

디지털 트윈을 위한 데이터 동기화 파이프라인 개발

오픈소스 URL :

https://github.com/1frbottom/DigitalTwin_PipeLine.git



2025학년도 2학기 SW 캡스톤디자인 경진대회

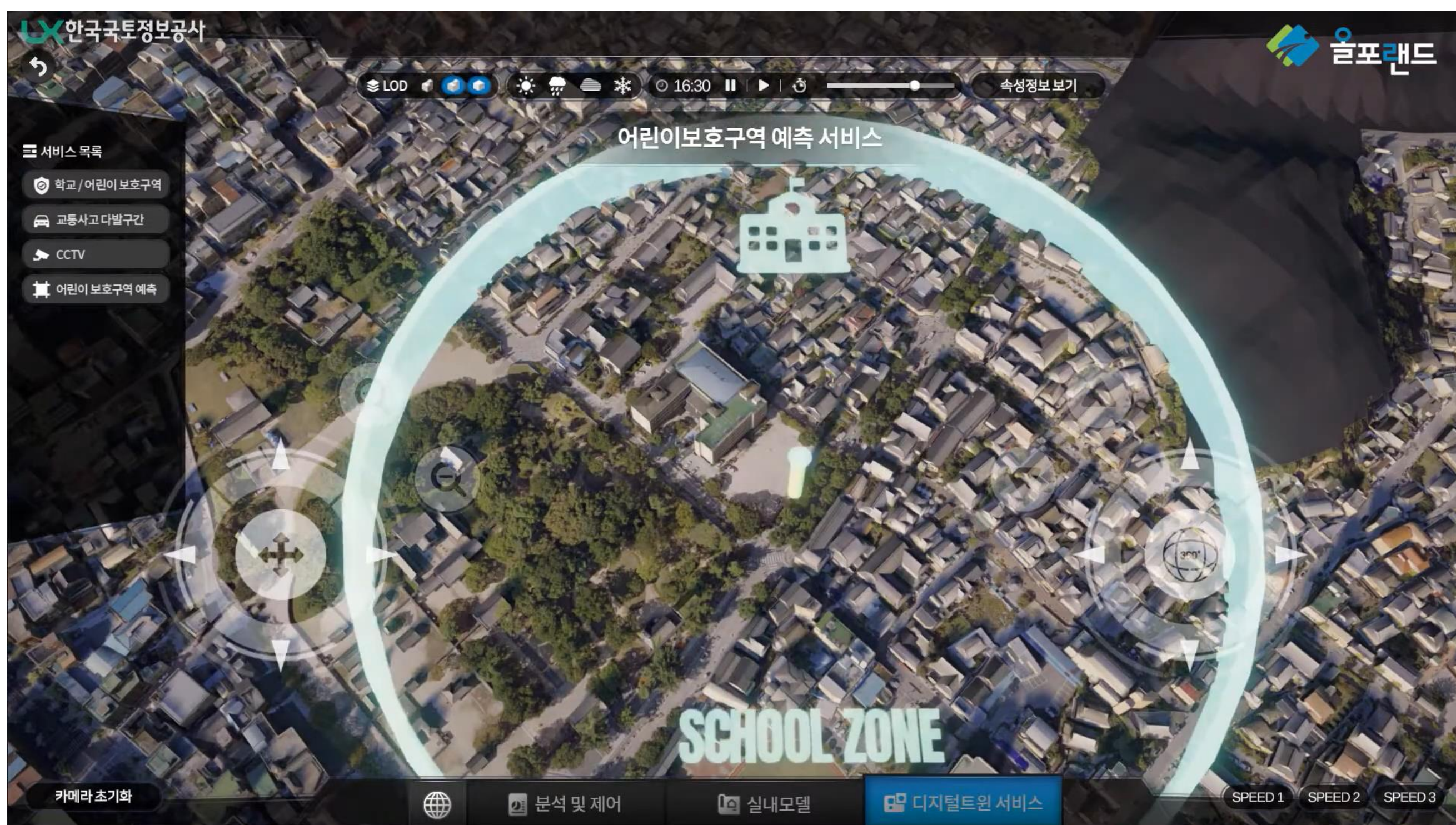
팀 명 베이즈

지도교수 김윤경

팀 원 김예린(IT정보공학과, 4), 심조운(컴퓨터공학과, 4), 정보경(IT지능정보공학과, 4), 최성원(분자생물학과, 3)

산업체 (주)올포랜드

개발 동기 및 목적



▲ (주)올포랜드 디지털트윈 기반 어린이보호구역 예측 서비스

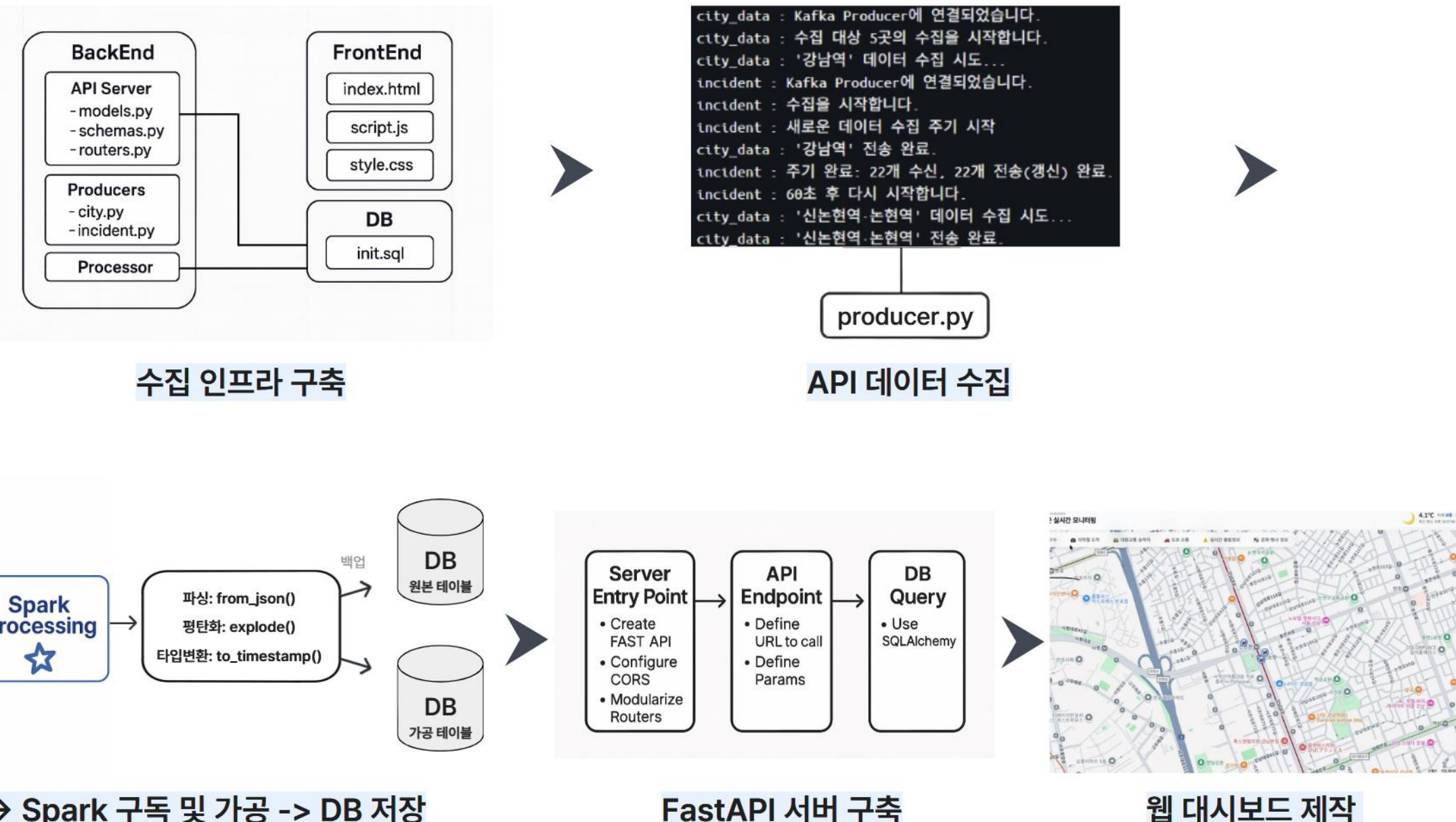
- 기존 디지털 트윈의 한계 극복 : 기존의 시각화 중심(Unity, Unreal 등) 에서 벗어나, 현실 세계의 변화를 실시간으로 반영하지 못하는 데이터 동기화 문제 해결.
- 실시간 데이터 동기화 파이프라인 구축 : 공공 데이터(인구, 교통, 환경)를 수집하여 DBMS로 지연을 최소화하여 적재&동기화하는 파이프라인 기술 개발.
- 실시간 도시 관제 실현 : 유동 인구와 교통량이 많은 강남역 등 주요 권역을 대상으로 실시간 모니터링 환경 제공.

주요 기술



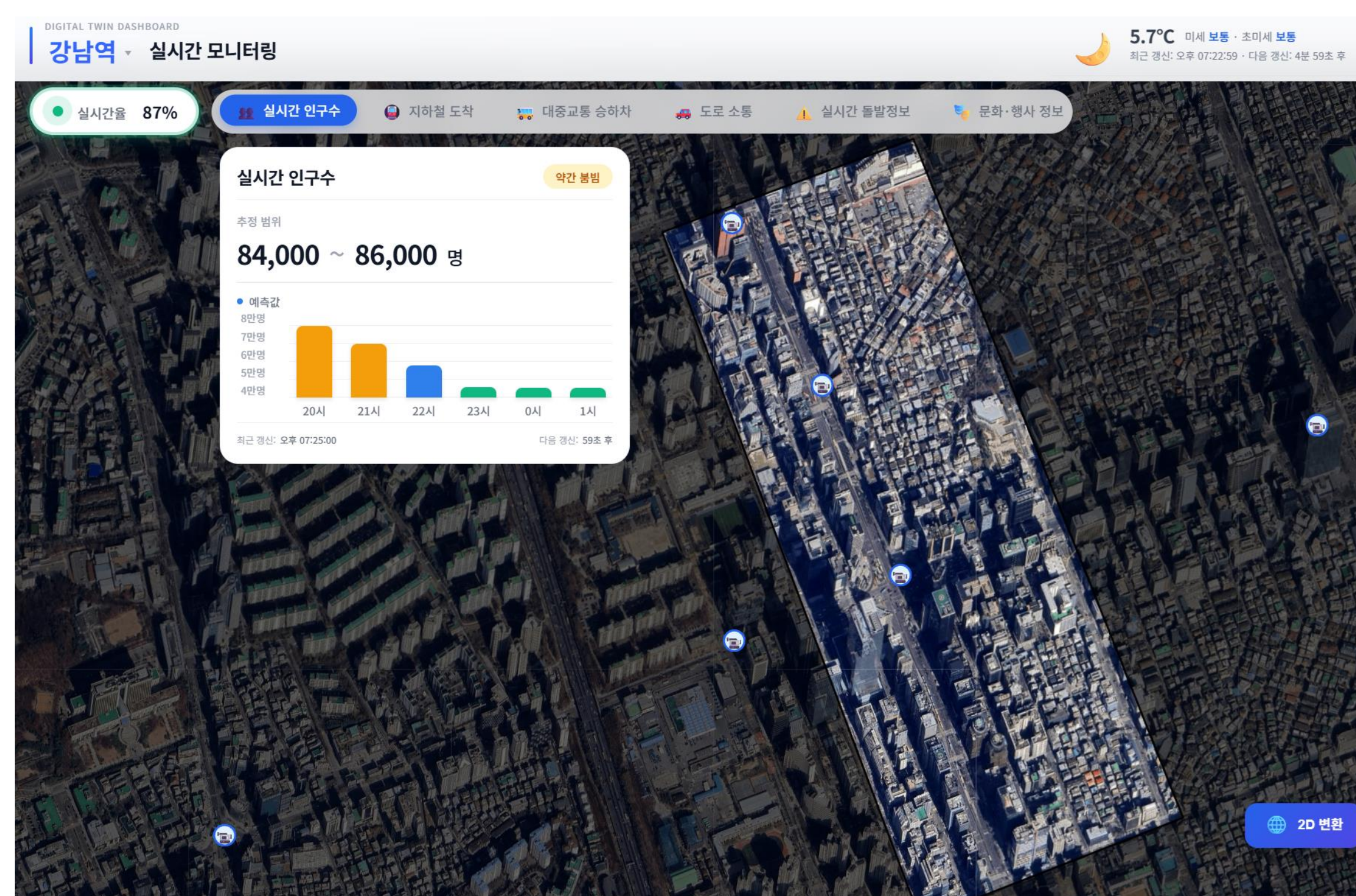
- Data Collection : Python (주기적 API 폴링 방식으로 안정적 데이터 수집).
- Message Queue : Apache Kafka (대용량 데이터 버퍼링 및 파이프라인 연계).
- Stream Processing : Apache Spark Structured Streaming (실시간 ETL, 데이터 정제 및 중복 제거).
- Database : PostgreSQL (시계열 데이터 저장, Trigger를 활용한 승하차 인원 자동 집계).
- Visualization : Google Maps API (2D/3D 지도, Polygon 혼잡도 표출), Chart.js.
- Infrastructure : Docker Compose (전체 시스템 컨테이너화 및 오케스트레이션).

개발 내용



- 이종 데이터 통합 수집 : 서울시 교통 정보 시스템 API를 활용하여 인구(실시간/예측), 교통(속도/정체), 지하철(도착정보), 기상, 문화행사 등 이질적인 데이터셋 통합.
- 실시간 동기화 시스템
 - 1) Kafka와 Spark Structured Streaming을 연동하여 1분/5분 단위의 데이터를 지연 없이 DB에 적재.
 - 2) 적재된 데이터들에서 추가적인 누적 및 집계를 통하여 최종적으로 시각화 단으로 전달.
- 디지털 트윈 대시보드
 - 1) 관제 구역 시각화 : Google Maps API를 활용해 모니터링 대상 권역들의 경계 구역을 폴리곤으로 시각화.
 - 2) 실시간 데이터 연동 UI
 - 인구 : 실시간 유동 인구 수치 및 혼잡도 단계를 텍스트 태그와 예측 그래프로 표현.
 - 교통 : 지하철 실시간 도착 정보와 도로 소통 단계를 게이지 UI로 표현.
 - CCTV : 주요 지점 좌표에 마커를 매핑하고, 클릭 시 실시간 영상 재생.

결과 및 분석



- 실시간성 확보 : 인구 및 돌발 정보 갱신 주기를 1분 내외로 단축 하여 현실과 가상 데이터 간의 시간차 최소화 성공.
- 데이터 가시성 향상 : 수치로만 존재하던 데이터를 지도 위 시각 정보(색상 기반 혼잡도, 그래프)로 변환하여 직관적인 상황 파악 지원.
- 시스템 안정성 : 수집(Producer), 처리(Spark), 저장(DB), 시각 화(Web) 모듈을 Docker 컨테이너로 분리하여 독립적인 운영 및 유지보수성 확보.
- 활용 가능성 : 단순 관제를 넘어, 시간대별 대중교통 이용 패턴 분석 및 인구 과밀 위험 지역 예측 등 도시 안전 관리 모델로 확장 가능.