

데이터 엔지니어란?	데이터 엔지니어는 데이터 분석가 및 데이터 사이언티스트가 데이터를 효율적으로 활용할 수 있도록 <b>데이터 수집, 변환, 저장, 배포 환경을 설계하고 최적화하는 역할을</b> 담당. 또한, 데이터 품질 유지 및 거버넌스를 고려하여 데이터 파이프라인을 구축하고 관리		
데이터 엔지니어의 업무	데이터 관련 환경 만들기 - 데이터 <u>저장, 처리 및 관리</u> 하는 <u>시스템</u> 을 계획, 구축 및 관리		
	데이터 저장 시스템	데이터 베이스 관리	SQL (MySQL, PostgreSQL), NoSQL (MongoDB, Cassandra) 등의 설계 및 최적화
		스토리지 시스템	클라우드 스토리지(AWS S3, Azure Blob, GCP Storage) 및 물리적 스토리지(SAN, NAS) 관리
		데이터 웨어하우스	데이터 웨어하우스 설계, 구축 및 운영 (Amazon Redshift, Google BigQuery, Snowflake 등)
	데이터 처리 시스템	대용량 데이터 처리	Hadoop, Spark, Python을 활용한 데이터 처리 및 최적화
		실시간 데이터 처리	Kafka, Flink, Storm 등을 활용한 실시간 스트리밍 데이터 처리
	데이터 파이프라인	데이터 흐름을 자동화	데이터 흐름을 자동화하여 데이터 수집, 변환, 저장, 분석 과정을 효율적으로 연결
	데이터 보안 시스템	데이터 보안 정책 및 관련 규정 준수	보안 시스템 구축 및 관리 (IDS/IPS, SIEM, 클라우드 보안 서비스 활용)
	데이터 관리 시스템	데이터 표준화 및 카탈로그 관리 메타데이터 관리 및 데이터 품질 모니터링 (Alation, DataHub 등)	
데이터 엔지니어 필요기술	프로그래밍 언어	Python, SQL, Java, Scala	
	데이터 저장 및 관리 능력	관계형 및 비관계형 데이터베이스 설계 및 최적화 데이터 백업, 보안 및 접근 관리 기술 클라우드 및 온프레미스 스토리지 활용능력	
	데이터 처리 및 파이프라인 구축 경험	대용량 데이터 처리 기술 : Apache Spark, Hadoop 실시간 데이터 처리 기술 : Kafka, Flink, Storm 데이터 파이프라인 및 ETL/ELT : Apache Airflow, dbt, Talend	
	데이터 보안 및 관리	데이터 보안 정책 및 규제 이해 보안 시스템 구축 및 관리 (클라우드 보안 서비스 활용)	
	데이터 모델링 및 설계	데이터 모델링 기법에 대한 이해 (정규화, 차원 모델링) 데이터 구조 설계 및 최적화	
참고	데이터 베이스 설계	사용자의 요구사항을 분석하여 데이터를 효율적으로 저장, 관리, 검색할 수 있도록 설계하는 것	
	데이터 구조 설계	데이터를 효율적으로 저장하고 처리하기 위한 데이터 요소의 논리적인 구조를 정의하는 과정	
이 자료는 Gemini의 말꼬리를 잡아가며 매우 많은 질의를 통해 얻어낸 답변을 정리한 내용입니다. - 리뷰는 ChatGPT 에게 받았습니다. Chat GPT가 괜찮대요.			