

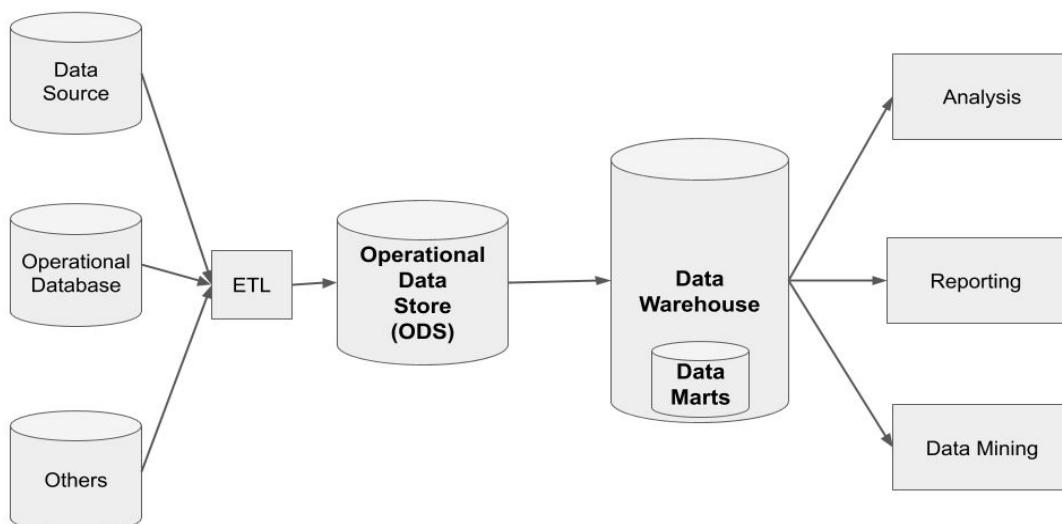
Data Warehousing

엑스트라이버 데이터 사이언스 지원자
안지용

데이터 웨어하우스란 사업에서 발생하는 대량의 데이터베이스를 축적해서 사업성 있는 인사이트 발견하고 의사결정을 돕는 기술이다. 데이터 웨어하우스는 사용자가 Query(질의)를 입력하여 데이터를 빠르고 쉽게 분석할 수 있도록 설계되어 있다. 데이터 분석과 보고서를 위해 설계된 BI 시스템의 핵심이기도 하다. 현재 항공사, 은행, 헬스케어, 정부 기관, 금융, 유통, 통신사, 호텔 부문 등에서 사용하고 있다. 주로 고객 정보나 사내 자원을 분석하여 매출과 이익을 높이는 수단으로 사용한다. 미래의 데이터 웨어하우스 분야는 규제 조건, 대규모 데이터베이스 구축, 멀티미디어 데이터 활용 등의 문제 해결이 다뤄질 것이다.

장 점	단 점
<ul style="list-style-type: none"> - 한 공간에서 중요한 데이터 접근 - Query(질의)에 해당하는 정보를 제공 - 원본 데이터를 전처리하여 시스템 효율화 - 분석과 보고서에 소요되는 시간 단축 - 시계열 분석이 가능하며 미래 예측에 용이 	<ul style="list-style-type: none"> - 비구조적 데이터에는 부적합 (숫자로 이뤄진(구조적) 데이터로 변환 작업 필요) - 데이터 웨어하우스 구축은 복잡하고 정해진 구조를 변경하기 어려움 - 비즈니스 비전이나 목적이 바뀌면 활용도가 낮아짐

데이터 원천(Data Source)에서 생성된 데이터는 ETL처리(데이터 추출,변형,저장하는 절차)되어 운영용 데이터 스토어(Operational Data Store, ODS)에 저장된다. 데이터 웨어하우스(Enterprise Data Warehouse)는 기업 내부에서 통합된 접근 방식(BI tool, SQL 등)으로 데이터를 활용하게 설계되어 있다. 사용자는 툴을 사용하여 데이터 웨어하우스나 데이터 마트(Data Mart)에서 의사 결정에 필요한 데이터 분석, 보고서 작성, 마이닝을 한다. 데이터 마트는 데이터 웨어하우스에서 특정 업무 분야에 맞게 분리한 데이터베이스다. 데이터 원천에서 바로 수집할 수도 있다.



<데이터 웨어하우스의 Data Flow>