시나리오:
당신은 수백만 명의 글로벌 사용자를 목표로 하는 소셜 미디어 스타트업의 데이터 엔지니어입니다. 이 서비스는 사용자의 '프로필 정보(ID, 이메일 등 정형 데이터)'와 '활동 로그(영상 시청 기록, '좋아요', 댓글 등 비정형 데이터)'를 모두 처리해야 합니다. 또한, 서비스가 갑자기 성장하더라도 안정적인 운영이 가능해야 하며, 수집된 데이터를 분석하여 사용자 맞춤형 콘텐츠 추천 모델을 개발해야 합니다.

문제:
위 시나리오를 바탕으로, 이 서비스에 필요한 데이터베이스 아키텍처를 설계하고 그 이유를 아래 요소들을 포함하여 종합적으로 서술하시오. (800자 이내)

1. 데이터베이스 유형 선택: 서비스의 각 기능(예: 사용자 프로필 관리, 활동 로그 수집)에 관계형 데이터베이스(RDB)와 비관계형 데이터베이스(NoSQL) 중 무엇을, 왜 사용해야 하는지 포함하라.
2. 시스템 환경 구성: 온프레미스(On-premise)가 아닌 클라우드(Cloud) 기반의 분산 시스템을 선택해야 하는 이유 2가지를 언급하고 간단히 설명하라.
3. 데이터 처리 시스템 분리: OLTP와 OLAP를 분리하여 구성해야 하는 이유를 설명하고, 이 두 시스템 간의 데이터 흐름(예: ETL)을 간략하게 제시하시오.

답:
이 서비스는 정형 데이터(사용자 프로필)와 비정형 데이터(활동 로그)를 모두 다루어야 하므로 이중 데이터베이스 아키텍처가 필요하다. 사용자 프로필은 스키마가 명확하고 무결성이 요구되므로 관계형 데이터베이스(RDB, 예: Cloud SQL/PostgreSQL)를 채택하여 안정적인 계정 관리와 트랜잭션 처리를 보장한다. 반면 활동 로그는 초당 수백만 건 이상 발생할 수 있고 스키마가 유연해야 하므로 비관계형 DB(NoSQL, 예: DynamoDB, Cassandra, 또는 데이터 레이크 기반 S3+Parquet)를 활용하여 확장성과 빠른 쓰기 성능을 확보한다.
시스템 환경은 클라우드 기반 분산 시스템으로 구성한다. 첫째, 클라우드는 자동 확장성을 제공하여 갑작스러운 사용자 증가에도 안정적으로 대응할 수 있다. 둘째, 멀티리전 가용성과 운영 편의성을 통해 글로벌 서비스 운영, 백업, 보안 패치 등을 관리형으로 처리해 인프라 부담을 줄인다.
또한 OLTP와 OLAP 시스템을 분리해야 한다. OLTP는 사용자 인증, 프로필 수정, 실시간 로그 저장 등 즉각적인 응답이 필요한 트랜잭션에 초점을 맞추고, OLAP은 대규모 로그를 기반으로 추천 모델 학습과 사용자 행동 분석을 수행한다. 이를 위해 이벤트 스트리밍(Kafka/Kinesis)이나 ETL 파이프라인(Airflow, Glue)을 통해 OLTP·NoSQL 데이터가 데이터 레이크 및 데이터 웨어하우스(BigQuery/Redshift)로 이관된다. 이렇게 함으로써 서비스는 실시간 안정성과 대규모 분석 가능성을 동시에 확보할 수 있다.