

데이터베이스



DATABASE



목차

1. 데이터베이스란?
2. 데이터베이스 관리 시스템(DBMS)
3. 데이터베이스 설계(개념적/논리적 설계)
4. 논리적 설계 – Key, 정규화

Next

물리적 설계(반정규화)

SQL 문법

데이터베이스(Database)

구조화된 정보 또는 데이터의 조직화된 모음

일반적으로 데이터베이스 관리 시스템(DBMS)에 의해 제어

연결된 애플리케이션과 함께 데이터와 DBMS를 하나로 묶어 데이터베이스 시스템이라고 하며
단축하여 데이터베이스라고 함

데이터베이스 유형(1)

- 관계형 데이터베이스: 열과 행이 있는 테이블 집합으로 구성. 정형 정보에 액세스하는 가장 효율적이고 유연한 방법 제공
- 객체 지향 데이터베이스: 객체 지향 데이터베이스의 정보가 객체 형태로 표현
- 분산 데이터베이스: 서로 다른 사이트에 위치한 둘 이상의 파일로 구성
- 데이터 웨어하우스: 데이터의 중앙 저장소. 빠른 쿼리 및 분석을 위해 특별히 설계된 데이터베이스
- NoSQL 데이터베이스: 비정형 및 반정형 데이터를 저장하고 조작



데이터베이스 유형(2)

기술 개발 접근 방식의 변화와 클라우드 및 자동화 같은 획기적 기술 발전이 데이터베이스를 새로운 방향으로 이끀.

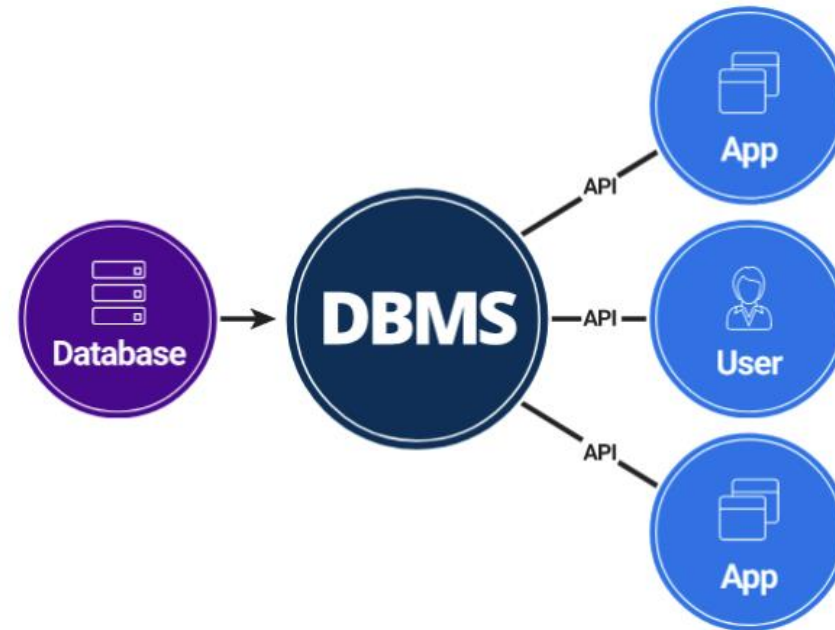
- 오픈 소스 데이터베이스: 소스 코드가 오픈 소스인 시스템. SQL, NoSQL 데이터베이스가 해당
- 클라우드 데이터베이스: private, public, hybrid 클라우드 컴퓨팅 플랫폼에 상주하는 정형, 비정형 데이터 모음
- 문서/JSON 데이터베이스: 행과 열이 아닌 JSON 형식으로 데이터를 저장하는 방식
- 자율 운영 데이터베이스: 클라우드를 기반으로 하며, 머신 러닝을 사용하여 데이터베이스 튜닝, 보안, 백업, 업데이트 및 기타 데이터베이스가 전통적으로 수행해 온 일상적인 관리 작업을 자동화

데이터베이스 관리 시스템(DBMS)

Database Management system의 준말

데이터를 한 곳에 모은 저장소를 만들고 그 저장소에 여러 사용자가 접근하여 데이터를 저장 및 관리 등의 기능을 수행하여 공유할 수 있는 환경을 제공하는 응용 소프트웨어 프로그램

공유 저장소(서버)를 구축하고 사용자들에게 접근 정보를 공유하여 데이터를 처리할 수 있는 인터페이스를 제공하고 복구 기능 과 보안성 기능 또한 제공한다.





데이터베이스 관리 시스템(DBMS) 기능

- 중복제어 기능: 동일한 데이터가 저장되는 것을 방지
- 접근 통제 기능: 사용자마다 다른 권한
- 편리한 인터페이스 제공
- 서로 다른 데이터 간 관계 표현
- 무결성 제약 조건을 정의, 검사
- 백업

데이터베이스 관리 시스템(DBMS) 종류

대표적으로 오라클, MySQL

MariaDB, Microsoft Access, Microsoft Excel 등 여러가지 DBMS가 존재하며 상황에 따라 다른 DBMS를 활용하며 각각의 DBMS의 내장함수들과 데이터 타입 등이 다르다.

데이터 모델(Data Model)

현실 세계의 정보를 인간과 컴퓨터가 이해할 수 있도록 추상화 하여 표현한 모델

요구조건 분석

요구사항, 범위 파악하여 외부 환경과 상호작용 분석을 통해 데이터에 대한 요구 분석

개념적 설계

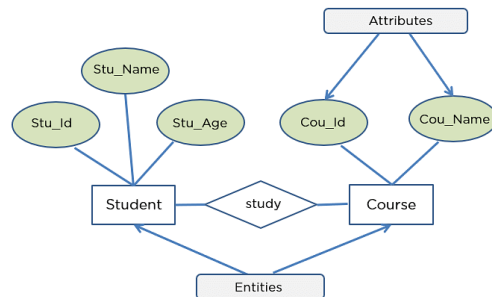
현실 세계에 대한 인식을 추상적, 개념적으로 표현하여 개념적 구조를 도출하는 데이터 모델
주요 산출물 : ER Diagram

논리적 설계

DBMS에 맞는 논리적 스키마를 설계하는 단계
논리적 설계 단계에서 정규화 수행
테이블을 설계하는 단계

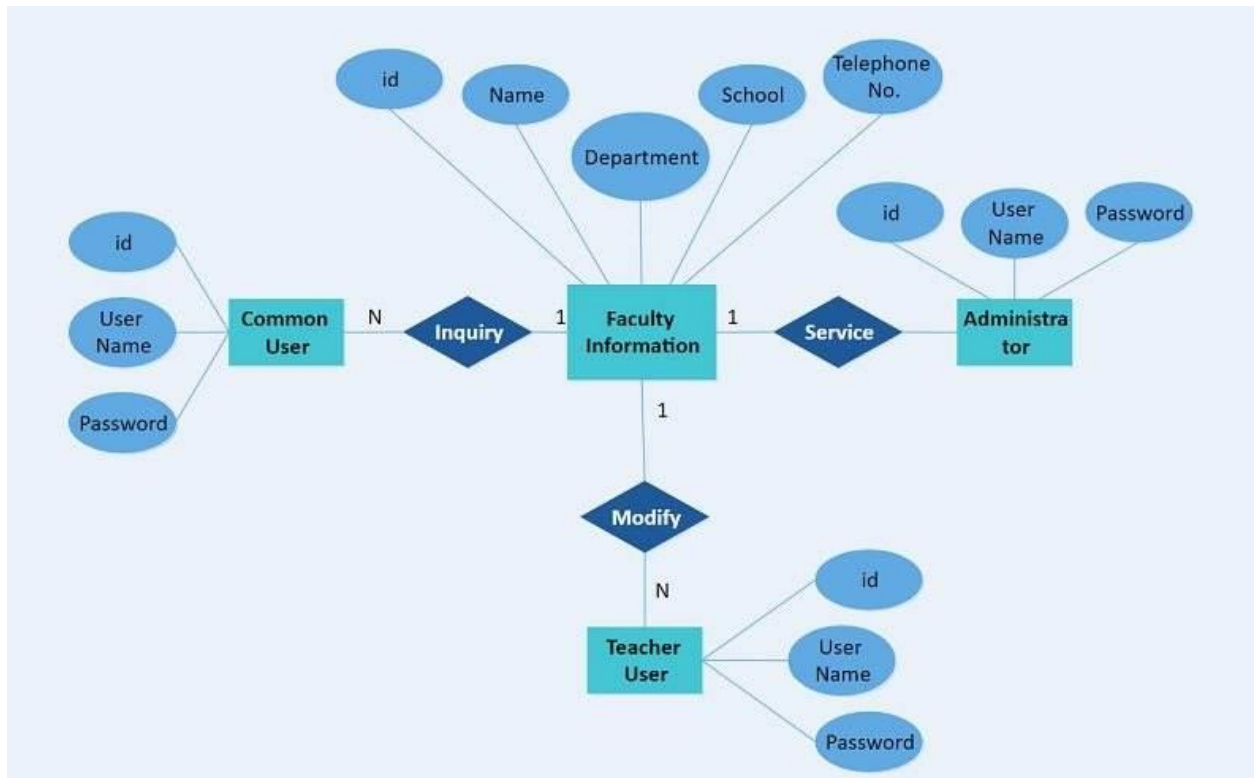
물리적 설계

논리 데이터 모델을 특정 DBMS의 특성 및 성능을 고려하여 물리적 스키마를 만드는 단계
테이블, 인덱스, 뷰 등 객체를 생성
성능 측면에서 반정규화 수행



개체-관계 모델(E-R Model)

현실 세계에 존재하는 데이터와 그들 간의 관계를 사람이 이해할 수 있는 형태로 명확하게 표현하기 위해 가장 널리 사용되고 있는 모델



Peter Chen Model

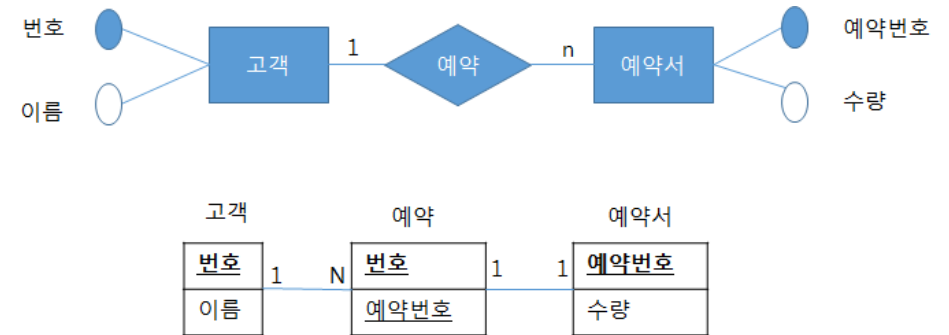
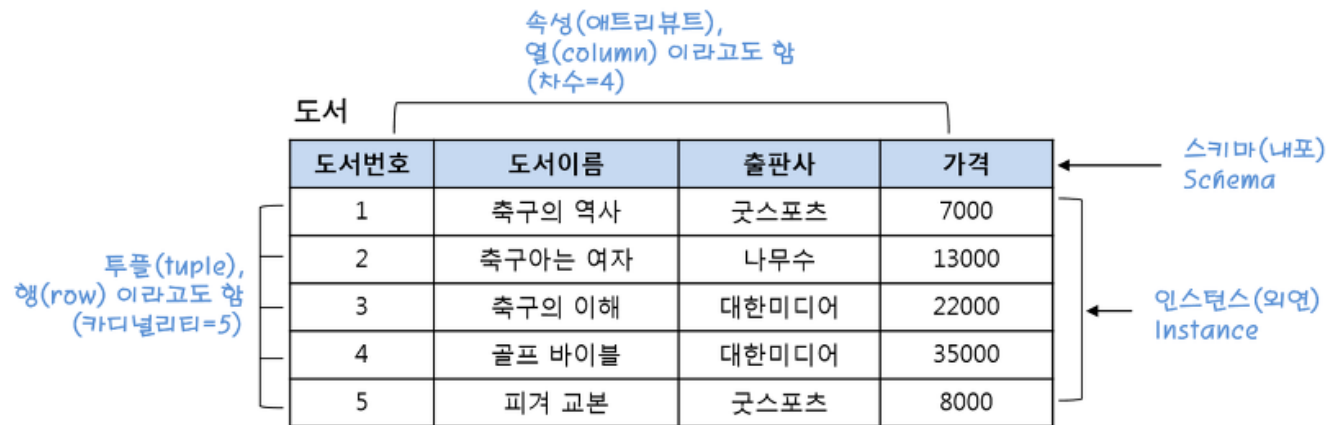
논리 데이터 모델

개념 모델로부터 업무 영역의 업무 데이터 및 규칙을 구체적으로 표현한 모델

관계 데이터 모델, 계층 데이터 모델, 네트워크 모델 존재

관계 데이터 모델(Relation Data Model)

데이터를 행과 열로 구성된 2차원 테이블 형태로 구성한 모델



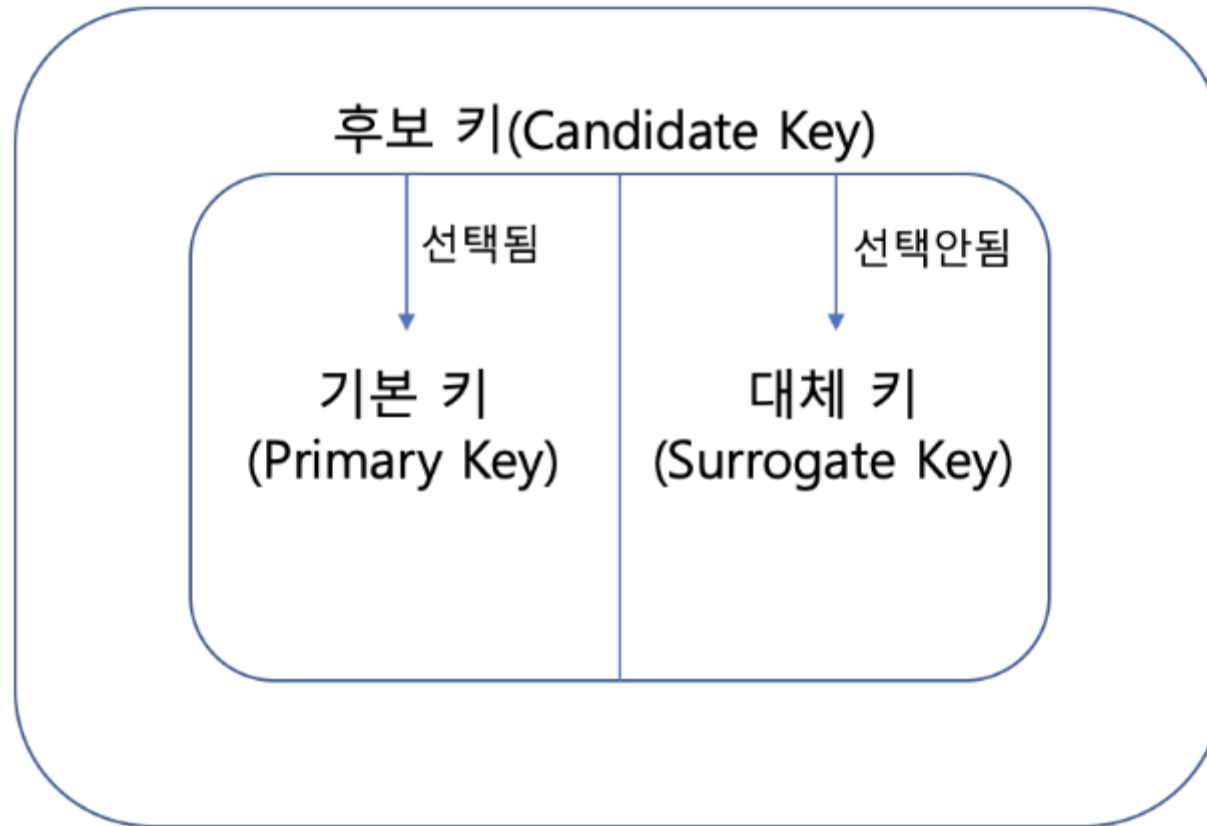
Key

데이터베이스에서 조건에 만족하는 튜플을 찾거나 순서대로 정렬할 때 다른 튜플들과 구별할 수 있는 유일한 기준이 되는 Attribute(속성)

Key의 종류

- 슈퍼 키(Super Key): 유일성을 만족하는 키
- 복합 키(Composite Key): 2개 이상의 속성(attribute)를 사용한 키
- 후보 키(Candidate Key): 유일성과 최소성을 만족하는 키. 기본키가 될 수 있는 후보이기 때문에 후보키라고 불림.
- 기본 키(Primary Key): 후보 키에서 선택된 키. NULL 값이 들어갈 수 없으며 기본키로 속성(Attribute)은 동일한 값이 들어갈 수 없다.
- 대체 키(Surrogate Key): 후보 키 중에 기본 키로 선택되지 않은 키
- 외래 키(Foreign Key): 어떤 테이블 간의 기본 키를 참조하는 속성. 테이블들 간의 관계를 나타내기 위해 사용

슈퍼 키(Super Key)



슈퍼 키(Super Key)

유일성을 만족하는 키

{학번}, {주민번호}, {이름+나이}, {학번+주민번호}

어떤 속성끼리 묶든 중복값이 안 나오고 서로 구별만 할 수 있으면 됨.

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

후보 키(Candidate Key)

슈퍼키들 중에서 속성이 최소한의 갯수

{이름+나이}: 최소성을 만족하지 못해서 후보키가 될 수 없음.

7조 테이블

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

기본 키(Primary Key)

후보키들 중에서 하나를 선택한 키

최소성과 유일성을 만족함

테이블에서 기본키는 오직 1개만 지정 가능

기본키는 NULL 값을 절대 가질 수 없고, 중복된 값을 가질 수 없음

7조 테이블

학번이 기본키

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

대체 키(Alternate Key)

후보키가 두 개 이상일 경우 그 중에서 어느 하나를 기본키로 지정하고 남은 후보키가 대체키

{학번}: 기본 키 ? {주민번호}: 대체 키

학번 기본 키가 없어지게 되면 주민번호는 없어진 기본 키로 대체 가능

7조 테이블

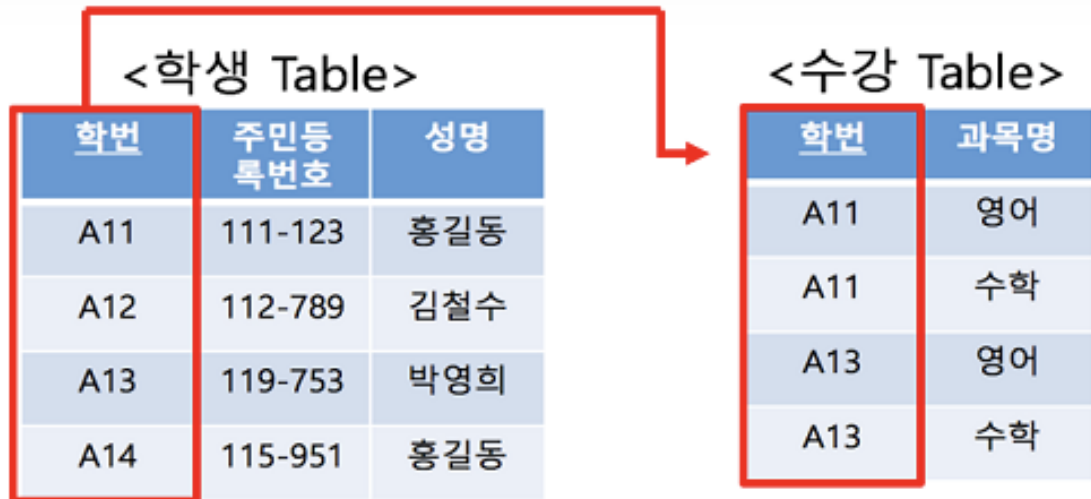
학번	주민번호	이름	나이
1	928888-8888888	김기범씨	27
2	929999-7777777	<u>추정범씨</u>	27
3	007777-6666666	이은빈씨	19
4	986666-5555555	<u>지혜리씨</u>	21

외래 키(Foreign Key)

참조될 속성 값은 테이블에서 기본키로 설정되어 있어야 함.

외래키는 참조되는 테이블의 기본키와 동일한 키 속성을 가짐.

외래 키 관계에서 부모 테이블이 삭제되면 자식 테이블을 먼저 삭제한 후 부모 테이블을 삭제해야 함.





정규화(Normalization)

목표: 테이블 간 중복된 데이터를 허용하지 않는 것 -> 무결성을 유지할 수 있음

이상 현상 방지를 위한 작업

이상현상(Anomaly)

테이블 내의 데이터들이 불필요하게 중복되어 테이블을 조작할 때 발생하는 데이터 불일치 현상

정규화를 통해 이상현상들을 해결할 수 있다.

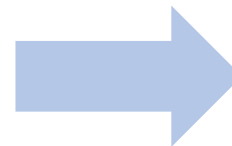
- 삽입 이상(insertion anomaly): 원하지 않는 자료가 삽입되거나, key가 없어 삽입하지 못하는(불필요한 데이터를 추가해야 삽입할 수 있음) 문제점
- 삭제 이상(deletion anomaly): 하나의 자료만 삭제하고 싶지만, 그 자료가 포함된 튜플 전체가 삭제됨으로 원하지 않는 정보 손실이 발생하는 문제점
- 갱신 이상(update anomaly): 일부만 변경하여 데이터가 불일치하는 모순, 또는 중복되는 튜플이 존재하게 되는 문제점

제 1 정규화

테이블의 컬럼이 원자값(Atomic Value)를 갖도록 테이블 분해

고객취미들(이름, 취미들)

이름	취미들
김연아	인터넷
추신수	영화, 음악
박세리	음악, 쇼핑
장미란	음악
박지성	게임



고객취미(이름, 취미)

이름	취미
김연아	인터넷
추신수	영화
추신수	음악
박세리	음악
박세리	쇼핑
장미란	음악
박지성	게임

제 2 정규화

테이블의 컬럼이 원자값(Atomic Value)를 갖도록 테이블 분해

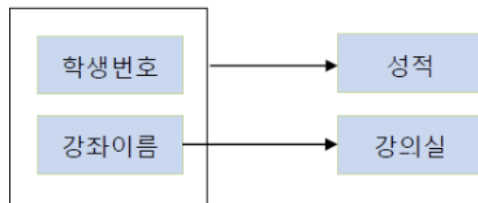
제 1 정규화를 진행한 테이블에 대해 **완전 함수 종속을 만족하도록** 테이블을 분해하는 것

*완전 함수 종속: 기본키의 부분집합이 결정자가 되면 안된다는 것

수강강좌

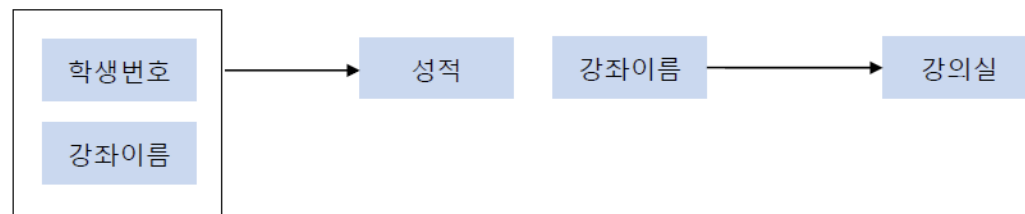
학생번호	강좌이름	강의실	성적
501	데이터베이스	공학관 110	3.5
401	데이터베이스	공학관 110	4.0
402	스포츠경영학	체육관 103	3.5
502	자료구조	공학관 111	4.0
501	자료구조	공학관 111	3.5

(학생번호, 강좌이름) → 성적
(강좌이름) → 강의실



수강

학생번호	강좌이름	성적
501	데이터베이스	3.5
401	데이터베이스	4.0
402	스포츠경영학	3.5
502	자료구조	4.0
501	자료구조	3.5



강의실

강좌이름	강의실
데이터베이스	공학관 110
스포츠경영학	체육관 103
자료구조	공학관 111

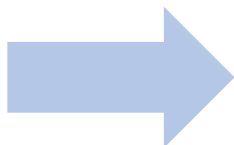
제 3 정규화

제 2 정규화를 진행한 테이블에 대해 **이행적 종속을 없애도록** 테이블을 분해하는 것

*이행적 종속: $A \rightarrow B$, $B \rightarrow C$ 가 성립할 때 $A \rightarrow C$ 가 성립되는 것

계절학기

학생번호	강좌이름	수강료
501	데이터베이스	20000
401	데이터베이스	20000
402	스포츠경영학	15000
502	자료구조	25000

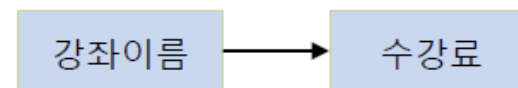
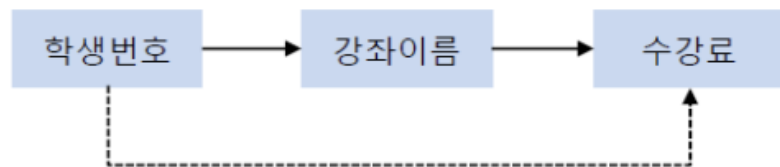


계절수강

학생번호	강좌이름
501	데이터베이스
401	데이터베이스
402	스포츠경영학
502	자료구조

수강료

강좌이름	수강료
데이터베이스	20000
스포츠경영학	15000
자료구조	25000



강좌가 변경되었을 경우, 수강료 따로 갱신 : 번거로움

BCNF 정규화

제 3 정규화를 진행한 테이블에 대해 모든 결정자가 후보키가 되도록 테이블을 분해하는 것

특강수강

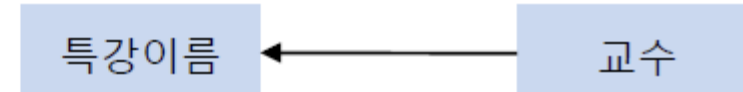
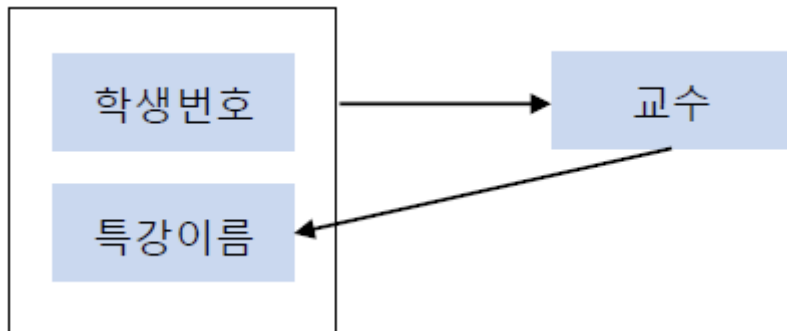
학생번호	특강이름	교수
501	소셜네트워크	김교수
401	소셜네트워크	김교수
402	인간과 동물	승교수
502	창업전략	박교수
501	창업전략	홍교수

특강신청

학생번호	교수
501	김교수
401	김교수
402	승교수
502	박교수
501	홍교수

특강교수

특강이름	교수
소셜네트워크	김교수
인간과 동물	승교수
창업전략	박교수
창업전략	홍교수



제 4 정규화

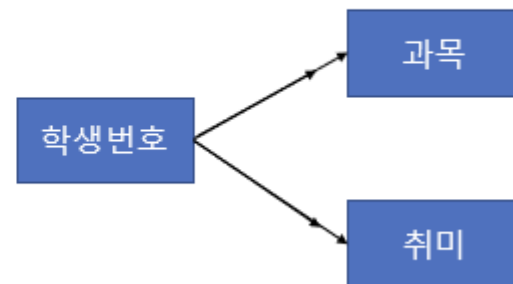
BCNF를 만족하고 다치 종속(Multi-valued Dependency)이 없어야 한다.

다치 종속

BCNF를 만족하고 다치 종속(Multi-valued Dependency)이 없어야 한다.

1. $A \twoheadrightarrow B$ 일 때 하나의 A 값에 여러 개의 B 값이 존재하면 다치 종속성을 가진다고 하고 $A \twoheadrightarrow B$ 라고 표시
2. 최소 3개의 칼럼이 존재
3. $R(A, B, C)$ 가 있을 때 A와 B 사이에 다치 종속성이 있을 때 B와 C가 독립적

학생번호	과목	취미
101	자바	노래
101	C++	게임
102	운영체제	노래
102	DB	게임



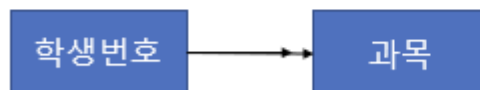
제 4 정규화

BCNF를 만족하고 다치 종속(Multi-valued Dependency)이 없어야 한다.

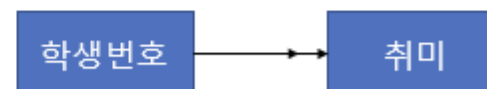
학생번호	과목	취미
101	자바	노래
101	자바	게임
101	C++	노래
101	C++	게임

[그림 2] 다치 종속성의 문제점

학생번호	과목
101	자바
101	C++
102	운영체제
102	DB



학생번호	취미
101	노래
101	게임
102	노래
102	게임



[그림 3] 4NF

제 5 정규화

중복을 제거하기 위해 분해할 수 있을 만큼 전부 분해하는 것
Project Join Normal Form(PJNF)라고 불린다.

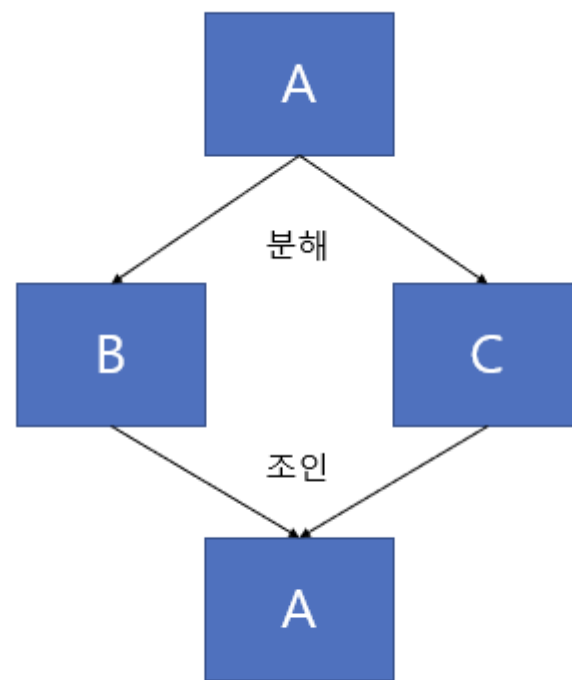
1. 4NF를 만족해야 한다.
2. **조인 종속(Join dependency)**이 없어야 한다.

*조인 종속: 하나의 릴레이션을 여러 개의 릴레이션으로 분해하였다가,
다시 조인했을 때 데이터 손실이 없고 필요 없는 데이터가 생기는 것

개발자	자격증
홍길동	정보처리기사
홍길동	빅데이터 분석기사
장길산	정보보안기사

자격증	언어
정보처리기사	C
빅데이터 분석기사	C++
정보보안기사	JAVA

개발자	언어
홍길동	C
홍길동	C++
장길산	JAVA



[그림 4] 조인 종속

