**1. (Big) Data**

빅데이터의 특징, 3V – Volume, Variety, Velocity

우리나라에서 다루는 데이터에는 빅데이터가 없다. (small data) 이번주 수업에는 정형데이터, 반정형데이터, 비정형데이터(비정형데이터가 다루기 어렵기에 가장 관심이 많다.) 수집, 정리, 분석을 할 것이다.

Unstructured Data : Structured Data = 80 : 20

data는 단순한 값을 말한다면 information은 유의미한 정보를 뜻한다. (data - 연도별 기온, information – 기온이 상승하는 이유)

**2. Structured Data**

**(1) 데이터베이스(Database)**

: data들을 어떻게 다뤄야 하는지 만들어 놓은 틀. 데이터베이스의 조건 – 통합(중복X, 무결성), 저장, 공유, 운영

DBMS : 데이터베이스를 관리해 주는 프로그램 ( DB 생성, 삽입, 수정, 삭제. 쿼리를 사용)

\*\* 가장 중요한 것은 DB Integrity!!! (=데이터베이스 무결성)

**(2) DBMS**

: 계층형 모델, 네트워크형 모델, **관계 데이터 모델**, 객체 관계 모델

관계 데이터 모델과 객체 관계 모델의 차이점 – 관계 데이터 모델의 Table은 객체 관계 모델의 객체가 된다. (우리는 관계 데이터 모델(sqlite)을 이용할 것이고, 객체 관계 모델처럼 이용하기 위해 ORM을 사용할 것이다. )

특징 - 데이터 안정성, 동시 접근성, **장애 허용성**, 데이터 무결성, 데이터 확장성, 데이터 보고서

(sqlite는 개인이 접근하므로 동시 접근성은 무시한다.)

용어 - Relation, Tuple, Attribute, Cardinality = # Tuples, Degree = # Atrributes, Domain, Primary Key(PK, 기본키, 각 튜플이 가지는 고유한 값)

ER 모델 - 데이터의 기본 조각이 되는 것들의 **관계**를 나타내는 그림

**(3) SQL**

SQL 명령 : DDL 데이터 정의 언어, DML 데이터 조작 언어, DCL 데이터 제어 언어 (grant, revoke)

Data Type : CHARACTER, VARCHAR, CLOB, BOOLEAN, BLOB, DATE, TIME, TIMESTAMP, SMALLINT, INTEGER, DECIMAL, NUMERIC, REAL, FLOAT, DOUBLE,,,

(LOB : Large OBject, VARCHAR : 가변길이), DBMS마다 처리 방법이 다르다.

LIKE 사용 할 때,

a% : a로 시작, %a : a로 끝, %or% : 중간에 or, \_r% : 한글자가 r , a\_%\_% :

limit와 order by,

update를 이용해서 data를 수정하기 보다는, order by를 이용해서 업데이트 최신순으로 정렬하고, limit를 이용해서 (0,1) 범위를 가져오면 가장 최근 값 한 개만 나온다.

Join,

a INNER JOIN(교집합), LEFT OUTER JOIN b ON a. a = b.a

RDBMS인 sqlite 이용.

다음 사이트의 게스트 하우스 DB에서 주어진 Easy, Medium, Hard Problems를 모두 풀어서 SQL 해답을 제출한다.

<https://sqlzoo.net/wiki/Guest_House>