

1. 데이터베이스 기본 개념

데이터베이스의 필요성

- 정보와 데이터
 - 데이터
 - 현실 세계에서 단순히 관찰하거나 측정하여 수집한 사실이나 값
 - 자료라고도 함
 - 정보
 - 데이터를 의사 결정에 유용하게 활용할 수 있도록 처리하여 체계적으로 조직한 결과물
 - 데이터에서 정보를 추출하는 과정 또는 방법을 정보 처리라고함
 - 데이터를 상황에 맞게 분석하거나 해석하여 데이터 간의 의미 관계를 파악하는 것
- 정보 시스템
 - 조직 운영에 필요한 데이터를 수집하여 저장해두었다가 의사 결정이 필요할 때 처리하여 유용한 정보를 만들어 주는 수단
 - 정보 시스템 안에서 데이터를 저장하고 있다가 필요할 때 제공하는 핵심 역할은 데이터베이스가 담당
 - 사용 목적에 따라 저장되는 데이터의 내용도 다르고 부르는 이름도 달라짐
 - 기업 경영 관리에 필요한 정보 시스템 -> 경영 정보 시스템(MIS)
 - 복합적이고 광범위한 의사 결정을 위해 사용되는 정보 시스템 -> 의사 결정 지원 시스템(DSS)

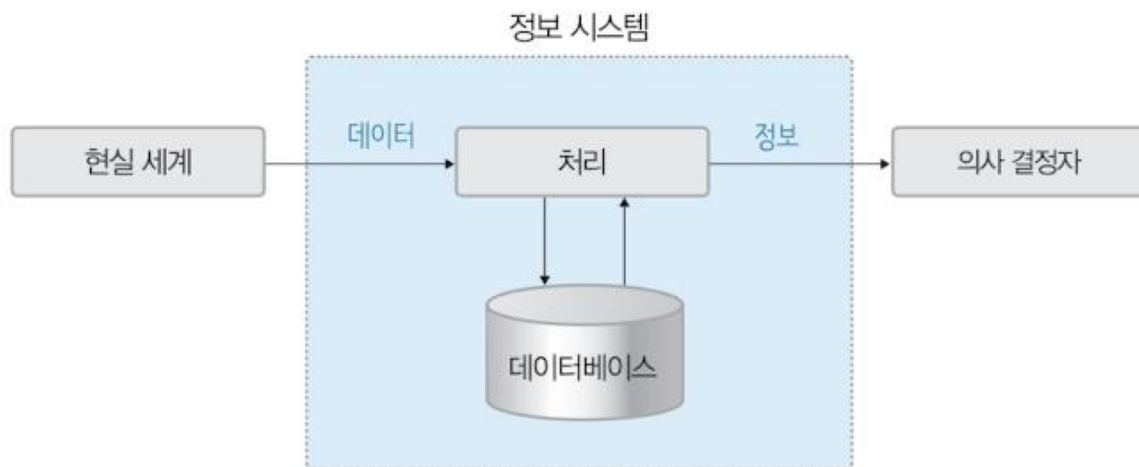


그림 1-3 정보 시스템의 역할과 구성

데이터베이스의 정의와 특징

- 데이터베이스는 쉽게 말해 관련있는 데이터를 모아두는 창고
- 일반적으로 정확한 의미는
 - 특정 조직의 여러 사용자가 **공유** 하여 사용할 수 있도록 **통합** 해서 **저장** 한 **운영** 데이터의 집합

데이터베이스 정의

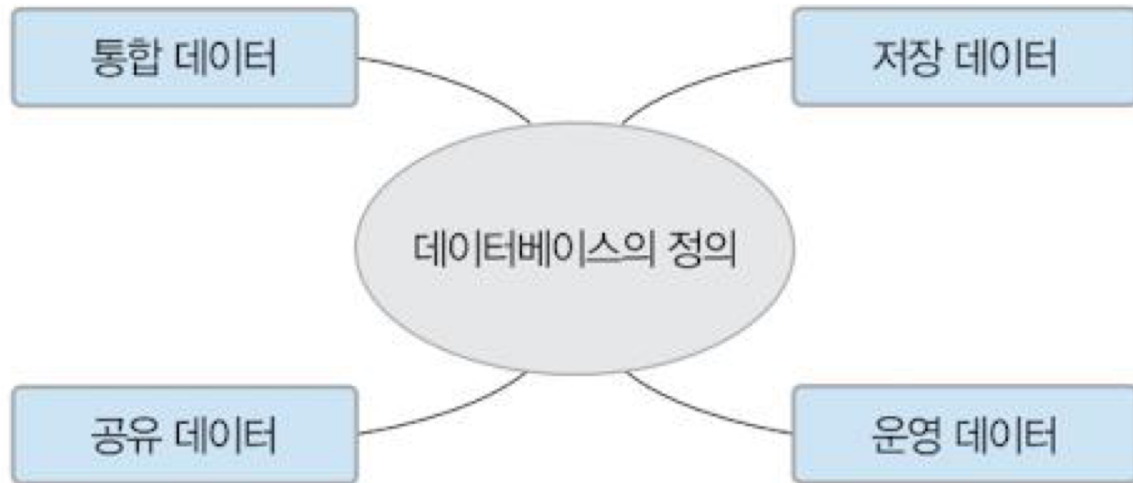


그림 1-4 데이터베이스의 정의

- 공유 데이터
 - 데이터베이스는 특정 조직의 여러 사용자가 함께 소유하고 이용할 수 있어야 하는 공용 데이터
 - 따라서 사용 목적이 다른 사용자들을 두루 고려하여 데이터베이스를 구성해야 함
- 통합 데이터
 - 데이터베이스는 데이터 중복성(같은 데이터가 여러개 존재하는 것)을 허용하지 않음
 - 관리하기 어려운 문제가 발생할 수 있기 때문
 - 하지만 효율성 때문에 중복을 의도적으로 허용하는 경우도 있으므로, 통합 데이터는 데이터의 중복을 최소화하고 통제가 가능한 중복만 허용하는 데이터라고 이해해야함
- 저장 데이터
 - 데이터베이스의 데이터는 주로 컴퓨터가 처리하므로, 컴퓨터가 접근할 수 있는 매체에 데이터베이스를 저장해야함
- 운영 데이터
 - 데이터베이스는 조직을 운영하고 조직의 주요 기능을 수행하기 위해 꼭 필요함
 - 지속적으로 유지해야 하는 데이터

데이터베이스 특징

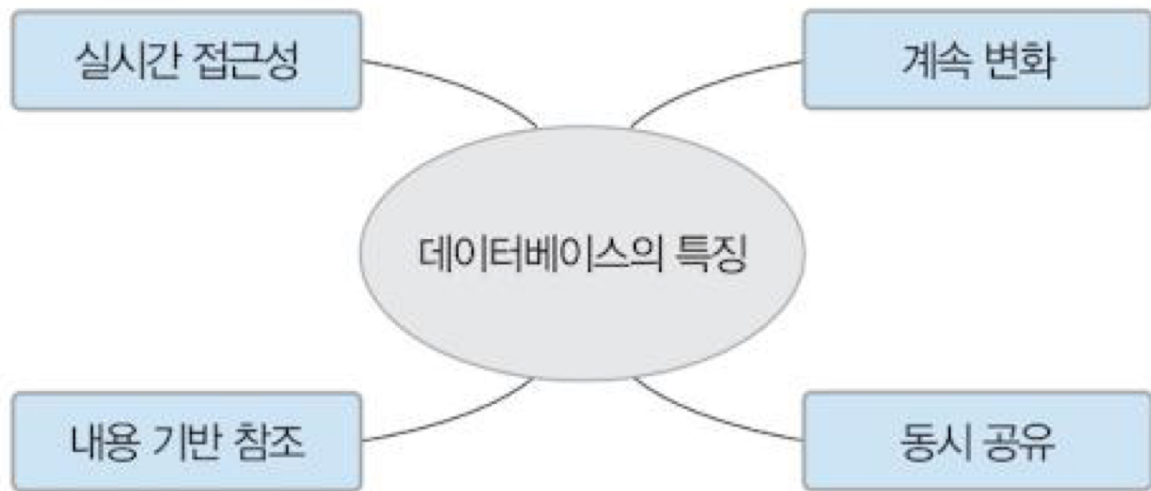


그림 1-5 데이터베이스의 특징

- 실시간 접근성
 - 데이터베이스는 사용자의 데이터 요구에 실시간으로 응답할 수 있어야 함
 - 대개 몇 초를 넘지 않는 시간 내에 데이터를 제공할 수 있어야 함
- 계속 변화
 - 데이터베이스는 현실 세계의 상태를 정확하게 반영해야 의미가 있음
 - 현실 세계는 끊임없이 변하므로 데이터베이스에 저장된 데이터도 계속 변해야함
 - 즉, 데이터베이스는 동적인 특징이 있어 데이터를 계속
 - 삽입(insert), 삭제(delete), 수정(update)하여 현재의 정확한 데이터를 유지해야함
- 동시 공유
 - 데이터베이스는 여러 사용자가 동시에 이용할 수 있는 동시 공유의 특징을 제공해야함
 - **단순히 여러 사용자가 함께 이용한다거나 시간 차를 두고 같은 데이터 영역을 함께 사용한다는 개념과는 다름**
 - 동시 공유는 여러 사용자가 서로 다른 데이터를 동시에 사용하는 것뿐 아니라, 같은 데이터터를 동시에 사용하는 것도 모두 지원한다는 의미
- 내용 기반 참조
 - 데이터베이스는 저장된 주소나 위치가 아닌 데이터의 내용, 즉 값으로 참조할 수 있음
 - 일반적인 컴퓨터처럼 데이터가 저장된 주소가 아닌 값으로 검색 가능
 - ex) 재고량이 100개 이상인 제품의 이름을 검색하시오
 - 이처럼 찾고자하는 데이터의 내용 조건만 제시하면 조건에 맞는 데이터가 서로 다른 위치에 저장되어 있어도 모두 검색 가능

데이터 과학 시대의 데이터

데이터의 유형

- 형태에 따른 데이터 분류

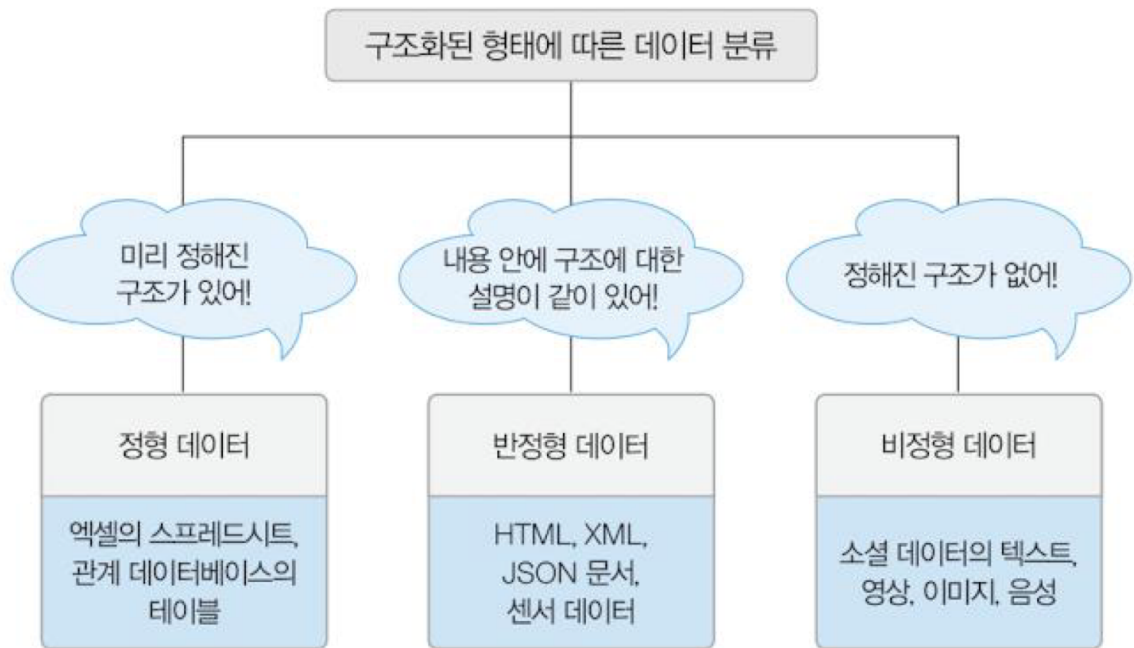


그림 1-7 형태에 따른 데이터 분류

- 정형 데이터
 - 구조화된 데이터, 즉 미리 정해진 구조에 따라 저장된 데이터
 - 엑셀의 스프레드시트, 관계 데이터베이스의 테이블
 - 데이터 구조에 대한 설명과 데이터 내용은 별도로 유지됨
 - 미리 정해진 데이터 구조를 **스키마**라고 함
- 반정형 데이터
 - 구조에 따라 저장된 데이터이지만 정형 데이터와 달리 데이터 내용 안에 구조에 대한 설명이 함께 존재함
 - 따라서 데이터 내용에 대한 설명, 즉 구조를 파악하는 파싱 과정이 필요
 - HTML, XML, JSON문서나 웹 로그, 센서 데이터등
 - 내용과 함께 설명된 데이터 구조를 스키마라고도 하지만 메타 데이터라고도 함
- 비정형 데이터
 - 정해진 구조가 없이 저장된 데이터
 - 소셜 데이터의 텍스트, 영상, 이미지, 음성, 워드, PDF문서등

- 특성에 따른 데이터 분류

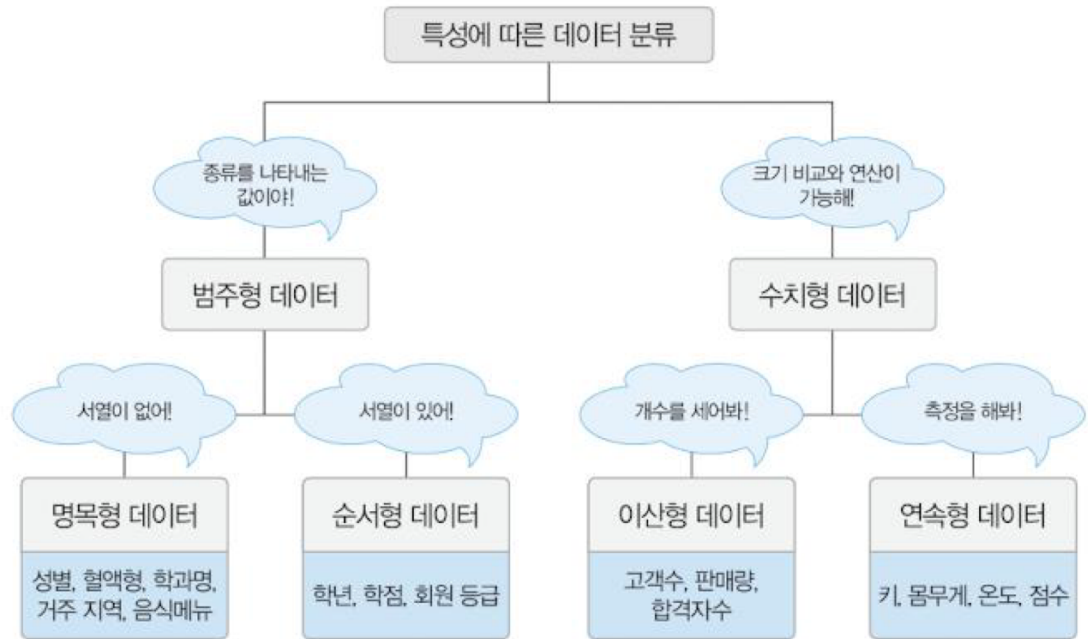


그림 1-11 특성에 따른 데이터 분류

- 범주형 데이터
 - 정성적 데이터
 - 범주로 구분할 수 있는 값
 - 즉, 종류를 나타내는 값을 가진 데이터
 - ex) 성별, 학년등
 - 명목형 데이터와 순서형 데이터로 나뉘어짐
 - 명목형 데이터
 - 서열이 없는 값을 가지는 데이터
 - ex) 성별, MBTI검사 결과등
 - 순서형 데이터
 - 순서, 즉 서열이 있는 값을 가지는 데이터
 - ex) 학년, 학점, 회원 등급등
 - 범주형 데이터는 대부분 문자 타입의 값으로, 양적 측면에서 크기 비교와 산술적인 연산이 가능하지 않음
 - 이러한 특성 때문에 질적 데이터라고도 함
- 수치형 데이터
 - 장량적 데이터
 - 양적 측면에서 크기 비교와 산술적인 연산이 가능한 숫자 값을 가진 데이터
 - 양적 데이터라고도 함
 - 이산형 데이터와 연속형 데이터로 나뉘어짐
 - 이산형 데이터
 - 단절된 숫자 값을 가지는 데이터

- ex) 고객 수, 판매량, 합격자 수등
- 보통 소수점이 없는 정수 타입의 값으로 표현
- 연속형 데이터
 - 연속적으로 이어진 숫자 값을 가지는 데이터
 - ex) 키, 몸무게, 온도, 점수등
 - 보통 소수점이 있는 실수 타입의 값으로 표현