



# KaDB

Kakao Developer Buddies

## CDC 중간 발표

# 목차

A table of contents

---

프로젝트 목표

선정 데이터 소개

사용 기술

요구사항 분석

서비스 아키텍처

# 팀원 소개

---



**윤건우**

컴퓨터공학과

201833937

PM, 자료조사



**강석훈**

컴퓨터공학과

201835605

발표



**김성민**

소프트웨어학과

201735815

PPT



**장현화**

컴퓨터공학과

201935334

자료조사



**채호기**

소프트웨어학과

201835535

PPT

# 프로젝트 목표



# 목표

---

1. 데이터베이스나 시스템에서 변경된 데이터를 감지하고 식별
2. 변경된 데이터가 각 데이터베이스에서 일치하는지 동기화 수행
3. 데이터의 변경 감지 및 식별에 대해 과정을 효율적으로 수행하여 성능을 최적화
4. 변경된 데이터에 대한 정보를 모니터링 하고 변경 사항을 추적하는 보고 제공



## 선정 데이터 소개



# 사용 예시

서울 특정 지역(석촌호수 인근)의 인구 밀집도 표시





# 서울시 실시간 도시데이터

## 서울 실시간 도시데이터란?

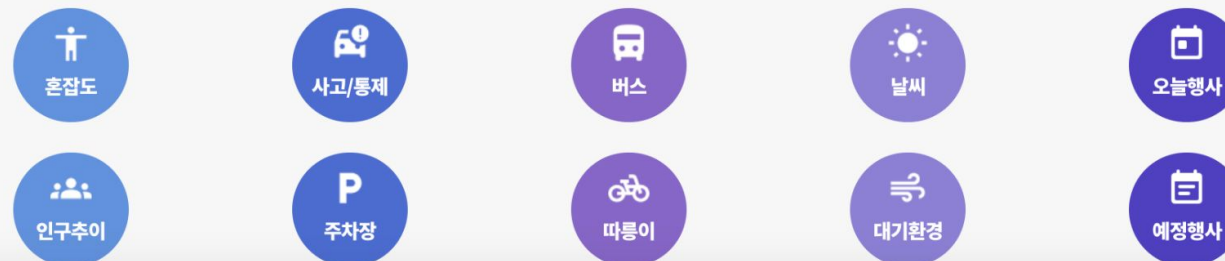
서울시 주요 113장소에 대한 분야별(실시간 인구, 도로소통, 대중교통, 날씨/환경, 문화행사) 실시간 정보들이 융합된 데이터

※ 각 분야의 데이터를 별도 수집/연계하지 않고 통합하여 활용 가능합니다.

서울 실시간 도시데이터 바로가기 >



서울 실시간 도시데이터





# 데이터 예시 - 샘플 코드

## Open API 샘플 코드

[!\[\]\(919a2cb85b99741a73c0c31a427236a8\_img.jpg\) 가이드 다운로드](#)[Java](#)[Javascript](#)[cURL](#)[Python](#)[Node.js](#)

```
/* Python3 샘플 코드 */  
  
import requests  
  
url = 'http://openapi.seoul.go.kr:8088/sample/xml/CardSubwayStatsNew/1/5/20220301'  
  
response = requests.get(url)  
print(response.content)
```

# 데이터 예시 - 요청 패러미터

## > 요청인자

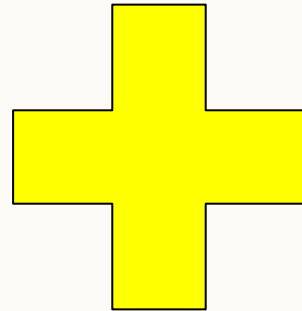
변수명	타입	변수설명	값설명
KEY	String(필수)	인증키	OpenAPI에서 발급된 인증키
TYPE	String(필수)	요청파일타입	xml:xml
SERVICE	String(필수)	서비스명	citydata
START_INDEX	INTEGER(필수)	요청시작위치	정수 입력 (페이징 시작번호 입니다: 데이터 행 시작번호)
END_INDEX	INTEGER(필수)	요청종료위치	정수 입력 (페이징 끝번호 입니다: 데이터 행 끝번호)
AREA_NM	STRING(필수)	핫스팟 장소명	장소명 or 장소코드 입력(서울시 주요 113장소 목록.xlsx 파일 참고)

# 데이터 예시 - 출력값

## > 출력값

No	출력명	출력설명
공통	list_total_count	총 데이터 건수 (정상조회 시 출력됨)
공통	RESULT.CODE	요청결과 코드 (하단 메세지설명 참고)
공통	RESULT.MESSAGE	요청결과 메시지 (하단 메세지설명 참고)
1	AREA_NM	핫스팟 장소명
2	AREA_CD	핫스팟 코드명
3	LIVE_PPLTN_STTS	실시간 인구현황
4	AREA_CONGEST_LVL	장소 혼잡도 지표
5	AREA_CONGEST_MSG	장소 혼잡도 지표 관련 메세지
6	AREA_PPLTN_MIN	실시간 인구 지표 최소값
7	AREA_PPLTN_MAX	실시간 인구 지표 최대값
8	MALE_PPLTN_RATE	남성 인구비율(남성)
9	FEMALE_PPLTN_RATE	여성 인구비율(여성)
10	PPLTN_RATE_0	0~10세 인구비율
11	PPLTN_RATE_10	10대 실시간 인구비율
12	PPLTN_RATE_20	20대 실시간 인구비율

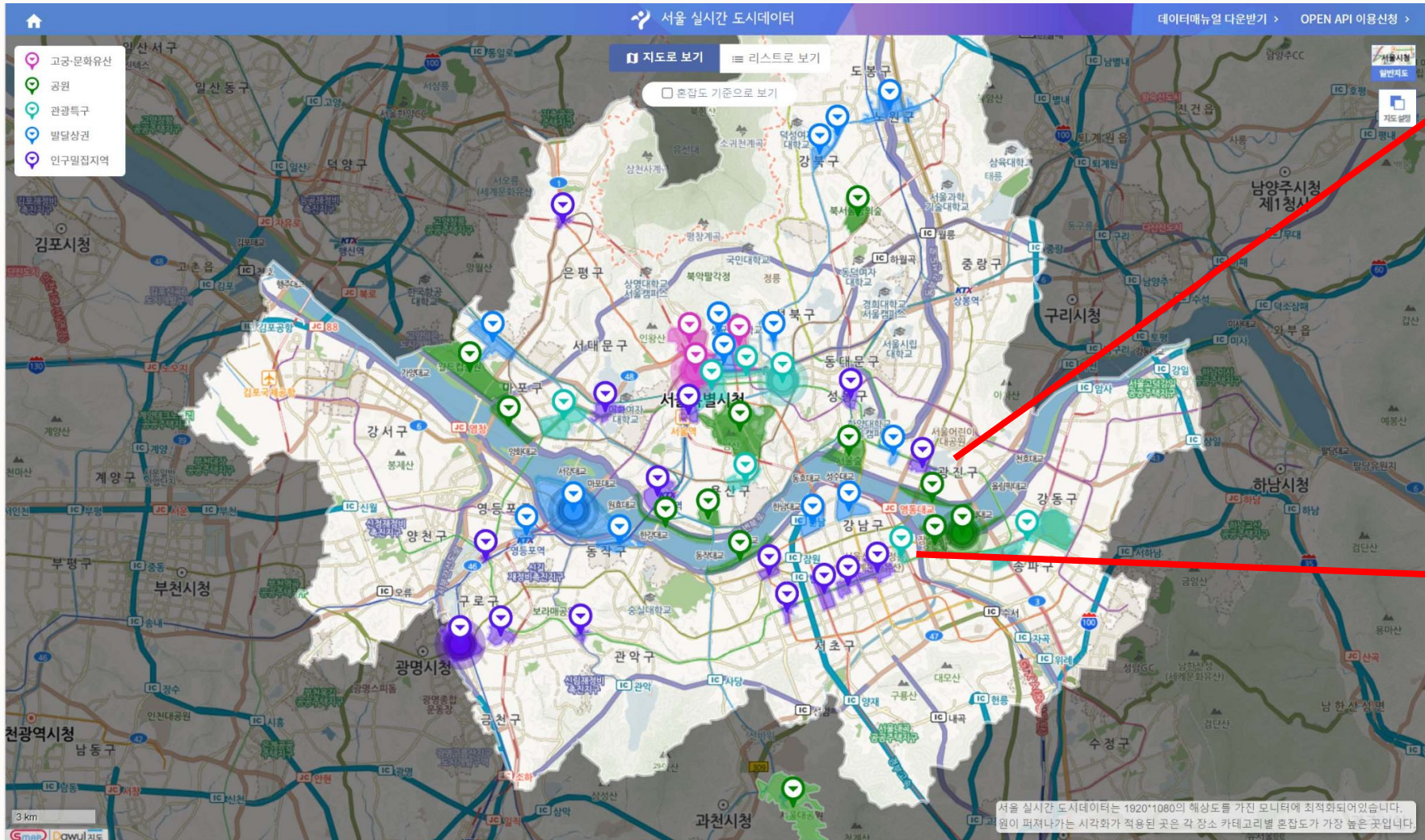
인구 밀집도



행사



# 기대효과



축제 지역 좌표 확인  
축제 지역 인근 혼잡도 파악

서울 명소들의 혼잡도 파악



# 사용 기술 소개



# 사용 기술 소개

---



# 사용 기술 소개

---



## Apache Airflow

워크플로우 자동화 및 스케줄링을 위한 오픈 소스 도구

데이터 처리, ETL(Extract, Transform, Load) 작업, 데이터 마이그레이션 등을 효율적으로 관리

복잡한 작업을 줄이고, 워크플로우의 스케줄링, 실행, 모니터링, 로깅 등을 관리

# 사용 기술 소개

---



**MySQL**

오픈 소스 관계형 데이터베이스 관리 시스템(RDBMS)

Oracle이 관리 및 지원

# 사용 기술 소개

---



## MongoDB

크로스 플랫폼 Document 지향 NoSQL DBMS

Binary JSON 사용



# 사용 기술 소개

---



## Apache Kafka

오픈소스 분산형 스트리밍 플랫폼

대용량의 데이터 스트림을 처리하고 저장 가능

실시간 로그 데이터 처리, 이벤트 소싱(Event Sourcing), 모니터링 및 경보 시스템, 지표 수집과 분석 등에 사용



## Debezium

Kafka Connect 기반의 오픈소스 플러그인

로그 기반의 CDC(Change Data Capture)기능을 지원

각 DB 테이블의 row 변화 발생시 이를 캡처하여 변경 사항에 대응할 수 있도록 하는 DB동기화 기능 지원



# elasticsearch

## Elastic Search

시간 분산 검색 및 분석 엔진

대량의 데이터를 신속하게 검색하고 분석 가능

로그 및 이벤트 데이터 분석, 실시간 알림 시스템, 비즈니스 인텔리전스(BI) 등에 사용

# 사용 기술 소개

---



## Pandas

데이터 조작 및 분석을 위한 **Python** 라이브러리

표 형식의 데이터나 다양한 형태의 데이터를 다룰 수 있게 해줌



## Seaborn

**Matplotlib** 기반의 **Python** 데이터 시각화 라이브러리

통계 그래픽을 그리기 위한 고수준 인터페이스를 제공



요구사항 분석



# 요구사항 명세

---

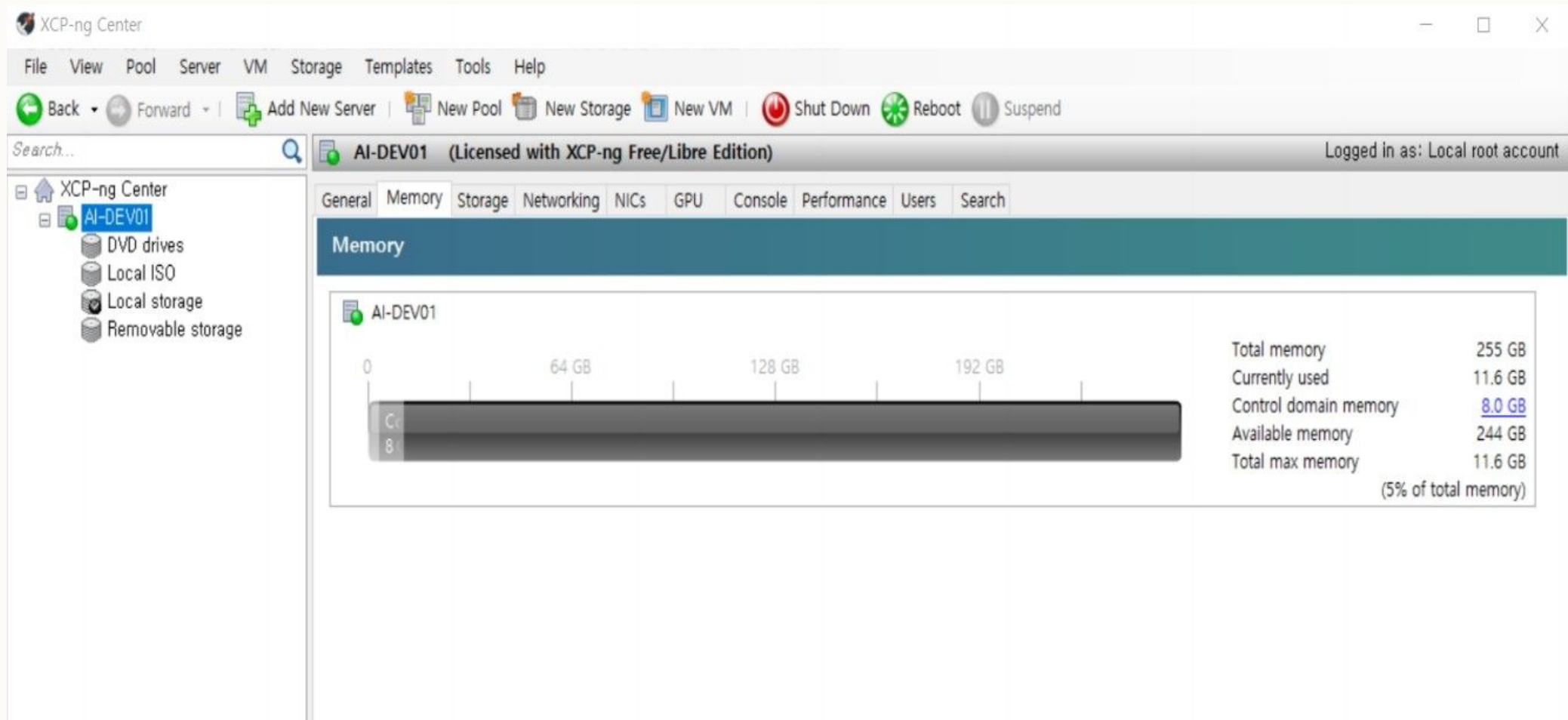
1. 행사 진행시에 해당 위치에 대한 복잡도를 표시할 수 있도록 한다.
2. 전처리한 데이터를 외부 사용자가 **API**를 활용해서 받을 수 있게 한다.

# 서비스 아키텍처



# 서비스 아키텍처

개발환경

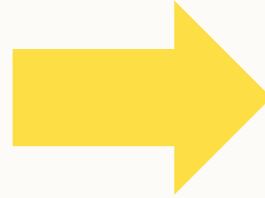


가천대학교 On-Premise Server

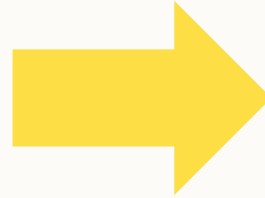
# 서비스 아키텍처

개발환경

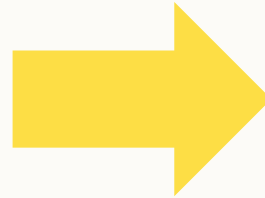
Source DB



CDC



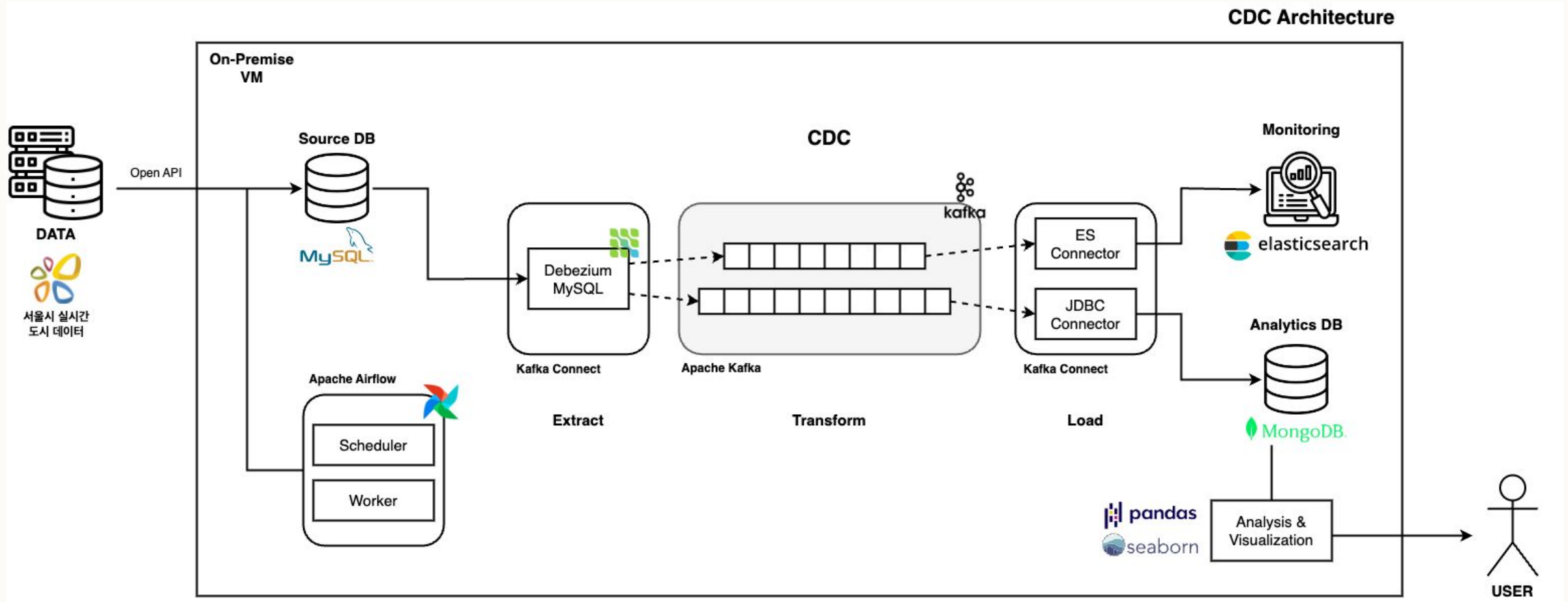
Target DB



# 서비스 아키텍처

아키텍처 구성

## ➔ 아키텍처 구성도

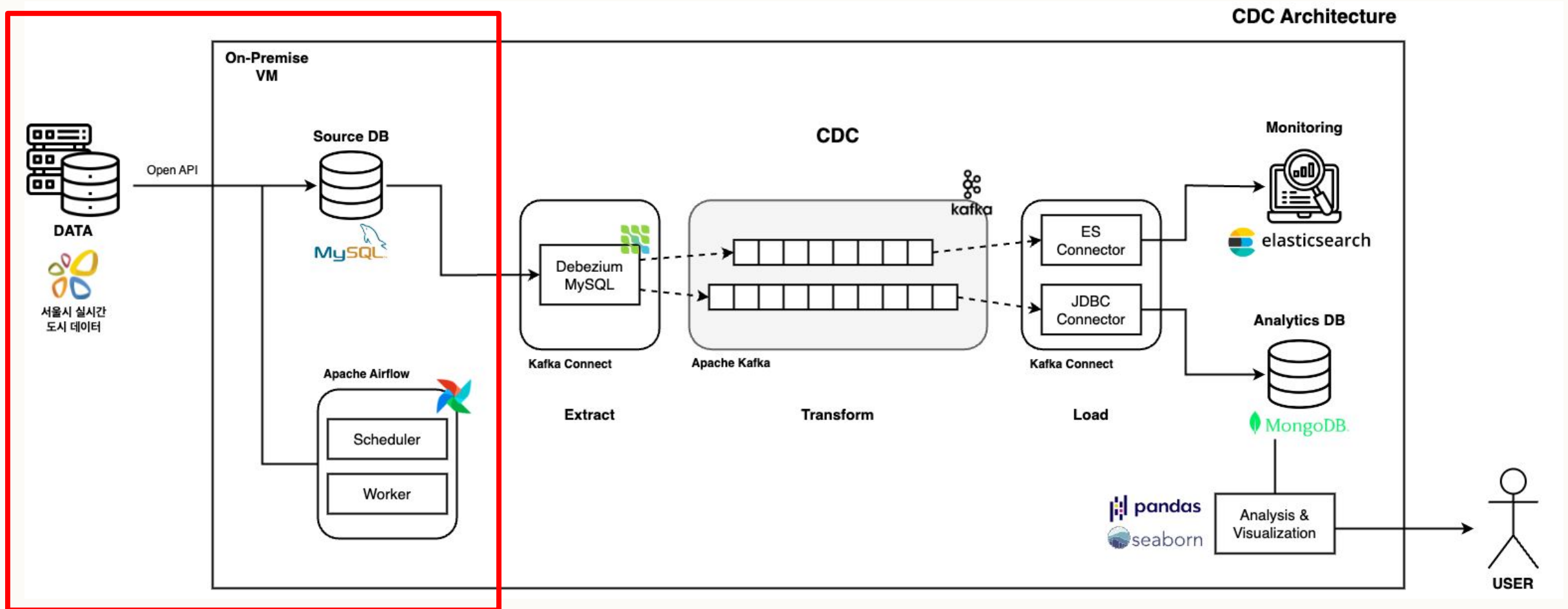




# 서비스 아키텍처

아키텍처 구성

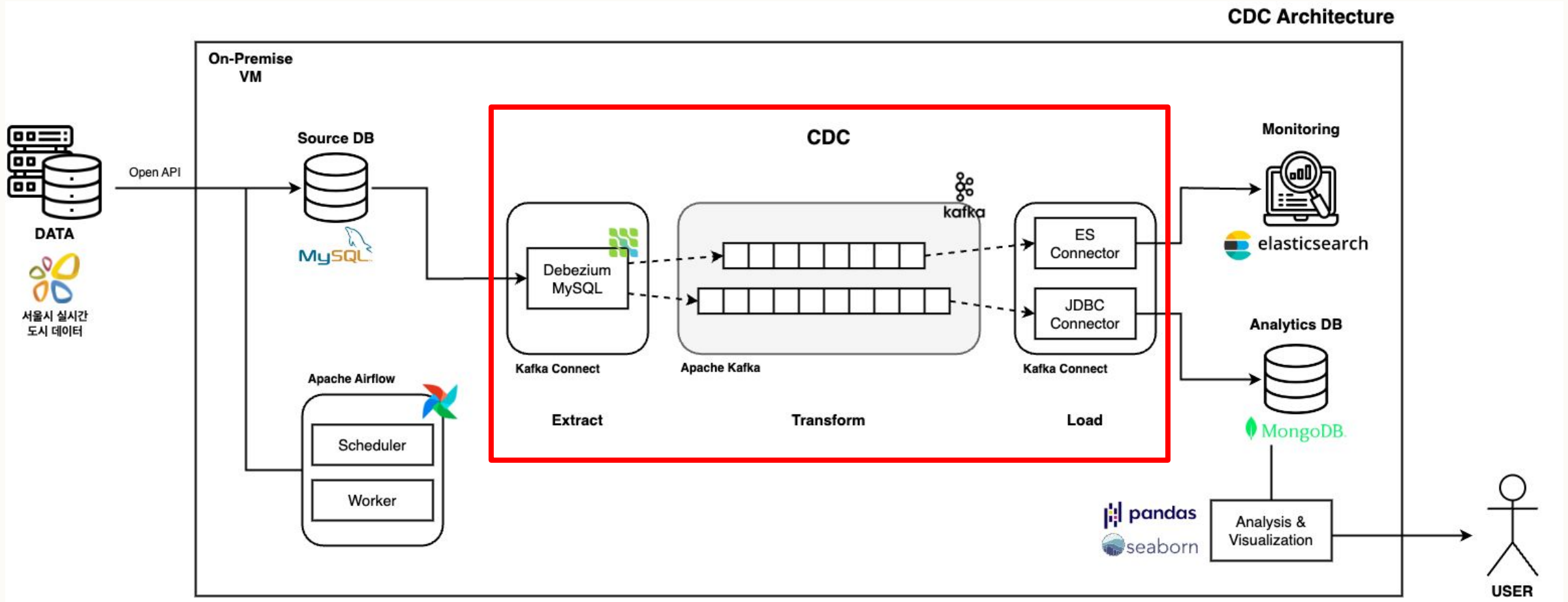
## ➡ 데이터 수집



# 서비스 아키텍처

아키텍처 구성

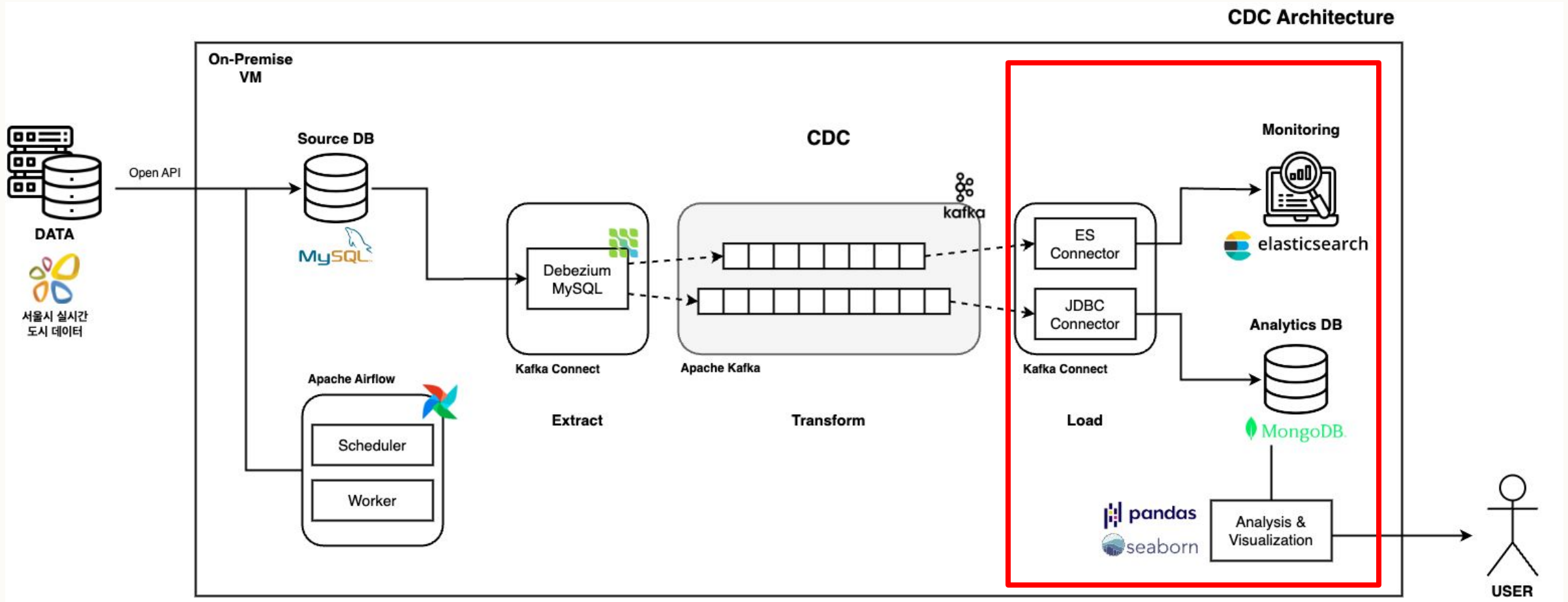
## ➔ CDC



# 서비스 아키텍처

아키텍처 구성

## ➔ 모니터링 및 인사이트 도출







Thanks

