

데이터베이스 개론

# 데이터 & 정보

---

- 데이터(data)

- 어떤 필요에 의해 수집했지만 아직 특정 목적을 위해 평가하거나 정제하지 않은 값이나 사실 또는 자료 자체를 의미.

- 정보(information)

- 수집한 데이터를 어떠한 목적을 위해 분석하거나 가공하여 가치를 추가하거나 새로운 의미를 이끌어 낼 수 있는 결과물.

- 효율적인 데이터 관리를 위한 조건

- 데이터를 **통합**하여 관리
- **일관된** 방법으로 관리
- 데이터 **누락 및 중복 제거**
- 여러 사용자(응용 프로그램 포함)가 **공동으로 실시간 사용** 가능

# 정보 처리

- 정보 처리(information processing)
  - 데이터에서 정보를 추출하는 과정 또는 방법.



- 정보 시스템(information system)
  - 조직 운영에 필요한 데이터를 수집하여 저장해두었다가 필요할 때 유용한 정보를 만들어 주는 수단.

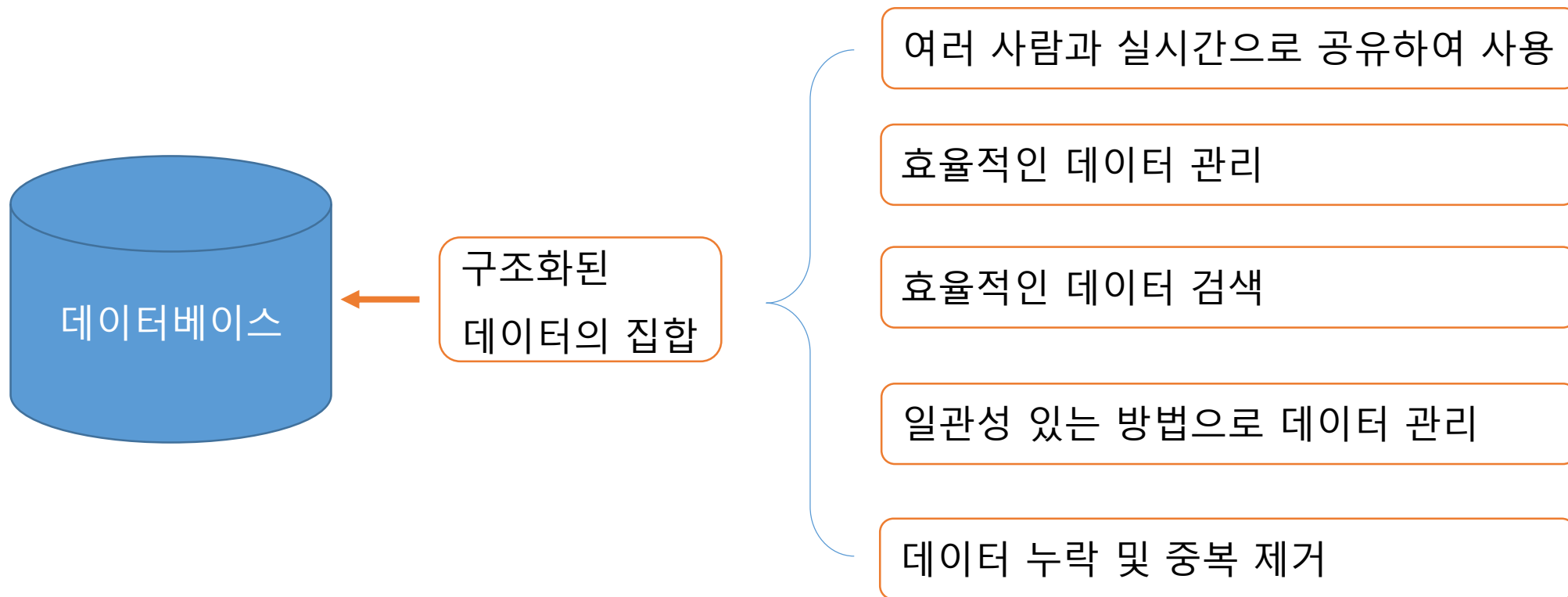
# 데이터의 분류

---

- 정형 데이터 (structured data)
    - 구조화된 데이터, 즉 미리 정해진 구조에 따라 저장된 데이터.
    - 예) 엑셀의 스프레드시트, 관계 데이터베이스의 테이블.
  - 반정형 데이터 (semi-structured data)
    - 구조에 따라 저장된 데이터이지만 데이터 내용 안에 구조에 대한 설명이 함께 존재.
    - 구조를 파악하는 파싱(parsing) 과정이 필요하고, 보통 파일 형태로 저장.
    - 예) 웹에서 데이터를 교환하기 위해 작성하는 HTML, XML, JSON 문서나 웹 로그, 센서 데이터 등.
  - 비정형 데이터 (unstructured data)
    - 정해진 구조가 없이 저장된 데이터.
    - 예) 소셜 데이터의 텍스트, 영상, 이미지, 워드나 PDF 문서와 같은 멀티미디어 데이터.
-

# 데이터베이스

---



# 데이터베이스 시스템의 구성

---

- 데이터베이스 관리 시스템(Database Management System, DBMS)
  - 사용자와 데이터베이스를 연결시켜주는 소프트웨어.
  - 연관성 있는 데이터들의 집합을 효율적으로 응용하기 위해 구성된 소프트웨어들의 집합체.
  - 방대한 양의 데이터를 편리하게 저장하고 효율적으로 관리하고 검색할 수 있는 환경을 제공하는 시스템 소프트웨어.
  - 데이터를 공유하여 정보의 체계적인 활용 가능.
  - 응용 프로그램과 데이터베이스의 중재자로서 모든 응용 프로그램들이 데이터베이스를 공용할 수 있게끔 관리해 주는 소프트웨어 시스템.
- 데이터베이스(database)
  - 기업이 지속적으로 유지 관리해야 하는 조직화된 정보의 집합.
  - 정보 시스템 안에서 데이터를 저장하고 있다가 필요할 때 제공하는 역할을 담당.
  - 특정 조직의 여러 사용자가 **공유**하여 사용할 수 있도록 **통합**해서 **저장**한 **운영** 데이터의 집합.
- 데이터 모델(data model)
  - 데이터가 저장되는 기법에 관한 내용으로, 눈에 보이지 않는 논리적인 개념.
  - 데이터가 저장되는 스타일을 나타냄.

# 데이터베이스 개념

---

- 공유 데이터(shared data)
    - 여러 사용자들이 서로 다른 목적으로 사용하는 공유 가능한 데이터.
  - 통합 데이터(integrated data)
    - 데이터베이스는 똑같은 데이터가 원칙적으로 중복되어 있지 않다는 것.
    - 데이터의 중복은 일반적으로 관리상의 복잡한 부작용을 초래.
  - 저장 데이터(stored data)
    - 컴퓨터가 접근할 수 있는 기억장치에 저장된 데이터.
    - 주로 하드디스크에 저장되어 관리됨.
  - 운영 데이터(operational data)
    - 존재 목적이 명확하고 유용성을 지니고 있는 데이터.
    - 즉, 단순히 데이터를 모아둔 개념이 아닌 병원 관리를 위한 데이터 구축과 같은 목적이 분명한 데이터.
-

# 데이터베이스 특징

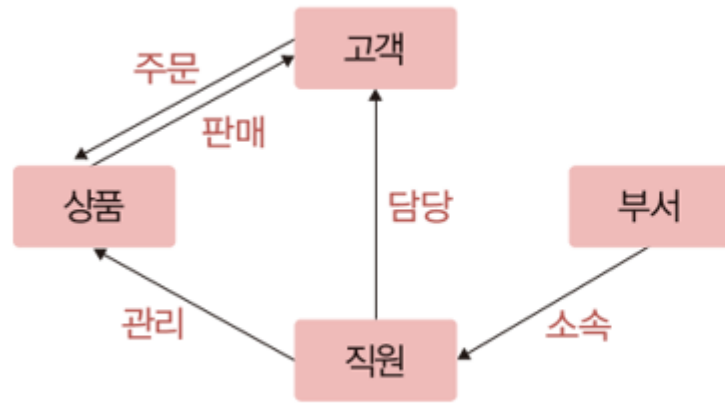
---

- 실시간 접근성(real-time accessibility)
  - 다수의 사용자의 요구에 대해서 처리 시간이 몇 초를 넘기지 말아야 한다는 의미.
- 지속적인 변화(continuous evolution)
  - 데이터베이스에 저장된 데이터는 지속적인 삽입, 삭제, 수정을 통해 최신의 정보가 정확하게 저장되어 처리되어야 함.
- 동시 공유(concurrent sharing)
  - 동일 데이터를 동시에 서로 다른 목적으로 사용할 수 있어야 함.
- 내용에 대한 참조(reference by content)
  - 데이터베이스 내에 있는 데이터 레코드들은 주소나 위치에 의해 참조되는 것이 아니라, 가지고 있는 값(내용)에 따라 참조해야 함.
  - 예) 재고량이 1,000개 이상인 제품의 이름 검색.

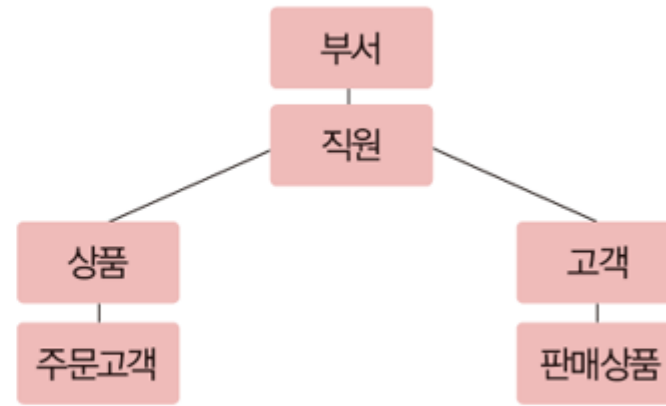


# 데이터 모델

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



(a) 네트워크 DBMS



(b) 계층 DBMS

# 데이터 모델

---

- 2세대 : 관계형 DBMS

- 관계형 DBMS : 데이터베이스를 테이블 형태로 구성

예) 오라클(Oracle), MS SQL 서버, 액세스(Access), 인포믹스(Informix), MySQL

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

예) 관계형 DBMS의 테이블

# 데이터 모델

---

- 3세대 : 객체지향 DBMS, 객체관계 DBMS
  - 객체지향 DBMS : 객체를 이용해 데이터베이스를 구성  
예) 오투(O2), 온투스(ONTOS), 젬스톤(GemStone)
  - 객체관계 DBMS : 객체 DBMS + 관계 DBMS

# 데이터 모델

---

- 4세대 : NoSQL / NewSQL DBMS

- NoSQL DBMS : 비정형 데이터를 처리하는데 적합하고 확장성이 뛰어남
  - ✓ 안정성과 일관성 유지를 위한 복잡한 기능을 포기
  - ✓ 데이터 구조를 미리 정해두지 않는 유연성
  - ✓ 확장성이 뛰어나 여러 대의 서버 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
  - ✓ 예) 구글 스패너(Spanner), 볼트DB(VoltDB), 누오DB(NuoDB)

# 데이터 모델

- DBMS의 발전 과정



# 관계형 데이터베이스 (RDBMS: Relational DataBase Management System)

- 가장 일반적인 형태의 DBMS – 작성과 이용이 비교적 쉽고 확장이 용이.
- 오라클(Oracle), 사이베이스(Sybase), 인포믹스(Infomix), MYSQL, Access, SQL Server.
- 관계형 데이터베이스 정보를 테이블 형태로 저장.
- 테이블은 2차원 형태의 표처럼 볼 수 있도록 로우(ROW:행)와 칼럼(COLUMN:열)으로 구성.

테이블 이름 : DEPT

DEPTNO	DNAME	LOC
10	ACCOUNTING	NEW YORK
20	RESEARCH	DALLAS
30	SALES	CHICAGO
40	OPERATIONS	BOSTON

- DEPARTMENT 테이블은 4개의 로우와 3개의 칼럼(부서번호:DEPTNO, 부서이름:DNAME, 지역:LOC)으로 구성된 테이블.

# 관계형 데이터베이스 (RDBMS: Relational DataBase Management System)

---

- 데이터 딕셔너리(Data Dictionary : DD)
  - 관계형 데이터베이스에서 객체를 정의하게 되면 그 객체가 가진 메타 데이터(metadata)의 정보가 저장되는 곳.
  - 사용자에게 의해서 추가, 삭제, 수정되지 못하며 오로지 오라클 시스템에 의해서만 가능.
- SQL(Structured Query Language)
  - 사용자와 관계형 데이터베이스를 연결시켜 주는 표준 검색 언어.
  - 데이터베이스에 저장된 데이터를 조회, 입력, 수정 삭제하는 등의 조작이나, 테이블을 비롯한 다양한 객체(시퀀스, 인덱스 등)를 생성 및 제어하는 역할.

# SQL(Structured Query Language)

---

- SQL의 종류

- 데이터 정의어(DDL)
  - 데이터베이스 관리자나 응용 프로그래머가 데이터베이스의 논리적 구조를 정의하기 위한 언어로서 데이터 사전(Data Dictionary)에 저장.
- 데이터 조작어(DML)
  - 데이터베이스에 저장된 데이터를 조작하기 위해 사용하는 언어로서 데이터 검색(Retrieval), 추가(Insert), 삭제>Delete), 갱신(Update) 작업 수행.
- 데이터 제어어(DCL)
  - 데이터에 대한 접근 권한 부여 등의 데이터베이스 시스템의 트랜잭션을 관리하기 위한 목적으로 사용되는 언어.



# SQL과 데이터 조회하기

유 형	명 령 문
DQL : Data Query Language(데이터 질의어)	SELECT(데이터 검색시 사용)
DML : Data Manipulation Language(데이터 조작어) - 데이터 변경 시 사용	INSERT(데이터 입력) UPDATE(데이터 수정) DELETE(데이터 삭제)
DDL : Data Definition Language(데이터 정의어) - 객체 생성 및 변경 시 사용	CREATE(데이터베이스 생성) ALTER(데이터베이스 변경) DROP(데이터베이스 삭제) RENAME(데이터베이스 객체이름 변경) TRUNCATE(데이터베이스 저장 공간 삭제)
TCL : Transaction Control Language(트랜잭션 처리어)	COMMIT(트랜잭션의 정상적인 종료처리) ROLLBACK(트랜잭션 취소)
DCL : Data Control Language(데이터 제어어)	GRANT(데이터베이스에 대한 일련의 권한 부여) REVOKE(데이터베이스에 대한 일련의 권한 취소)

# 오라클 접속하기

- 데이터베이스 사용자는 오라클 계정(Account)이라는 용어와 같은 의미로 사용.
- 오라클을 설치하면 한 개 이상의 데이터베이스 권한을 갖는 디폴트(기본적인) 사용자가 존재.
- 오라클에서 제공되는 사용자 계정.

사용자 계정	설 명
SYS	오라클 Super 사용자 계정이며 데이터베이스에서 발생하는 모든 문제들을 처리할 수 있는 권한을 가지고 있다.
SYSTEM	오라클 데이터베이스를 유지보수 관리할 때 사용하는 사용자 계정이며, SYS 사용자와 차이점은 데이터베이스를 생성할 수 있는 권한이 없으면 불완전 복구를 할 수 없다.
SCOTT	처음 오라클을 사용하는 사용자의 실습을 위해 만들어 놓은 연습용 계정이다.
HR	이 역시 오라클에 접근할 수 있도록 샘플로 만들어 놓은 사용자 계정이다.

# 오라클 접속하기

---

## ➤ Command 환경에서 SQL\*Plus 로그인

- 데이터베이스 접속을 시도하면 오라클 데이터베이스를 사용할 수 있는 사용자인지를 검증하기 위해서 사용자 계정과 암호를 묻게 됨.

형식	SQLPLUS 사용자계정/암호
----	------------------

- 오라클을 설치하게 되면 기본적으로 생성되는 사용자 계정 중에서 hr을 제공. hr 계정의 패스워드는 tiger.

예	sqlplus hr/tiger
---	------------------

- 실습에 사용할 scott 사용자 계정.

예	sqlplus scott/tiger
---	---------------------

---