



Normalization

≡ 태그	DB
≡ 주차	1주차



목차

0. Key

- 0-1. 슈퍼 키(Super Key)
- 0-2. 후보 키(Candidate Key)
- 0-3. 기본 키(Primary Key)
- 0-4. 대체 키(Alternate Key)
- 0-5. 외래 키(Foreign Key)

1. Normalization

- 1-1. 특징
- 1-2. 장단점
 - 1) 장점
 - 2) 단점

2. Normalization의 단계

- 2-1. 제 1정규화(1NF)
- 2-2. 제 2정규화(2NF)
- 2-3. 제 3정규화(3NF)
- 2-4. BCNF(Boyce-Codd Normal Form)

0. Key

- 무언가를 식별하는 고유한 식별자(identifier)
- DB에서 조건에 만족하는 Relationship의 행을 찾거나 순서대로 정렬할 때 다른 행과 구별할 수 있는 유리한 기준이 되는 속성의 집합

0-1. 슈퍼 키(Super Key)

- 각 행을 유일하게 식별할 수 있는 **하나 이상의 속성들의 집합**

- 유일성만 만족하면 슈퍼 키가 될 수 있음

7조 테이블

학번	주민번호	이름	나이
1	928888-8888888	김기범씨	27
2	929999-7777777	<u>추정범씨</u>	27
3	007777-6666666	이은빈씨	19
4	986666-5555555	<u>지혜리씨</u>	21



유일성

하나의 키로 특정 행을 바로 찾아낼 수 있는 고유한 데이터 속성

0-2. 후보 키(Candidate Key)

- 각 행을 유일하게 식별할 수 있는 최소한의 속성들의 집합
- 유일성과 최소성을 동시에 만족해야 함

7조 테이블

학번	주민번호	이름	나이
1	928888-8888888	김기범씨	27
2	929999-7777777	<u>추정범씨</u>	27
3	007777-6666666	이은빈씨	19
4	986666-5555555	<u>지혜리씨</u>	21

0-3. 기본 키(Primary Key)

- 후보 키 중 하나를 선택한 키로, 최소성과 유일성을 만족하는 속성
- 테이블에서 오직 1개만 지정 가능
- Not Null, Distinct

0-4. 대체 키(Alternate Key)

- 후보 키가 두 개 이상일 경우, 어느 하나를 기본 키로 지정하고 남은 후보키

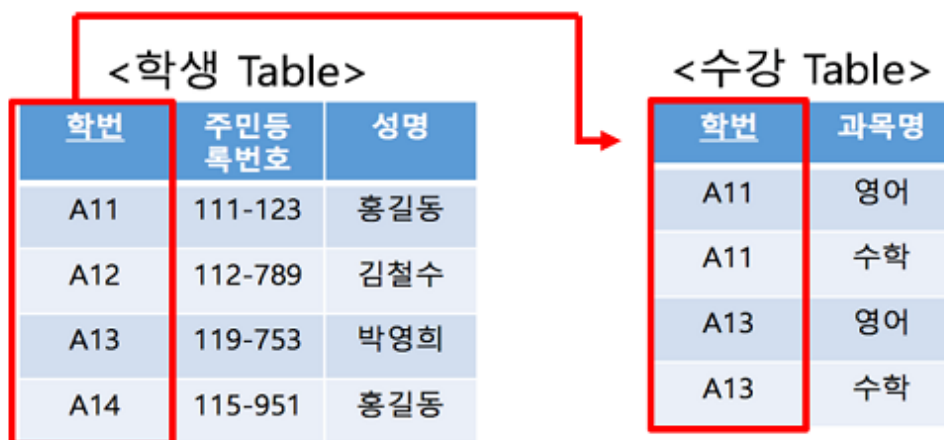
7조 테이블

학번	주민번호	이름	나이
1	928888-8888888	김기범씨	27
2	929999-7777777	추정범씨	27
3	007777-6666666	이은빈씨	19
4	986666-5555555	지혜리씨	21

0-5. 외래 키(Foreign Key)

- 다른 테이블의 데이터를 참조할 때 없는 값을 참조할 수 없도록 하는 제약
- 참조되는 열은 해당 테이블의 기본 키여야 함
- 참조되는 열과 참조하는 열은 동일한 키 속성을 가짐
- 참조되는 부모 테이블이 자식 테이블보다 먼저 삭제될 수 없음
- 기본 형태:

칼럼이름 칼럼타입 FOREIGN KEY (외래키) REFERENCES 부모테이블명(참조)



```
SQL> CREATE TABLE CHILD(
  2 ID NUMBER NOT NULL PRIMARY KEY,
  3 NUM NUMBER,
  4 CONSTRAINT FK_PID FOREIGN KEY (ID) REFERENCES PARENT(ID));
```

테이블이 생성되었습니다.

외래키 지정

1. Normalization

- *이상 현상(Anomaly) 예방을 위한 설계 과정
- 이상 현상이 있는 Relation을 분해하여 이상 현상 제거



이상 현상(Anomaly)

DB에서 일관성이 없거나 잘못된 데이터가 발생하는 현상. DB에 중요한 문제를 발생시킬 수 있다.

1. **삽입 이상(Insertion Anomaly)**: 데이터 삽입 시 의도와 다른 값들도 삽입
2. **삭제 이상>Delete Anomaly**: 데이터 삭제 시 의도와 다른 값들도 연쇄 삭제
3. **갱신 이상(Update Anomaly)**: 속성값 갱신 시 일부 튜플만 갱신되어 모순 발생

1-1. 특징

- 데이터 중복을 제거하고 데이터의 저장을 논리적으로 구조화
- 정보의 일관성과 정확성(데이터 무결성) 유지 가능
- DB 용량 감소

1-2. 장단점

1) 장점

- DB 변경 시 이상 현상(Anomaly) 제거 가능
- 데이터 및 데이터 형의 추가 등으로 인한 확장성이 좋음
- DB와 연동된 응용 프로그램에 최소한의 영향 만을 줌

2) 단점

- Relation의 분해로 인해 Relation간에 JOIN 연산이 많아짐
- query에 대한 응답 시간이 느려질 **수도** 있음
- 데이터를 처리할 때 속도가 빨라질 수도, 느려질 수도 있음



JOIN이 많이 발생하여 성능 저하가 나타나면 **반정규화(De-normalization)** 적용 가능!



반정규화(Denormalization)

복잡한 SQL 문장을 단순화하기 위해 적용하는 모델링 실무 기술. 데이터 무결성을 보장할 수 있는 방법을 고려한 후에 적용해야 한다. 정규화와는 *Trade-Off* 관계.

2. Normalization의 단계

2-1. 제 1정규화(1NF)

- 각 Row가 유일한 키로 식별되는 단계
- 즉, 각 Column이 원자값(Atomic Value)를 갖도록 테이블을 분해하는 것
- 제 1정규화는 아래 규칙을 만족해야 함
 1. 각 Column이 하나의 속성(Attribute)만을 가져야 한다.
 2. 하나의 Column은 같은 종류나 타입의 값을 가져야 한다.
 3. 각 Column이 유일한 이름을 가져야 한다.
 4. 칼럼의 순서가 상관 없어야 한다.

1NF가 적용되지 않은 테이블

이름	나이	수강 과목
홍길동	20	C, C++

이름	나이	수강 과목
이순신	21	JAVA
이 산	22	DB, 운영체제

1NF를 적용한 테이블

이름	나이	수강 과목
홍길동	20	C
홍길동	20	C++
이순신	21	JAVA
이 산	22	DB
이 산	22	운영체제

- 데이터 중복이 크게 줄어 공간 효율성 향상
- 데이터 무결성 유지에 도움
- 데이터의 수정, 삭제, 삽입 연산에 이상 문제 방지
- 데이터 구조가 단순 명확해짐

2-2. 제 2정규화(2NF)

- 제 1정규화를 거친 후 *부분적 종속성(Partial Dependency)을 제거하는 단계
- 즉, *완전 함수 종속을 만족하도록 테이블을 분해해야 함
- 제 2정규화는 아래 규칙을 만족해야 함
 1. 제1정규화를 만족해야 한다.
 2. 모든 Column이 부분적 종속이 없어야 한다. (\leftrightarrow 완전 함수 종속을 만족해야 한다)



*부분 함수 종속

Relation에서 종속자가 기본 키가 아닌 다른 속성에 종속되거나, 기본 키를 구성하는 여러 속성들의 부분집합 중 일부분에만 종속되어 있는 것

*

완전 함수 종속

모든 Column이 기본 키에 완전히 종속되는 것
(기본 키(Primary Key)의 부분집합이 결정자가 되어서는 안됨!)

1NF만 적용된 테이블

이름	나이	수강 과목
홍길동	20	C
홍길동	20	C++
이순신	21	JAVA
이 산	22	DB
이 산	22	운영체제

2NF를 적용한 테이블 (2개의 테이블로 분해)

이름	나이
홍길동	20
이순신	21
이 산	22

이름	수강 과목
홍길동	C
홍길동	C++
이순신	JAVA
이 산	DB
이 산	운영체제

- 각 테이블이 각각의 역할에 집중하여 관리됨
- 중복이 줄어 데이터 무결성 유지
- 전체적인 데이터 관리 및 유지보수 향상

2-3. 제 3정규화(3NF)

- 제 2정규화를 거친 후 ***이행적 종속성(Transitive Dependency)**을 제거하는 단계
- 기본 키가 아닌 모든 Column이 기본 키에만 직접적으로 종속되어야 함
- 제 3정규화는 아래 규칙을 만족해야 함
 - 제 2정규화를 만족해야 한다.
 - 기본 키를 제외한 속성들 간의 이행 종속성이 없어야 한다.



*이행적 종속성

$A \rightarrow B \ \& \ B \rightarrow C \Rightarrow A \rightarrow C$ 가 성립하는 것 (삼단논법?)

제 2정규화만 만족하는 테이블

고객 번호	이름	등급	할인율
1	김싸피	GOLD	20%
2	이싸피	VIP	30%
3	박싸피	BRONZE	10%
4	정싸피	GOLD	20%

제 3정규화를 적용한 테이블

고객 번호	이름	등급	등급	할인율
1	김싸피	GOLD	GOLD	20%
2	이싸피	VIP	VIP	30%
3	박싸피	BRONZE	BRONZE	10%
4	정싸피	GOLD		

- 각 테이블은 원자적인 값을 가지게 되고 데이터 중복이 줄어들음
- 이상 현상을 피할 수 있으며 유지 보수가 훨씬 용이해짐

2-4. BCNF(Boyce-Codd Normal Form)

- 제 3정규화를 강화한 버전
- 제 3정규화를 만족하면서 모든 결정자가 후보키 집합에 속한 정규형
- BCNF는 아래 규칙을 만족해야 함
 1. 제 3정규화를 만족해야 한다.
 2. 모든 결정자가 후보키 집합에 속해야 한다.

제 3정규화만 만족하는 테이블

학번	과목명	담당 교수
100	수학	P1
100	컴퓨터	P2
200	수학	P1
300	컴퓨터	P3
400	컴퓨터	P4

- 후보키: (학번, 과목명), (학번, 담당 교수)
 - '담당 교수' 만으로는 후보키가 될 수 없지만, 과목명을 정할 수 있으므로 결정자가 됨.
 - 즉, '담당 교수'가 '과목명'을 결정하므로, 문제 발생!
- 아래 문제점 발생
 1. **삽입 이상:** 새로운 교수가 특정 과목을 담당한다는 새로운 정보를 추가할 수 없음. 적어도 한 명 이상의 수강 학생 필요
 2. **삭제 이상:** 학번 100이 '컴퓨터' 과목을 취소하면, P2가 컴퓨터 과목을 담당한다는 정보도 삭제된다.
 3. **갱신 이상:** P1의 과목이 변경되면 P1인 행을 모두 찾아 변경시켜주어야 한다.

BCNF를 만족하는 테이블

학번	담당 교수	담당 교수	과목명
100	P1	P1	수학
100	P2	P2	과학
200	P1	P3	과학
300	P3	P4	과학
400	P4		



1NF, 2NF, 3NF, 그리고 BCNF까지가 standard로 여겨지고 있다.



Reference

[DataBase] 정규화(1NF, 2NF, 3NF, BCNF)

정규화(Normalization) 정규화란 이상현상이 있는 릴레이션을 분해하여 이상현상을 없애는 과정이다. 이상현상이 존재하는 릴레이션을 분해하여 여러 개의 릴레이션을 생성하

<https://superohinsung.tistory.com/111>



[DB] 8. 정규형 (1NF, 2NF, 3NF, BCNF)

[목차] 1. 제1 정규형 2. 제2 정규형 3. 제3 정규형 4. BCNF 정규형이란 정규화된 결과를 말하며, 제1 정규형, 제2 정규형, 제3 정규형, BCNF, 제4 정규형, 제5 정규형이

<https://rebro.kr/160>

