

데이터의 구조

스키마

테이블

ID(PK)	학생ID	이름	성적	출석률	시험 불참 여부	동아리 활동 여부
1	레코드 S001	Michael	A+	60	0	1
2	S002	Jackson	A0	63	1	1
3	S003	Mike	A+	63	1	1
4	S004	Brian	A+	99	0	1
5	S005	Young	A+	69	0	0
6	S006	Haley	A+	79	1	1
7	S007	Kelly	B0	81	1	1
8	S008	Christine	A0	96	1	1
9	S009	Eevee	B+	83	1	1

- 스키마는 데이터베이스의 구조를 정의하는 큰 틀, 같은 주제를 갖고 있는 관계를 규정하는 역할
- 테이블은 데이터를 실제로 저장하는 구조로 데이터를 추가/조회/수정/삭제 할 수 있음
- 각 행은 **관측값**, 각 열은 **변수**, 각 셀은 **값**
- 각 행은 고유한 기본값 (Primary Key)이 있으며 다른 테이블과 연결할때 (Join) 외래키 (Foreign Key)를 사용함

데이터의 형식

1. 데이터베이스 형식 변경

- MySQL 에서 TEXT로 저장된 데이터가 PostgreSQL에서는 VARCHAR로 변환될 수 있음

2. 파일 형식 변환

- CSV, XML등 다양한 형식으로 저장된 데이터가 있으나 최근에는 JSON 형식을 요구함

3. 날짜와 시간 형식

- YYYY-MM-DD 형식을 MM/DD/YYYY 으로 변경 하거나 시간대를 고려한 UTC 시간으로 변환

4. 숫자 형식과 소수점 처리

- 소수점을 처리하는 데이터를 정수로 변환하거나 소수점 자리를 변환해야함

5. 문자 인코딩

- UTF-8로 저장된 데이터를 이전할 때 ISO-8859-1 인코딩을 요구하는 경우도 있음 (ASCII, EUC-JP/KR, GB2312)

6. 데이터 계층 구조와 변화

- 관계형 구조를 NoSQL 구조로 변환할때는 비정형 구조나 JSON 문서 형태로 저장 되서 데이터 이전이 매우 복잡