

투빅스 24기 박현서

1. 무엇을 수집할 것인지에 따라 다르다. 예를 들어 사용자 프로필과 같이 표형태로 관리가 되고 있는 곳은 관계형 데이터베이스를 이용하면 되고, 활동 로그 수집과 같은 서비스는 NoSQL을 사용하면 된다.
2. 클라우드 기반을 선택해야하는 이유는 크게 두가지로 정리할 수 있다. 먼저, 클라우드 기반은 사용자가 갑자기 많아져도 서버를 자동으로 늘려 서비스 중단 없이 비상상황에 대응 가능하기 때문이다. 두번째로, 데이터가 여러 지역에 자동으로 복사 및 저장되므로 장애가 생겨도 복구가 빠르기 때문이다.
3. OLTP(Online Transaction Processing): 로그인, 회원가입, 좋아요 누르기 등 실시간으로 사용자와 상호작용해야하는 데이터 처리. 빠른 응답과 정확한 트랜잭션이 중요함. ->RDB 저장

OLAP(Online Analytical Processing): 수백만 명의 활동 로그를 모아 분석 및 추천 학습 모델에 활용함. 대용량 집계와 복잡한 쿼리에 최적화된 데이터 웨어하우스가 적합

이 둘을 분리해야하는 이유는, 실시간 서비스와 분석 기능을 같은 DB에서 진행하게 되면 서로 자원을 차지하기 때문에 성능이 저하될 위험이 있기 때문이다.

또한 이 둘의 저장구조도 다르다. OLTP는 빠른 입력과 조회를 위한 행 중심 저장, OLAP는 대규모 집계와 분석 최적화를 위한 열 중심 저장구조 방식을 따른다. 따라서 이 둘을 분리해서 데이터와 목적에 맞게 사용할 필요가 있다.

OLTP 흐름: 사용자 행동 -> API 서버 -> OLTP DB -> 사용자 화면에 즉시 반영

OLAP 흐름: OLTP/로그 DB, ETL 및 스트리밍 파이프라인 -> 데이터 웨어하우스 -> 분석/모델링 -> 결과를 다시 OLTP로 전송