# 데이터베이스 개론 3판

1장 데이터베이스 기본 개념 2장 데이터베이스 관리 시스템

차용두교수





# 교수 소개

영진전문대학교 컴퓨터정보계열 AI클라우드프로그래밍과 교수 평생직업교육 TF 팀장

이름: 차용두

연락처: 053-940-5654

이메일: storm77@yju.ac.kr

# 교재 정보







데이터베이스 개론 🚥

김연희 지음

연습문제 해답은 제공하지 않습니다

👭 하빛이키데미

- 도서명 : IT CookBook, 데이터베이스 개론(3판)
- 저자 : 김연희
- 출판사 : 한빛아카데미(주)
- 페이지 / 정가: 552p / 29,000원
- 실습 파일 : <a href="http://www.hanbit.co.kr/src/4577">http://www.hanbit.co.kr/src/4577</a>

# 주요 내용 요약





#### 1 데이터베이스 기초 이론 (1~3장)

1장에서 데이터베이스를, 2장에서 DBMS를 소개합니다. 3장에서는 이들을 조합한 데이터베이스 시스템을 소개합니다.

#### ② 데이터 모델과 연산 (4~6장)

4장에서 데이터 모델링의 개념과 데이터 모델의 역할을 알아봅니다. 그리고 5장에서 핵심 데이터 모델인 관계 데이터 모델의 전반을, 6장에서 관계 데이터의 주요 연산을 살펴봅니다.

#### ❸ 데이터베이스 언어 SQL (7장)

SQL의 주요 기능을 소개한 후, 테이블 생성과 데이터 검색 및 조작을 위해 SQL로 질의문을 작성하는 방법을 알아봅니다.

#### 4 데이터베이스 설계 (8~9장)

데이터베이스 설계의 중요성과 목표를 소개하고, 두 가지 주요 설계 방법을 다룹니다. 8장에서는 E-R 모델과 릴레이션 변환 규칙을 이용한 설계 방법을, 9장에서는 정규화를 이용한 설계 방법을 알아봅니다.

#### **⑤** 데이터베이스 관리 (10~11장)

10장에서 다양한 회복 기법과 병행 수행 시 발생할 수 있는 문제를 해결하는 병행 제어 기법에 대해 알아봅니다. 11장에서는 데이터베이스 보안을 유지하기 위해 SQL을 이용해 권한을 부여하고 취소하는 방법을 알아봅니다.

#### **⑥** 데이터베이스 응용 기술 (12~13장)

12장에서 관계 데이터베이스와 다른 특성을 가진 객체지향·객체관계·분산·멀티미디어 데이터베이스를 소개합니다. 13장에서는 데이터 과학과 빅 데이터의 관련성을 알아보고 빅데이터 관련 기술을 소개합니다.

#### **7** 데이터베이스 활용 (부록)

오라클을 이용해 데이터베이스를 실제로 구축하는 방법을 알아봅니다. 그리고 책 전반의 이론을 적용할 수 있는 간단한 프로젝트를 소개합니다.

# 강의 계획표



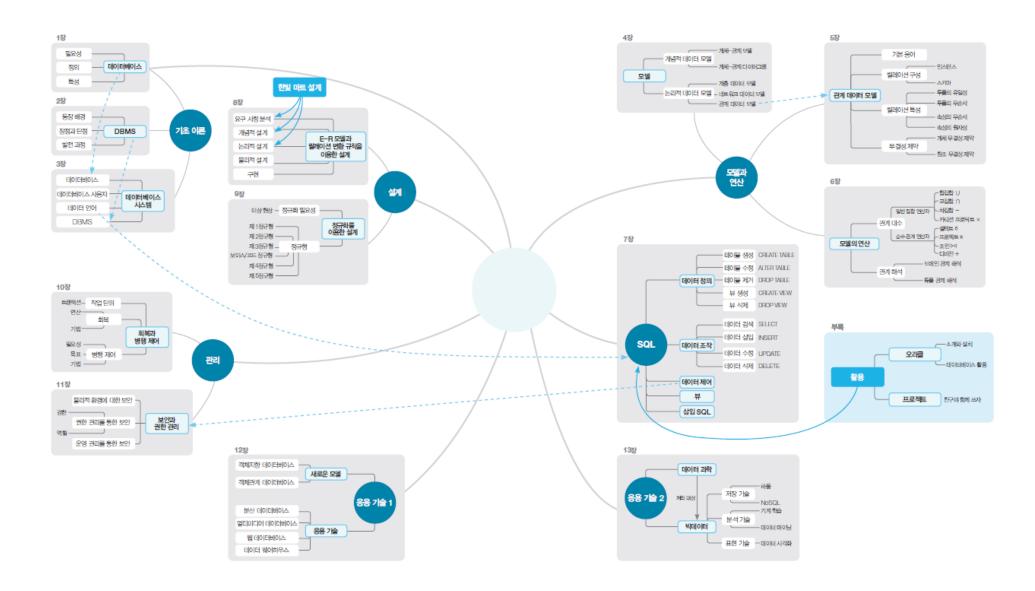


주	해당 장/주제	주제
1	1장, 2장	데이터베이스와 데이터베이스 관리 시스템의 개념
2	3장	3단계 데이터베이스 구조와 데이터베이스 시스템의 구성 요소
3	4장	데이터 모델링의 개념과 개체-관계 모델을 이용한 모델링
4	5장	관계 데이터 모델의 개념과 릴레이션, 키, 무결성 제약조건
5	6장	관계 대수의 기본 연산자를 이용한 질의 표현
6	7장	SQL의 데이터 정의 기능을 이용한 질의문
7	7장	SQL의 데이터 조작 기능을 이용한 질의문
8	필기/실기	중간고사
9	7장	뷰의 개념과 필요성, 삽입 SQL의 특성
10	8장	데이터베이스 설계의 각 단계별 설명과 설계 방법
11	9장	정규화의 필요성과 정규화 방법
12	10장	트랜잭션의 개념과 특성, 장애의 유형과 회복 기법
13	10장	트랜잭션 스케줄의 개념과 병행 제어 기법
14	11장	SQL을 이용한 권한과 역할의 부여와 취소
15	필기/실기	기말고사

# 이 책의 내용 흐름도







# 3판에서 달라진 부분





- ◆ 빅데이터에 대한 풍부한 설명을 담았으며 특히 빅데이터 표현 기술 내용을 추가했습니다.
- ◆ 2020년에 개편된 정보처리기사·데이터 분석 전문가 및 준전문가·전산직 공무원·정보관리기술사 시험의 최신 기출 경향에 맞게 연습문제를 보강하였습니다.
- ◆ 부록을 통해, 최신 버전의 오라클로 데이터베이스를 직접 구축해 볼 수 있게 안내하였습니다.
- ◆ 뿐만 아니라 설치하지 않고 웹에서 사용할 수 있는 Live SQL을 활용하는 방법도 담았습니다.
- ◆ 이전 정보처리기사 및 공무원 시험 기출 문제는 PDF로 제공합니다.

# 디지털 인재 100만명 양성











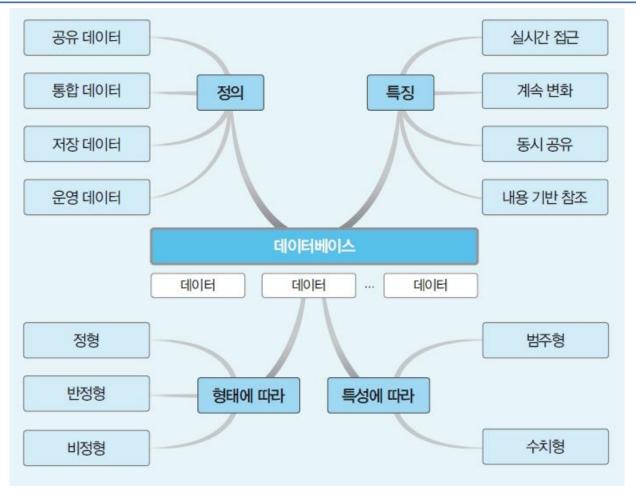
# 1장 데이터베이스 기본 개념

- 01 데이터베이스의 필요성
- 02 데이터베이스의 정의와 특징
- 03 데이터 과학 시대의 데이터

# 학습목표







- 데이터와 정보의 차이를 이해한다.
- 데이터베이스의 필요성을 알아본다.
- 데이터베이스의 정의에 숨겨진 의미와 주요 특징을 이해한다.
- 형태와 특성에 따른 데이터 분류 방법을 알아본다.





### ◆ 데이터와 정보

- 데이터(data)
  - 현실 세계에서 단순히 관찰하거나 측정하여 수집한 사실이나 값
- 정보(information)
  - 의사 결정에 유용하게 활용할 수 있도록 **데이터를 처리한 결과물**





### ◆ 데이터와 정보



그림 1-1 데이터와 정보의 이해 : 원유와 가공 우유







■ 데이터에서 정보를 추출하는 과정 또는 방법

한빛 인터넷 쇼핑몰 주문 내역

주문 번호	주문 일자	제품명	판매 금액
1	2022-01-10	냉장고	50만 원
2	2022-02-12	세탁기	30만 원
3	2022-03-03	세탁기	30만원
4	2022-04-05	에어컨	70만 원
5	2022-05-15	에어컨	80만 원
6	2022-06-19	에어컨	70만 원
7	2022-07-07	에어컨	70만 원
8	2022-08-12	냉장고	40만 원
9	2022-10-11	청소기	10만 원
10	2022-12-27	전자레인지	15만원

데이터

정보

저	품별	종	판미	바악

제품	총 판매액	
에어컨	290만 원	
냉장고	90만 원	
세탁기	60만 원	
전자레인지	15만 원	
청소기	10만 원	

정보 처리



그림 1-2 정보 처리의 예





### ◆ 정보 시스템과 데이터베이스

- 정보 시스템(information system)
  - 조직 운영에 필요한 데이터를 수집하여 저장해두었다가 필요할 때 유용한 정보를 만들어 주는 수단
- 데이터베이스
  - 정보 시스템 안에서 데이터를 저장하고 있다가 필요할 때 제공하는 역할 담당





### ◆ 정보 시스템과 데이터베이스

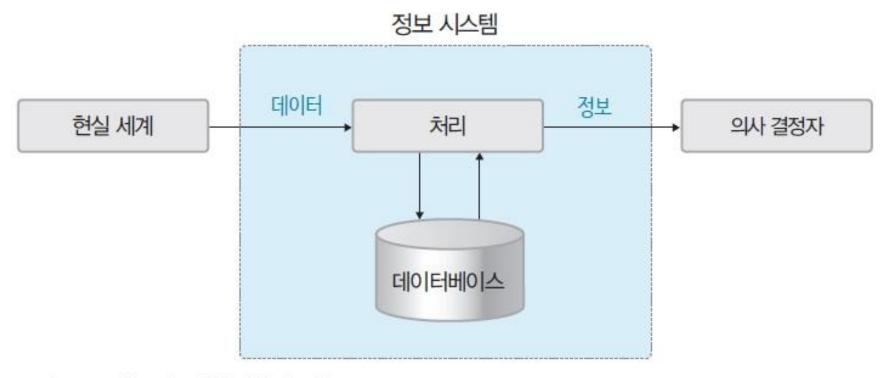


그림 1-3 정보 시스템의 역할과 구성

# 02 데이터베이스의 정의와 특징





- ◆ 데이터베이스(DB; DataBase) 정의
  - 특정 조직의 여러 사용자가 <mark>공유</mark>하여 사용할 수 있도록 **통합**해서 **저장**한 운영 데이터의 집합

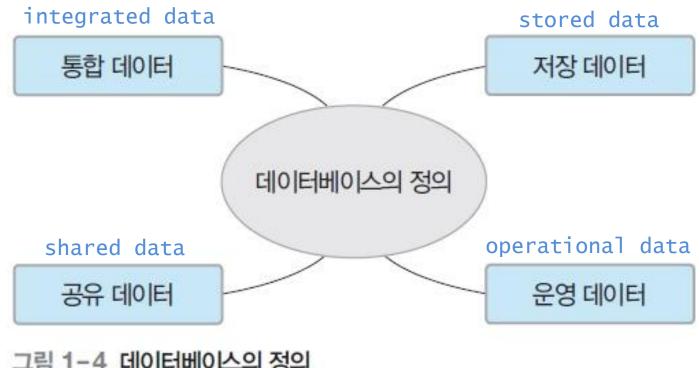


그림 1-4 데이터베이스의 정의

# 02 데이터베이스의 정의와 특징 - 정의





### ◆ 공유 데이터

■ 특정 조직의 여러 사용자가 함께 소유하고 이용할 수 있는 공용 데이터

### ◆통합 데이터

■ 최소의 중복과 통제 가능한 중복만 허용하는 데이터

### ◆ 저장 데이터

■ 컴퓨터가 접근할 수 있는 매체에 저장된 데이터

### ◆운영 데이터

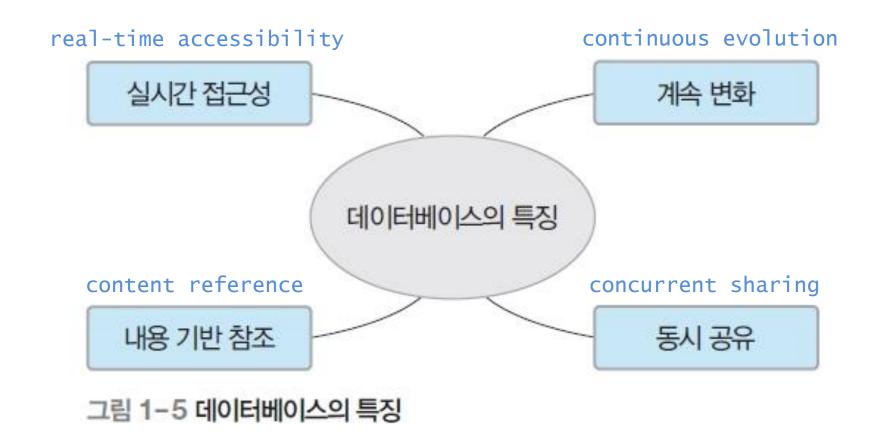
■ 조직의 주요 기능을 수행하기 위해 지속적으로 꼭 필요한 데이터

# 02 데이터베이스의 정의와 특징





### ◆ 데이터베이스의 특징



# 02 데이터베이스의 정의와 특징 - 특징





### ◆ 실시간 접근

■ 사용자의 데이터 요구에 실시간으로 응답

### ◆계속 변화

■ 데이터의 계속적인 삽입, 삭제, 수정을 통해 현재의 정확한 데이터를 유지

### ◆ 동시 공유

■ 서로 다른 데이터의 동시 사용뿐만 아니라 같은 데이터의 동시사용도지원

### ◆ 내용 기반 참조

- 데이터가 저장된 주소나 위치가 아닌 내용으로 참조 가능
- •예) 재고량이 1,000개 이상인 제품의 이름을 검색하시오.

# 02 데이터베이스의 정의와 특징







그림 1-6 데이터베이스의 이용





### ◆ 형태에 따른 데이터 분류

- 정형 데이터
- 반정형 데이터
- 비정형 데이터

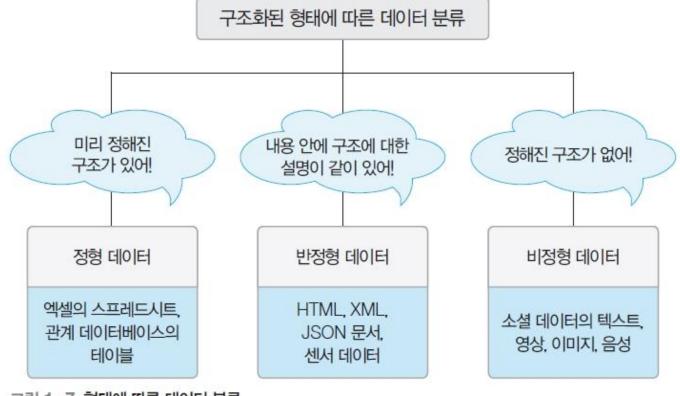


그림 1-7 형태에 따른 데이터 분류





### ◆ 정형 데이터(structured data)

- 구조화된 데이터, 즉 미리 정해진 구조에 따라 저장된 데이터
  - 데이터 구조에 대한 설명과 데이터 내용은 별도로 유지됨
- 예) 엑셀의 스프레드시트, 관계 데이터베이스의 테이블

A	А	В	С	D
1	일자	배송 업체	배송 건수	전일대비 상승률
2	2022-03-02	빠르다 택배	100	0%
3	2022-03-02	한빛 택배	200	10%
4	2022-03-02	안전 택배	50	3%
5	2022-03-02	당일 택배	30	-10%

그림 1-8 정형 데이터의 예





### ◆ 반정형 데이터(semi-structured data)

- 구조에 따라 저장된 데이터이지만 데이터 내용 안에 구조에 대한 설명이 함께 존재
- 구조를 파악하는 파싱(parsing) 과정이 필요, 보통 파일 형태로 저장
- 예) 웹에서 데이터를 교환하기 위해 작성하는 HTML, XML, JSON 문서나 웹 로그, 센서 데이터 등

```
{
 "이름": "오형준",
 "나이": 23,
 "성별": "남"
}
```

〈친구정보〉 〈이름〉오형준〈/이름〉 〈나이〉23〈/나이〉 〈성별〉남〈/성별〉 〈/친구정보〉

(a) JSON

(b) XML





### ◆ 비정형 데이터(unstructured data)

- 정해진 구조가 없이 저장된 데이터
- 예) 소셜 데이터의 <u>텍스트, 영상, 이미지, 워드나 PDF 문서와 같은 멀티미디어 데이터</u>

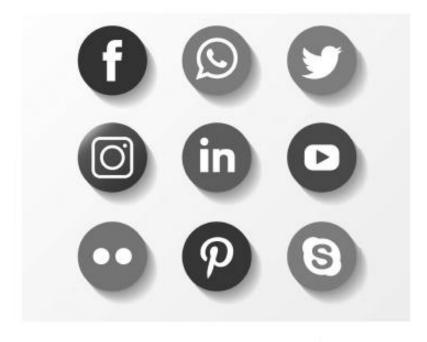


그림 1-10 비정형 데이터의 예(Designed by S.salvador / Freepik)

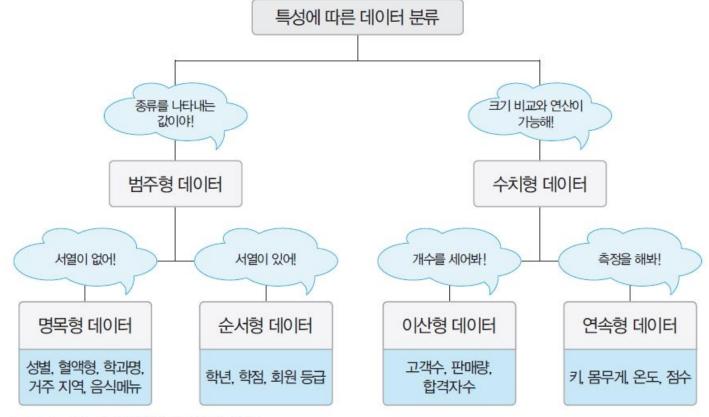




### ◆특성에 따른 데이터 분류

- 범주형 데이터
- 수치형 데이터

통계적 관점에서 데이터 특성에 따라 적합한 분석 방법을 선택하기 위해 데이터 분석 분야에서 주로 활용







## ◆ 범주형 데이터(categorical data)

- 범주로 구분할 수 있는 값, 즉 종류를 나타내는 값을 가진 데이터
- 크기 비교와 산술적인 연산이 가능하지 않아 질적 데이터라고도 함
- 명목형 데이터(nominal data)
  - 순서, 즉 서열이 없는 값을 가지는 데이터
  - 예) 성별, 혈액형, 학과명, 거주 지역, 음식 메뉴, MBTI 검사 결과
- 순서형 데이터(ordinal data)
  - 순서, 즉 서열이 있는 값을 가지는 데이터
  - 예) 학년, 학점, 회원 등급





### ◆ 범주형 데이터

명목형 데이터 서열이 없는 값을 가지는 데이터 선열이 있는 값을 가지는 데이터 서열이 있는 값을 가지는 데이터

그림 1-12 범주형 데이터





### ◆ 수치형 데이터(numerical data)

- <u>크기 비교와 산술적인 연산</u>이 가능한 숫자 값을 가진 데이터
- <u>양적 데이터</u>라고도 함
- 이산형 데이터(discrete data)
  - 개수를 셀 수 있는 띄엄띄엄 단절된 숫자 값을 가지는 데이터
  - 예) 고객 수, 판매량, 합격자 수
- 연속형 데이터(continuous data)
  - 측정을 통해 얻어지는 연속적으로 이어진 숫자 값을 가지는 데이터
  - 예) 키, 몸무게, 온도, 점수





### ◆ 수치형 데이터

이산형 데이터 단절된 숫자 값을 가지는 데이터 연속적으로 이어진 숫자 값을 가지는 데이터

그림 1-13 수치형 데이터





### ◆ 정성적 데이터와 정량적 데이터

- 정성적 데이터(qualitative data)
  - 좁은 의미로는 범주형 데이터
  - 사람의 주관적인 생각과 평가를 기술한 비정형 데이터
- 정량적 데이터(quantitative data)
  - 좁은 의미로는 수치형 데이터
  - 객관적인 측정을 통해 수치나 도형, 기호 등으로 표현한 정형 데이터

정량적 데이터에 비해 정성적 데이터가 저장 및 처리 측면에서 더 큰 비용이 드는 경우가 많음

# 학습내용

### 1. 데이터베이스의 필요성

- ① 데이터와 정보
- ② 정보 시스템과 데이터베이스

### 2. 데이터베이스의 정의와 특징

- ① 데이터베이스의 정의 : 공유, 통합, 저장, 운영
- ② 데이터베이스의 특징 : 실시간 접근, 계속 변화, 동시 공유, 내용 참조

### 3. 데이터 과학 시대의 데이터

- ① 형태에 따른 데이터 분류 : 정형, 반정형, 비정형
- ② 특성에 따른 데이터 분류 : 범주형(명목형, 순서형), 수치형(이산형, 연속형)





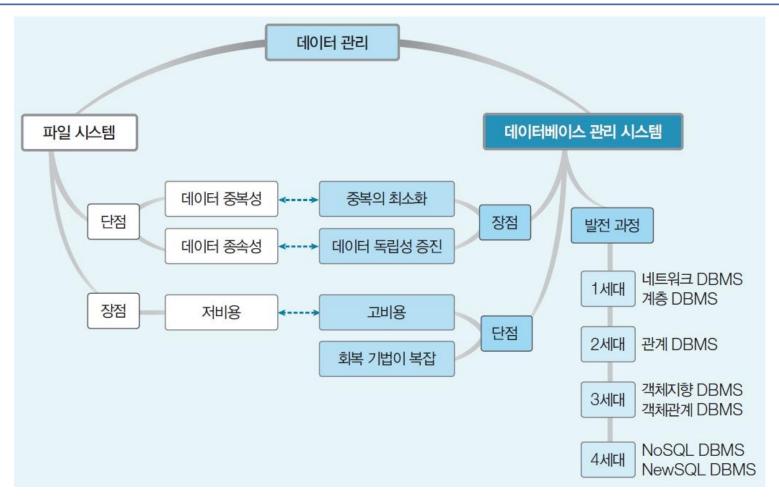
# 2장 데이터베이스 관리 시스템

- 01 데이터베이스 관리 시스템의 등장 배경
- 02 데이터베이스 관리 시스템의 정의
- 03 데이터베이스 관리 시스템의 장·단점
- 04 데이터베이스 관리 시스템의 발전 과정

# 학습목표







- 파일 시스템의 문제점과 데이터베이스 관리 시스템의 필요성을 알아본다.
- 데이터베이스 관리 시스템의 필수 기능을 살펴본다.
- ㆍ 데이터베이스 관리 시스템의 장단점을 알아본다.
- 데이터베이스 관리 시스템의 발전 과정을 살펴본다.

## 01 데이터베이스 관리 시스템의 등장 배경





### ◆ 파일 시스템(file system)

- 데이터를 파일로 관리하기 위해 <u>파일을 생성·삭제·수정·검색</u>하는 기능을 제공하는 소프트웨어
- 응용 프로그램별로 필요한 데이터를 별도의 파일로 관리함

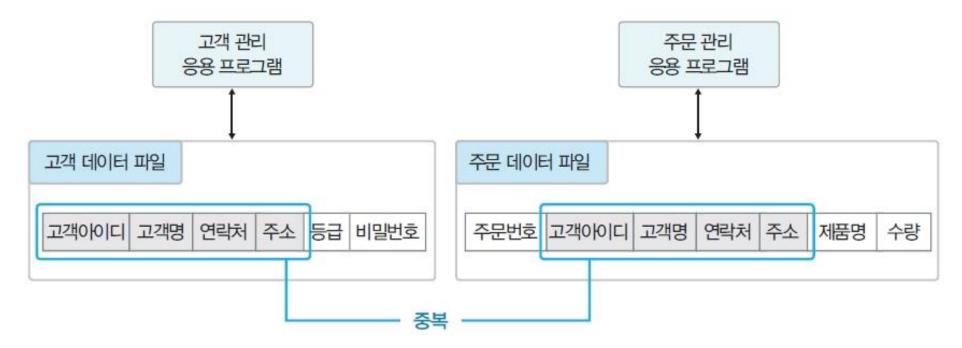


그림 2-1 파일 시스템에서의 데이터 관리

## 01 데이터베이스 관리 시스템의 등장 배경





### ◆ 파일 시스템의 문제점

- 같은 내용의 데이터가 여러 파일에 중복 저장된다.
- 응용 프로그램이 데이터 파일에 종속적이다.
- 데이터 파일에 대한 동시 공유, 보안, 회복 기능이 부족하다.
- 응용 프로그램을 개발하기 쉽지 않다.

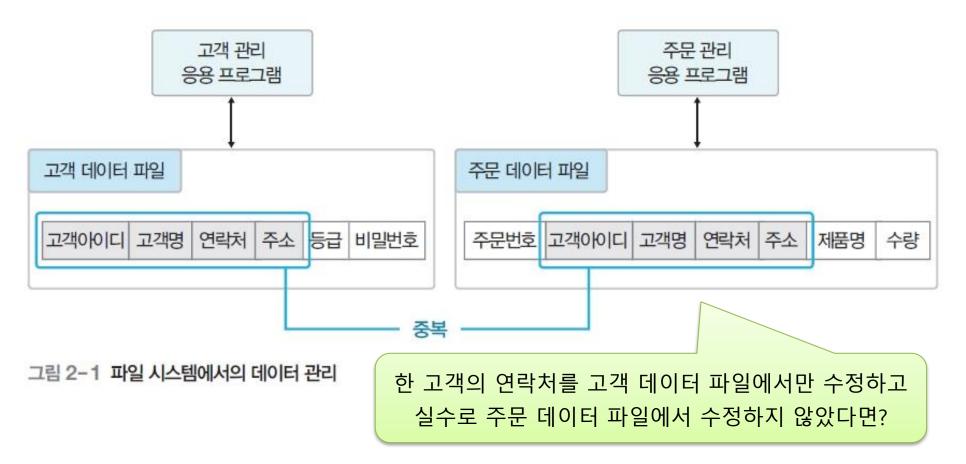
# 01 데이터베이스 관리 시스템의 등장 배경





### ◆ 파일 시스템의 주요 문제점

- 같은 내용의 데이터가 여러 파일에 중복 저장된다 → 데이터 중복성
  - 저장 공간의 낭비는 물론 데이터 일관성과 데이터 무결성을 유지하기 어려움



### 01 데이터베이스 관리 시스템의 등장 배경





#### ◆ 파일 시스템의 주요 문제점

■ 데이터 중복성 문제를 해결하는 방법은?

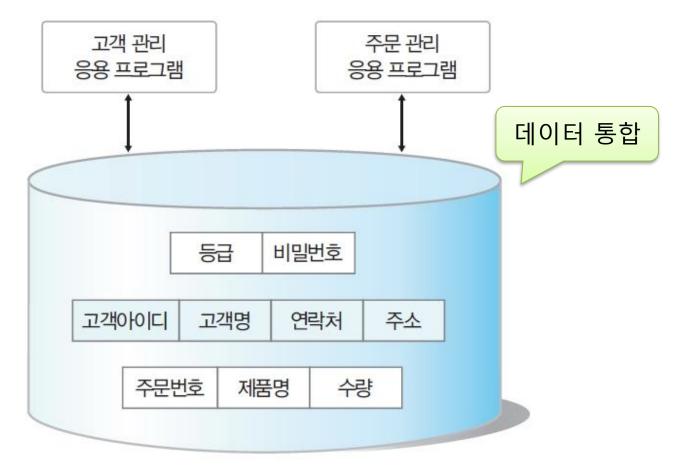


그림 2-2 파일 시스템의 데이터 중복성 문제를 해결하는 1차 방안

### 01 데이터베이스 관리 시스템의 등장 배경





#### ◆ 파일 시스템의 주요 문제점

- 응용 프로그램이 데이터 파일에 종속적이다 → 데이터 종속성
  - 사용하는 파일의 구조를 변경하면 응용 프로그램도 함께 변경해야 함

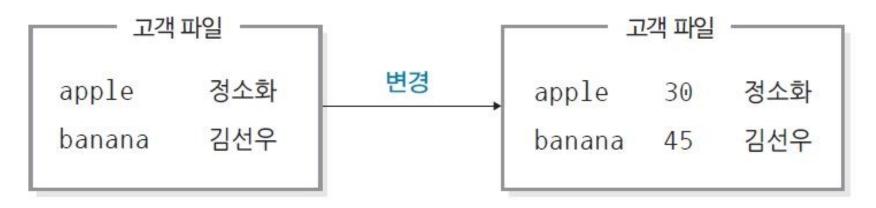


그림 2-3 파일 구조 변경 예

### 02 데이터베이스 관리 시스템의 정의





### ◆ 데이터베이스 관리 시스템

- DBMS(DataBase Management System)
- 파일 시스템의 문제를 해결하기 위해 제시된 소프트웨어
- 조직에 필요한 데이터를 데이터베이스에 통합하여 저장하고 관리함

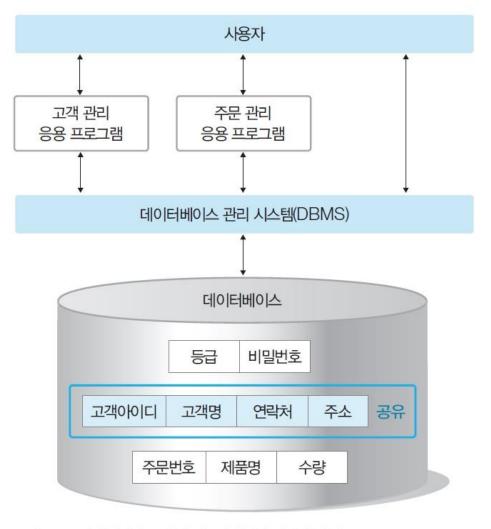


# 02 데이터베이스 관리 시스템의 정의





### ◆ 데이터베이스 관리 시스템에서의 데이터 관리



### 02 데이터베이스 관리 시스템의 정의





#### ◆ 데이터베이스 관리 시스템의 주요 기능

정의 기능 데이터베이스 구조를 정의하거나 수정할 수 있다.

조작 기능 데이터를 삽입·삭제·수정·검색하는 연산을 할 수 있다.

제어 기능 데이터를 항상 정확하고 안전하게 유지할 수 있다.

그림 2-6 데이터베이스 관리 시스템의 주요 기능





#### ♦장점

- 데이터 중복을 통제할 수 있다.
  - 데이터베이스에 데이터를 통합하여 관리하므로 데이터 중복 문제 해결
- 데이터 독립성이 확보된다.
  - 응용 프로그램 대신 데이터베이스에 접근하고 관리하는 모든 책임 담당
    - → 응용 프로그램과 데이터베이스 사이에 독립성이 확보됨
- 데이터를 동시 공유할 수 있다.
  - 동일한 데이터를 여러 응용 프로그램이 공유하여 동시 접근할 수 있게 지원
    - → 동시 접근 제어 기술 보유
- 데이터 보안이 향상된다.
  - 중앙 집중식으로 데이터를 관리하므로 효율적인 접근 제어 가능
    - 권한이 없는 사용자의 접근, 허용되지 않은 데이터와 연산에 대한 요청 차단





### ♦장점

- 데이터 무결성을 유지할 수 있다.
  - 데이터 삽입·수정 등의 연산이 수행될 때마다 유효성을 검사하여 데이터 무결성(정확성)을 유지
- 표준화할 수 있다.
  - 데이터베이스 관리 시스템이 정한 표준화된 방식을 통해 데이터베이스에 접근
- 장애 발생 시 회복이 가능하다.
  - 데이터 일관성과 무결성을 유지하면서 장애 발생 이전 상태로 데이터를 복구하는 회복 기능 지원
- 응용 프로그램 개발 비용이 줄어든다.
  - 파일 시스템을 사용할 때보다 데이터 관리 부담이 줄어 응용 프로그램 개발 비용 및 유지 보수 비용이 줄어듦





### ♦ 단점

- 비용이 많이 든다.
  - 별도 구매 비용이 들고, 동시 사용이 허용되는 사용자 수에 따라 가격 증가
- 백업과 회복 방법이 복잡하다.
  - 장애 발생의 원인과 상태를 정확히 파악하기 어렵고 회복 방법도 복잡함
- 중앙 집중 관리로 인한 취약점이 존재한다.
  - 데이터베이스나 데이터베이스 관리 시스템에 장애가 발생하면 전체 시스템의 업무 처리가 중단됨
  - 데이터베이스 의존도가 높은 시스템일수록 가용성과 신뢰성에 치명적임



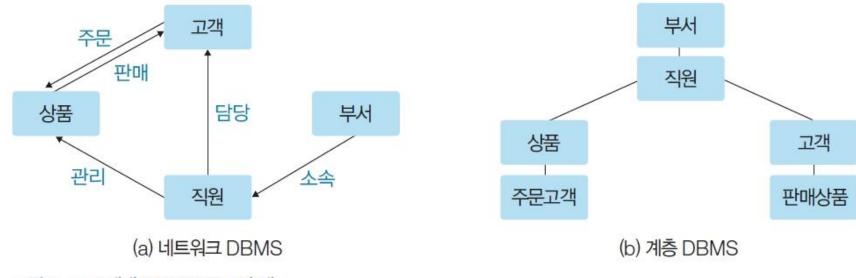


#### 장점 단점 □ 데이터 중복을 통제할 수 있다 □ 비용이 많이 든다 □ 데이터 독립성이 확보된다 □ 백업과 회복 방법이 복잡하다 □ 데이터를 동시 공유할 수 있다 □ 중앙 집중 관리로 인한 취약점이 존재한다 □ 데이터 보안이 향상된다 □ 데이터 무결성을 유지할 수 있다 표준화할 수 있다 □ 장애 발생 시 회복이 가능하다 □ 응용 프로그램 개발 비용이 줄어든다 그림 2-7 데이터베이스 관리 시스템의 장점과 단점





- ◆ 1세대 : 네트워크 DBMS, 계층 DBMS
  - 네트워크 DBMS : 데이터베이스를 그래프 형태로 구성
    - 예) IDS(Integrated Data Store)
  - 계층 DBMS : 데이터베이스를 트리 형태로 구성
    - 예) IMS(Information Management System)







#### ◆ 2세대 : 관계 DBMS

- 관계 DBMS : 데이터베이스를 테이블 형태로 구성
  - 예) 오라클(Oracle), MS SQL 서버, 액세스(Access), 인포믹스(Informix), MySQL, 마리아DB(MariaDB)

아이디	비밀번호	0름	연락처	주소	적립금
apple	1234	정소화	02-111-1111	서울시 마포구	1000
banana	9876	김선우	02-222-2222	경기도 부천시	500

그림 2-9 관계 DBMS의 테이블 예: 고객 테이블





- ◆ 3세대: 객체지향 DBMS, 객체관계 DBMS
  - 객체지향 DBMS : 객체를 이용해 데이터베이스를 구성
    - 예) 오투(O2), 온투스(ONTOS), 젬스톤(GemStone)
  - 객체관계 DBMS : 객체지향 DBMS + 관계 DBMS
    - 예) 관계 DBMS 제품들이 객체지향 기능을 지원하면서 객체관계 DBMS로 분류되기도 함(오라클이 대표적)





#### ◆ 4세대: NoSQL • NewSQL DBMS

- NoSQL DBMS : 비정형 데이터를 처리하는데 적합하고 확장성이 뛰어남
  - 안정성과 일관성 유지를 위한 복잡한 기능 포기
  - 데이터 구조를 미리 정해두지 않는 유연성을 가짐
  - 여러 대의 컴퓨터에 데이터를 분산하여 저장하고 처리하는 환경에서 주로 사용
  - 예) 몽고디비(MongoDB), H베이스(HBase), 카산드라(Cassandra), 레디스(Redis), 네오포제이(Neo4j), 오리엔트DB(OrientDB) 등
- NewSQL DBMS: 관계 DBMS의 장점 + NoSQL의 확장성 및 유연성
  - 정형 및 비정형 데이터를 안정적이고 빠르게 처리 가능
  - 예) 구글 스패너(Spanner), 볼트DB(VoltDB), 누오DB(NuoDB)





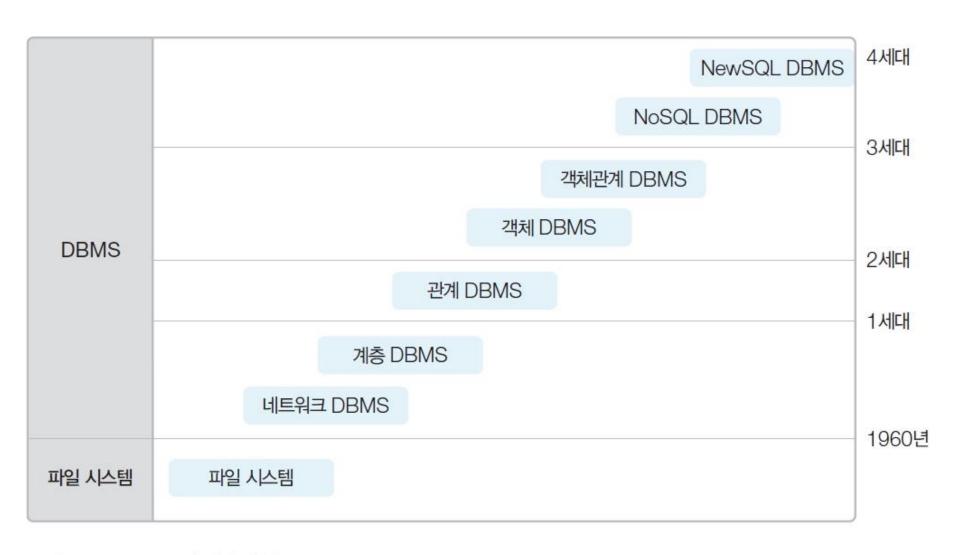


그림 2-10 DBMS의 발전 과정

# 학습내용

- 1. DBMS 등장 배경
- 2. DBMS의 정의
- 3. DBMS의 장단점
- 4. DBMS의 발전과정
  - ① 1세대 DBMS : 네트워크, 계층 DBMS
  - ② 2세대 DBMS : 관계 DBMS
  - ③ 3세대 DBMS : 객체지향, 객체관계 DBMS
  - ④ 4세대 DBMS : NoSQL, NewSQL DBMS

다음 시간에는

# 3장 데이터베이스 시스템

에 대해 학습해 보겠습니다.

