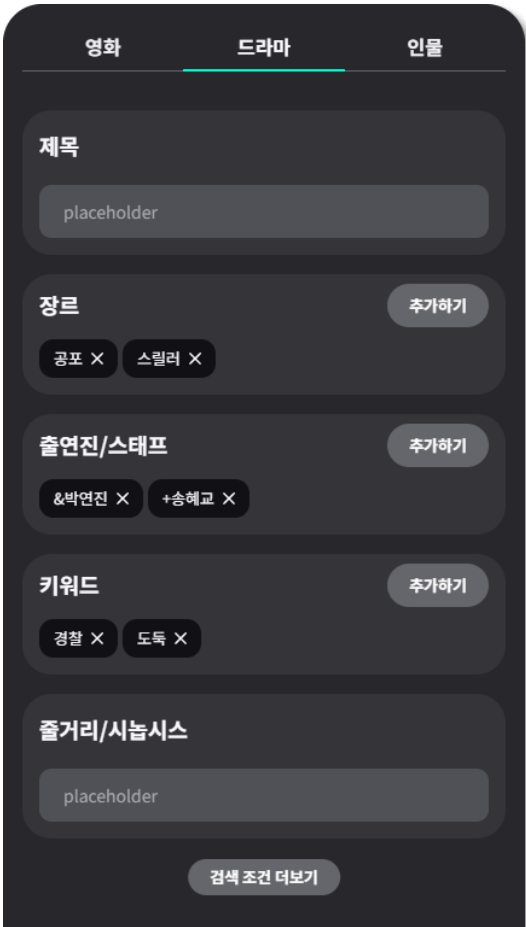


영화, 드라마 정보 API 제공
위키 프로젝트로, 사용자가 직접 정보 업데이트
필터링 기능이 제공되지 않고, 연도별 정렬이 없어 불편함

DMZ

국내 드라마 · 영화의 새로운 지도

가장 방대한 국내 드라마 · 영화 · 인물 검색



한눈에 보는
미디어 - 인물 관계 지도



대상

영화/드라마 작품 및 인물(배우/제작진) 검색을 원하는 업계 관계자
미디어를 연구하는 전문 연구원
영화/드라마를 검색하고 싶은 일반 사용자

기능

아카이브 - 영화/드라마 작품 및 인물 검색
지도 - 시각적으로 작품/인물 관계를 탐색

기대 효과

다양한 필터링을 이용해 작품과 인물 검색 제공
영화 및 드라마를 시각적으로 표현하고 직관적으로 탐색하는 사용자 경험 제공
국내 유일 제작진/배우 통합 검색 서비스 제공
대한민국 미디어 작품을 아카이빙해 문화 유산 보관, 모든 사용자에게 공개

DMZ

02 서비스

한눈에 보는
미디어 & 인물 정보

국내 미디어
데이터 검색 시스템

지도로 찾는
미디어 - 인물 관계



아카이브



지도

한눈에 보는
미디어 & 인물 정보

국내 미디어
데이터 검색 시스템

지도로 찾는
미디어 - 인물 관계



아카이브



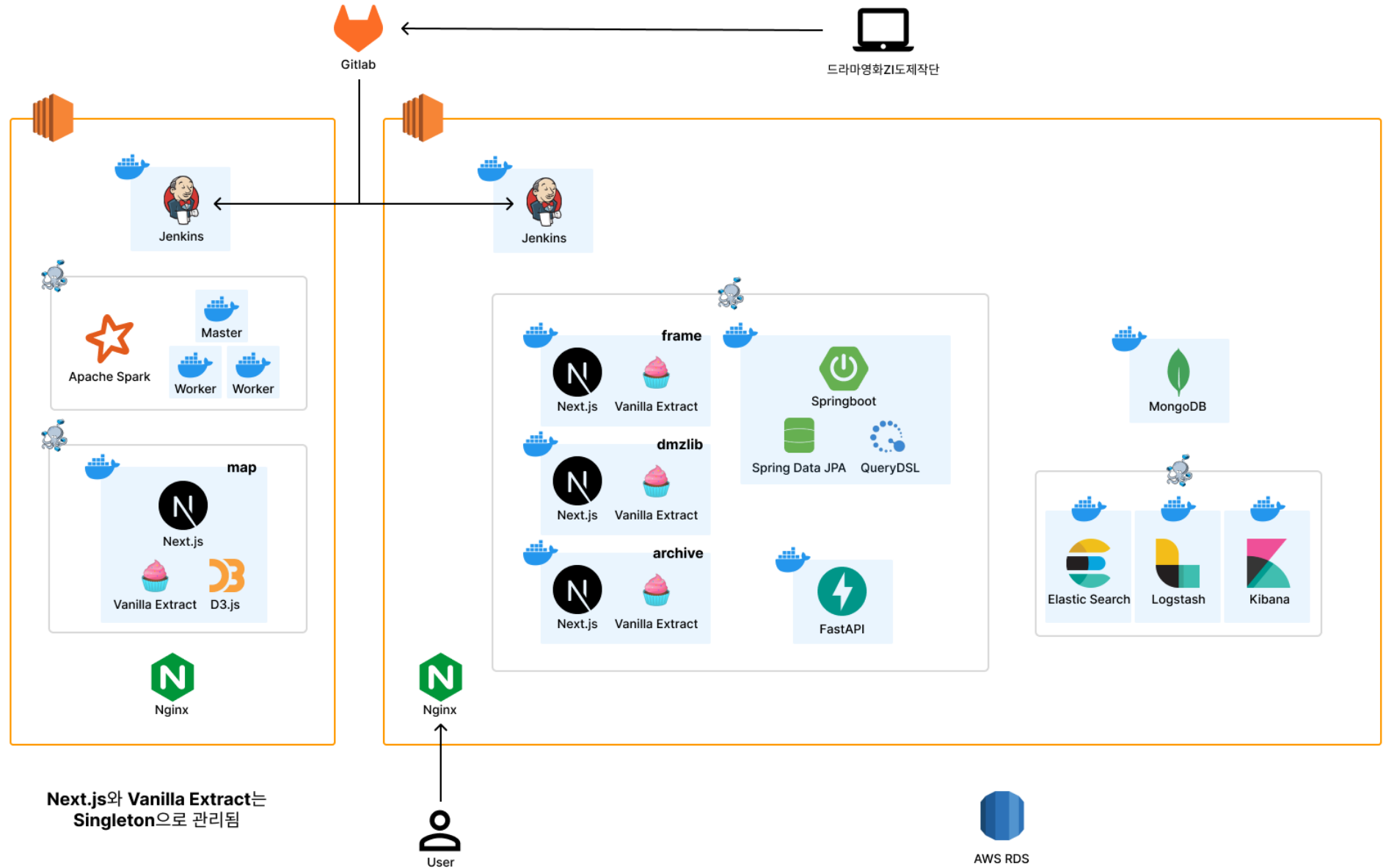
지도

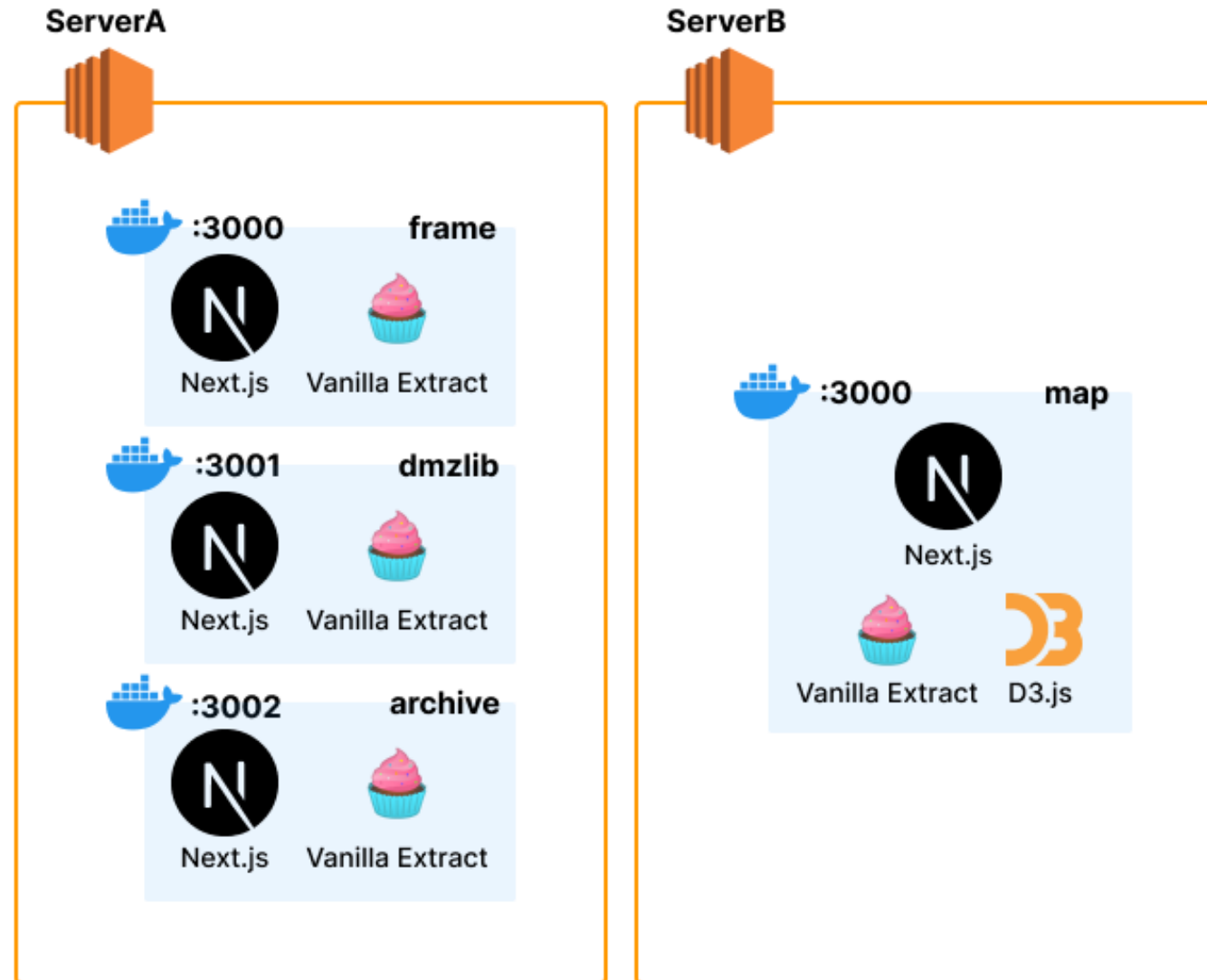
DMZ

03 개발

03-1 아키텍처

시스템 아키텍처





수집 항목

작품 – 한국 영화 및 드라마
인물 – 배우 및 제작진

데이터 수집처 후보

한국영상자료원 – 한국 영화 및 영화인 데이터 제공
TMDB – 한국 영화 및 드라마, 인물(배우/제작진) API 제공
위키 – 위키피디아, 나무위키, 방송사 공식 홈페이지

기술 및 기능 요구사항

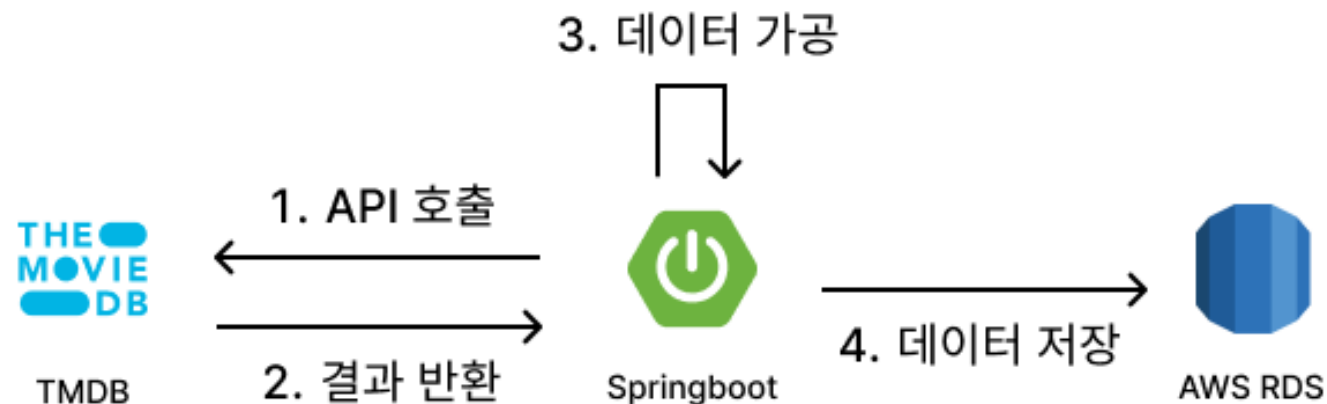
서비스 마감 기한 엄수 – 서비스 규모에 맞게 데이터 수집 및 정제
지속적 업데이트 – 새로운 인물 및 작품 공개 시 서비스에 데이터 업데이트

데이터 수집 실험

크롤링 – 한국영상자료원/위키를 이용해 정보 수집 및 전처리 수행
API를 사용한 파이프라인 구축 – TMDB에서 제공하는 API 이용
실험 진행 – 2023.09.06 ~ 2023.09.08

실험 결과

크롤링 – 별도의 크롤링 서버 구축 필요, 개월 단위의 업무 예상
API를 사용한 파이프라인 구축 – 빠르게 데이터 수집 가능



DMZ

03 개발 - 아카이브

기술 및 기능 요구사항

개발 마감 기한 엄수 – 2023-10-06

도메인 기술 적용 – 분산 기술

문제 사항

분산 기술을 경험한 개발자가 **없음**

분산 기술 학습 및 구현에 **시간이 많이 걸림**

데이터 검색에 최적화된 기술을 선정하기 **힘듦**

해결 방안

두 가지 기술을 이용한 **동시 개발 진행**

QueryDSL을 이용한 개발 – **분산 X**, 빠른 개발 **O**

Elasticsearch를 이용한 개발 – 분산**O**, **기한 엄수 모름**

SpringBoot를 이용한 검색 API 서버 개발

QueryDSL을 이용해 복잡한 필터링도 검색 가능하게 함

3회 이상의 최적화 - 커버링 인덱스 적용, Entity → DTO 조회, 조인 테이블 최소화

237% 빨라진 검색 속도 개선으로 성능 향상에 성공 - 28.980초 → 8.587초

동적 쿼리

총 233,737 rows

28.980초 소요

Entity → DTO 조회

단순 조회 시 발생하는 성능 이슈 발견

필요한 컬럼만 DTO로 만들어 조회

대량 데이터의 조회에 최적화됨

커버링 인덱스

인덱스 내부에서

SELECT 절 내부 항목을 빠르게 조회

조인 테이블 최소화

필터링 유무에 따른 불필요한 테이블 조인 현상 발견

필요한 테이블만 사용해 검색에 사용

커버링 인덱스: 쿼리를 충족시키는 데 필요한 모든 데이터를 갖는 인덱스

ElasticSearch와 FastAPI를 이용한 검색 API 서버 개발

작품/인물 검색에 최적화된 인덱스 설계, 빠른 검색이 가능하도록 구현
Elasticsearch 사용으로 분산 기술 적용, 서비스 확장에도 유연하게 대처
계획 - Logstash로 지속적 데이터 업데이트 기능 수행

Data Migration

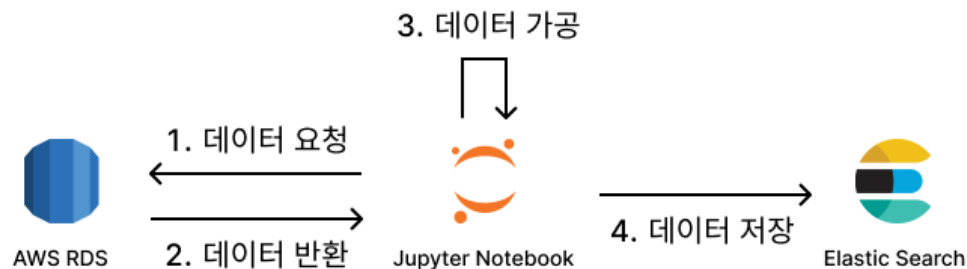
RDS에 저장된 데이터를 ES에 저장
RDMBS 형태의 데이터를 schemaless하게 변경

검색 쿼리 최적화

키워드, 줄거리 및 필터링 적용한 검색 가능

인덱스 설계

검색 최적화를 위한 8개의 인덱스 설계



DMZ

03 개발 – 지도 관계 계산 / 렌더링

기술 및 기능 요구사항

관계 계산 – 작품/작품, 작품/인물 간의 유사도 계산

지도 렌더링 – 약 30,000명 이상의 인물, 12,500여개의 작품을 지도 위에 표현

문제 사항

계산 복잡도 – 모든 작품/작품, 작품/인물 간의 관계를 만들면 **억 단위의 데이터가 생성됨**

계산 시간 – 전체 관계 계산 시 **약 1시간 소요** (작품/작품)

렌더링 – **N일 이상 소요**, 사용자 진입 시 지도 렌더링을 수행하면 **이용 불가**

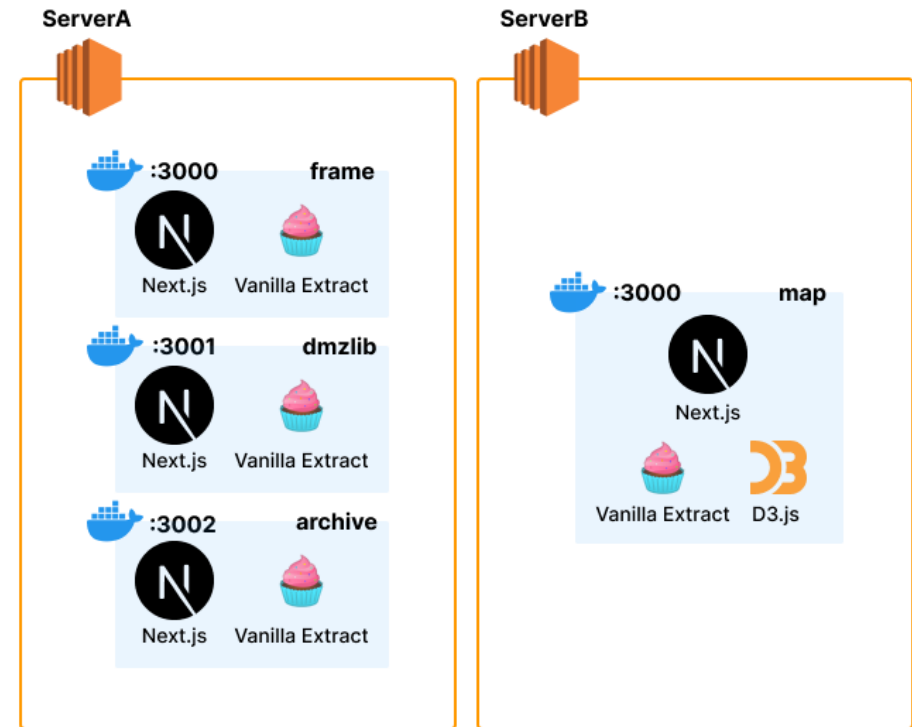
해결 방안

사전 관계 계산 – 대용량 데이터 연산에 최적인 Spark를 이용해 모든 관계를 사전에 계산

빠른 데이터 접근 – 계산된 관계를 MongoDB에 저장해 데이터 접근 속도 향상

모듈 페더레이션 적용 – MSA 형식을 프론트엔드 개발에 도입 (MFA)

2. 저장된 데이터를 활용해 그래프 그리기



Frontend



Backend



Database



Infra



Cooperation



Data



DMZ

04 마무리



DMZ

인프라: 정예지
데이터 수집: 김경륜
Elasticsearch: 김예진

아카이브

SpringBoot Server: 김경륜
FastAPI Server: 김예진

프론트: 박성준, 유승윤

지도

NextJS Server: 유승윤
Spark: 민병기

프론트: 유승윤

DMZ

국내 드라마 · 영화의 새로운 지도