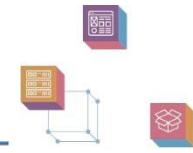




데이터베이스 기본 개념

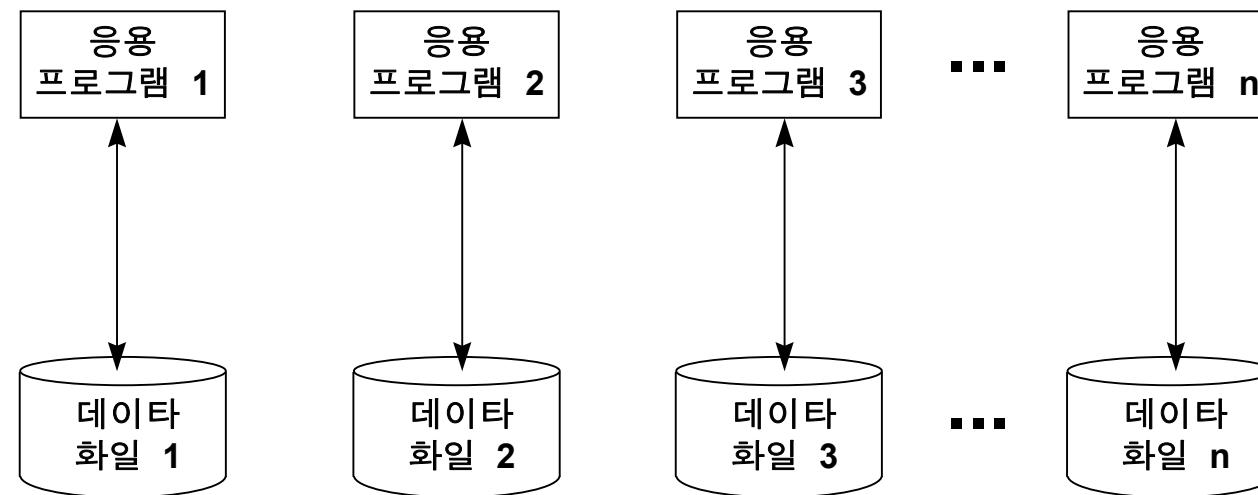
- 01 데이터베이스의 필요성
- 02 데이터베이스의 정의와 특징
- 03 데이터 과학 시대의 데이터

파일과 데이터베이스

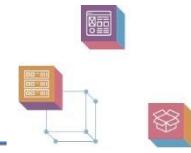


◆ 파일 처리 시스템

- 각 응용 프로그램이 개별적인 파일 처리 루틴을 작성
- 응용 프로그래머에 부담

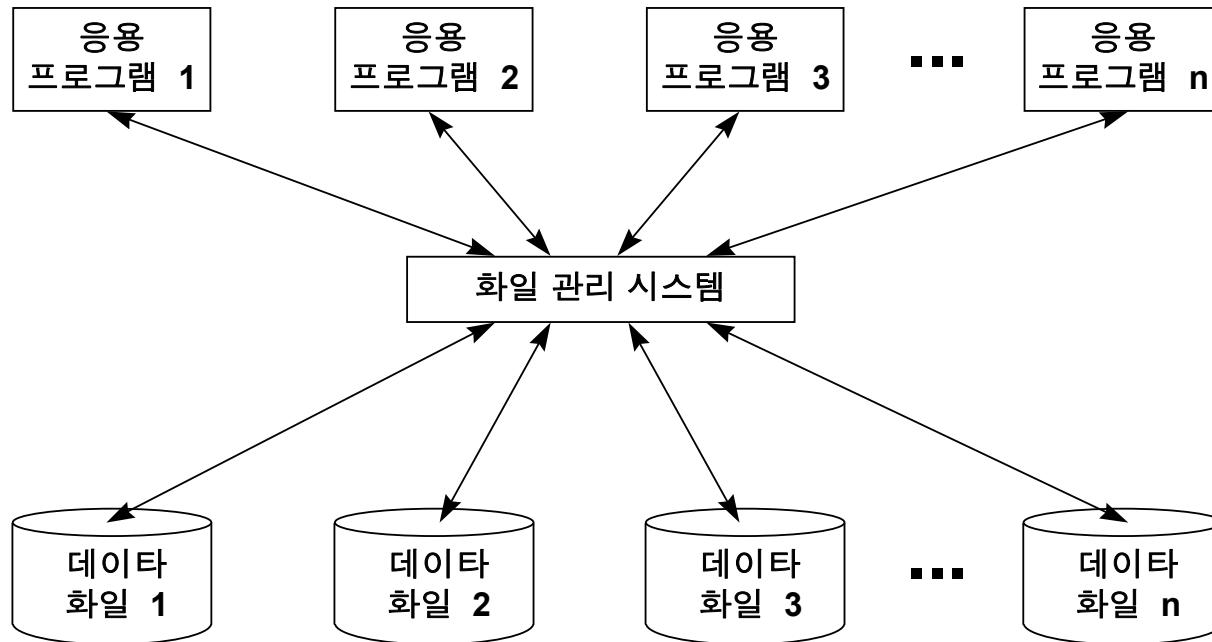


파일과 데이터베이스

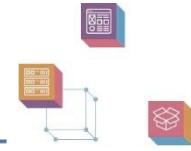


◆ 파일 관리 시스템(file management system)

- 파일 처리 루틴 공유
- 데이터 파일의 공용은 없음



파일/파일 시스템 단점



- ◆ 중복성: 같은 내용의 데이터를 여러 파일에 중복 저장
 - 급여 파일

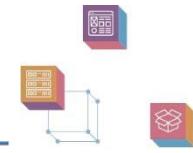
교수 번호	이름	학과	호봉	봉급	공제액	지급액	주민 등록번호	주소
-------	----	----	----	----	-----	-----	---------	----

- 인사 파일

교수 번호	이름	학과	호봉	주민 등록 번호	연구실 번호	주소	경력
-------	----	----	----	----------	--------	----	----

====> 저장 공간의 낭비, 데이터 모순성(data inconsistency) 야기.

파일/파일 시스템 단점



◆ 데이터 종속성

- 응용 프로그램(application programs)과 데이터(data) 간의 상호 의존 관계
- 데이터의 구성 방법이나 접근 방법의 변경 시 관련 응용 프로그램도 같이 변경

DBMS

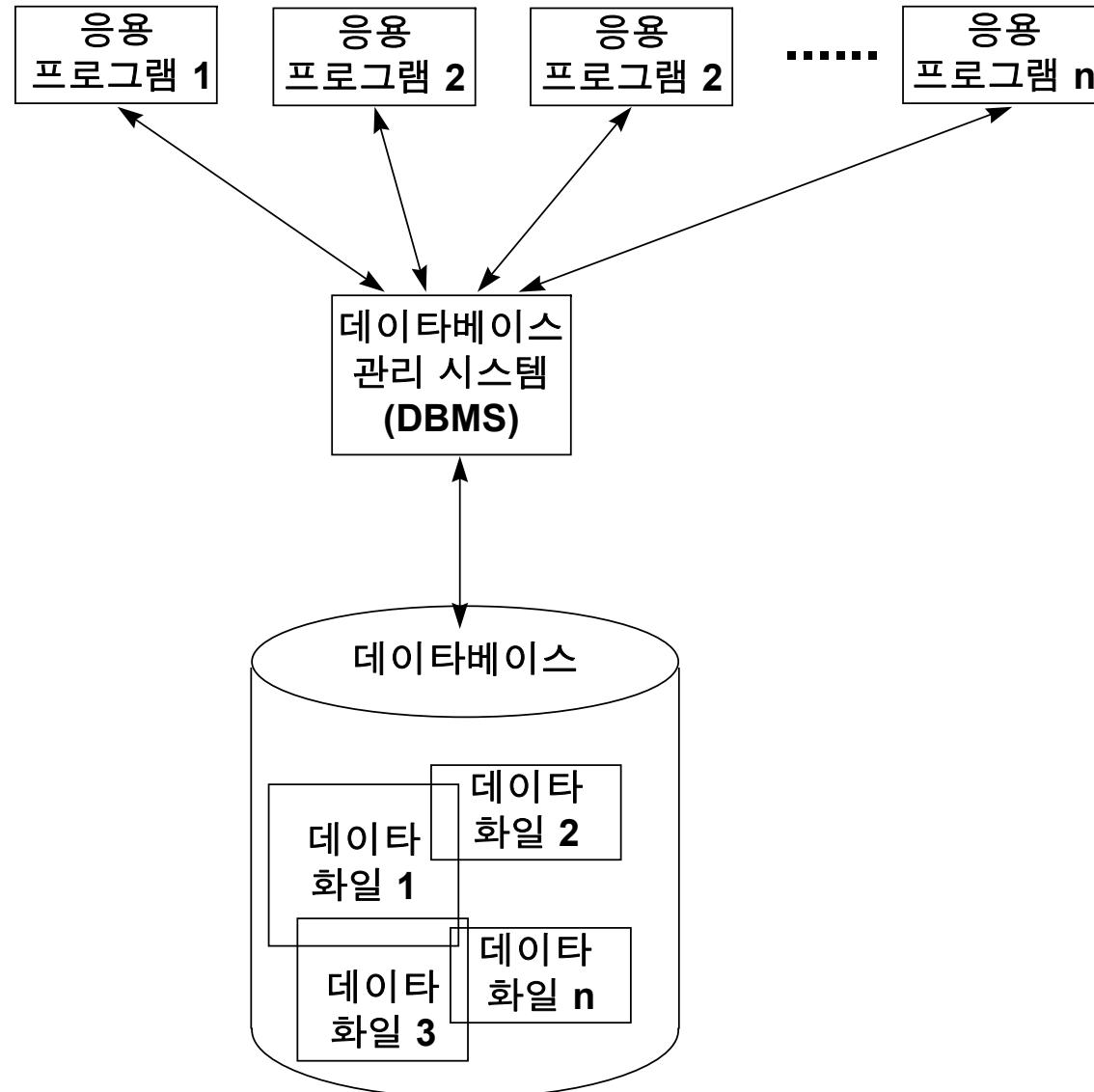
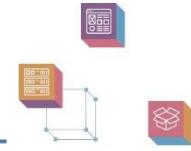


◆ 데이터베이스 관리 시스템(DBMS: database management system)

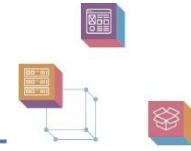
- 응용 프로그램과 데이터의 중재자로서 모든 응용 프로그램들이 **데이터베이스**(database)를 공용할 수 있게 관리해 주는 소프트웨어 시스템

- 기능
 - 데이터의 정의(data definition), 표현, 저장
 - 데이터 조작(data manipulation)
 - 보안, 회복, 공용 제어, 무결성 기법 등을 포함한 데이터 제어(data control)
 - 사용자 인터페이스(user interface) 제공

DBMS



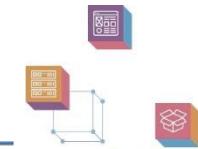
01 데이터베이스의 필요성



◆ 데이터와 정보

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

01 데이터베이스의 필요성



◆ 데이터와 정보

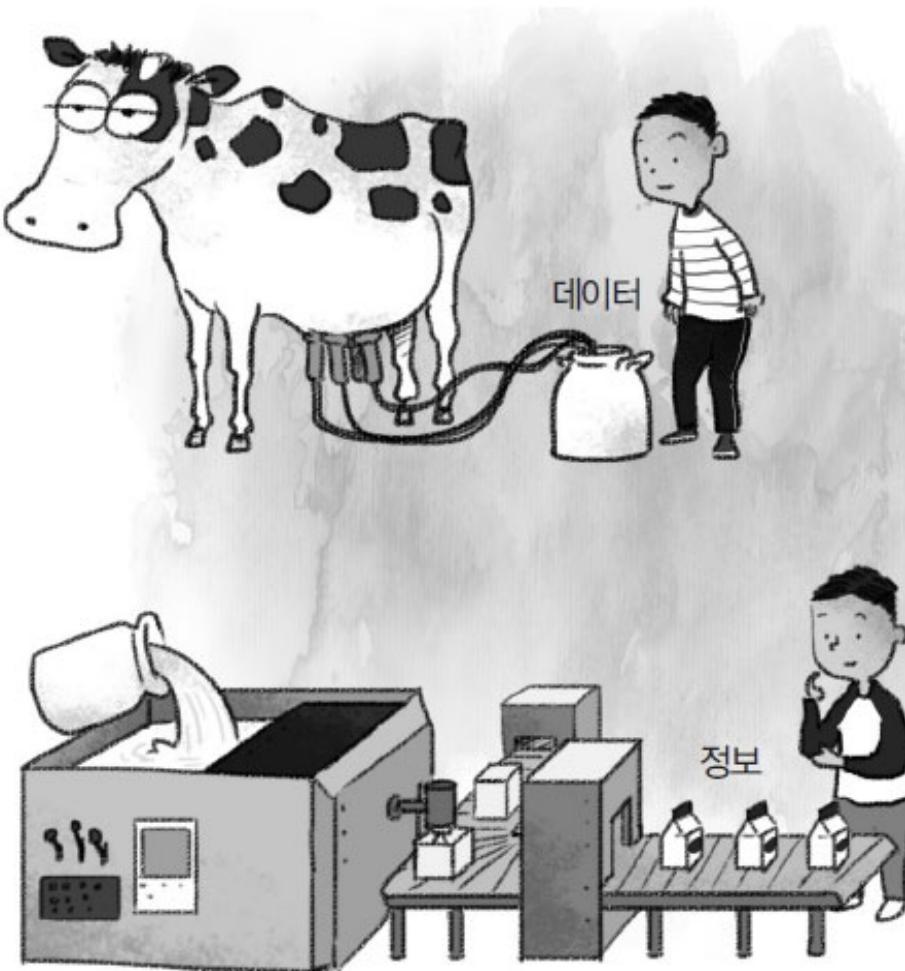
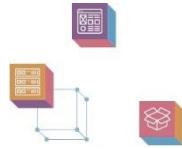


그림 1-1 데이터와 정보의 이해 : 원유와 가공 우유

01 데이터베이스의 필요성



◆ 정보 처리(information processing)

- 데이터에서 정보를 추출하는 과정 또는 방법

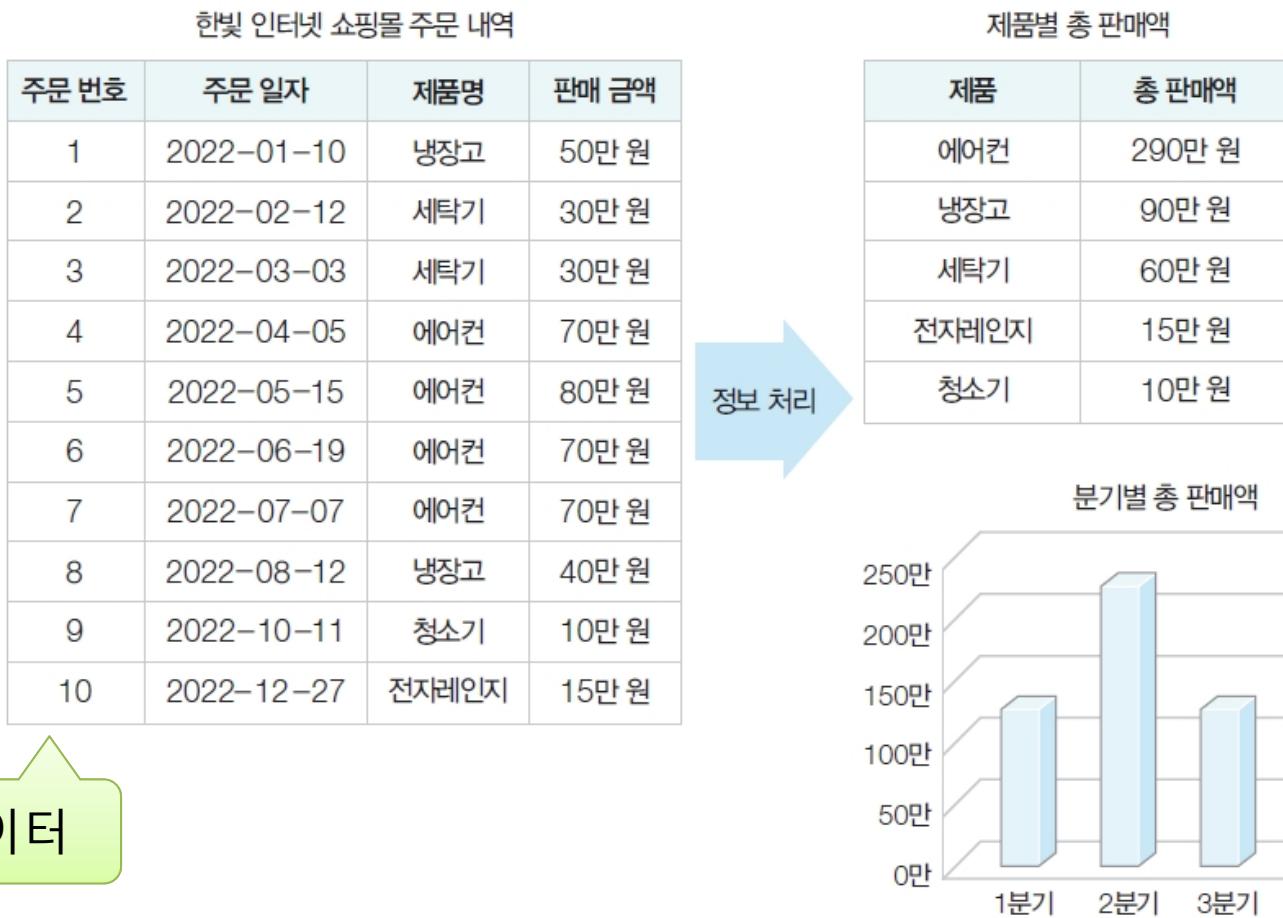
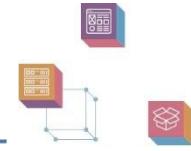


그림 1-2 정보 처리의 예

01 데이터베이스의 필요성



◆ 정보 시스템과 데이터베이스

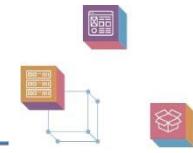
■ 정보 시스템(information system)

- 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

01 데이터베이스의 필요성



◆ 정보 시스템과 데이터베이스

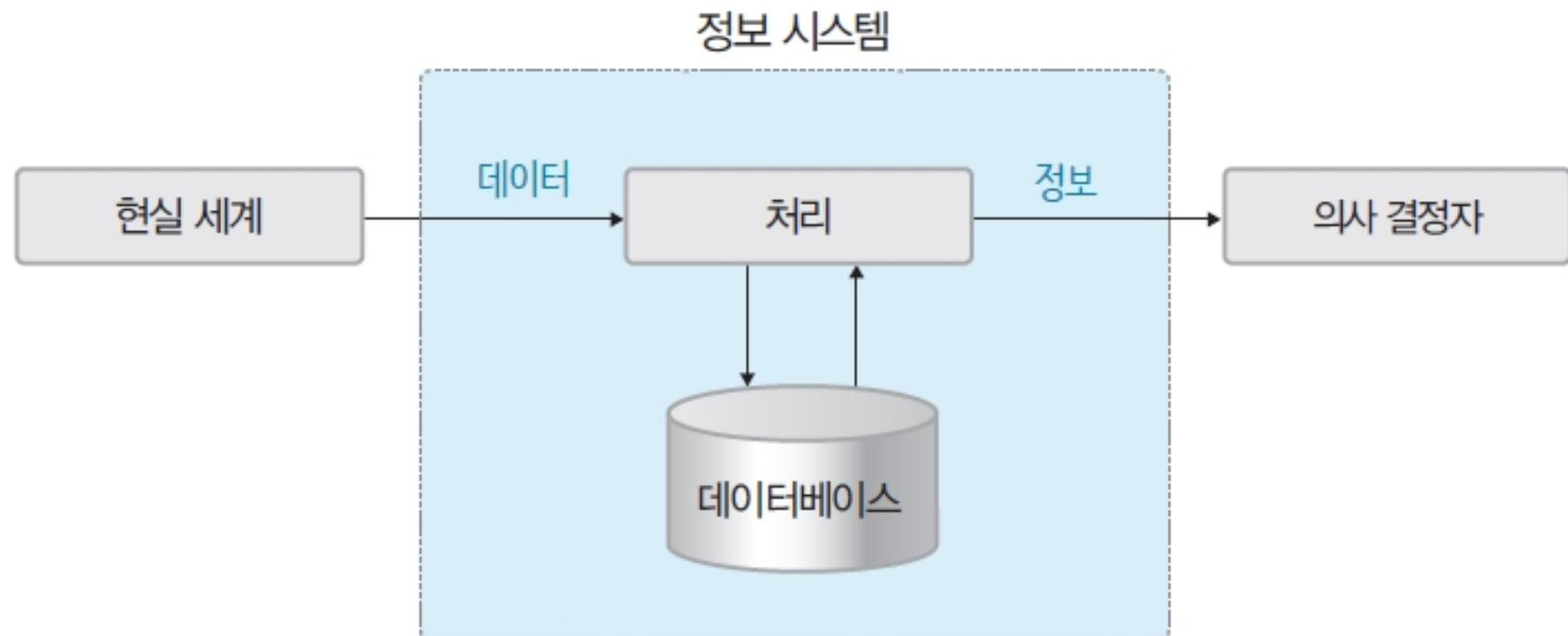
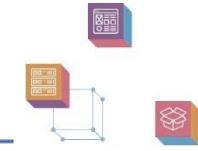


그림 1-3 정보 시스템의 역할과 구성

02 데이터베이스의 정의와 특징



◆ 데이터베이스(DB; DataBase)

- 특정 조직의 여러 사용자가 **공유**하여 사용할 수 있도록 **통합**해서 **저장**한 **운영** 데이터의 집합

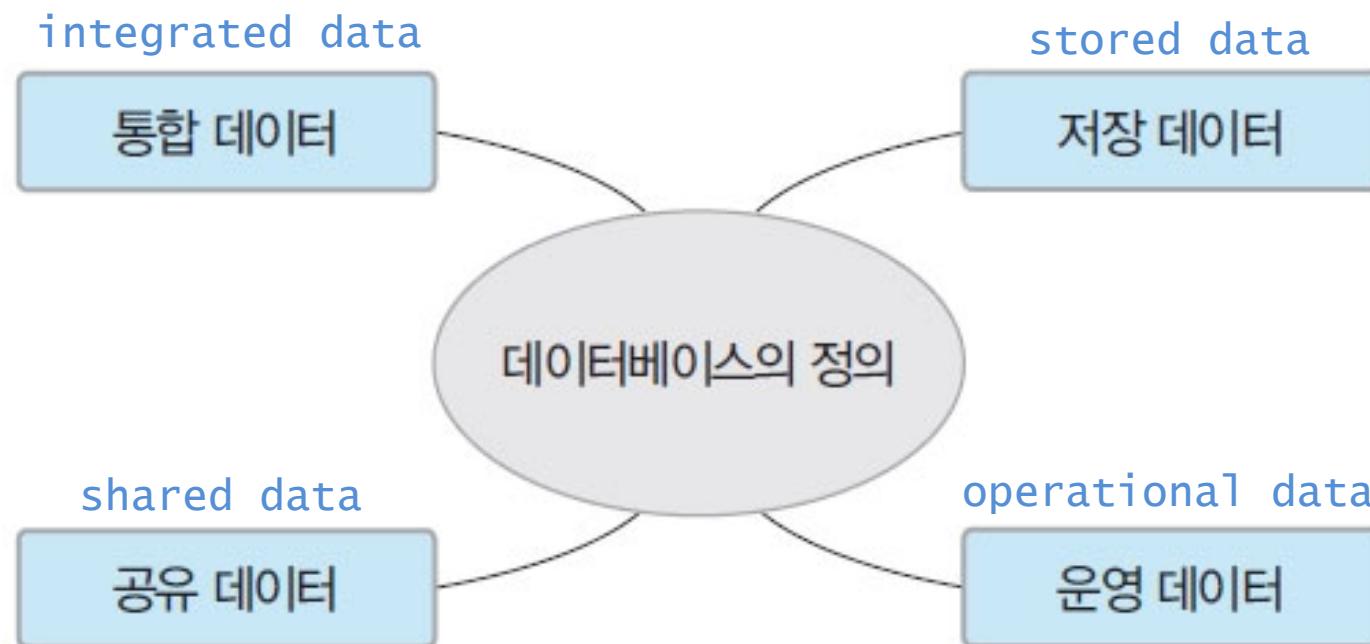
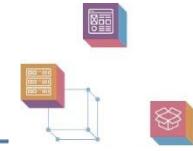


그림 1-4 데이터베이스의 정의

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

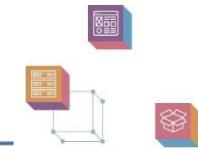
◆ 저장 데이터

- 컴퓨터가 접근할 수 있는 매체에 저장된 데이터

◆ 운영 데이터

- 조직의 주요 기능을 수행하기 위해 지속적으로 꼭 필요한 데이터

02 데이터베이스의 정의와 특징



◆ 데이터베이스의 특징

real-time accessibility

실시간 접근성

continuous evolution

계속 변화

데이터베이스의 특징

content reference

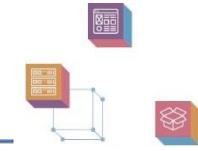
내용 기반 참조

concurrent sharing

동시 공유

그림 1-5 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 동시 사용도 지원

◆ 내용 기반 참조

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

02 데이터베이스의 정의와 특징

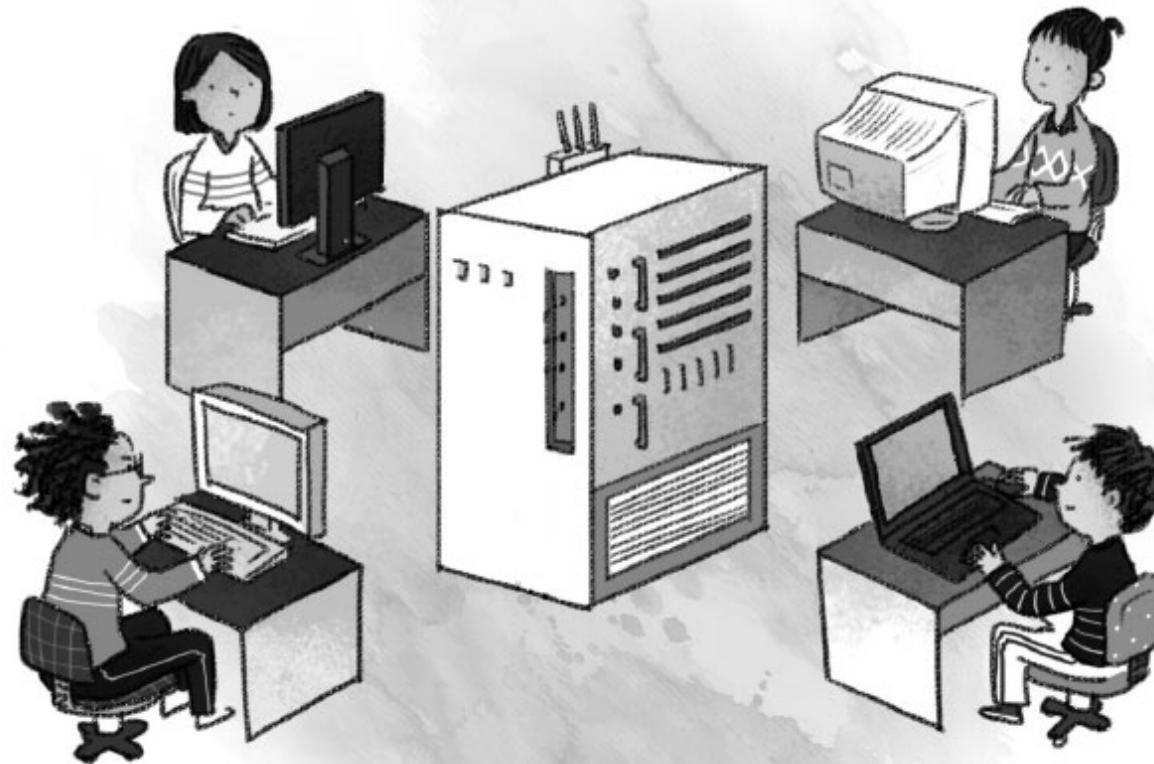
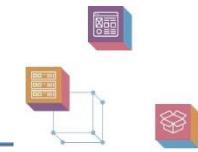
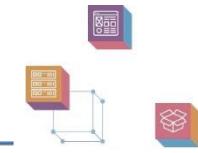


그림 1-6 데이터베이스의 이용

03 데이터 과학 시대의 데이터



◆ 형태에 따른 데이터 분류

- 정형 데이터
- 반정형 데이터
- 비정형 데이터

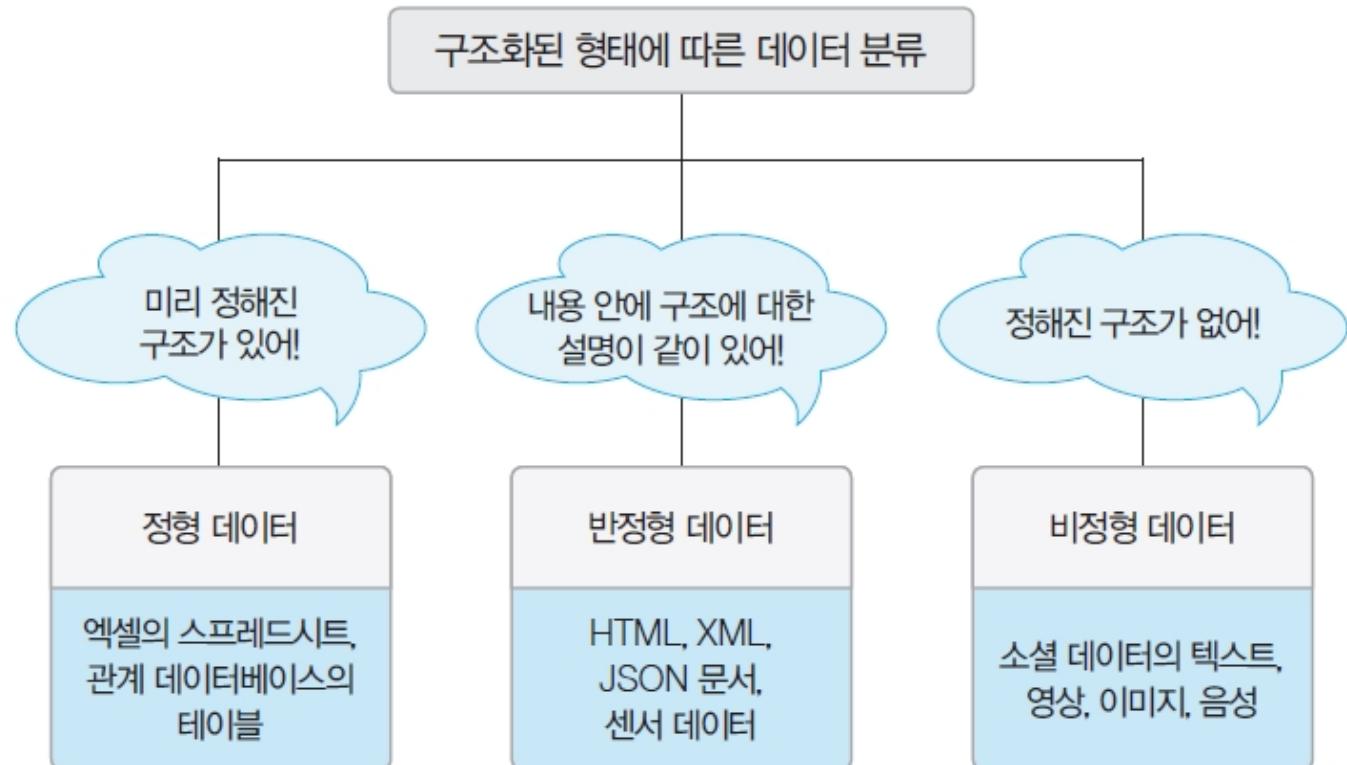
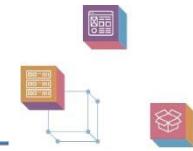


그림 1-7 형태에 따른 데이터 분류

03 데이터 과학 시대의 데이터



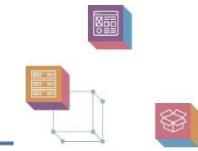
◆ 정형 데이터(structured data)

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

	A	B	C	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

03 데이터 과학 시대의 데이터



◆ 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"  
}
```

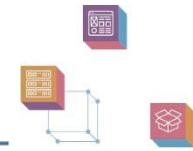
(a) JSON

```
<친구정보>  
    <이름> 오형준 </이름>  
    <나이> 23 </나이>  
    <성별> 남 </성별>  
</친구정보>
```

(b) XML

그림 1-9 반정형 데이터의 예

03 데이터 과학 시대의 데이터



◆ 비정형 데이터(unstructured data)

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

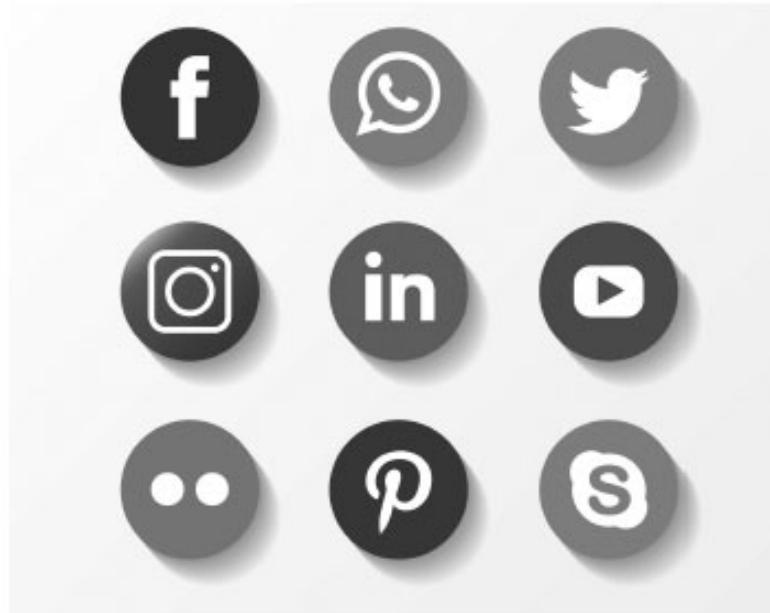
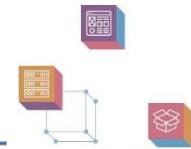


그림 1-10 비정형 데이터의 예(Designed by S.salvador / Freepik)

03 데이터 과학 시대의 데이터



◆ 특성에 따른 데이터 분류

- 범주형 데이터
- 수치형 데이터

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

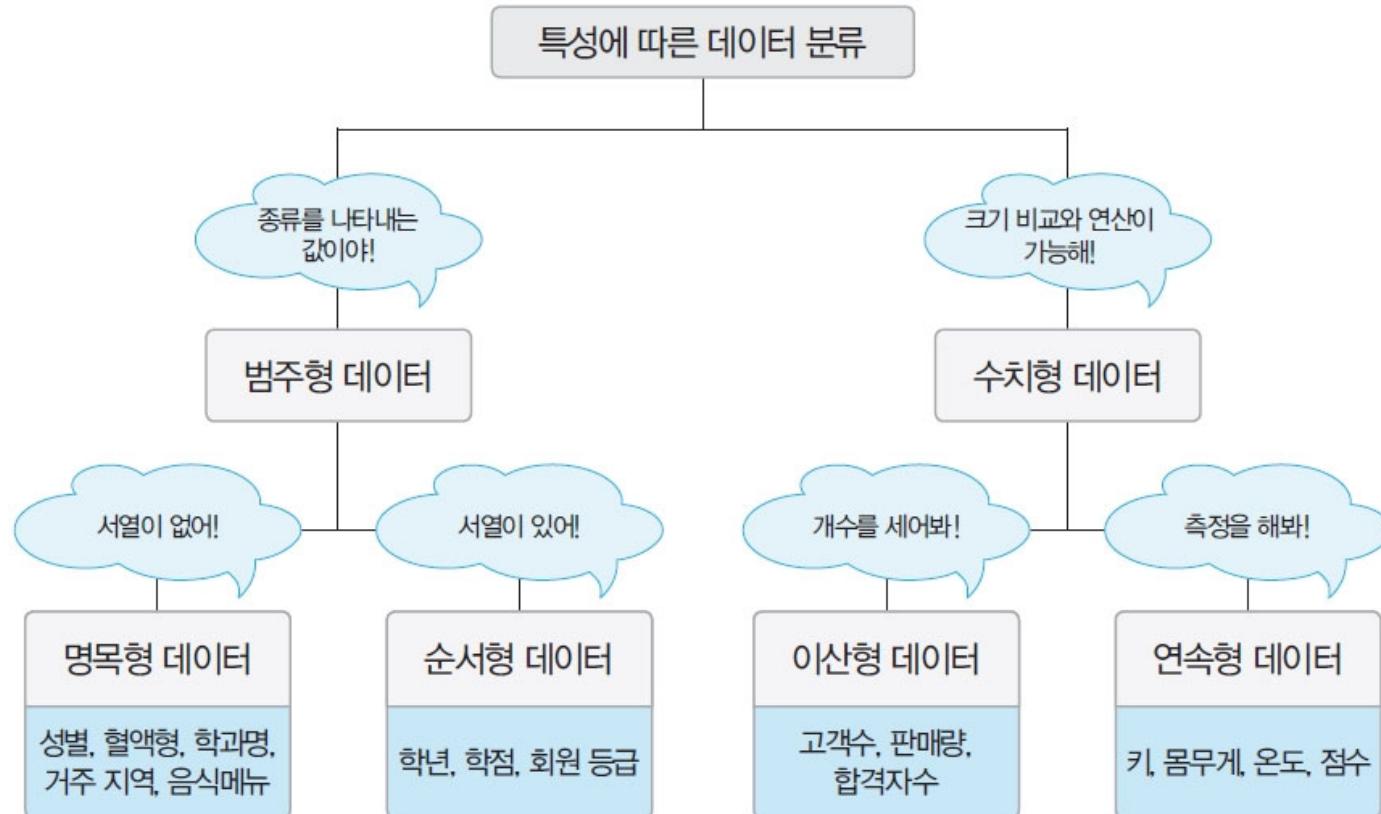
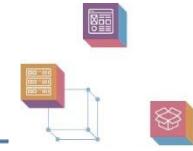


그림 1-11 특성에 따른 데이터 분류

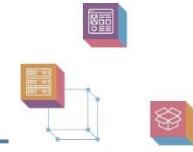
03 데이터 과학 시대의 데이터



◆ 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

03 데이터 과학 시대의 데이터



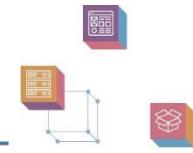
◆ 범주형 데이터

명목형 데이터 서열이 없는 값을 가지는 데이터

순서형 데이터 서열이 있는 값을 가지는 데이터

그림 1-12 범주형 데이터

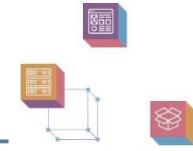
03 데이터 과학 시대의 데이터



◆ 수치형 데이터(numerical data)

- 크기 비교와 산술적인 연산이 가능한 숫자 값을 가진 데이터
- 양적 데이터라고도 함
- 이산형 데이터(discrete data)
 - 개수를 셀 수 있는 띠엄띄엄 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

03 데이터 과학 시대의 데이터



◆ 수치형 데이터

이산형 데이터

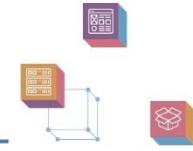
단절된 숫자 값을 가지는 데이터

연속형 데이터

연속적으로 이어진 숫자 값을 가지는 데이터

그림 1-13 수치형 데이터

03 데이터 과학 시대의 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

IT@COOKBOOK 데이터베이스 개론(3판)

[강의교안 이용 안내]

- 본 강의교안의 저작권은 한빛아카데미㈜에 있습니다.
- 이 자료를 무단으로 전제하거나 배포할 경우 저작권법 136조에 의거하여 최고 5년 이하의 징역 또는 5천만원 이하의 벌금에 처할 수 있고 이를 병과(併科)할 수도 있습니다.