

김현승, 「디지털 스토리텔링」, 한국전통문화대학교 수업 자료, 2024.

# 데이터베이스 개론

---

김현승

한국학중앙연구원 디지털인문학연구소

2024.04.30.

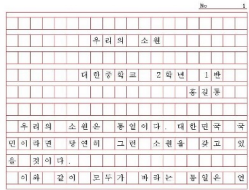


01

# 데이터베이스 구축의 필요성

# Contents

## 저장 매체의 변화



**종이**  
시대 : ??~??  
저장용량 : ??장, ??권



**자기 테이프**  
시대 : ~60년대  
저장용량 : 1.44 MB



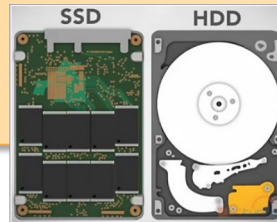
**천공 카드(punched card)**  
시대 : ~70년대  
저장용량 : 1.44 MB



**플로피 디스크**  
시대 : 80~90년대  
저장용량 : 1.44 MB



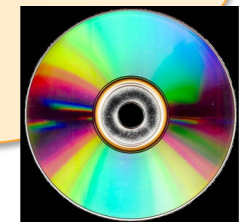
**클라우드 드라이브**  
시대 : 2010~현재  
저장용량 : ???(무한대?)



**반도체 드라이브**  
시대 : 2005~현재  
저장용량 : 500MB~??TB



**하드디스크**  
시대 : 1990~현재  
저장용량 : 10MB~??TB



**CDROM/DVD**  
시대 : 1990~현재?  
저장용량 : 300MB~2.5T

# Contents

## 데이터베이스 구축의 필요성

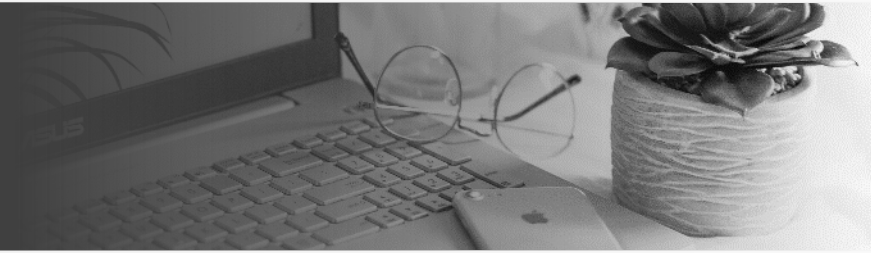


- 데이터 중복 최소화
- 중복 최소화로 데이터 저장 공간 절약
- 데이터 공유 가능
- 일관성, 무결성, 보안성 유지
- 최신 데이터 유지
- 데이터의 논리적, 물리적 독립성 확보
- 쉬운 데이터 접근



# Contents

## 데이터베이스 구축의 필요성



우리가 살고 있는 시대에는 개인과 조직들이 스스로 원하는 업무를 컴퓨터와 네트워크를 통해 처리하고 있습니다. 과거에 종이 서류와 필기 등 사람들의 수작업으로 의해 이루어진 대부분의 작업들이 컴퓨터와 네트워크를 기반으로 움직이는 데이터에 의해 처리됩니다.

실생활 속에서 필요한 각종 정보를 담은 데이터를 컴퓨터에 반영하여 처리하는 시스템을 정보 시스템이라고 합니다. 이 정보 시스템을 이용하게 됨으로써 예전에는 수작업으로 장부를 기록하고, 재고를 기록하던 단순한 일들을 요즘은 컴퓨터를 이용해 프로그램과 데이터로 처리하게 되었습니다.

정보 시스템을 이용할 경우 처리 속도를 단축시킬 수 있음은 물론이고 단순 업무에서 해방될 수 있어서 인간은 좀 더 가치로운 곳에 자신의 능력을 투자할 수 있게 되었습니다.

이영걸, 『SQL Server 2008』, 길벗, 2009, 14쪽.



# Contents

## 데이터베이스 구축의 필요성



디지털 인문학을 연구하는 시각에서, SQL을 다룰 수 있다고 하는 것이 의미하는 바가 무엇인지 떠올려 본다면 아마도 디지털 인문학이 인문지식 데이터의 기반 위에서 연구되고 활용되기 때문일 것입니다.

한국연구재단의 2014년 인문학대중화 사업 중 디지털 인문학 분야의 “조선시대 표류 노드 시각망 구축”, 2015년 토대연구지원사업인 “수신사 및 조사시찰단 자료 DB 구축” 등의 사업들도 모두 인문지식 데이터를 체계적으로 수집 관리하고, 다양한 목적으로 그것을 분석, 종합할 수 있는 데이터베이스의 기술적 도움이 있었기 때문에 가능했습니다.

김현 외 2명, 『디지털 인문학 입문』, 한국외국어대학교 지식출판부, 2016.





02

## 데이터베이스 개론



02

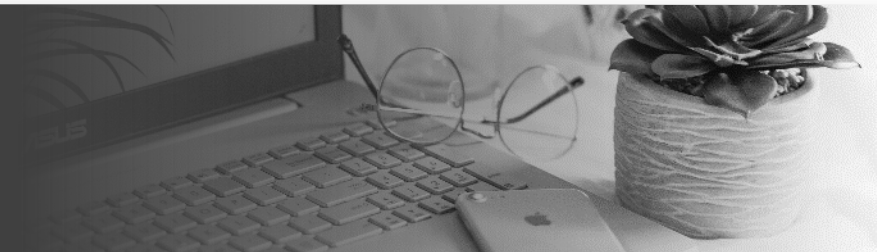
## 디지털베이스 개론

# Part 1 데이터베이스란?

데이터베이스의 정의와 특징

# Contents

## 데이터베이스란?

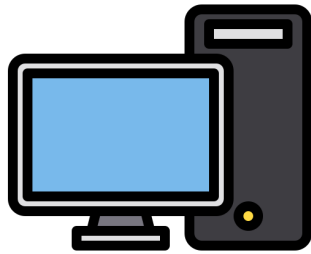


**데이터베이스(database, DB)**는 여러 사람이 공유할 목적으로 통합 관리하기 위해 논리적으로 연관된 데이터를 모아 일정한 형태로 저장해 놓은 것을 의미한다.

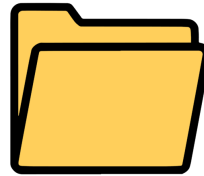
즉, 여러 시스템 또는 사용자들이 공용(shared)할 목적으로 통합(integrated), 저장(stored)한 **데이터의 집합**이다.

# Contents

## 데이터베이스란?



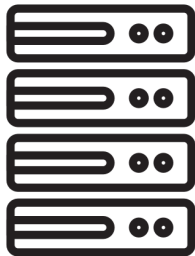
컴퓨터



폴더



파일



서버



데이터베이스



테이블

# Contents

## 데이터베이스의 정의: ISOS



정의	설명
통합된 데이터 Integrated Data	데이터 중복을 최소화한 데이터이다.
저장된 데이터 Stored Data	컴퓨터가 접근할 수 있는 저장 매체에 저장된 데이터이다.
운영 데이터 Operational Data	조직의 고유한 업무를 수행하는 데 반드시 필요한 데이터이다.
공용 데이터 Shared Data	여러 응용 시스템이 공동으로 소유하고 유지하는 데이터이다.

# Contents

## 데이터베이스의 특징: R1C3



정의	설명
실시간 접근성 Real Time Accessibility	사용자 질의에 실시간 응답으로 처리한다.
지속적인 변화 Continuous Evolution	삽입, 삭제, 수정 작업을 하여 항상 최신 데이터를 동적으로 유지한다.
동시 공유 Concurrent Sharing	목적이 서로 다른 여러 사용자가 동시에 원하는 데이터를 공유한다.
내용에 의한 참조 Content Reference	데이터베이스에 있는 데이터를 참조할 때 레코드의 주소나 위치가 아니라 사용자가 요구하는 데이터 내용을 참조한다.

# Contents

## 데이터베이스의 장점/단점



장점	단점
<ul style="list-style-type: none"><li>▪ 데이터 중복 최소화</li><li>▪ 중복 최소화로 데이터 저장 공간 절약</li><li>▪ 데이터 공유 가능</li><li>▪ 일관성, 무결성, 보안성 유지</li><li>▪ 최신 데이터 유지</li><li>▪ 데이터의 논리적, 물리적 독립성 확보</li><li>▪ 쉬운 데이터 접근</li></ul>	<ul style="list-style-type: none"><li>▪ 데이터베이스 전문 인력 필요</li><li>▪ 시스템 운영 비용 부담</li><li>▪ 데이터 백업 및 복구 어려움</li><li>▪ 시스템의 복잡함</li><li>▪ 대용량 디스크로 액세스가 집중되면서 과부하 발생</li></ul>



02

## 디지털베이스 개론

## Part 2 데이터베이스의 종류

계층형 데이터베이스

네트워크형 데이터베이스

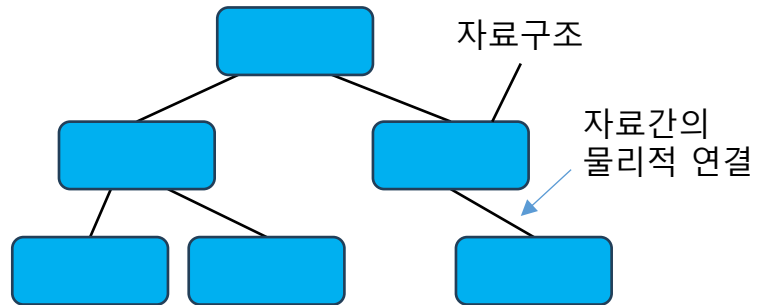
키-값 데이터베이스

관계형 데이터베이스

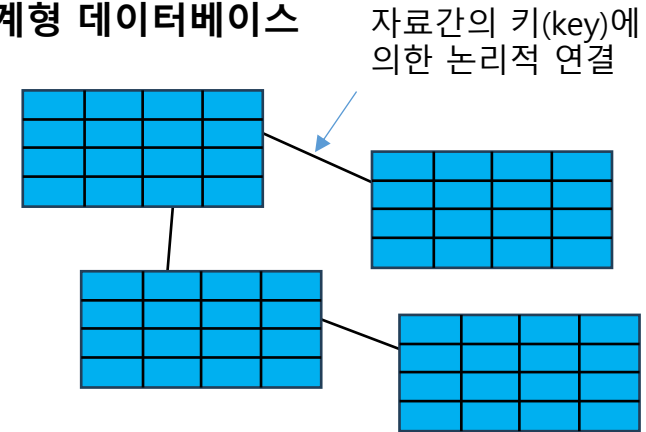
# Contents

## 데이터베이스(DB) 종류

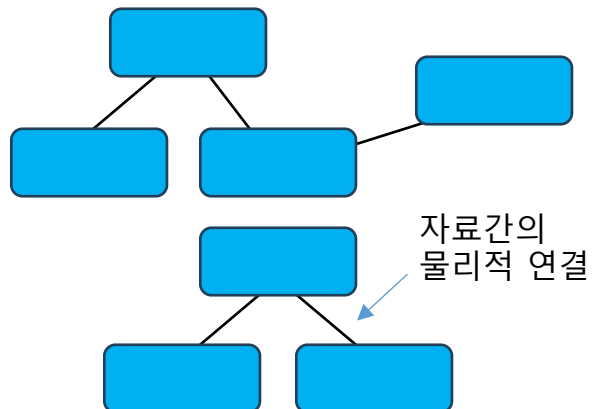
### 계층형 데이터베이스



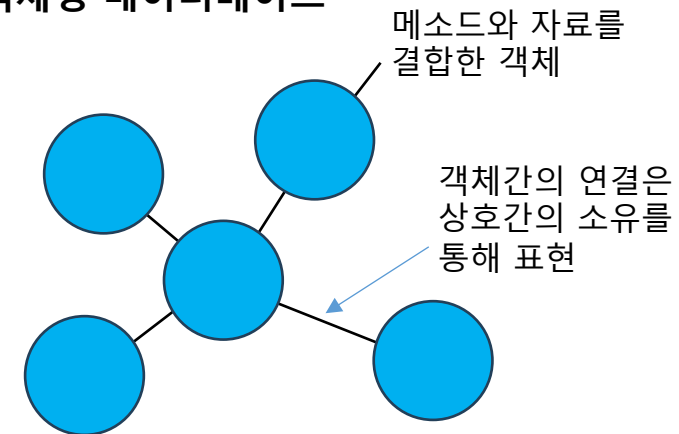
### 관계형 데이터베이스



### 네트워크형 데이터베이스



### 객체형 데이터베이스

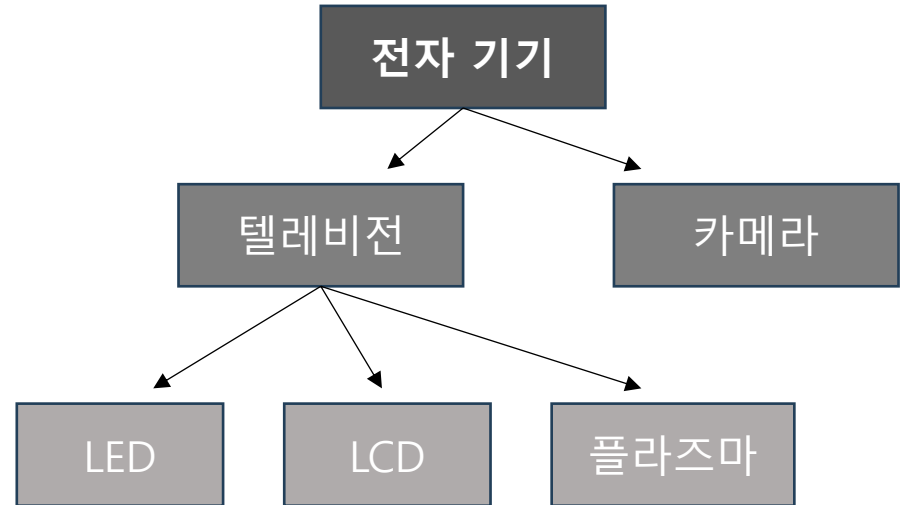


# Contents

## DB 종류: 계층형 데이터베이스

### 계층형(Hierarchical) 데이터베이스

- 처음으로 등장한 DB 개념. 1960년대에 시작.
- 부모와 자식 관계인 트리(tree) 구조.



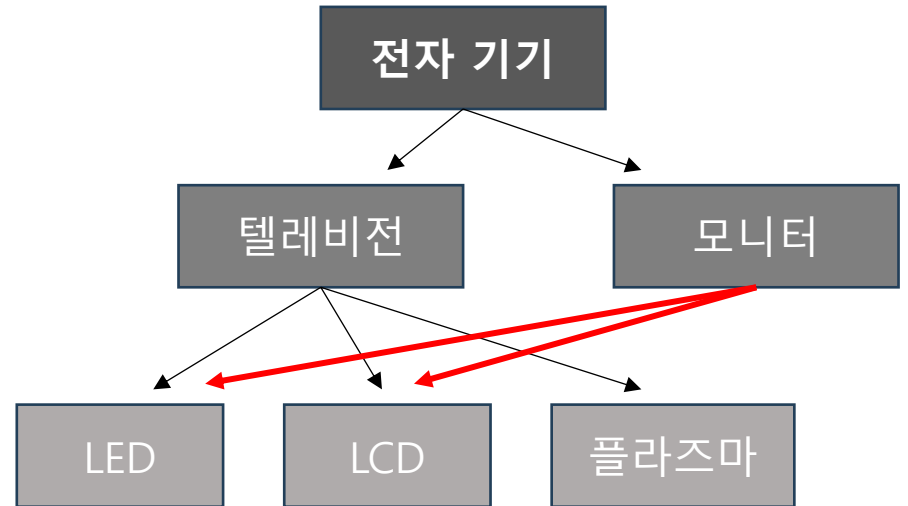
✓ 레코드(record) : 데이터베이스 테이블(table)에서 저장되어 있는 값들의 모임. 행(row), 튜플(tuple).

# Contents

## DB 종류: 계층형 데이터베이스

### 계층형(Hierarchical) 데이터베이스

- 계층형 데이터베이스는 데이터 중복이 발생하기 쉬우며, 데이터는 상하 종속 관계이므로 초기에 이 방식을 채택하면 이후 프로세스를 변경하기 어려워서 현재는 거의 사용하지 않는다.



# Contents

## DB 종류: 네트워크형 데이터베이스

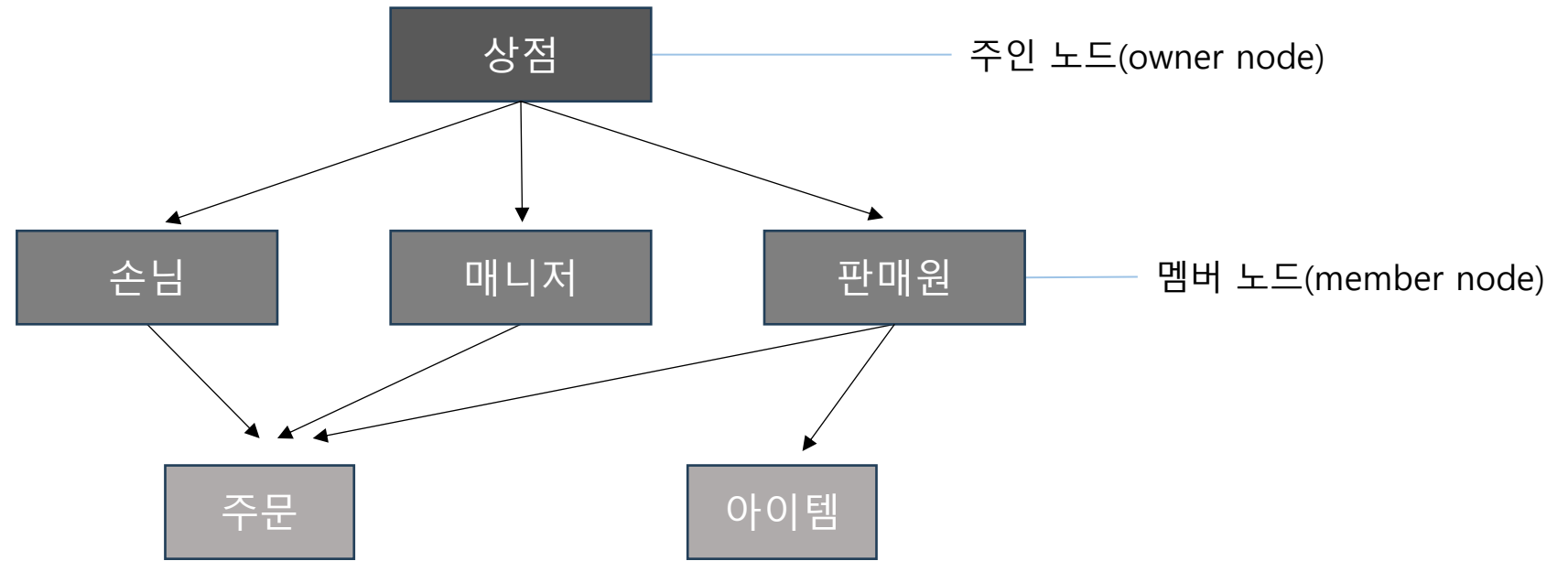
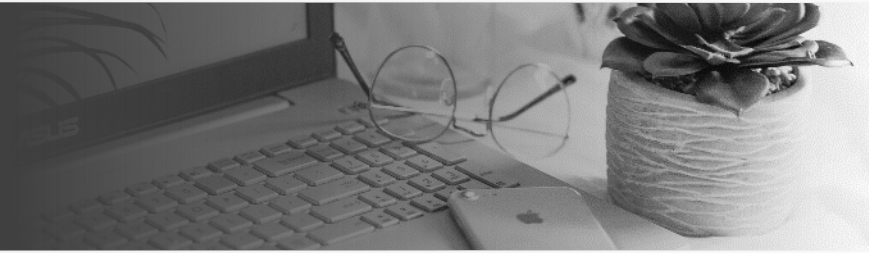


### 네트워크형(Network) 데이터베이스

- 데이터를 노드(node)로 표현한 모델.
- 계층형 DBMS의 문제점을 개선하기 위해 1970년대에 등장.
- 노드(node)는 기본적으로 컴퓨터 네트워크를 이루는 기초 단위로, 네트워크상에 있으며 서로 대등한 관계이다. 네트워크형 데이터베이스는 계층형 데이터베이스의 단점인 데이터 중복 문제, 상하 종속 관계를 해결했지만 구조가 복잡하여 변경, 운영하기 어렵고 데이터 종속성 문제까지 있다. 네트워크형 DB를 잘 활용하려면 프로그래머가 모든 구조를 이해해야만 프로그램 작성이 가능하다는 단점이 있다. 이것 역시 지금은 거의 사용하지 않는 형태이다.

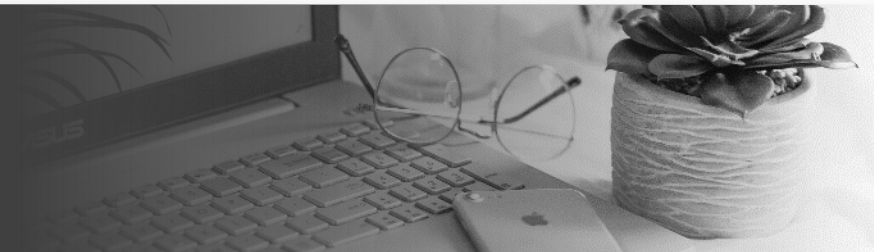
# Contents

## DB 종류: 네트워크형 데이터베이스



# Contents

## DB 종류: 키-값 데이터베이스



### 키-값(key-value) 데이터베이스

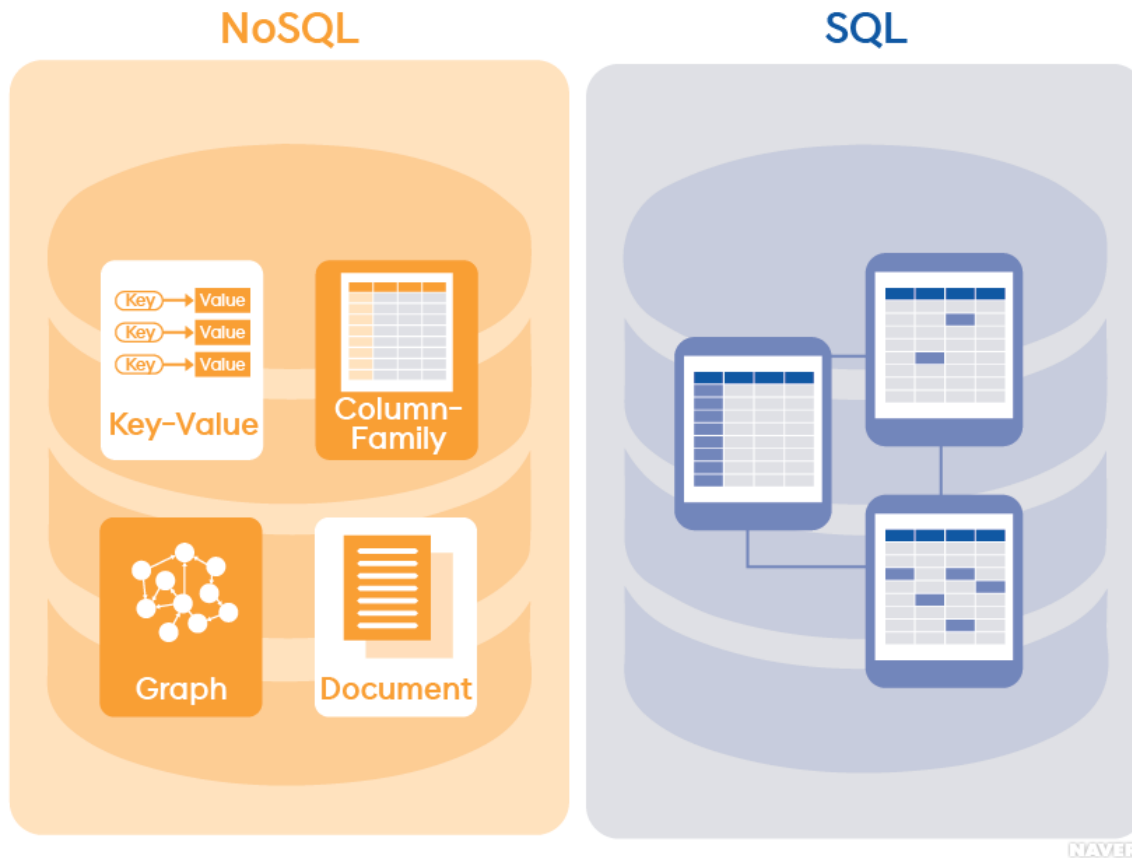
- NoSQL의 한 종류.
- 키-값을 일대일 대응해 데이터를 저장
- 데이터 중복이 발생하며 비정형 데이터 저장에 유리.

✓ NoSQL(Not Only SQL) : 관계형 데이터 모델을 사용하지 않는 데이터베이스

# Contents

## NoSQL (Not Only SQL)

- NoSQL 데이터베이스는 분산 환경에서 단순 검색 및 추가 작업을 위한 키 값을 최적화할 수 있어 지연(latency)과 처리율(throughput)이 우수하다. 그리고 대규모 확대가 가능한 수평적인 확장성을 가져 대규모의 데이터를 유연하게 처리할 수 있다.



“노에스큐엘”, 『용어로 알아보는 우리시대 DATA』, 네이버 지식백과. ①

# Contents

## DB 종류: 키-값 데이터베이스



### 키-값(key-value) 데이터베이스

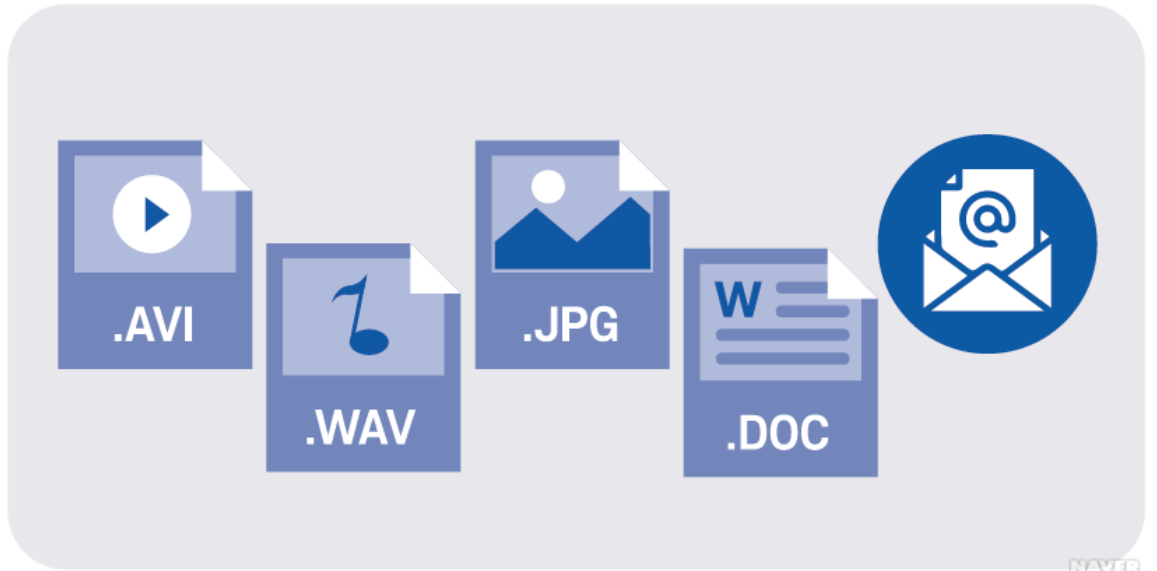
- NoSQL의 한 종류.
- 키-값을 일대일 대응해 데이터를 저장
- 데이터 중복이 발생하며 비정형 데이터 저장에 유리.

✓ 비정형 데이터(unstructured data) : 정의된 구조가 없이 정형화되지 않은 데이터

# Contents

## 비정형 데이터 (非定型- , unstructured data)

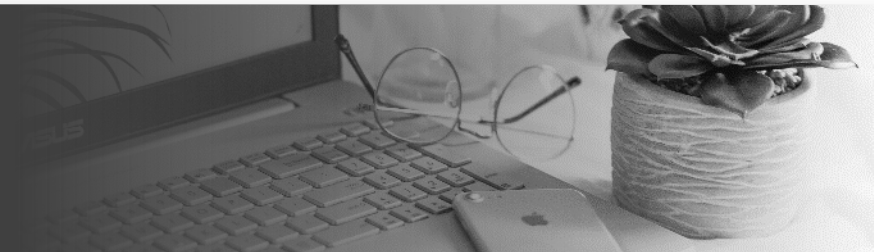
- 비정형 데이터는 데이터 구조가 없어 비정형 데이터 자체만으로는 내용에 대한 질의 처리(query processing)를 할 수 없다.
- 대표적인 비정형 데이터에는 보고서(문서), 메일 본문, 동영상 파일, 오디오 파일, 사진 등이 있다.



“비정형 데이터”, 『용어로 알아보는 우리시대 DATA』, 네이버 지식백과. ①

# Contents

## DB 종류: 키-값 데이터베이스



### 키-값(key-value) 데이터베이스

- 키-값 데이터베이스는 관계형 데이터베이스와 함께 가장 많이 사용한다.
- 키와 값 한 쌍으로 데이터를 저장하는 비관계형 데이터베이스 유형이다.
- 고유한 식별자로 사용하며 단순한 객체에서 복잡한 집합체에 이르기까지 무엇이든 키와 값이 될 수 있다.
- 특징은 스키마 없이 작동한다는 점이다. 따라서 데이터 구조를 미리 정의할 필요가 없으며 시간이 지나더라도 언제든지 바꿀 수 있으므로 비정형 데이터를 쉽게 저장할 수 있다.

✓ 스키마(schema) : 데이터베이스의 구조와 제약 조건에 대하여 전반적인 명세를 기술한 것.

# Contents

## DB 종류: 관계형 데이터베이스



### 관계형(Relational) 데이터베이스

- 현재 실무에서 가장 많이 사용하는 데이터베이스 종류이다.
- 테이블(table)이라는 최소 단위로 구성되며, 이 테이블은 하나 이상의 열(column)과 행(row)으로 이루어져 있다.

열(column) 속성(attribute) 필드(field)			

테이블(table)

행(row)  
튜플(tuple)  
레코드(record)

# Contents

## DB 종류: 관계형 데이터베이스



테이블

행(row)



행(row)



이름	전화번호	이메일	주소
김현승	010-3067-4837	koreadressdhedu@gmail.com	김포시



열(column)

열(column)

열(column)

열(column)

# Contents

## DB 종류: 관계형 데이터베이스

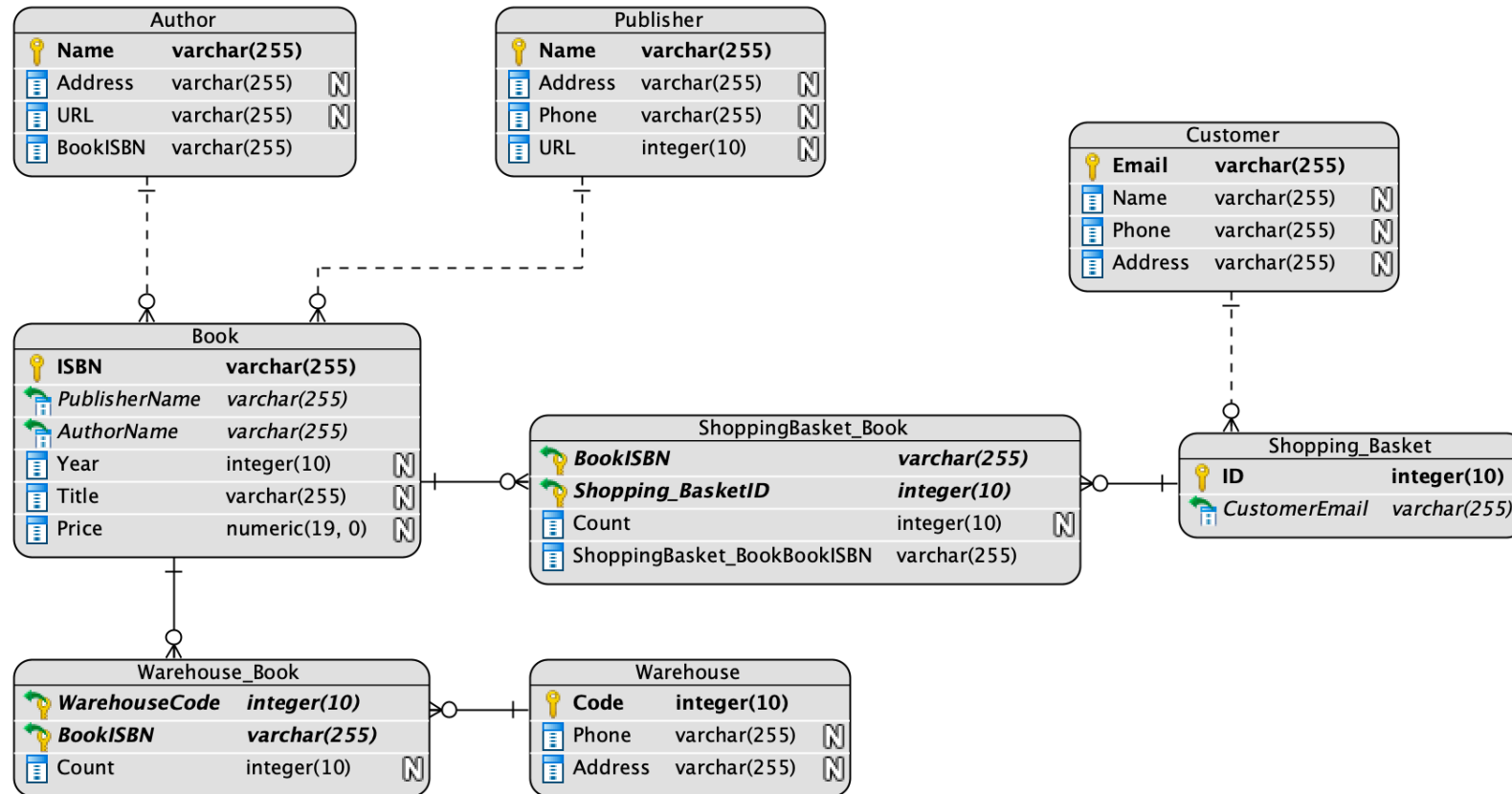


### 데이터베이스의 구성 요소

구성 요소	설명
열 (column)	각 열은 고유한 이름을 가지며 자신만의 타입을 가지고 있다. 열은 필드(field) 또는 애트리뷰트(attribute) 라고도 한다.
행 (row)	관계된 데이터의 묶음을 의미하며 한 테이블의 모든 행은 같은 수의 열을 가지고 있다. 행은 튜플(tuple) 또는 레코드(record) 라고도 한다.
테이블 (table)	행과 열 값들의 모음을 나타내는 것으로, 도메인 특성에 따라 데이터를 논리적으로 그룹화하여 모아 놓은 것이다.

# Contents

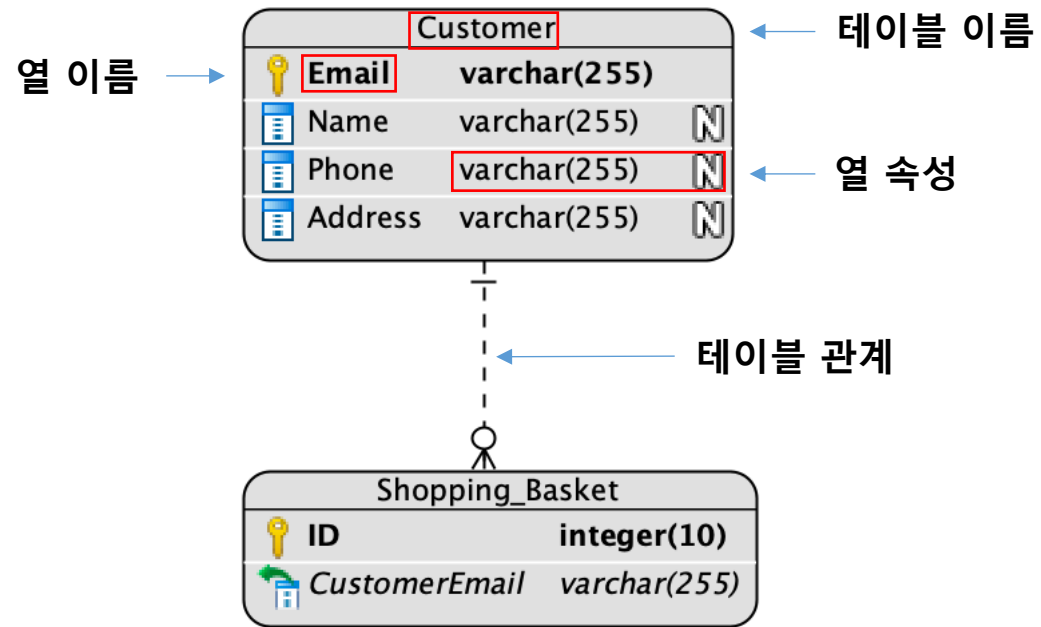
## DB 종류: 관계형 데이터베이스



관계형 데이터베이스의 ERD(Entity Relationship Diagram)

# Contents

## DB 종류: 관계형 데이터베이스



ERD(Entity Relationship Diagram) 구성 요소의 명칭

# Contents

## 관계형 데이터베이스 예시



### 주문

주문_번호	회원_이름	회원_주소	주문_상품	배송_주소
100	김현승	김포	디지털_인문학_입문	김포
101	이정보	서울	SQL_수업	서울
102	김현승	김포	SQL_수업	김포

# Contents

## 관계형 데이터베이스 예시



### 주문

주문_번호	회원_이름	회원_주소	주문_상품	배송_주소
100	김현승	<del>김포</del> 부여	디지털_인문학_입문	김포
101	이정보	서울	SQL_수업	서울
102	김현승	<del>김포</del> 부여	SQL_수업	김포

# Contents

## 관계형 데이터베이스 예시



회원

회원_번호	회원_이름	회원_주소
1000	김현승	김포
1001	이정보	서울
1002	김현승	김포

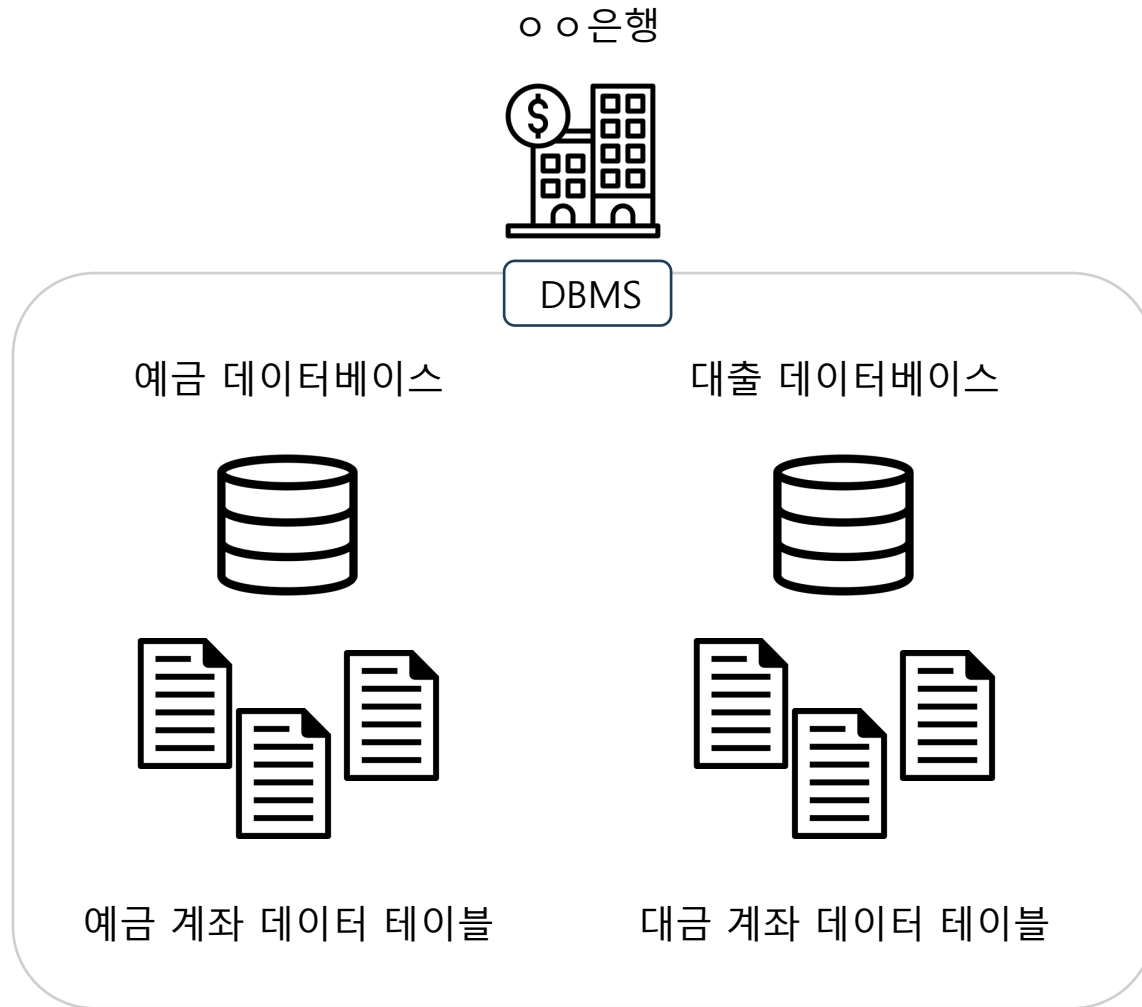
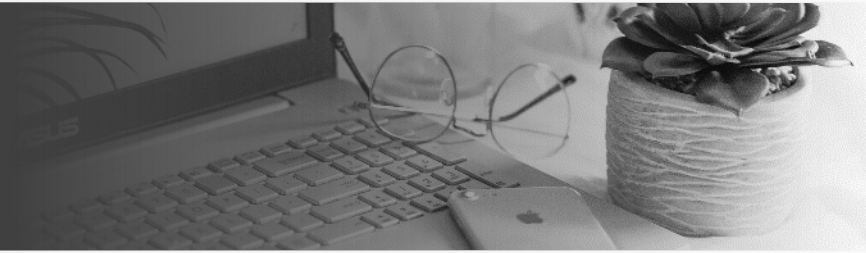
주문


주문_번호	회원_번호	주문_상품	배송_주소
100	1000	디지털_인문학_입문	김포
101	1001	SQL_수업	서울
102	1000	SQL_수업	김포





# Contents

## 관계형 데이터베이스 예시



 예금 계좌 주인

 은행 담당 직원

 입금  출금 인터넷 뱅킹

 ATM 기기

# Contents

## 관계형 데이터베이스 제품



제품 이름	상용 유무	특징
SQL Server	상용	Microsoft에서 개발한 RDBMS
Oracle Database	상용	Oracle에서 개발한 RDBMS
DB2	상용	IBM에서 개발한 RDBMS
PostgreSQL	오픈 소스	버클리 대학교에서 개발한 RDBMS
MySQL	오픈 소스	오픈 소스 커뮤니티에서 개발한 RDBMS
SQLite	오픈 소스	오픈 소스 커뮤니티에서 개발한 RDBMS로 embeded system에 주로 사용

## Part 3

# 관계형 데이터베이스의 키(key)

02

디지털베이스 개론

후보 키 / 기본 키 / 대체 키

외래 키 / 슈퍼 키

# Contents

## 후보 / 기본 / 대체 키(key)

### 후보 키 (Candidate Key)

: 한 릴레이션의 특정 튜플을 유일하게 구별할 수 있는 애트리뷰트들의 집합. 릴레이션에 있는 모든 튜플에 대해 유일성(uniqueness)과 최소성(minimality)을 만족시켜야 한다.



### 기본 키 (Primary Key)

: 후보 키 중에서 데이터베이스 관리자가 선택한 주 키(Main Key). Null 값을 가질 수 없음. Null은 정보 부재를 명시적으로 표시하는 특수한 데이터 값. 기본 키로 정의된 애트리뷰트에는 동일한 값이 중복 저장될 수 없다.

### 대체 키(Alternate Key)

: 기본 키를 제외한 나머지 후보 키.

# Contents

## 후보 / 기본 / 대체 키(key)



학번	이름	주민등록번호	전화번호
20240001	김인문	051124-3456789	010-3434-6789
20240002	이정보	050311-4567891	010-4012-7891
20240003	박문화	050802-4891234	010-2088-4891
20240004	정박사	050430-3789123	010-3034-9123

- ✓ 이름은 중복이 가능하기에 학번, 주민등록번호, 전화번호가 기본키로 설정이 가능하다.
- ✓ 기본키 설정이 중요하다. 중복과 NULL 이 있어서는 안된다.

# Contents

## 외래 키(Foreign key)



### 외래 키 (Foreign Key)

: 관계를 맺고 있는 릴레이션의 기본 키에 해당하는 애트리뷰트. 외래 키로 지정된 애트리뷰트는 참조된 릴레이션의 기본 키에 없는 값을 가질 수 없다.

# Contents

## 외래 키(Foreign key)



학번	이름	주민등록번호	전화번호
20240001	김인문	051124-3456789	010-3434-6789
20240002	이정보	050311-4567891	010-4012-7891
20240003	박문화	050802-4891234	010-2088-4891
20240004	정박사	050430-3789123	010-3034-9123

학번	수업번호	수업명
20240001	245015	데이터베이스 개론
20240002	245022	디지털 인문학 입문
20240003	245030	인문학적 사유와 철학
20240004	245032	한국 복식 구성

# Contents

## 슈퍼 키(Super key)



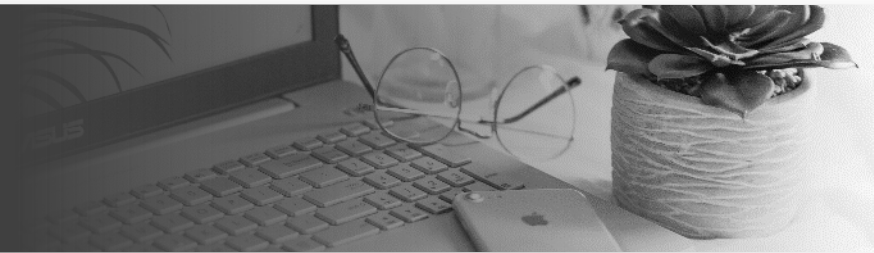
### 슈퍼 키 (Super Key)

: 두 개 이상의 애트리뷰트의 집합으로 구성되는 키. 유일성(uniqueness)은 만족시키지만 최소성(minimality)은 만족시키지 않는 애트리뷰트의 집합.

학번	이름	주민등록번호	전화번호
20240001	김인문	051124-3456789	010-3434-6789
20240002	이정보	050311-4567891	010-4012-7891
20240003	박문화	050802-4891234	010-2088-4891
20240004	정박사	050430-3789123	010-3034-9123

# Contents

## 슈퍼 키(Super key)



**후보 키** : 학번, 주민등록번호, 전화번호

**기본 키** : 후보 키 중 선택된 것

**대체 키** : 후보 키 중 선택하지 않은 나머지

**슈퍼 키** : 학번 + 이름 / 학번 + 이름 + 주민등록번호 /.... 등

- ✓ 유일성 : 하나의 키로 특정 행을 바로 찾아낼 수 있는 것(후보 / 기본 / 슈퍼 키)
- ✓ 최소성 : 레코드를 식별하는 데 꼭 필요한 속성으로 구성되어 있는가.

학번	이름	주민등록번호	전화번호
20240001	김인문	051124-3456789	010-3434-6789
20240002	이정보	050311-4567891	010-4012-7891
20240003	박문화	050802-4891234	010-2088-4891
20240004	정박사	050430-3789123	010-3034-9123



03

# DBMS

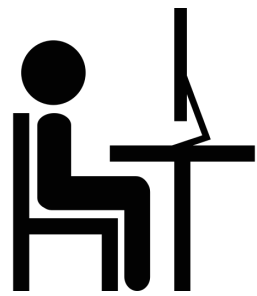
(DataBase Management System)

# Contents

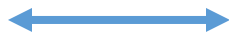
## DBMS(DataBase Management System)



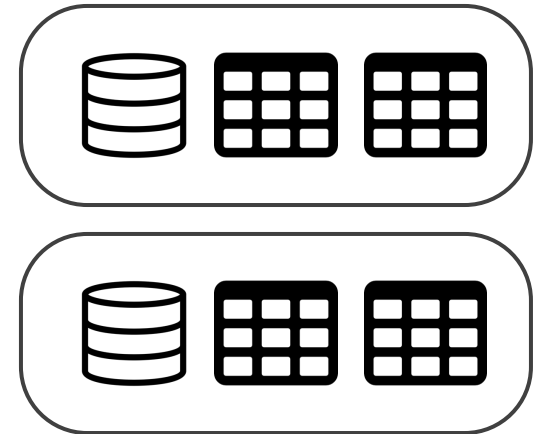
**데이터베이스 관리 시스템(DataBase Management System, DBSM)**는 데이터 입력, 수정, 삭제 등의 기능을 제공하는 별도의 소프트웨어다.



사용자



DBMS



데이터베이스

# Contents

## DBMS(DataBase Management System)

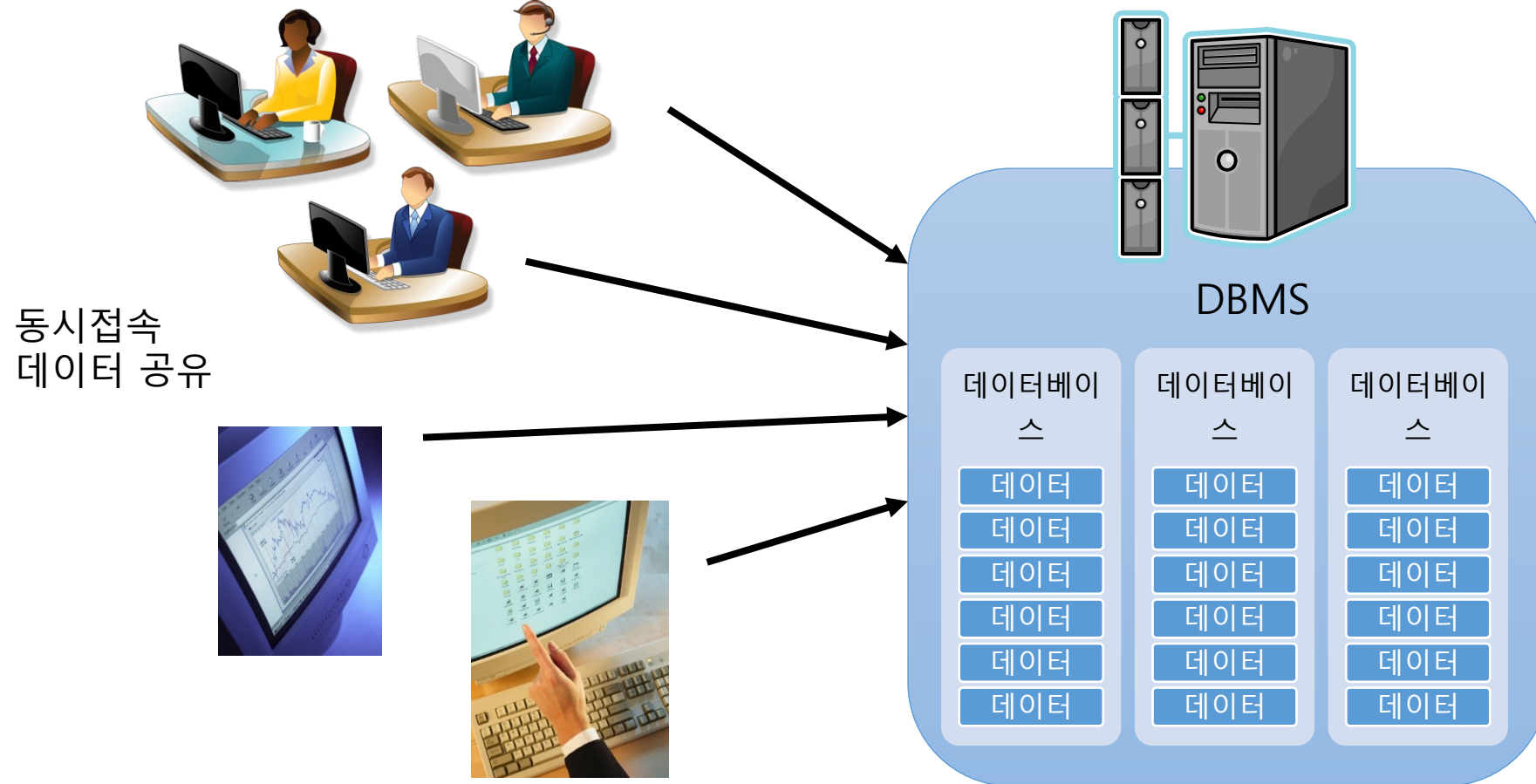


DB  $\neq$  DBMS

가끔 구분하지 않고 사용하는 경우가 있는데, 혼동하지 않도록 한다.

# Contents

## DBMS(DataBase Management System)



# Contents

## DBMS(DataBase Management System)



문서 작성



표 계산



사진 편집



데이터베이스 사용





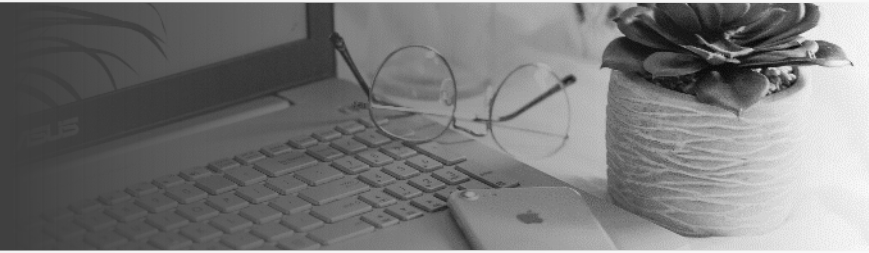
04

# SQL

(Structured Query Language)

Contents

DBMS에서 사용하는 언어 SQL

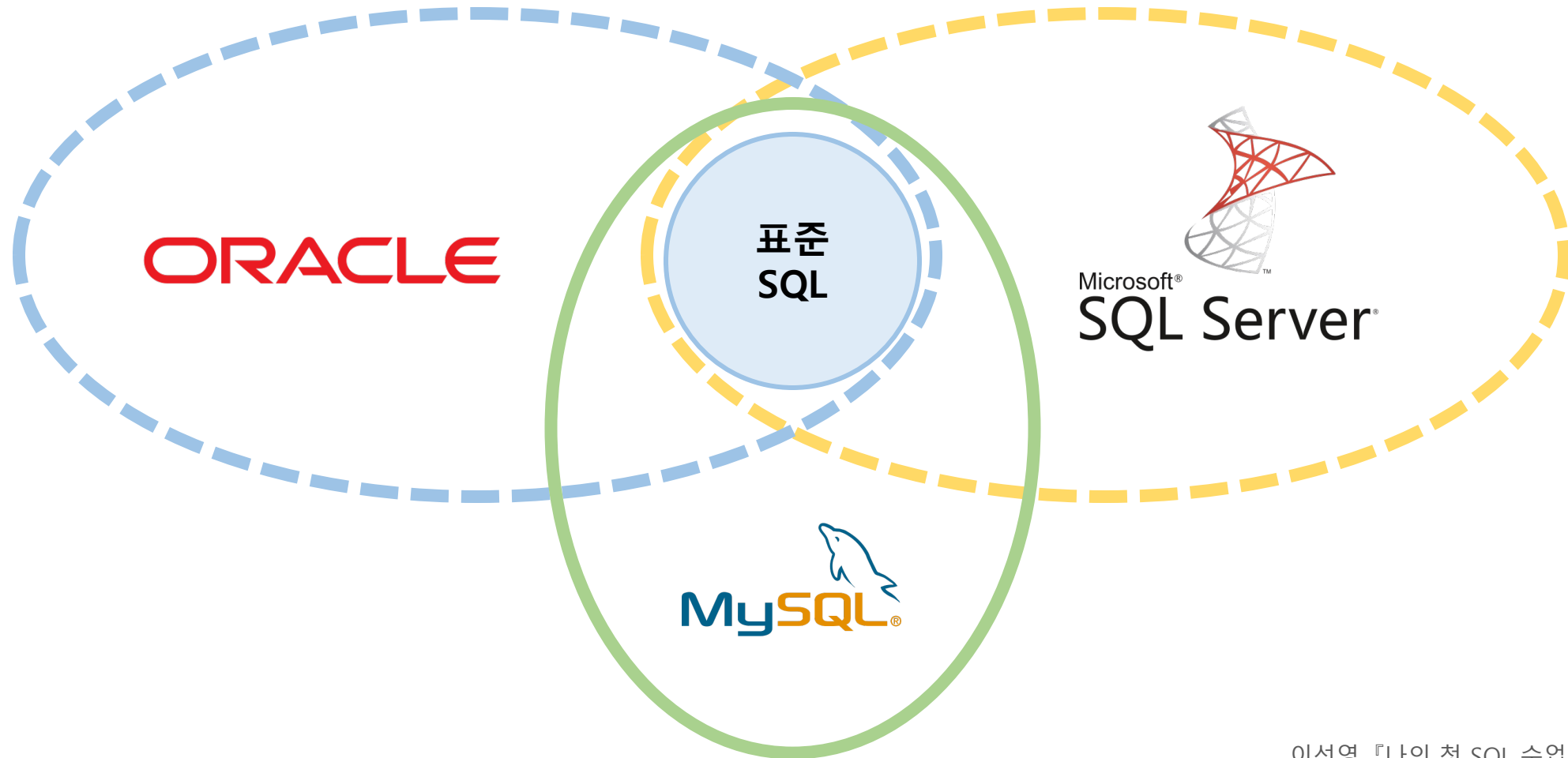


# Structured Query Language

DBMS를 사용할 때, 데이터베이스에 접근할 수 있는 데이터베이스 언어

# Contents

## DBMS(DataBase Management System)



# References

## 참고문헌



- 김현 · 임영상 · 김바로, 『디지털 인문학 입문』, 한국외국어대학교 지식출판원, 2016. : [https://dh.aks.ac.kr/Edu/wiki/index.php/디지털\\_인문학\\_입문](https://dh.aks.ac.kr/Edu/wiki/index.php/디지털_인문학_입문)
- 김현, 「인문정보 데이터베이스」, 한국학중앙연구원 한국학대학원 수업 자료, 2020.
- 김바로, 「인문정보 데이터베이스」, 한국학중앙연구원 한국학대학원 수업 자료, 2022.
- 한국학중앙연구원 한국학대학원 인문정보학과 SQL 스터디 자료, 2020.
- 강성욱, 『Do it! MySQL로 배우는 SQL 입문』, 이지스퍼블리싱, 2024.
- 이선영, 『나의 첫 SQL 수업』, 시대인, 2023.
- 우재남, 『혼자 공부하는 SQL』, 한빛미디어, 2021.
- 이영걸, 『SQL Server 2008』, 길벗, 2009.
- "노에스큐엘", 『용어로 알아보는 우리시대 DATA』, 네이버 지식백과. : <https://terms.naver.com/entry.naver?cid=59277&docId=6559347&categoryId=69439>
- "비정형 데이터", 『용어로 알아보는 우리시대 DATA』, 네이버 지식백과. : <https://terms.naver.com/entry.naver?cid=59277&docId=6559387&categoryId=69439>