

CHAPTER 01

데이터베이스 기본 개념

Database Fundamental Concepts

박상돈 조교수

대전대학교 컴퓨터공학과



목차

01

데이터베이스의 필요성

데이터와 정보, 정보 처리, 정보 시스템

02

데이터베이스의 정의와 특징

DB의 정의, 4대 특징, CRUD 개념

03

데이터 과학 시대의 데이터

형태별·특성별 데이터 분류

01

데이터베이스의 필요성

Why Do We Need Databases?

데이터와 정보



데이터 (Data)

현실 세계에서 단순히 관찰하거나 수집한 **사실 (Fact)**이나 **값(Value)**

- 아직 가공되지 않은 원시 상태의 자료
- 그 자체로는 의미를 가지기 어려움
- 숫자, 문자, 기호 등 다양한 형태

예시: "서울", "35", "2025-02-26"



정보 (Information)

데이터를 의사 결정에 유용하게 활용할 수 있도록 **처리·조직화한 결과물**

- 데이터에 맥락과 의미를 부여한 것
- 의사 결정에 직접 활용 가능
- 분석·가공 과정을 통해 생성됨

예시: "서울의 평균기온은 35°C로 폭염"

핵심: 데이터는 원재료, 정보는 가공된 제품 — 데이터 → 처리(가공) → 정보

정보 처리 (Information Processing)

정의: 데이터(Data)에서 정보(Information)를 추출하는 과정 또는 방법

데이터를 상황에 맞게 분석·해석하여 데이터 간의 의미 관계를 파악하는 것



예시

데이터: 서울, 35°C, 2025-08-01, 맑음 → 정보: "2025년 8월 1일 서울 기온 35°C, 폭염 경보 발령 필요"

정보 시스템과 데이터베이스



정보 시스템 (Information System)

조직 운영에 필요한 데이터를 수집·저장·가공·분배하는 체계적인 시스템

의사 결정 지원 시스템 (DSS)

경영 관리자의 의사 결정을 지원하기 위한 정보를 제공하는 시스템

경영 정보 시스템 (MIS)

기업의 경영 관리에 필요한 의사 결정용 정보 시스템. 재무, 인사, 생산, 마케팅 등 다양한 분야에 활용



데이터베이스 계층 구조

DB 데이터베이스 (Database)
데이터의 저장소

DBMS DB 관리 시스템
데이터를 관리하는 소프트웨어

DBS 데이터베이스 시스템
DB + DBMS + 사용자 + 응용 프로그램

02

데이터베이스의 정의와 특징

Definition and Characteristics of Database

데이터베이스(DB)의 정의

특정 조직의 여러 사용자가 공유하여 사용할 수 있도록 통합해서 저장한 운영 데이터의 집합



공유 데이터 Shared Data

특정 조직의 여러 사용자가 함께 소유하고 이용할 수 있는 공용 데이터. 각 부서와 사용자가 자신의 목적에 맞게 활용 가능.



통합 데이터 Integrated Data

데이터의 중복을 최소화한 통합된 데이터 집합. 불필요한 중복은 제거하고 통제 가능한 중복만 허용함.



저장 데이터 Stored Data

컴퓨터가 접근할 수 있는 저장 매체(디스크, SSD 등)에 저장된 데이터. 물리적으로 영구 보존이 가능해야 함.



운영 데이터 Operational Data

조직의 주요 기능을 수행하기 위해 지속적으로 유지·관리해야 하는 필수 데이터. 임시 데이터는 해당하지 않음.

데이터베이스의 특징 (1/2)



실시간 접근 (Real-Time Accessibility)

사용자의 데이터 요구에 실시간으로 빠르게 응답할 수 있어야 합니다.

- 질의(Query)에 대해 즉각적인 결과를 반환
- 온라인 트랜잭션 처리(OLTP) 환경에서 핵심적인 요구사항
- 수 초 이내의 응답 시간 보장이 필수적



계속 변화 (Continuous Evolution)

데이터베이스는 동적(Dynamic) 특성을 가지며, 항상 최신의 정확한 데이터 상태를 유지해야 합니다.

CRUD — 소프트웨어 기본 데이터 처리 요구사항:

C (Create): 새로운 데이터 삽입 | R (Read): 데이터 검색·조회

U (Update): 기존 데이터 수정 | D (Delete): 불필요한 데이터 삭제

데이터베이스의 특징 (2/2)



동시 공유 (Concurrent Sharing)

여러 사용자가 동시에 같은 데이터 또는 서로 다른 데이터를 사용할 수 있습니다.

- 동시성 제어(Concurrency Control)를 통해 데이터 일관성 유지
- 트랜잭션 격리 수준(Isolation Level)으로 충돌 방지
- 다중 사용자 환경에서 데이터 무결성 보장이 핵심 과제



내용으로 참조 (Content Reference)

데이터의 물리적 주소가 아닌, 데이터의 내용(Content)으로 검색합니다.

- 사용자가 찾고자 하는 조건을 제시하면 해당 데이터를 모두 검색 가능
- 데이터가 물리적으로 분산 저장되어 있어도 논리적으로 통합 검색
- SQL의 WHERE 절이 대표적인 내용 기반 참조 방식

*예: SELECT * FROM 학생 WHERE 학과 = '컴퓨터공학과'*

03

데이터 과학 시대의 데이터

Data in the Age of Data Science

형태에 따른 데이터 분류

데이터는 구조화된 형태에 따라 3가지로 구분됩니다.



정형 데이터 Structured

미리 정해진 스키마(Schema)에 따라 행과 열로 구성된 데이터. 관계형 데이터베이스, 스프레드시트 등.

RDBMS, CSV, Excel



반정형 데이터 Semi-Structured

고정된 스키마 없이 데이터 내부에 구조(메타데이터)가 포함된 데이터. 유연한 구조 변경이 가능.

HTML, XML, JSON, 센서 데이터



비정형 데이터 Unstructured

정해진 구조가 없는 자유 형태의 데이터. 전체 데이터의 약 80%를 차지하며 빠르게 증가 중.

텍스트, 이미지, 영상, 음성, PDF

정형 데이터 (Structured Data)

미리 정해진 구조(Schema)에 따라 저장된 데이터 — 행(Row)과 열(Column)로 구성

예시: 학생 정보 테이블

학번	이름	학과	학년	평점
20210001	김철수	컴퓨터공학	3	3.8
20210002	이영희	전자공학	2	4.2
20210003	박민준	컴퓨터공학	4	3.5

특징

- 고정된 스키마 보유
- SQL로 검색·조작 가능
- 데이터 무결성 보장
- 관계형 DB에서 관리
- 트랜잭션 지원

대표 시스템:

Oracle, MySQL, PostgreSQL, MS SQL Server

반정형 데이터 (Semi-Structured Data)

데이터 내용 안에 구조(메타데이터)에 대한 설명이 함께 존재하는 데이터

XML 예시

```
<학생목록>
  <학생>
    <학번>20210001</학번>
    <이름>김철수</이름>
    <학과>컴퓨터공학</학과>
  </학생>
</학생목록>
```

JSON 예시

```
{
  "학생": [
    {
      "학번": "20210001",
      "이름": "김철수",
      "학과": "컴퓨터공학"
    }
  ]
}
```

비정형 데이터 (Unstructured Data)

정해진 구조 없이 저장되는 데이터 – 전체 데이터의 약 80% 이상을 차지



텍스트

SNS 게시물, 이메일, 뉴스 기사, 리뷰, 웹 문서



이미지

사진, 의료 영상(X-ray, CT), 위성 이미지, 그래픽



영상/음성

유튜브 동영상, 팟캐스트, CCTV 영상, 음악 파일



문서 파일

PDF, Word, PowerPoint 등 비구조적 포맷

비정형 데이터의 급증 → 빅데이터 기술(Hadoop, Spark, NoSQL)과 AI/ML 기반 분석의 중요성 대두

특성에 따른 데이터 분류

데이터의 특성에 따라 범주형 데이터와 수치형 데이터로 구분합니다.



범주형 데이터 (Categorical / Qualitative Data)

종류를 나타내는 값을 가진 데이터 – 질적/정성적 데이터(Qualitative Data)라고도 함

명목형 데이터 (Nominal)

서열(순서)이 없는 값을 가지는 데이터

- 단순히 범주를 구분하는 용도
- 값 간의 크기 비교나 순서 비교 불가능
- 산술 연산이 의미 없음

예시: 혈액형(A, B, O, AB), MBTI, 성별,
주거 지역, 전공, 국적

순서형 데이터 (Ordinal)

서열(순서)이 있는 값을 가지는 데이터

- 값 간의 순서 비교 가능
- 그러나 값 간의 간격이 일정하지 않음
- 산술 연산이 적합하지 않음

예시: 학년(1~4), 학점(A~F), 회원 등급
(골드>실버>브론즈), 만족도(상/중/하)

핵심: 범주형 데이터는 연산이 어려움 – "1학년 + 2학년 = ?" → 의미 없는 연산

수치형 데이터 (Numerical / Quantitative Data)

크기 비교와 산술 연산이 가능한 데이터 — 양적/정량적 데이터(Quantitative Data)라고도 함

이산형 데이터 (Discrete)

단절된(불연속적) 숫자 값을 가지는 데이터

- 주로 개수(Count)를 세는 데이터
- 정수 값만 취할 수 있는 경우가 많음
- 값 사이에 중간값이 존재하지 않음

예시: 일일 판매량, 합격자 수, 교통사고 건수, 가족 수, 수강 인원

연속형 데이터 (Continuous)

연속적으로 이어진 숫자 값(실수)을 가지는 데이터

- 측정(Measurement)을 통해 얻는 데이터
- 두 값 사이에 무한한 중간값 존재
- 소수점 이하 값이 의미를 가짐

예시: 키(172.5cm), 몸무게(68.3kg), 온도(36.5°C), 속도(km/h), 주가

핵심: 수치형 데이터는 연산 가능 — 평균, 합계, 표준편차 등 통계 분석에 직접 활용



감사합니다

Thank You

박상돈 조교수
대전대학교 컴퓨터공학과