



어떤 어려움이 있으신가요?  
상황에 맞는 판례를 찾아드립니다.

**PANGARM**

**판가름**

## 판가름이란?

---

사용자들에게 상황에 맞는 판례를 제공하는 서비스

## 핵심 기능

---

### 맞춤 판례 서비스:

사용자가 상황을 설명하면, 우리의 시스템이 이를 분석하여 해당 상황에 관련된 판례를 추천합니다. 사용자의 상황을 이해하고 필요한 정보를 제공하여 법률적 문제를 해결하는데 도움을 줍니다.

### 맞춤 뉴스 서비스:

사용자의 검색 및 활동 기록을 기반으로 개인화된 뉴스 추천을 제공합니다. 이를 통해 사용자는 자신의 관심사에 맞는 최신 뉴스를 발견할 수 있습니다.



---

어떤 기능을 제공하지?

# 페르소나

## 정보

이름 : 김관우

역할 : 대학생

나이 : 21세

직업 : 편의점 아르바이트생

특징 : 법을 잘 모름, MZ답게 글을 잘 안 읽음

## 문제 상황

관우는 대학생으로서 학업과 병행하면서 편의점에서 아르바이트를 하고 있습니다. 최근에는 근로 시간과 급여 지급, 휴게 시간 등과 관련된 노동 규정에 대해 의문을 가지고 있습니다.

## 목표

관우는 편의점 아르바이트를 하면서 자신의 노동 권리를 보호하고 싶어 합니다. 그는 근로 시간, 급여, 휴게 시간 등의 문제에 대한 법적인 지식을 습득하여 자신의 권리를 지키고 불공정한 처우를 방지하고 싶어 합니다.



# 판례에서는 어떤 기능을 제공할까?

## 서술형 판례 검색

사용자가 본 판례들을 제공합니다.

## 판례 요약본 제공

이해하기 쉽게 요약된 판례를 제공합니다.

## 판례 상세본 제공

상세한 판례를 제공합니다.

## 판례 북마크 제공

원하는 판례를 북마크해 후에 찾아볼 때 편합니다.

## 판례 히스토리 제공

사용자가 본 판례들을 제공합니다.

## 검색 히스토리 제공

사용자가 검색한 히스토리를 제공합니다.



# 뉴스에서는 어떤 기능을 제공할까?

## 최신 뉴스 제공

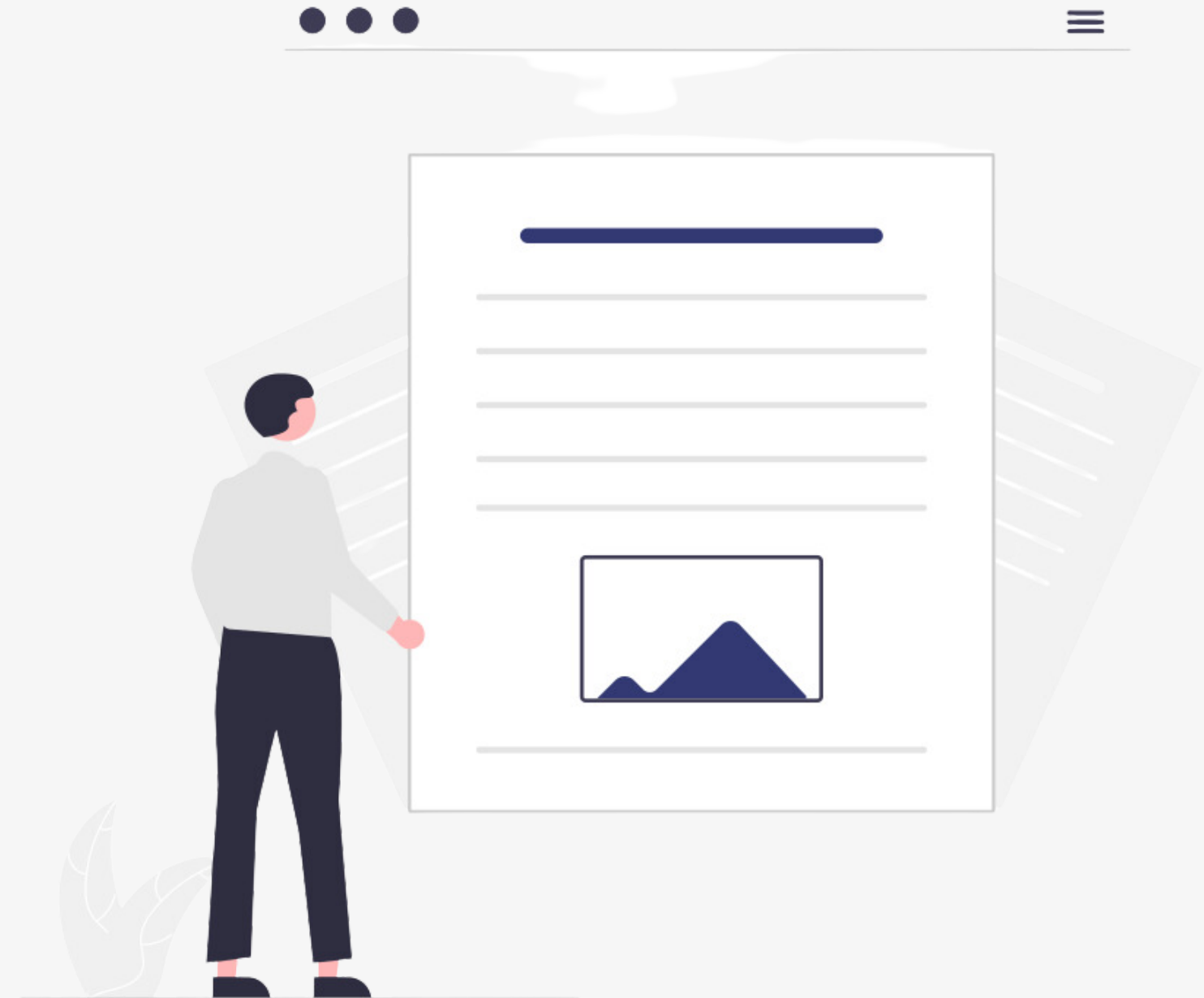
법 관련 최신 뉴스들을 제공합니다.

## 카테고리 별 구독

카테고리 별 뉴스를 제공합니다.

## 뉴스레터 구독 서비스

구독한 뉴스들을 메일로 보내줍니다.



---

# 시연 영상

# 시연





---

어떻게 기능을 제공하지?  
검색

# 검색 기능

## Semantic Search

단어와 구문의 의미를 해석하는 검색 엔진 기술

## Ko-sentence-transformers

한국어 버전의 자연어 처리 모델

---

## Keyword Search

## Elastic Search

분산형 RESTful 검색 및 분석 엔진

---

## 대용량 처리

## Hadoop & Spark

대량의 자료를 처리할 수 있는 큰 컴퓨터 클러스터에서 동작하는 분산 응용 프로그램을 지원하는 프레임워크

# 자연어는 어떻게 처리하는가?

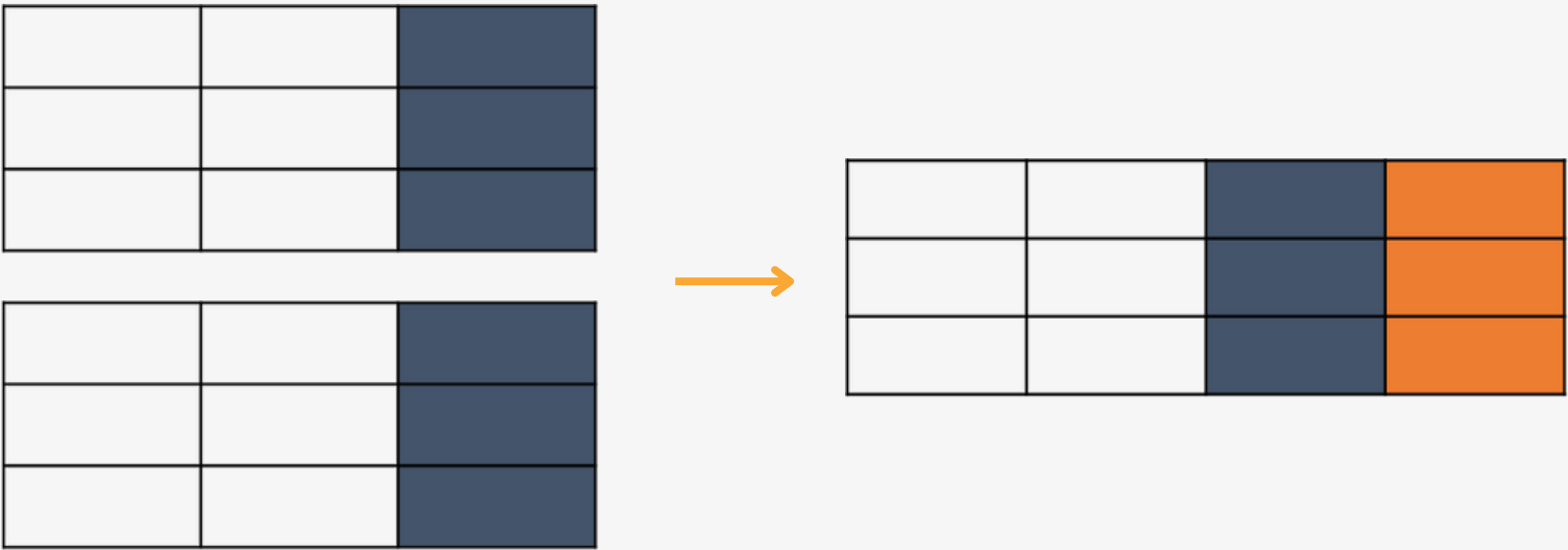
## 문자열 토큰화

우리는 인공지능을 만들기 위해서 공부한다.  
→ 우리/는/인공지능/을/만들/기/위해서/공부/한다.

## 벡터화

우리/는/인공지능/을/만들/기/위해서/공부/한다. →


## 유사도 비교

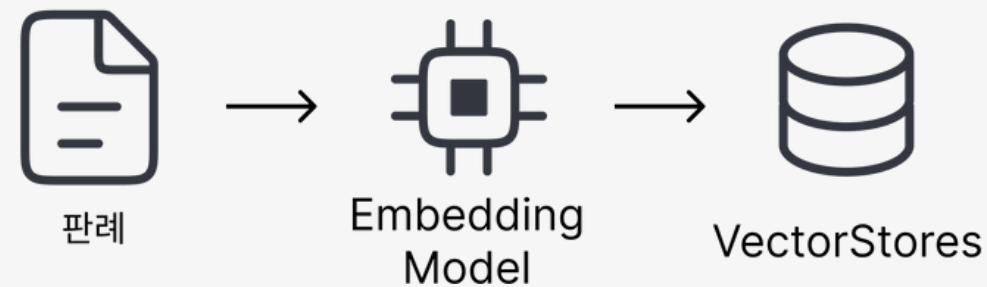


---

# 트러블 슈팅 검색

# 검색기능 (기존 방식)

## 판례 저장 흐름도



이미지 분석에서 사전학습 모형을 사용하는 것처럼 텍스트 데이터를 미리 학습한 사전학습 모델

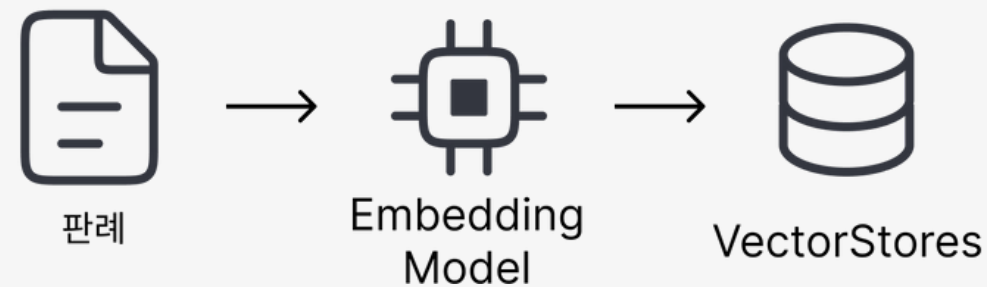
### 특징

모델 사이즈가 상당히 큼.(기본적으로 GPU 레벨에서 작동)  
자연어에는 강하지만 **특정 영역(법, 의료)에서 약함.**

**병목현상 및 유사도 측정 문제 발생**

# 검색기능 (1차 해결)

## 판례 저장 흐름도



### KO-SENTENCE-TRANSFORMERS

KORNLU 데이터셋으로 한국어 사전학습 모델을 파인튜닝한 모델

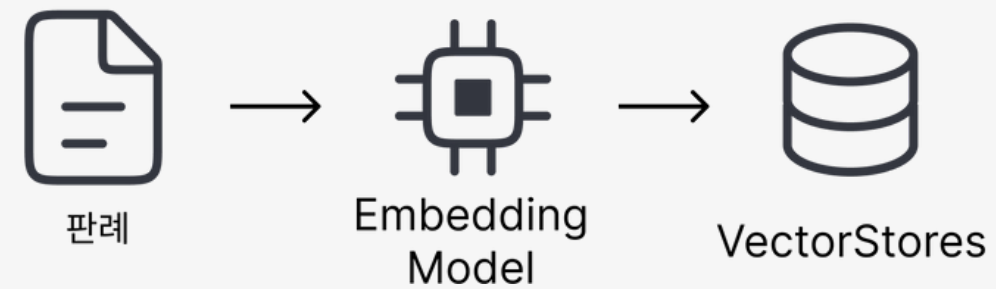
#### 특징

CPU 레벨에서 고속 처리 가능.

병목현상 및 유사도 문제 해결

# 검색기능 (1차 해결)

## 판례 저장 흐름도



**VECTORSTORES**

DATABASE

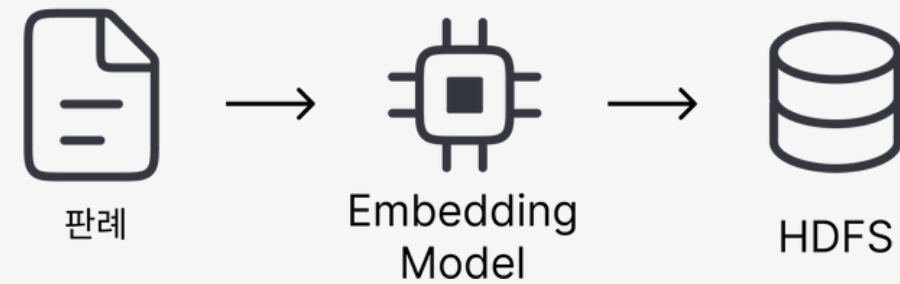
### 특징

관계형 데이터 베이스와 같은 특징(스케일 아웃 어려움)

판례 데이터를 저장하는데 한계가 존재

# 검색기능 (2차 해결)

## 판례 저장 흐름도



### HDFS

하둡이 실행되는 파일을 관리해주는 시스템

### 특징

데이터를 분산 저장하는데 용이(빅데이터 저장에 용이)  
슬레이브 노드의 일부가 고장 나도 데이터 손실을 방지 가능

**분산 저장으로 해결**



# 검색기능 (3차 해결)

## 응답속도 문제

8만개 MAP-REDUCE(유사도 계산 및 정렬) 대략 30초 소요

# 검색기능 (3차 해결)

## 응답속도 문제

8만개 MAP-REDUCE(유사도 계산 및 정렬) 대략 30초 소요

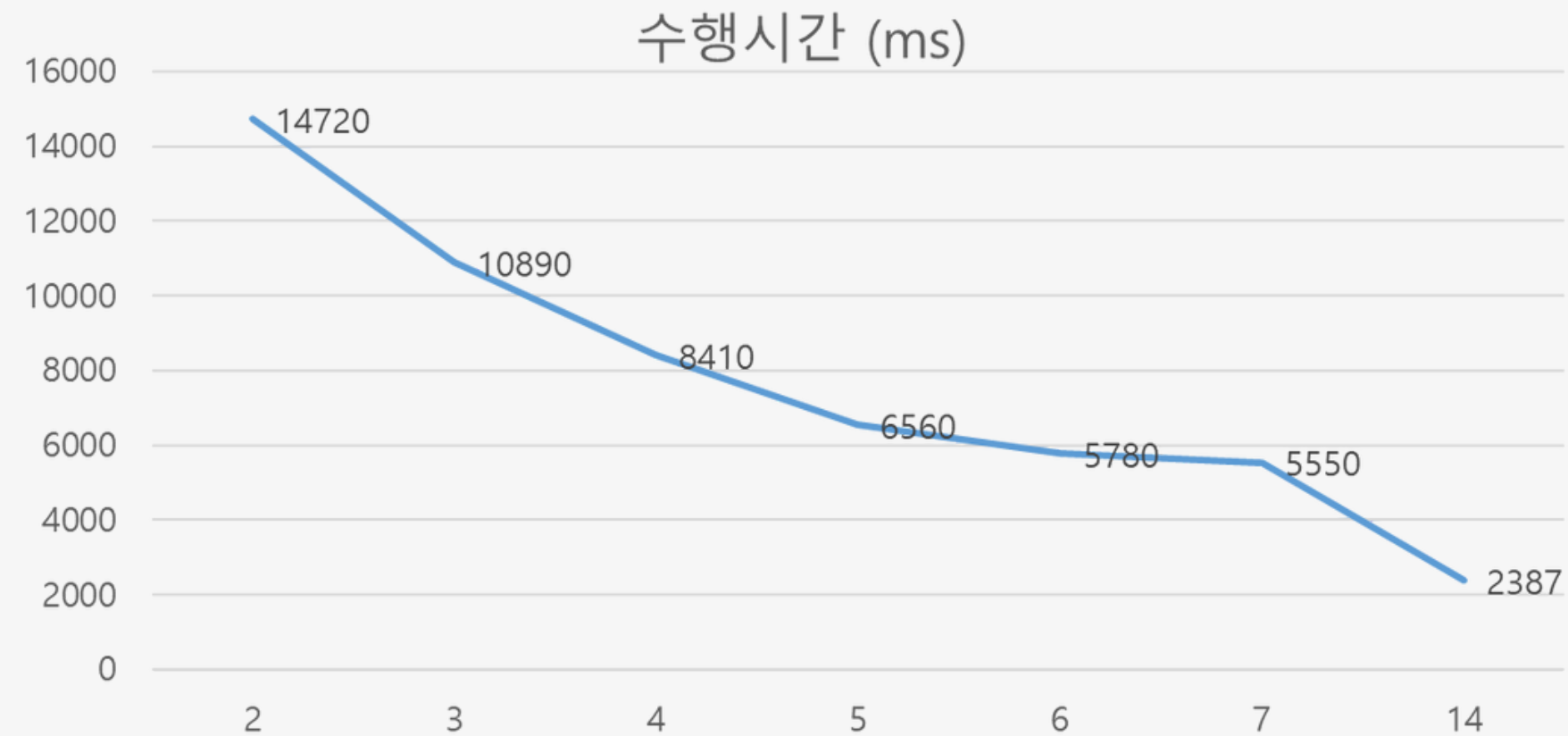


서버 2대 -> 14대 증축

# 검색기능 (3차 해결)

## 응답속도 문제

8만개 MAP-REDUCE(유사도 계산 및 정렬) 대략 30초 소요

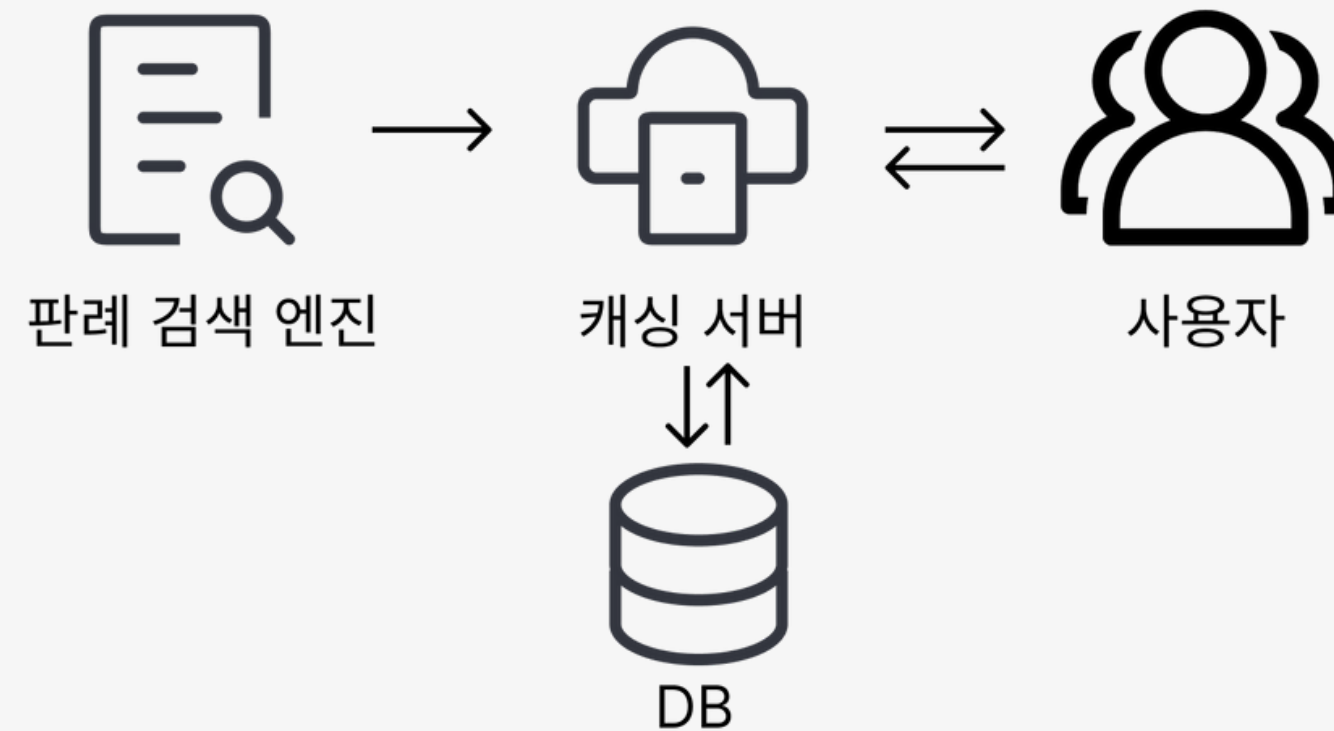


서버 2대 -> 14대 증축  
30 -> 2.3초 단축(1200% 단축)

# 검색기능 (4차 해결)

**아직 배고프다!**

캐싱 서버 도입



재 요청 시

2.3초 -> 0.5초

# 검색기능 (5차 해결)

한번 더!

TANSTACK QUERY 도입



판례 검색 엔진



캐싱 서버



사용자



DB

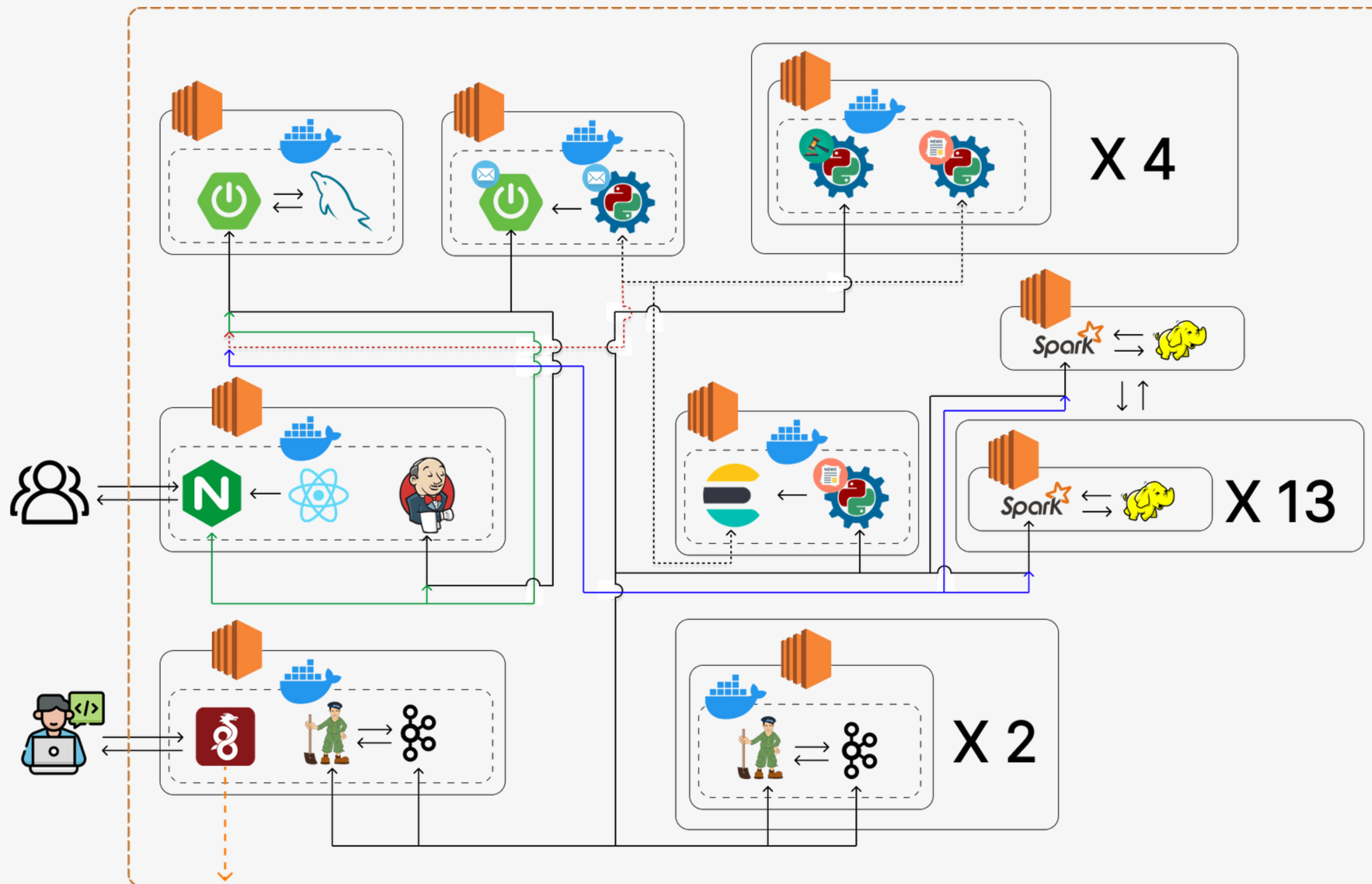
같은 요청이 있을 때 서버로부터 데이터를 다시 가져오지 않고 캐시된 데이터를 사용

재 요청 시  
서버 요청 X

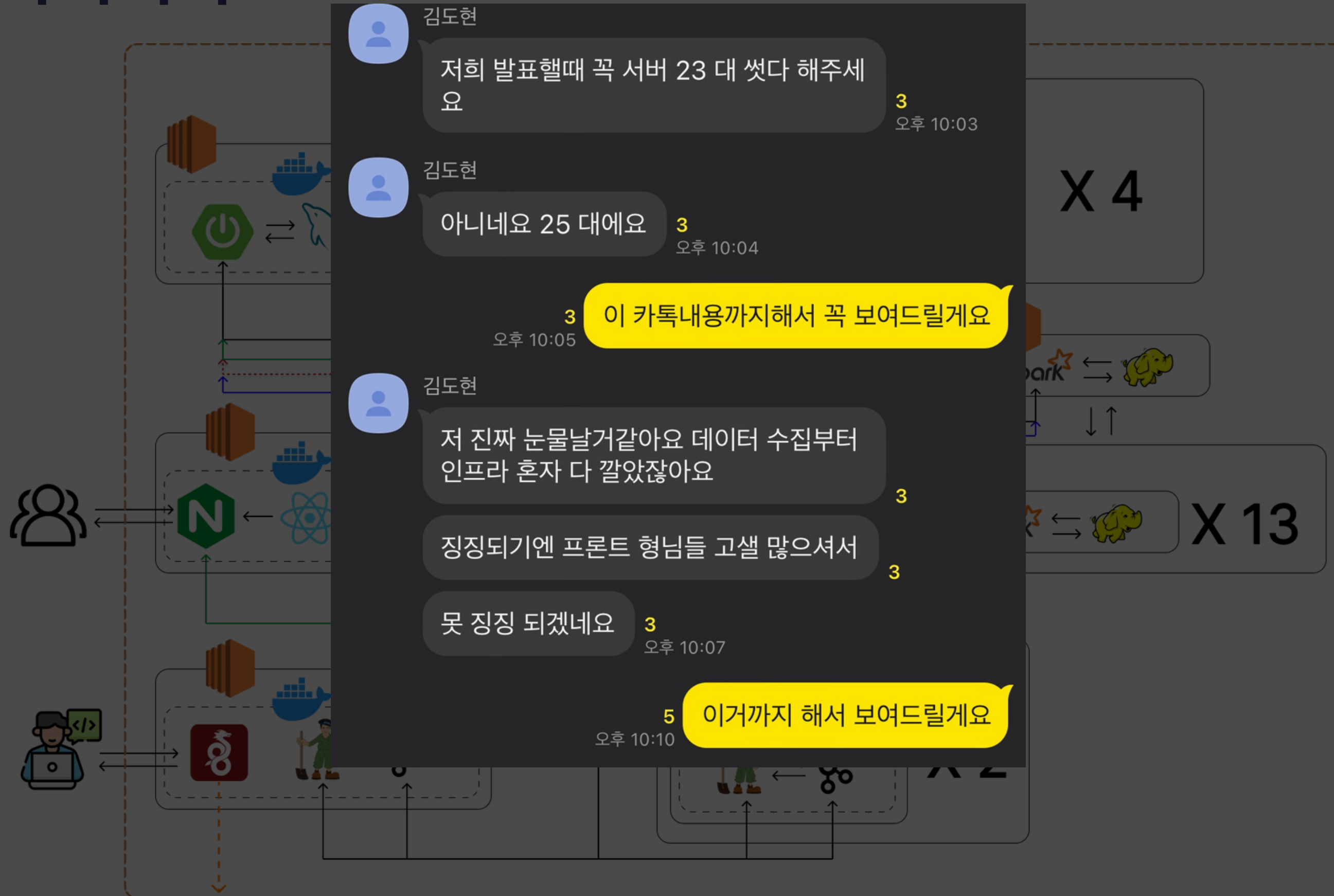
---

# 산출물

# 시스템 아키텍처



# 시스템 아키텍처





감사합니다.  
궁금한 점이 있으신가요?

판가름  
그럴수도 있조(A509)