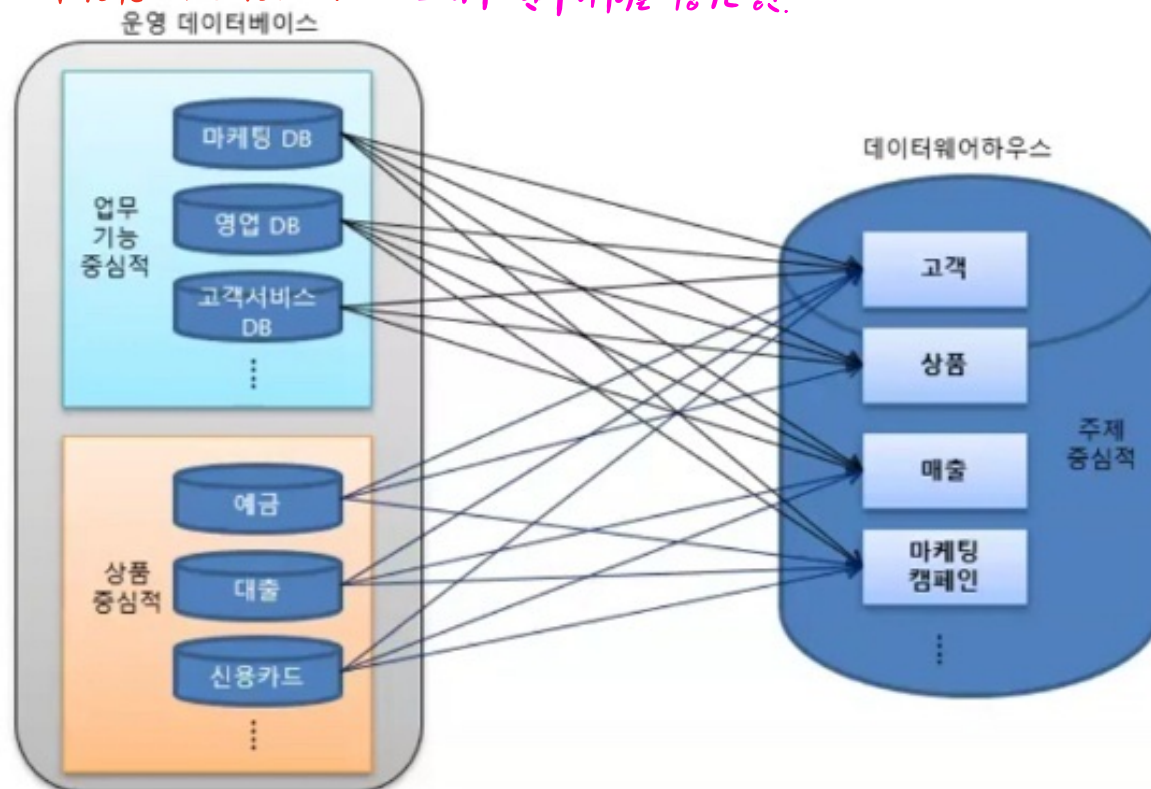


데이터웨어하우스

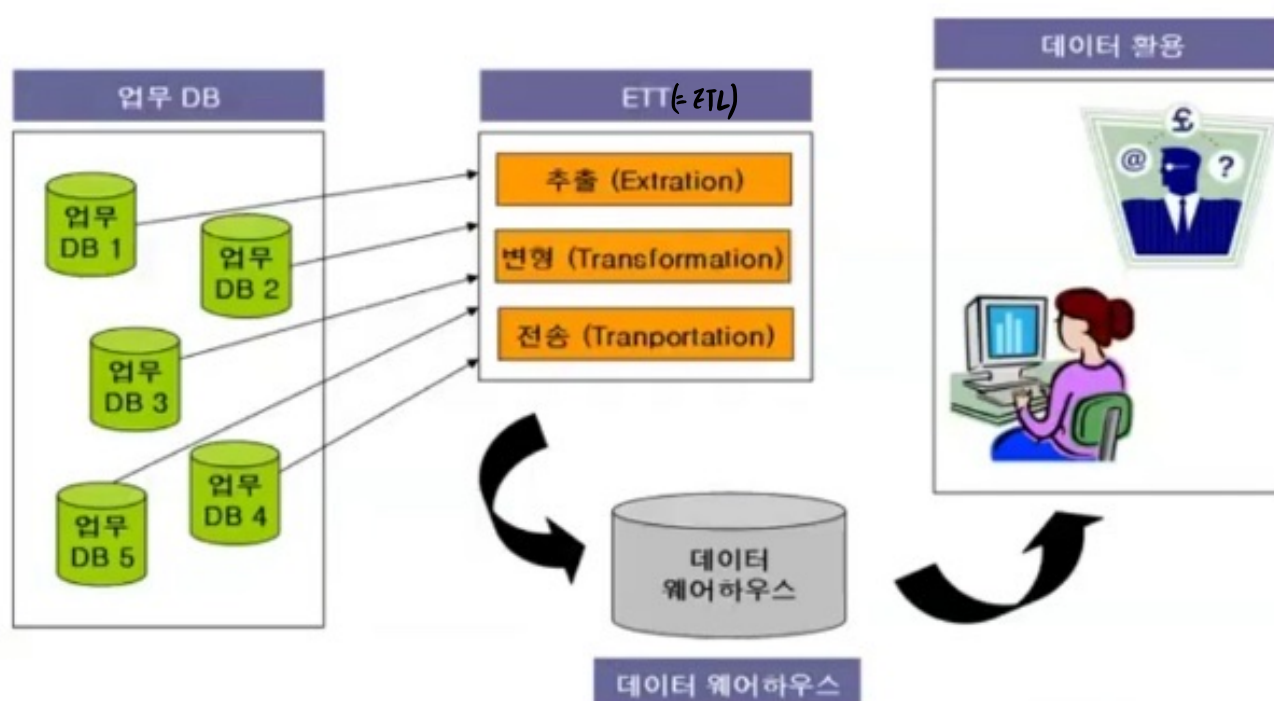
- 데이터의 주제지향성 (subject orientation)
 - DB : 업무기능 중심 - DW : 특정주제에 따른 분류
 - ↑ 현재 사용중인 데이터를 저장하는 곳
 - ↑ '과거 ~ 현재' 데이터를 저장하는 곳.



40

데이터웨어하우스

- 데이터웨어하우징 (Data Warehousing)
 - 데이터웨어하우스를 구축하고 활용해나가는 일련의 과정을 의미



Source: 김형수외, 고객관계관리전략, 2009

42

데이터웨어하우스

- 데이터마트(data mart):

- 데이터웨어하우스의 일부분
- 특정 사용자 집단이 사용할 수 있도록 특정 초점을 가지고 요약된 조직 데이터의 일부분
- 초점은 일반적으로 단일 주제 영역이나 업무 영역에 맞춤

- 데이터마트와 데이터웨어하우스 차이

- 데이터의 용량과 사용자 규모에서 차이가 있을 뿐 ETL도구, 운영 데이터베이스 및 데이터베이스관리시스템(DBMS), 데이터 애플리케이션 (분석도구) 등의 기본 구성요소는 동일

[개념정리]-Data Mart, Data Warehouse, Data Lake

🕒 1 minute read

데이터를 활용하기 위해서는 먼저 데이터를 수집해야하고, 그 데이터를 저장해야한다.



그리고 데이터를 저장하는 저장소, 데이터베이스(Database, DB)를 가르는 3가지 스타일이 있다.

데이터베이스를 가르는 3가지 스타일: **Data Mart**, **Data Warehouse**, **Data Lake**

수집되는 데이터 특성 및 데이터 활용방식에 따라 다르게 사용되며, 오늘은 이 DB 스타일 간 특징/차이점을 정리해보고자 한다.

Data Mart(DM, 데이터 마트)

- Data Mart(이하 DM)은 현업에서 데이터를 활용하는 담당자가 데이터를 활용하기 위한 형태 ~~격의 그대로~~ ^로 데이터를 쌓아두는 공간이다.
 ↗ WAS에서 호출되는 테이블들의 집합.
- 각 현업에서 사용하는 업무단위처럼 상세한 단위로 데이터를 저장하고, 사용한다.
- DM은 현업 담당자(소비자)가 필요한 데이터(상품)을 직접 골라담아 소비할 수 있는 공간(소매점..마트!)이라고 할 수 있다.

그리고 데이터 소매점(DM)에는 데이터를 공급하는 도매점(DW)이 있다.

Data Warehouse(DW, 데이터 웨어하우스)

①

- Data Warehouse(이하 DW)는 데이터 소매점(DM)에 공급하게 될 데이터를 다양한 원천에서 수집하여 주제별로 저장하는 데이터 도매점과 같은 공간이다.
- 즉 다양한 원천에서 발생하는 데이터를 소비자에게 전달하기 전에 통합하여 저장하는 공간이라고 할 수 있다.
- 그런데 다양한 원천에서 발생하는 데이터는 발생 지점의 환경에 따라 서로 다른 구조(스키마)나 데이터에 사용된 용어(동음이의어/이음동의어)등의 문제가 야기된다.
- DW는 이러한 문제들을 전사적 관점에서 고려하여 다양한 데이터를 통합저장하는 구조로 설계되고 구축된다.

②

단, DW 스타일에서는 통합되는 과정에서 데이터를 정제하는 비용과 버려지는 데이터가 발생한다. 이러한 점이 Data Lake의 등장야 야기되는 지점이다.

Data Lake(DL, 데이터 레이크)

- (Data Lake(이하 DL)은 다양한 원천을 하나의 통합된 형태로 정제한 DW와 달리) 다양한 원천을 그대로 가져와 저장하여 다양성을 보존하는 스타일을 가지고 있다.

즉, DL은 원천에서 발생한 데이터를 다양한 형태 그대로 한 곳에 저장하는 공간이라고 할 수 있다.

- 내 가설에 따르면 DL은 크게 3가지가 맞물려서 등장한 것으로 생각한다. – Data Science 붐에 따른 기존과 차별화된 분석 니즈(한 가지 관점으로 통합된 DW 대신 다양한 가정을 추론할 수 있도록 설계된 구조가 요구됨) – IoT 센서데이터와 SNS(소셜 네트워크 서비스)등에서 발생하는 비정형 데이터의 범람 – Hadoop 분산형 저장구조를 통한 대용량 데이터의 저비용(효율적) 저장 기술 도래

