

데이터베이스 정규화

데이터들의 중복을 최대한 제거하여 이상 현상(Anomaly)을 방지 하는 기술

이상현상

-갱신 이상(Modification Anomaly) : 반복된 데이터 중에 일부를 갱신 할 시 데이터의 불일치

어떤 값을 업데이트 할 때 그 속성의 다른 속성값들 과의 불일치가 발생하는 현상

| 학번 | 과목명 | 성적 | 이름 |
|-----|-----|----|-----|
| 100 | DB | 92 | 홍길동 |
| 101 | 운체 | 90 | 김길동 |
| 100 | 운체 | 83 | 홍길동 |
| 102 | DB | 92 | 최길동 |

→ ‘신길동’ 으로 변경한 경우(1,3번째 행 중 1번째 행만)
3번째 레코드에 있는 속성의 내용과 불일치 발생하게 된다.
이런 문제를 해결하기 위해서는 각각의 속성값들을 다 변경

-삽입 이상(Insertion Anomaly) : 새 데이터를 삽입할 때 의도치 않은 데이터가 삽입됨으로써 생기는 데이터의 불일치

| 학번 | 과목명 | 성적 | 이름 |
|-----|-----|----|-----|
| 100 | DB | 92 | 홍길동 |
| 101 | 운체 | 90 | 김길동 |
| 100 | 운체 | 83 | 홍길동 |
| 102 | DB | 92 | 최길동 |
| 103 | DB | | 고길동 |

→ 새로운 학생에 대한 정보를 입력할 때 학번, 과목명, 이름은 작성했지만
아직 시험을 보지 않았기 때문에 성적을 입력하지 못하는 경우,
불필요한 정보를 함께 입력하지 않는 한 테이블에 입력할 수 없다.

-삭제 이상(Deletion Anomaly) : 데이터를 삭제할 때 의도치 않은 데이터까지 삭제됨으로써 생기는 데이터의 불일치

| 학번 | 과목명 | 성적 | 이름 |
|-----|-----|----|-----|
| 100 | DB | 92 | 홍길동 |
| 101 | 운체 | 90 | 김길동 |
| 100 | 운체 | 83 | 홍길동 |

→ 과목명을 삭제하는 경우 어떤 과목의 성적인지를 확인할 수 없게됨
(연쇄삭제가 발생)

무결성

- 1. 개체 무결성 : 모든 테이블이 기본 키(primary key)를 가져야 하고, 기본 키로 선택된 필드는 고유한 값(중복x)을 가져야 하며, 빈 값은 허용하지 않는다.
- 2. 참조 무결성 : 외래 키 값은 Null이거나 참조 릴레이션의 기본 키 값과 동일해야 한다. 즉 릴레이션은 참조할 수 없는 외래 키 값을 가질 수 없다.

정규화 과정

제 1정규화

- 1) 테이블에 속한 모든 도메인이 원자 값(Atomic Value)만으로 이루어 져야 한다.
- 2) 기본 키(primary key)를 사용하여 관련 데이터의 **각 집합을 고유하게 식별**할 수 있어야한다. = pk가 중복된 부분이 있으면 테이블 나누기

학생수강성적

| 학생번호 | 학생이름 | 주소 | 학과 | 학과사무실 | 강좌이름 | 강의실 | 성적 |
|------|------|----------|------|--------|--------|---------|-----|
| 501 | 박지성 | 영국 맨체스터 | 컴퓨터과 | 공학관101 | 데이터베이스 | 공학관 110 | 3.5 |
| 401 | 김연아 | 대한민국 서울 | 체육학과 | 체육관101 | 데이터베이스 | 공학관 110 | 4.0 |
| 402 | 장미란 | 대한민국 강원도 | 체육학과 | 체육관101 | 스포츠경영학 | 체육관 103 | 3.5 |
| 502 | 추신수 | 미국 클리블랜드 | 컴퓨터과 | 공학관101 | 자료구조 | 공학관 111 | 4.0 |
| 501 | 박지성 | 영국 맨체스터 | 컴퓨터과 | 공학관101 | 자료구조 | 공학관 111 | 3.5 |

 : 기본 키

제 2정규화(부분함수 종속 제거)

- 1) 기본 키(primary key)에 대해 모든 컬럼이 완전 함수적 종속을 만족한다.(특정 컬럼만 종속될 경우 2정규화 위배)
- 2) 복합키 일 때만, 제 2정규형의 조건이 된다. ~> 복합키가 아닌 경우 제 3정규형을 확인!!

제 3정규화(이행적 함수 종속 제거)

- 1) 테이블이 제 2정규화가 되었다.
- 2) 기본 키(primary key)가 아닌 속성(Attribute)들은 기본 키에만 의존해야 한다.

기본릴레이션(비정규형)

학생수강성적

| 학생번호 | 학생이름 | 주소 | 학과 | 학과사무실 | 강좌이름 | 강의실 | 성적 |
|------|------|----------|------|--------|--------|---------|-----|
| 501 | 박지성 | 영국 맨체스타 | 컴퓨터과 | 공학관101 | 데이터베이스 | 공학관 110 | 3.5 |
| 401 | 김연아 | 대한민국 서울 | 체육학과 | 체육관101 | 데이터베이스 | 공학관 110 | 4.0 |
| 402 | 장미란 | 대한민국 강원도 | 체육학과 | 체육관101 | 스포츠경영학 | 체육관 103 | 3.5 |
| 502 | 추신수 | 미국 클리블랜드 | 컴퓨터과 | 공학관101 | 자료구조 | 공학관 111 | 4.0 |
| 501 | 박지성 | 영국 맨체스타 | 컴퓨터과 | 공학관101 | 자료구조 | 공학관 111 | 3.5 |

 :기본키



| 학생번호 | 학생이름 | 주소 | 학과 | 학과사무실 |
|------|------|----------|------|--------|
| 501 | 박지성 | 영국 맨체스타 | 컴퓨터과 | 공학관101 |
| 401 | 김연아 | 대한민국 서울 | 체육학과 | 체육관101 |
| 402 | 장미란 | 대한민국 강원도 | 체육학과 | 체육관101 |
| 502 | 추신수 | 미국 클리블랜드 | 컴퓨터과 | 공학관101 |

학생&학과 테이블

| 학생번호 | 강좌이름 | 강의실 | 성적 |
|------|--------|--------|-----|
| 501 | 데이터베이스 | 공학관110 | 3.5 |
| 401 | 데이터베이스 | 공학관110 | 4.0 |
| 402 | 스포츠경영학 | 체육관103 | 3.5 |
| 502 | 자료구조 | 공학관111 | 4.0 |
| 501 | 자료구조 | 공학관111 | 3.5 |

강좌&성적 테이블

학생&학과 테이블

| 학생번호 | 학생이름 | 주소 | 학과 | 학과사무실 |
|------|------|----------|------|--------|
| 501 | 박지성 | 영국 맨체스타 | 컴퓨터과 | 공학관101 |
| 401 | 김연아 | 대한민국 서울 | 체육학과 | 체육관101 |
| 402 | 장미란 | 대한민국 강원도 | 체육학과 | 체육관101 |
| 502 | 추신수 | 미국 클리블랜드 | 컴퓨터과 | 공학관101 |

학생테이블

| 학생번호 | 학생이름 | 주소 | 학과 |
|------|------|----------|------|
| 401 | 김연아 | 대한민국 서울 | 컴퓨터과 |
| 402 | 장미란 | 대한민국 강원도 | 체육학과 |
| 501 | 박지성 | 영국 맨체스타 | 체육학과 |
| 502 | 추신수 | 미국 클리블랜드 | 컴퓨터과 |

학과테이블

| 학과 | 학과사무실 |
|-------|--------|
| 컴퓨터학과 | 공학관101 |
| 체육학과 | 체육관101 |

제 1정규화 과정 : 현재 테이블은 도메인이 원자 값으로 구성 되어있다.

제 2정규화 과정 : 기본키에 대해 모든 컬럼은 완전 함수 종속 해야 한다.

학생번호, 학과를 통해 학생이름, 주소를 알 수 있다.(즉, 완전 함수 종속)

하지만, 학과사무실 같은 경우는 학과를 통해서만 알 수 있다.(부분 함수 종속)

~> 이러한 부분함수종속을 제거하기 위해 테이블을 나눠준다.

제 3정규화 과정 : 기본 키가 아닌 속성들은 오로지 기본 키에만 의존

강좌테이블

강좌&성적 테이블

| 학생번호 | 강좌이름 | 강의실 | 성적 |
|------|--------|--------|-----|
| 501 | 데이터베이스 | 공학관110 | 3.5 |
| 401 | 데이터베이스 | 공학관110 | 4.0 |
| 402 | 스포츠경영학 | 체육관103 | 3.5 |
| 502 | 자료구조 | 공학관111 | 4.0 |
| 501 | 자료구조 | 공학관111 | 3.5 |

성적 테이블

| 성적번호 | 학생번호 | 강좌이름 | 성적 |
|------|------|--------|-----|
| 1 | 401 | 데이터베이스 | 4.0 |
| 2 | 402 | 스포츠경영학 | 3.5 |
| 3 | 501 | 데이터베이스 | 3.5 |
| 4 | 501 | 자료구조 | 3.5 |
| 5 | 502 | 자료구조 | 4.0 |

강좌 테이블

| 강좌이름 | 강의실 |
|--------|--------|
| 데이터베이스 | 공학관110 |
| 스포츠경영학 | 체육관103 |
| 자료구조 | 공학관111 |

- 제 1정규화 과정 : 현재 테이블은 도메인이 원자 값으로 구성 되어있다.(성적테이블에서 기본 키 추가)
- 제 2정규화 과정 : 기본키에 대해 모든 컬럼은 완전 함수 종속 해야 한다.
학생번호, 강좌이름을 통해 성적을 알 수 있다.(즉, 완전 함수 종속)
하지만, 강의실 같은 경우는 강좌이름으로만 알 수 있다(부분 함수 종속)
~> 이러한 부분함수종속을 제거하기 위해 테이블을 나눠준다.
- 제 3정규화 과정 : 기본 키가 아닌 속성들은 오로지 기본 키에만 의존

정규화를 통해 얻어진 테이블

학생테이블

| 학생번호 | 학생이름 | 주소 | 학과 |
|------|------|----------|------|
| 401 | 김연아 | 대한민국 서울 | 컴퓨터과 |
| 402 | 장미란 | 대한민국 강원도 | 체육학과 |
| 501 | 박지성 | 영국 맨체스타 | 체육학과 |
| 502 | 추신수 | 미국 클리블랜드 | 컴퓨터과 |

성적테이블

| 성적번호 | 학생번호 | 강좌이름 | 성적 |
|------|------|--------|-----|
| 1 | 401 | 데이터베이스 | 4.0 |
| 2 | 402 | 스포츠경영학 | 3.5 |
| 3 | 501 | 데이터베이스 | 3.5 |
| 4 | 501 | 자료구조 | 3.5 |
| 5 | 502 | 자료구조 | 4.0 |

학과테이블(reference먼저 생성)

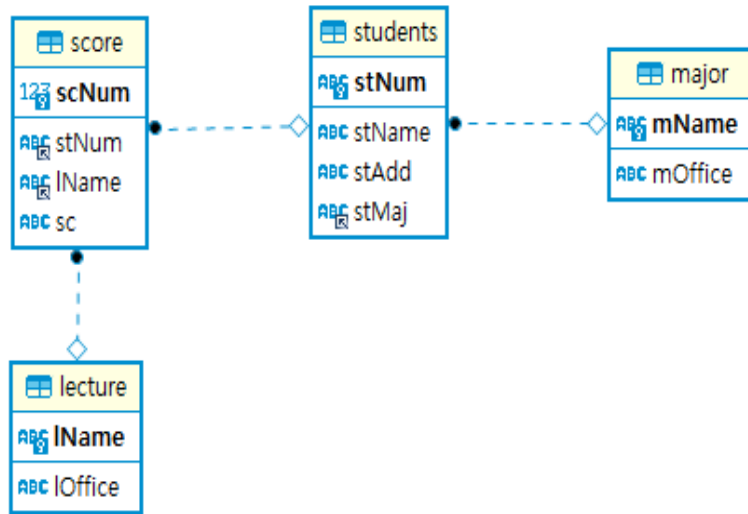
| 학과 | 학과사무실 |
|-------|--------|
| 컴퓨터학과 | 공학관101 |
| 체육학과 | 체육관101 |

강좌테이블(reference먼저 생성)

| 강좌이름 | 강의실 |
|--------|--------|
| 데이터베이스 | 공학관110 |
| 스포츠경영학 | 체육관103 |
| 자료구조 | 공학관111 |

정규화를 통해 분리한 테이블

E-R diagram



학생 테이블

| | stNum | stName | stAdd | stMaj |
|---|-------|--------|----------|-------|
| 1 | 401 | 김연아 | 대한민국 서울 | 컴퓨터학과 |
| 2 | 402 | 장미란 | 대한민국 강원도 | 체육학과 |
| 3 | 501 | 박지성 | 영국 맨체스터 | 체육학과 |
| 4 | 502 | 추신수 | 미국 클리블랜드 | 컴퓨터학과 |

학과 테이블

| | mName | mOffice |
|---|-------|---------|
| 1 | 체육학과 | 체육관101 |
| 2 | 컴퓨터학과 | 공학관101 |

성적 테이블

| | scNum | stNum | lName | sc |
|---|-------|-------|--------|-----|
| 1 | 1 | 401 | 데이터베이스 | 4.0 |
| 2 | 2 | 402 | 스포츠경영학 | 3.5 |
| 3 | 3 | 501 | 데이터베이스 | 3.5 |
| 4 | 4 | 501 | 자료구조 | 3.5 |
| 5 | 5 | 502 | 자료구조 | 4.0 |

강좌 테이블

| | lName | lOffice |
|---|--------|---------|
| 1 | 데이터베이스 | 공학관110 |
| 2 | 스포츠경영학 | 체육관103 |
| 3 | 자료구조 | 공학관111 |

Inner Join

학생테이블

| | ABC stNum | ABC stName | ABC stAdd | ABC stMaj |
|---|-----------|------------|-----------|-----------|
| 1 | 401 | 김연아 | 대한민국 서울 | 컴퓨터학과 |
| 2 | 402 | 장미란 | 대한민국 강원도 | 체육학과 |
| 3 | 501 | 박지성 | 영국 맨체스터 | 체육학과 |
| 4 | 502 | 추신수 | 미국 클리블랜드 | 컴퓨터학과 |

학과테이블(reference 먼저 생성)

| | ABC mName | ABC mOffice |
|---|-----------|-------------|
| 1 | 체육학과 | 체육관101 |
| 2 | 컴퓨터학과 | 공학관101 |

뷰 테이블 생성

```
create view view01 as select s.stNum, s.stName, s.stAdd, s.stMaj, m.mOffice from students s inner join major m on s.stMaj = m.mName;

select * from view01 v ;
```

| | ABC stNum | ABC stName | ABC stAdd | ABC stMaj | ABC mOffice |
|---|-----------|------------|-----------|-----------|-------------|
| 1 | 402 | 장미란 | 대한민국 강원도 | 체육학과 | 체육관101 |
| 2 | 501 | 박지성 | 영국 맨체스터 | 체육학과 | 체육관101 |
| 3 | 401 | 김연아 | 대한민국 서울 | 컴퓨터학과 | 공학관101 |
| 4 | 502 | 추신수 | 미국 클리블랜드 | 컴퓨터학과 | 공학관101 |

성적테이블

| | 123 scNum | ABC stNum | ABC lName | ABC sc |
|---|-----------|-----------|-----------|--------|
| 1 | 1 | 401 | 데이터베이스 | 4.0 |
| 2 | 2 | 402 | 스포츠경영학 | 3.5 |
| 3 | 3 | 501 | 데이터베이스 | 3.5 |
| 4 | 4 | 501 | 자료구조 | 3.5 |
| 5 | 5 | 502 | 자료구조 | 4.0 |

강좌테이블

| | ABC lName | ABC lOffice |
|---|-----------|-------------|
| 1 | 데이터베이스 | 공학관110 |
| 2 | 스포츠경영학 | 체육관103 |
| 3 | 자료구조 | 공학관111 |

뷰 테이블 생성

```
create view view02 as select s.scNum, s.stNum, s.lName, l.lOffice, s.sc from score s inner join lecture l on s.lName = l.lName;

select * from view02 v ;
```

| | 123 scNum | ABC stNum | ABC lName | ABC lOffice | ABC sc |
|---|-----------|-----------|-----------|-------------|--------|
| 1 | 1 | 401 | 데이터베이스 | 공학관110 | 4.0 |
| 2 | 2 | 402 | 스포츠경영학 | 체육관103 | 3.5 |
| 3 | 3 | 501 | 데이터베이스 | 공학관110 | 3.5 |
| 4 | 4 | 501 | 자료구조 | 공학관111 | 3.5 |
| 5 | 5 | 502 | 자료구조 | 공학관111 | 4.0 |


```
select v1.stNum, v1.stName, v1.stAdd, v1.stMaj, v1.mOffice, v2.scNum, v2.lName, v2.lOffice, v2.sc from view01 v1 inner join view02 v2 on v1.stNum = v2.stNum;
```

| | ABC stNum | ABC stName | ABC stAdd | ABC stMaj | ABC mOffice | 123 scNum | ABC lName | ABC lOffice | ABC sc |
|---|-----------|------------|-----------|-----------|-------------|-----------|-----------|-------------|--------|
| 1 | 501 | 박지성 | 영국 맨체스터 | 체육학과 | 체육관101 | 3 | 데이터베이스 | 공학관110 | 3.5 |
| 2 | 401 | 김연아 | 대한민국 서울 | 컴퓨터학과 | 공학관101 | 1 | 데이터베이스 | 공학관110 | 4.0 |
| 3 | 402 | 장미란 | 대한민국 강원도 | 체육학과 | 체육관101 | 2 | 스포츠경영학 | 체육관103 | 3.5 |
| 4 | 501 | 박지성 | 영국 맨체스터 | 체육학과 | 체육관101 | 4 | 자료구조 | 공학관111 | 3.5 |
| 5 | 502 | 추신수 | 미국 클리블랜드 | 컴퓨터학과 | 공학관101 | 5 | 자료구조 | 공학관111 | 4.0 |

||

학생수강성적

| 학생번호 | 학생이름 | 주소 | 학과 | 학과사무실 | 강좌이름 | 강의실 | 성적 |
|------|------|----------|------|--------|--------|---------|-----|
| 501 | 박지성 | 영국 맨체스터 | 컴퓨터과 | 공학관101 | 데이터베이스 | 공학관 110 | 3.5 |
| 401 | 김연아 | 대한민국 서울 | 체육학과 | 체육관101 | 데이터베이스 | 공학관 110 | 4.0 |
| 402 | 장미란 | 대한민국 강원도 | 체육학과 | 체육관101 | 스포츠경영학 | 체육관 103 | 3.5 |
| 502 | 추신수 | 미국 클리블랜드 | 컴퓨터과 | 공학관101 | 자료구조 | 공학관 111 | 4.0 |
| 501 | 박지성 | 영국 맨체스터 | 컴퓨터과 | 공학관101 | 자료구조 | 공학관 111 | 3.5 |

조인을 통해 비정규형 형태의 테이블로 변환