

온프레미스란?

온프레미스(On-Premise) 환경은 서버, 네트워크, 저장장치 등을 회사 내부에 직접 구축하고 운영하는 방식입니다.

즉, 모든 시스템을 직접 설치하고, 직접 관리하는 방식입니다.

쉽게 말하면“클라우드 없이, 내 건물 안에 컴퓨터 방(서버실)을 두고 직접 돌리는 것”이라고 생각하면 됩니다.

16. 데이터 수집 전처리에서 스케줄 관리란 어떤 스케줄인가?

데이터 파이프라인의 스케줄 관리란 수집·전처리 작업의 실행 시점과 주기를 관리하는 기능을 말한다.

예를 들어, 매시간 또는 매일 자동으로 데이터 수집과 변환 작업을 실행하도록 예약하거나, 실시간 파이프라인의 트리거 조건을 설정하는 방식이다.

이를 통해 데이터 수집 및 적재 작업이 계획된 일정에 따라 안정적으로 수행되도록 함으로써 데이터 최신성을 유지하고 운영 편의성을 높인다.

17. 벡터 임베딩 모델 설정 기준은?

벡터 임베딩 모델을 설정할 때는 다음 기준을 고려한다:

- **업무·도메인 특성:** 처리하려는 데이터 유형(예: 뉴스, 대화, 과학문서 등)과 작업 유형(검색, 분류, 추천 등)에 적합한 모델을 선택한다.
- **언어 및 데이터 규모:** 주로 다룰 언어와 사용 가능한 학습/추론 데이터 크기에 따라 대형 모델 또는 소형 경량 모델을 결정한다.
- **라이선스 비용, 컴퓨팅 자원 고려:** 오픈소스 여부, 상용 서비스 비용, 클라우드 비용이나 하드웨어 구축 비용 등을 검토하여 제약 조건에 맞게 선택한다.
- **성능 검증:** 후보 모델을 실제 데이터로 벤치마크 테스트하여 유사도 정확도, 응답 시간, 메모리 사용량 등을 평가한다.

18. RAG에서 접근권한에 따라 답변은 어떻게 출력되는가?

RAG 시스템은 사용자 권한에 맞춰 답변에 참고되는 문서를 제한한다.

구체적으로, 문서를 색인할 때 해당 문서에 접근 가능한 사용자나 역할 정보를 함께 저장하고, 사용자가 쿼리하면 그 사용자의 권한으로 접근 가능한 문서만 검색 대상으로 한다.

따라서 최종 생성된 답변은 오직 사용자가 허가된 문서 내용에 기반하며, 권한이 없는 정보는 포함되지 않는다.

19. 통합 수집 프로세스에서 정확도 점검은 어떻게 수행되는가?