

투빅스 정규세션 6주차 DB&MLOps 과제 제출

24기 김준수

사용자 프로필(ID, 이메일, 비밀번호 등)은 구조가 고정되어 있고 데이터 간 관계(친구, 팔로워 등)가 중요하므로 RDB를 사용한다. RDB를 사용하면 트랜잭션 관리와 정합성이 보장되어 사용자 정보의 신뢰성을 유지할 수 있기 때문이다. 반면 영상 시청 기록, '좋아요', 댓글 등은 양이 방대하고 구조가 일정하지 않으므로 NoSQL를 사용한다. NoSQL은 스키마 제약이 없고 수평적 확장이 쉬워 급격한 트래픽 증가에도 대응할 수 있다.

시스템을 클라우드 기반의 분산 시스템을 사용하는 이유 첫째는 사용량 증가 시 자동으로 자원을 확장할 수 있어 예측 불가능한 트래픽 폭주에 유연하다. 둘째, 전 세계 사용자에게 빠른 응답을 제공하기 위해 지역별 데이터센터를 활용할 수 있다.

또한 OLTP와 OLAP를 분리하여 구성해야 하는 이유는 서비스 운영에 필요한 빠른 트랜잭션 처리와 대규모 데이터 분석은 성격이 다르기 때문이다. OLTP에서 발생한 데이터를 주기적으로 ETL 과정을 거쳐 데이터 웨어하우스로 전송하면, 분석 시스템이 이를 기반으로 추천 모델을 학습하고 다시 서비스에 반영할 수 있다.