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. (800 자 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.

   이 서비스의 데이터베이스 아키텍처는 정형 데이터와 비정형 데이터를 함께 처리하는 것이기 때문에, 관계형 데이터베이스(RDB)와 비관계형 데이터베이스(NoSQL)를 혼합해 사용하는 것이 효율적이다. 사용자 프로필, 이메일, 결제 내역처럼 정확성과 일관성이 중요한 데이터는 스키마가 명확하고 트랜잭션 관리가 필요하기 때문에 RDB 를 이용하는 것이 좋다. 하지만, 사용자의 활동 로그나 좋아요, 댓글, 시청 기록처럼 빠르게 쌓이고 구조가 일정하지 않은 데이터는 문서형 NoSQL 을 사용하는 것이 더 좋다.

2. 시스템 환경 구성: 온프레미스(On-premise)가 아닌 클라우드(Cloud) 기반의 분산 시스템을 선택해야 하는 이유 2 가지를 언급하고 간단히 설명하라.

   시스템에서 온프레미스보다 클라우드 기반 분산 환경을 선택해야 하는 첫 번째 이유는, 클라우드는 오토스케일링과 글로벌 리전을 지원하여 트래픽 폭주나 해외 사용자 증가에도 유연하게 대응할 수 있다는 것이다. 두번째로, 클라우드는 백업, 장애 복구, 모니터링이 자동화된 관리형 서비스를 통해 인프라 운영 부담을 줄이고 개발 효율을 높일 수 있다.

3. 데이터 처리 시스템 분리: OLTP 와 OLAP 를 분리하여 구성해야 하는 이유를 설명하고, 이 두 시스템 간의 데이터 흐름(예: ETL)을 간략하게 제시하시오.

   OLTP 는 실시간 사용자 요청(회원가입, 로그인 등)을 빠르게 처리하고, OLAP 은 대용량 로그 데이터를 분석해 추천 모델을 학습하는 데 사용된다. 이것을 위해서 서비스 데이터는 RDB 와 NoSQL 에 저장된 뒤, ETL 파이프라인을 통해 정제되어 데이터 웨어하우스로 이동한다. 이 구조를 통해 운영 성능과 분석 효율을 동시에 확보할 수 있기 때문에 OLTP 와 OLAP 를 분리하여 구성해야 한다.