



1. 상세 개념적 모델링

1-1 상세 개념적 데이터 모델링

- ☐ 기본 개념적 데이터 모델링
 - 실세계의 현상을 있는 그대로 표현하는 것
- ☐ 상세 개념적 데이터 모델링
 - 실체와 속성을 추가하거나 제거하여 실체 간의 관계를 명확히 하는 단계
- ☐ 상세 개념적 데이터 모델링 단계를 수행하는 절차와 방법을 정규화라고 함

2. 정규화 (Normalization)

2-1 정규화를 하는 이유

- ☐ 데이터 중복성 제거
- ☐ 데이터 모형 단순화
- ☐ 속성 배열 상태 검증
- ☐ 실체와 속성의 누락 여부 검증
- ☐ 데이터 모형의 안정성 유지

2-2 데이터의 논리적 표현

- ☐ 관계 스키마 (relational scheme)의 설계
 - 관계 모델을 이용하여 어떻게 실세계를 정확히 표현할 것인가?
 - 속성 (attribute), 개체 (entity), 관계 (relationship) 파악
 - 관련된 속성들을 릴레이션으로 묶음
 - ✓ 데이터 종속성 : 속성들간의 관계성
 - ✓ 효율적인 데이터 조작
 - ✓ 데이터의 일관성 유지
 - 변칙적 성질의 예방
 - ✓ 이상 (anomaly)

2-3 이상 (anomaly)

- ☐ 릴레이션에 어떤 데이터 조작을 가한 경우, 릴레이션의 데이터 무결성을 위배하는 경우가 발생 → 이상 (anomaly)
- ☐ 이상 (anomaly) 유발 사항
 - 단일 값이 아닌 속성
 - 다중 의미를 가지는 속성
 - 추출 (파생) 속성
 - 인스턴스의 반복된 출현

2-3 이상 (anomaly)

- ☐ 단일 값이 아닌 속성
 - 속성이 유일한 값을 가지지 않는 경우 문제가 발생

사원번호	이름	나이	성별	주소	전화번호
100	홍길동	22	남	서울	011-555-7777
200	장길산	26	남	대구	051-888-8888, 011-888-7777
300	나애리	28	여	부산	053,999,5555
400	제갈량	30	남	광주	010-7777-6666

- 해결 방법
 - 전화번호 개수만큼 속성 추가
 - child entity 생성

2-3 이상 (anomaly)

☐ 다중 의미를 가지는 속성

- 하나의 속성에 두 개 이상의 의미를 담으려고 할 때 발생

회원번호	이름	나이	가입 및 탈퇴 날짜
100	홍길동	22	2007년 12월 30일
200	장길산	26	2001년 01월 01일
300	나애리	28	2009년 02월 03일
400	제갈량	30	2003년 10월 10일

- 해결 방법
 - 속성 상태 추가 (탈퇴, 가입)
 - 탈퇴날짜, 가입날짜 속성 추가

2-3 이상 (anomaly)

☐ 추출 (파생) 속성

- 데이터의 원본이 아닌 복제본이거나 계산된 값이 될 경우 발생

회원번호	이름	나이	생년월일
100	홍길동	29	1981년 02월 12일
200	장길산	26	1984년 01월 01일
300	나애리	28	1982년 02월 03일
400	제갈량	30	1980년 10월 10일

- 나이 값이 관리 목적이라면 추출 (파생) 속성이라고 볼 수 없음

2-3 이상 (anomaly)

데이터 중복을 최소화하는 것은
데이터 모델링에서 가장 중요한
요소

인스턴스의 반복 출현

- 속성의 잘못된 배치나 선정, Entity의 잘못된 분리로 인해 인스턴스 반복 발생

오피스텔

오피스텔 번호	건물 이름	건물주	등기날짜
100	두꺼비 빌딩	두꺼비	2003/01/01
200	개구리 빌딩	개구리	2004/08/15
300	서강중 빌딩	서강중	2009/03/16

방

오피스텔 번호	방 번호	층수	방주인 이름	건물 주소
100	103	1	홍길동	대구 북구 두꺼비동 111 두꺼비 빌딩
100	201	2	장길산	대구 북구 두꺼비동 111 두꺼비 빌딩
200	1101	1	나애리	서울 강남구 개구리동 222 개구리 빌딩
300	2201	2	제갈량	부산 진구 서강중동 333 서강중 빌딩

2-3 이상 (anomaly)

- 데이터의 불필요한 중복은 릴레이션을 조작할 때 곤란한 현상을 발생시킴

수강 릴레이션

SNO	CNO	GRA	YEAR
100	C413	A	4
100	E412	A	4
200	C123	B	3
300	C312	A	1
300	C324	C	1
300	C413	A	1
400	C312	A	4
400	C324	A	4
400	C413	B	4
400	E412	C	4
500	C312	B	2

2-3 이상 (anomaly) (계속)

☐ 삭제이상 (deletion anomaly)

- 200번 학생이 'c123'의 등록을 취소
⇒ 3학년이라는 정보도 함께 삭제됨
- 연쇄 삭제(triggered deletion)에 의한 정보의 손실

SNO	CNO	GRA	YEAR
100	C413	A	4
100	E412	A	4
200	C123	B	3
300	C312	A	1
300	C324	C	1
300	C413	A	1
400	C312	A	4
400	C324	A	4
400	C413	B	4
400	E412	C	4
500	C312	B	2

2-3 이상 (anomaly) (계속)

☐ 삽입이상 (insertion anomaly)

- 600번 학생이 2학년이라는 사실을 삽입
⇒ 어떤 과목을 등록하지 않는 한 삽입이 불가능 (SNO, CNO 기본 키)
- 원하지 않는 정보의 강제 삽입

SNO	CNO	GRA	YEAR
100	C413	A	4
100	E412	A	4
200	C123	B	3
300	C312	A	1
300	C324	C	1
300	C413	A	1
400	C312	A	4
400	C324	A	4
400	C413	B	4
400	E412	C	4
500	C312	B	2

600 2

2-3 이상 (anomaly) (계속)

☐ 갱신이상 (update anomaly)

- 400번 학생의 학년을 4에서 3으로 변경
⇒ 학년이 400인 4개의 튜플 모두를 갱신시켜야 함
- 중복 데이터의 일부 갱신으로 정보의 모순성 (inconsistency) 발생

SNO	CNO	GRA	YEAR
100	C413	A	4
100	E412	A	4
200	C123	B	3
300	C312	A	1
300	C324	C	1
300	C413	A	1
400	C312	A	4
400	C324	A	4
400	C413	B	4
400	E412	C	4
500	C312	B	2

2-3 이상 (anomaly) (계속)

☐ 이상의 원인

- 속성들 간에 존재하는 여러 종속관계를 하나의 릴레이션에 표현

☐ 이상의 해결

- 속성들 간의 종속관계를 분석하여 여러 개의 릴레이션으로 분해
⇒ 정규화 (normalization)

2-3 이상 (anomaly) (계속)

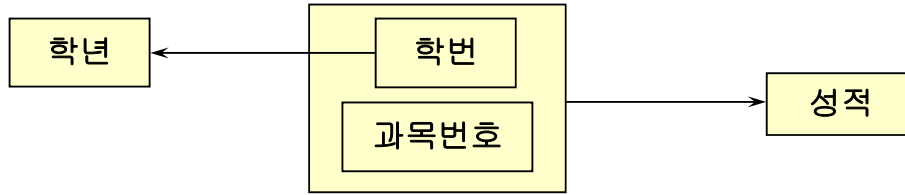
- ☐ 스키마 설계 : 데이터베이스의 논리적 설계
 - 속성들과 이들의 제약조건 (종속성) 들을 수집
 - 수집된 결과를 명시된 제약조건에 따라 여러 개의 릴레이션으로 분할
⇒ 스키마 변환 (schema transformation)
- ☐ 스키마 변환의 원리
 - 정보의 무손실
 - 데이터의 중복성 감소
 - 분리의 원칙

2-4 함수 종속 (FD : Functional Dependency)

- ☐ 어떤 릴레이션 R에서, 속성 X의 값 각각에 대해 속성 Y의 값이 하나만 연관
⇔ 속성 Y는 속성 X에 함수 종속 : $X \rightarrow Y$
- ☐ 속성 X는 Y를 (함수적으로) 결정. 즉, X는 Y의 결정자(determinant)
- ☐ X, Y는 복합 속성일 수 있음
- ☐ 릴레이션 R에서 속성 X가 키이면, R의 모든 속성 Y에 대해 $X \rightarrow Y$ 성립
- ☐ 함수종속 $X \rightarrow Y$ 의 경우
 - 속성 X가 반드시 키 (유일 값) 라는 것을 요건으로 하지 않음
즉, X(X) 값이 하나 이상의 튜플값으로 존재 가능

2-4 함수 종속 (FD : Functional Dependency) (계속)

☐ 수강 릴레이션 (기본키 : 학번, 과목번호)



(학번, 과목번호) → 성적

학번 → 학년

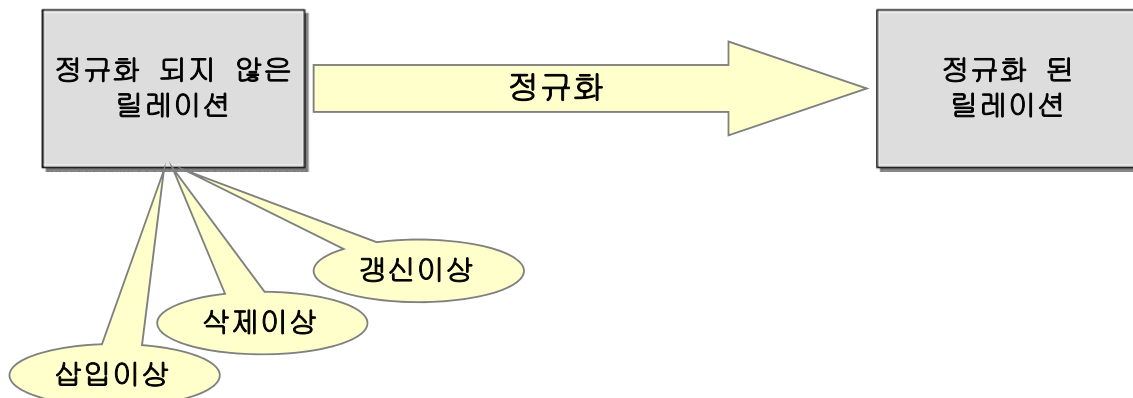
- (학년)은 (학번)에 **완전 함수 종속**
그러나, {학번, 과목번호}에는 **부분 함수 종속**
- (성적)은 {학번, 과목번호}에 **완전 함수 종속**

2-5 정규형 개요

☐ 정규형 (Normal Form) 의 본질

- 어떤 상황에서도 바람직한 모델을 유지하도록 하는 것

☐ 정규화 과정



2-5 정규형 개요

정규화 장점

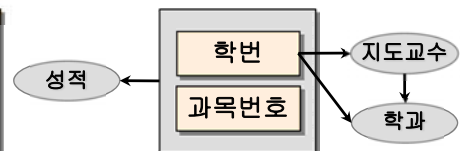
- 중복된 값이 줄어듬
- 널 값이 줄어듬
- 복잡한 코드로 데이터 모델을 보완할 필요가 없음
- 클러스터드 인덱스 최대 이용 가능
- 효율적인 테이블 생성 가능

2-5 제1정규형 (1NF:First Normal Form)

- 릴레이션에 포함되는 속성의 값이 다중값(multi value)을 지원하지 않고, 원자값(atomic value)만을 지원

수강지도 (학번, 지도교수, 학과, 과목번호, 성적)

학번	지도교수	학과	과목번호	성적
100	P1	컴퓨터	C413	A
100	P1	컴퓨터	E412	A
200	P2	전기	C123	B
300	P3	컴퓨터	C312	A
300	P3	컴퓨터	C324	C
300	P3	컴퓨터	C413	A
400	P1	컴퓨터	C312	A
400	P1	컴퓨터	C324	A
400	P1	컴퓨터	C413	B
400	P1	컴퓨터	E412	C



2-5 제1정규형 (1NF:First Normal Form) (계속)

☐ 1NF에서의 이상

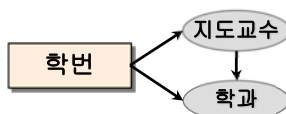
- 삽입이상
 - 500번 학생의 지도교수가 P4라는 사실만을 삽입할 경우 어떤 교과목을 등록하지 않는 한 삽입 불가능
- 삭제이상
 - 200번 학생이 C123의 등록을 취소하여 이 튜플을 삭제할 경우 지도교수가 P2라는 정보까지 손실됨
- 갱신이상
 - 400번 학생의 지도교수를 P1에서 P3로 변경할 경우 학번이 400인 4개 튜플의 지도교수 값을 모두 P3로 변경해야 함

2-6 제2정규형 (2NF:Second Normal Form)

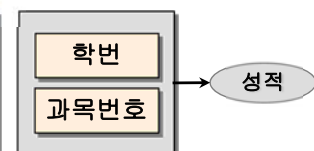
- ☐ 릴레이션이 제 1 정규형의 조건을 만족하고, 릴레이션의 기본키가 아닌 속성들이 기본키에 완전히 함수적으로 종속

지도 (학번, 지도교수, 학과) 수강 (학번, 과목번호, 성적)

학번	지도교수	학과
100	P1	컴퓨터
200	P2	전기
300	P3	컴퓨터
400	P1	컴퓨터



학번	과목번호	성적
100	C413	A
100	E412	A
200	C123	B
300	C312	A
300	C324	C
300	C413	A
400	C312	A
400	C324	A
400	C413	B
400	E412	C



2-6 제2정규형 (2NF:Second Normal Form) (계속)

☐ 2NF (지도 릴레이션)에서의 이상

- 삽입이상
 - 어떤 지도교수가 특정 학과에 속한다는 사실의 삽입 불가능
- 삭제이상
 - 300번 학생의 투플을 삭제하면 지도교수 p3가 컴퓨터공학과에 속한다는 정보 손실
- 갱신이상
 - 지도교수 p1의 소속이 컴퓨터공학과에서 전자과로 변경된다면 학번이 100과 400번인 두개의 투플을 모두 변경하여야 함

☐ 2NF 이상의 원인 : 이행적 함수 종속이 존재

2-6 제2정규형 (2NF:Second Normal Form) (계속)

☐ 이행적 함수 종속 (TD, Transitive Dependency)

- $A \rightarrow B$ 와 $B \rightarrow C \Rightarrow A \rightarrow C$
(즉, 애트리뷰트 c는 애트리뷰트 A에 이행적 함수 종속)

☐ 2NF 이상의 해결

- 프로젝션으로 릴레이션 분해 (이행적 함수 종속을 제거)
 $\Rightarrow 3NF$

2-7 제3정규형 (3NF:Third Normal Form)

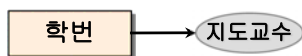
- ☐ 릴레이션이 제 2 정규형의 조건을 만족하고, 기본키가 아닌 모든 속성들이 기본키에 대하여 이행적 함수 종속(transitive FD)이 아니어야 함

학생지도 (학번, 지도교수)

학번	지도교수
100	P1
200	P2
300	P3
400	P1

지도교수학과 (지도교수, 학과)

지도교수	학과
P1	컴퓨터
P2	전기
P3	컴퓨터



2-7 제3정규형 (3NF:Third Normal Form) (계속)

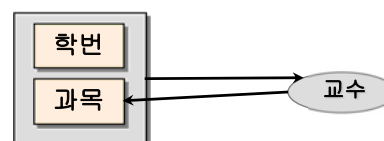
☐ 3NF의 약점

한 학생은 하나 이상의 전공을 신청, 교수는 하나의 전공과목 지도

- 복수의 후보키를 가지고 있고,
- 후보키들이 복합 속성들로 구성되고,
- 후보키들이 서로 중첩되는 경우에는 적용할 수가 없음

학번	과목	교수
100	프로그래밍	P1
100	자료구조	P2
200	프로그래밍	P1
200	자료구조	P3
300	자료구조	P3
300	프로그래밍	P4

수강과목 (학번, 과목, 교수)



2-7 제3정규형 (3NF:Third Normal Form) (계속)

3NF (수강과목 릴레이션)에서의 이상

- 삽입이상
 - 교수 p5가 자료구조를 담당한다는 사실의 삽입은 학번 (수강 학생) 이 있어야 가능
- 삭제이상
 - 100번 학생이 자료구조를 취소하여 튜플을 삭제하면 p2가 담당교수라는 정보도 삭제됨
- 갱신이상
 - p1이 프로그래밍 과목 대신 자료구조를 담당하게 되면 p1이 나타난 모든 튜플을 변경하여야 함

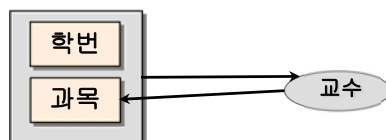
⇒ 원인 : 교수가 결정자이지만 후보키가 아님

2-8 보이스/코드 정규형 (BCNF:Boyce/Codd Normal Form)

릴레이션이 제 3 정규형의 조건을 만족하고, 릴레이션의 모든 결정자가 기본 키 (또는 후보키) 이어야 함

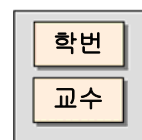
수강과목 (학번, 과목, 교수)

학번	과목	교수
100	프로그래밍	P1
100	논리회로	P2
200	프로그래밍	P1
200	자료구조	P3
300	자료구조	P3
300	프로그래밍	P4



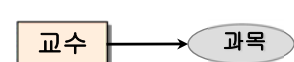
수강교수 (학번, 교수)

학번	교수
100	P1
100	P2
200	P1
200	P3
300	P3
300	P4



과목교수 (교수, 과목)

교수	과목
P1	프로그래밍
P2	논리회로
P3	자료구조
P4	프로그래밍



2-9 제4정규형 (4NF:Fourth Normal Form)

다치 종속 (MVD, Multi-Valued Dependency)

- 주어진 릴레이션이 최소 3개 이상의 속성들을 가지고, 그 중에서 2개 이상이 다중 값 속성
- $R(A,B,C)$ 인 경우 $A \twoheadrightarrow B$ 이고, $A \twoheadrightarrow C$ 이면서 B 와 C 는 서로 독립적인 속성일 경우

2-9 제4정규형 (4NF:Fourth Normal Form) (계속)

비정규 릴레이션 (과목목록 릴레이션)

과목 (C)	교수 (P)	교재 (T)
파일처리	$\begin{Bmatrix} P1 \\ P2 \end{Bmatrix}$	$\begin{Bmatrix} T1 \\ T2 \end{Bmatrix}$
데이터베이스	P3	$\begin{Bmatrix} T3 \\ T4 \\ T5 \end{Bmatrix}$

정규 릴레이션 (개설과목 릴레이션)

과목 (C)	교수 (P)	교재 (T)
파일처리	P1	T1
파일처리	P1	T2
파일처리	P2	T1
파일처리	P2	T2
데이터베이스	P3	T3
데이터베이스	P3	T4
데이터베이스	P3	T5

2-9 제4정규형 (4NF:Fourth Normal Form) (계속)

개설 교과목에서의 변경 이상

- 교수 P4가 데이터베이스를 담당한다는 정보 삽입시 3개의 교재에 대한 튜플을 삽입해야 함

BCNF 이상의 원인

- 즉, 과목은 교수나 교재의 값 하나를 결정하는 것이 아니라 값의 집합 (set of values)을 결정

2-9 제4정규형 (4NF:Fourth Normal Form) (계속)

과목 (C)	교수 (P)	교재 (T)
파일처리	P1	T1
파일처리	P1	T2
파일처리	P2	T1
파일처리	P2	T2
데이터베이스	P3	T3
데이터베이스	P3	T4
데이터베이스	P3	T5

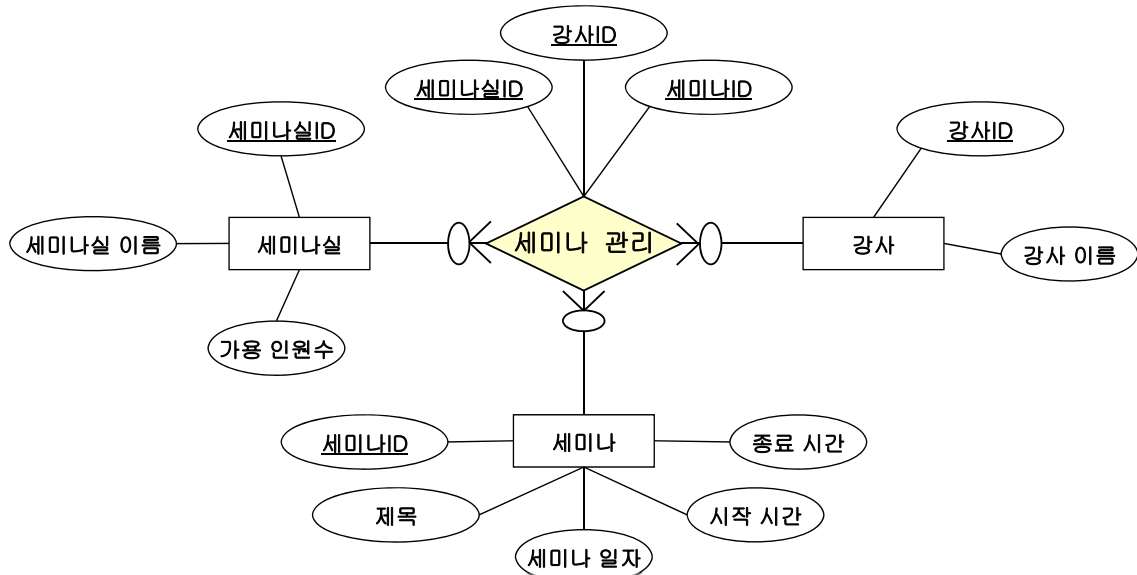
MVD 과목 ->> 교수 | 교재

과목 (C)	교수 (P)
파일처리	P1
파일처리	P2
데이터베이스	P3

과목 (C)	교재 (T)
파일처리	T1
파일처리	T2
데이터베이스	T3
데이터베이스	T4
데이터베이스	T5

2-9 제4정규형 (4NF:Fourth Normal Form) (계속)

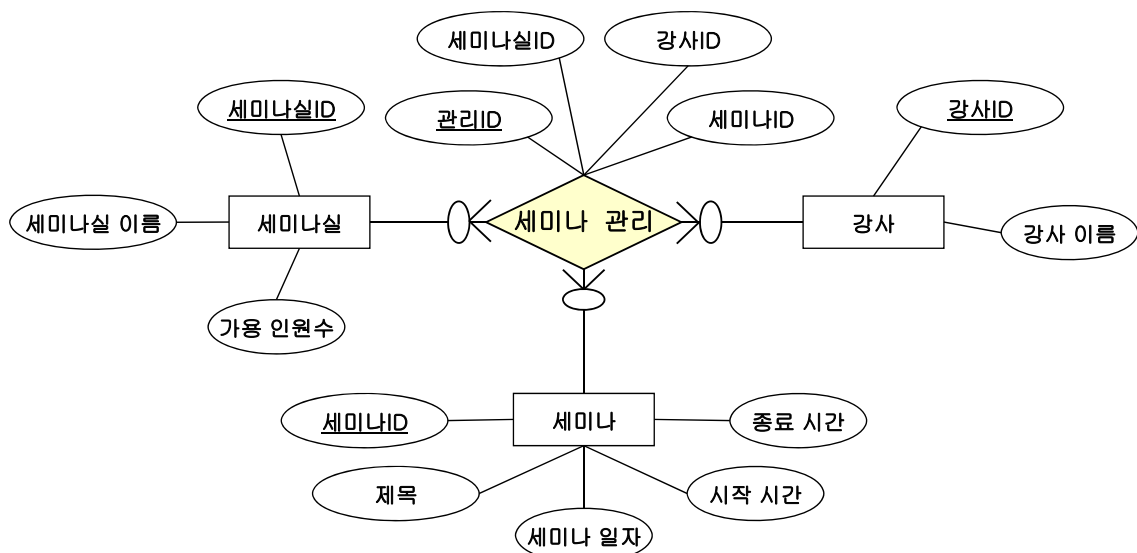
3항 이상의 릴레이션 관계



- 동일한 날짜에 동일한 강사가 동일한 세미나 제목으로 강의할 경우?

2-9 제4정규형 (4NF:Fourth Normal Form) (계속)

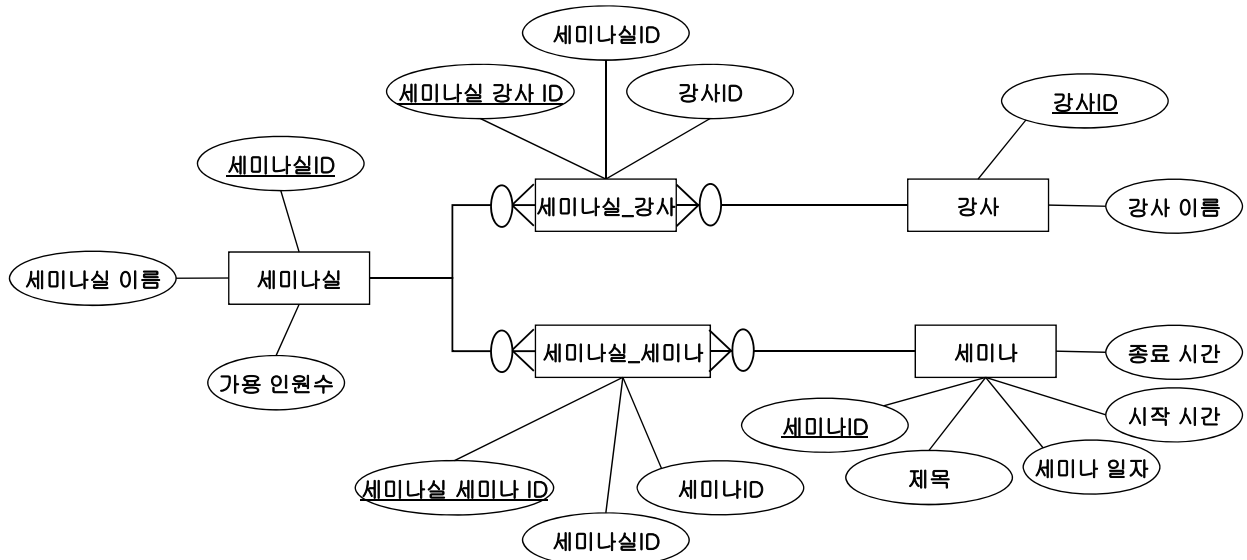
3항 이상의 릴레이션 관계 (계속)



- 다치 종속 해결?

2-9 제4정규형 (4NF:Fourth Normal Form) (계속)

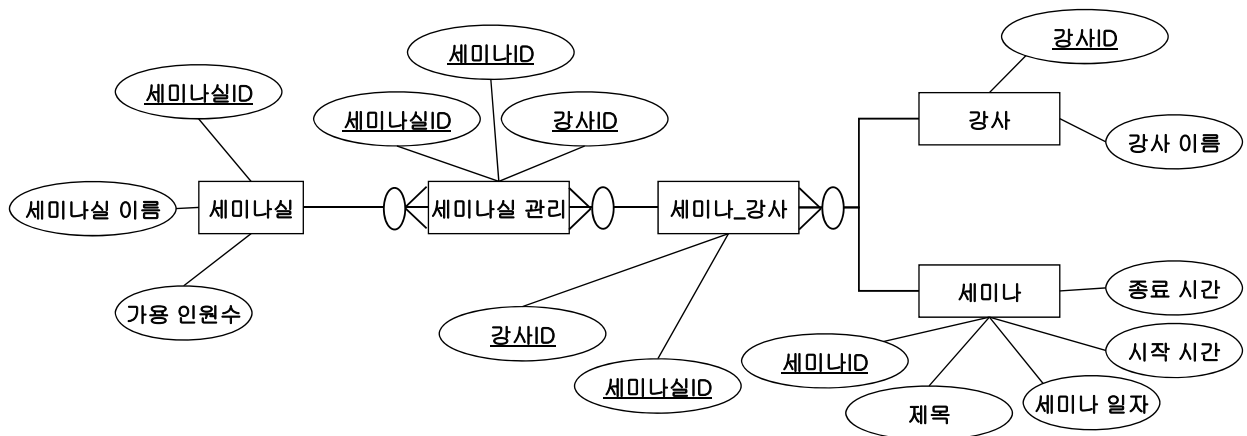
3항 이상의 릴레이션 관계 (계속)



- 다치종속을 해결했다고 좋은 모델일까?

2-9 제4정규형 (4NF:Fourth Normal Form) (계속)

3항 이상의 릴레이션 관계 (계속)

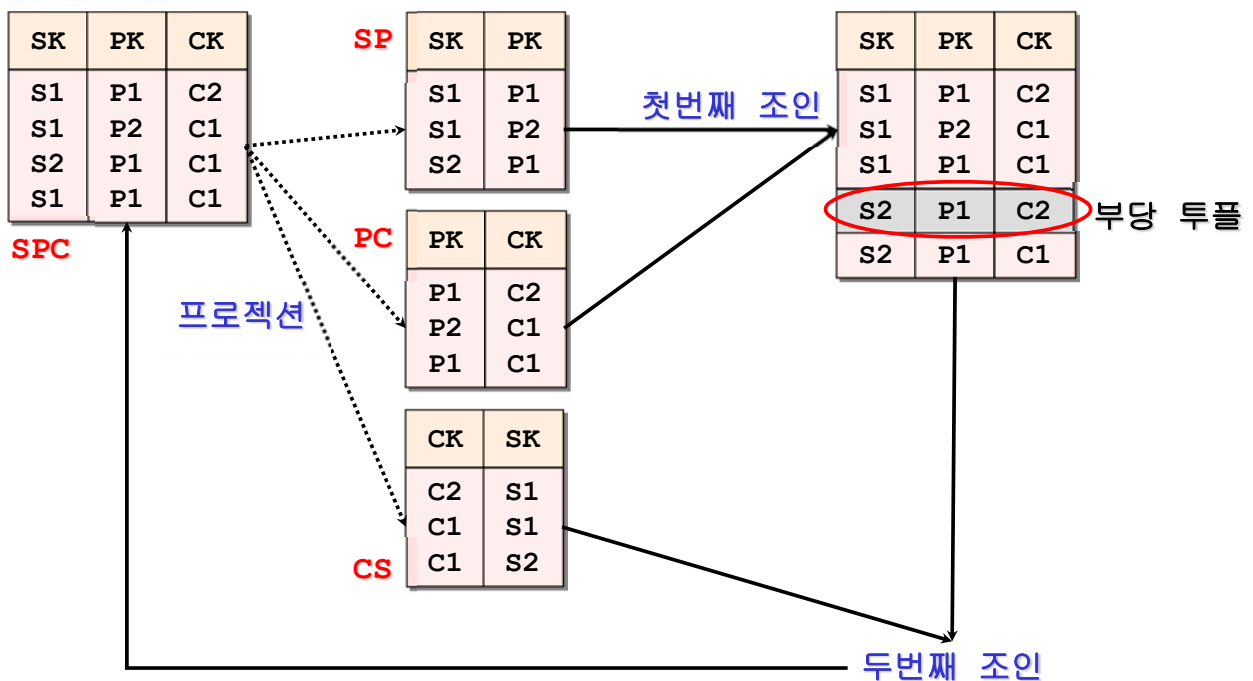


2-10 제5정규형 (5NF:Fifth Normal Form)

릴레이션 SPC (4NF)

- SPC를 **프로젝션**하여 세 개의 SP, PC, CS를 생성
- 세 개의 릴레이션 SP, PC, CS를 **조인**해서는 SPC의 재생성이 가능하나 그 어느 두 개의 조인만으로는 재생성 불가능

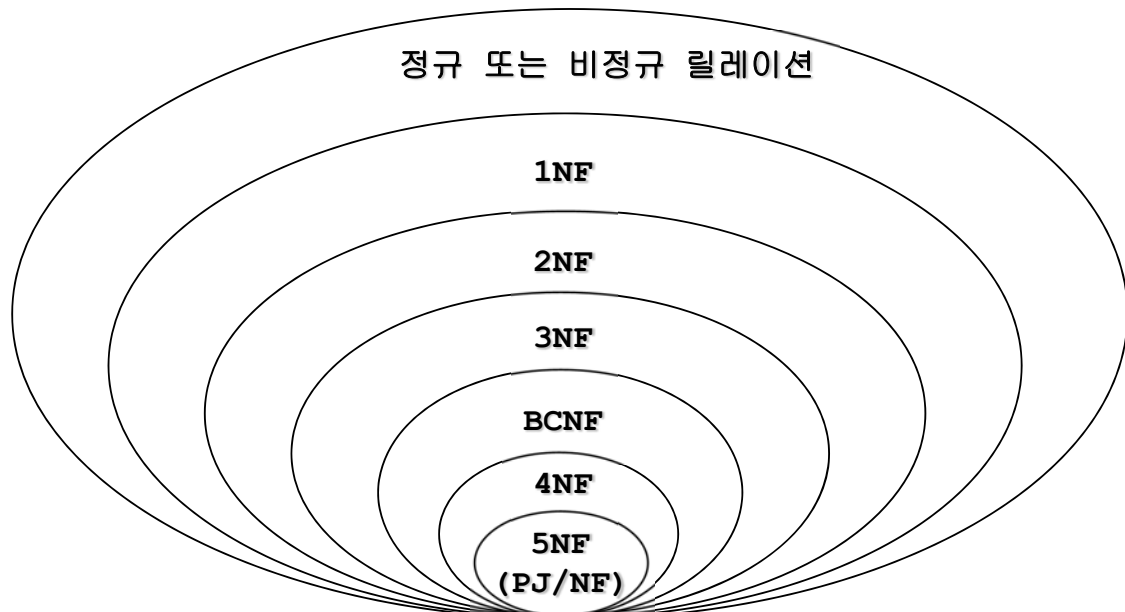
2-10 제5정규형 (5NF:Fifth Normal Form) (계속)



2-11 정규화 과정 (무손실 분해)



2-12 정규형들간의 관계



2-12 정규형들간의 관계 (계속)

- ☐ 릴레이션의 정규화는 실제 데이터 값이 아니라 개념적인 측면에서 다루어져야 함
- ☐ 실제 정규화 과정은 정규형의 순서와 다를 수 있음
- ☐ 현실적으로 모든 릴레이션을 반드시 5NF에 속하도록 분해할 필요는 없음
- ☐ 학생주소 (학번, 이름, 주소, 전화번호) : 5NF이 아님

기본키 : 학번

FD : 전화번호 → 주소



학생전화 (학번, 이름, 전화번호) : 5NF

전화주소 (전화번호, 주소) : 5NF

⇒ 이름, 전화번호, 주소는 분리하지 않고 사용하는 것이 편리하므로 위의 5NF으로의 분해는 무의미

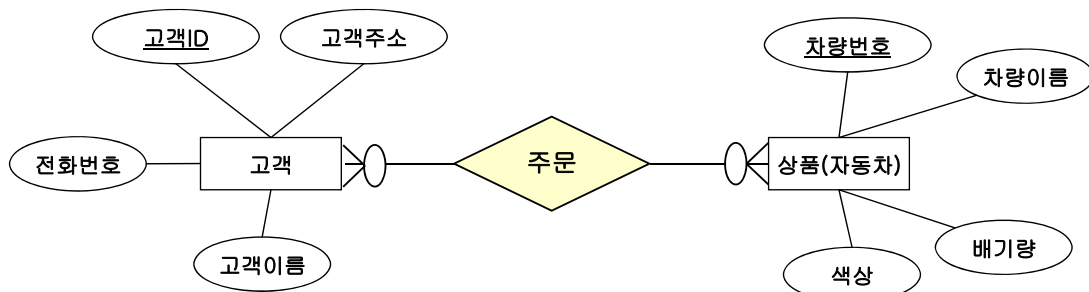
2-13 정규화의 장점과 단점

- ☐ 정규화 정도가 높은 경우
 - 유연한 데이터 구축 가능
 - 데이터의 높은 정확성
 - 물리적 접근 복잡
 - 길이가 짧은 데이터 생성
- ☐ 정규화 정도가 낮은 경우
 - 개체 (Entity) 수가 줄어듦

2-14 과도한 정규화 작업을 피해야 하는 경우

- ☐ 대상의 개체 (Entity)가 검색만으로 사용되는 경우
- ☐ 다른 개체의 속성과 함께 사용되며 그 빈도가 적은 경우
- ☐ 소규모의 데이터를 저장하는 경우이거나 드물게 변경되는 경우
- ☐ 중복되는 데이터의 변경이 거의 발생하지 않고 기억 공간이 비교적 적은 경우

2-15 N:M 관계의 해소



2-15 N:M 관계의 해소 (계속)

테이블 명 : 고객

컬럼 명	고객ID	고객이름	고객주소	전화번호
키 형태	PK			
견본 데이터	HONG	홍길동	서울	222-2222
	PARK	박찬호	대전	111-1111
	PSR	박세리	대구	333-3333

테이블 명 : 상품

컬럼 명	차량번호	차량이름	색상	배기량	고객ID
키 형태	PK				FK
견본 데이터	1	소나타	파랑	2000	HONG
	1	소나타	파랑	2000	PARK
	2	티코	노랑	800	PSR
	3	그랜저	보라	3000	HONG

2-15 N:M 관계의 해소 (계속)

테이블 명 : 고객

컬럼 명	고객ID	고객이름	고객주소	전화번호	차량번호
키 형태	PK				FK
견본 데이터	HONG	홍길동	서울	222-2222	1
	HONG	홍길동	서울	222-2222	3
	PARK	박찬호	대전	111-1111	1
	PSR	박세리	대구	333-3333	2

테이블 명 : 상품

컬럼 명	차량번호	차량이름	색상	배기량
키 형태	PK			
견본 데이터	1	소나타	파랑	2000
	2	티코	노랑	800
	3	그랜저	보라	3000

2-15 N:M 관계의 해소 (계속)

테이블 명 : 고객

컬럼 명	고객 ID	고객 이름	고객 주소	전화번호
키 형태	PK			
건본 데이터	HONG	홍길동	서울	222-2222
	PARK	박찬호	대전	111-1111
	PSR	박세리	대구	333-3333

테이블 명 : 상품

컬럼 명	차량 번호	차량 이름	색상	배기량
키 형태	PK			
건본 데이터	1	소나타	파랑	2000
	2	티코	노랑	800
	3	그랜저	보라	3000

테이블 명 : 계약

컬럼 명	고객 ID	차량 번호	계약일	계약금
키 형태	PK, FK	PK, FK		
건본 데이터	HONG	1	20000601	1000
	PARK	1	20000602	2000
	PSR	2	20000610	1600
	HONG	3	20000620	1500

3. 기타 모델링 기법

3-1 순환 관계형 데이터 모델

- ☐ 하나의 개체 (Entity) 에서 스스로 관계를 갖는 경우
- ☐ 자기 자신과 관계를 맺음으로 해서 각 레코드 간의 관계를 정의

테이블 명 : 사원

컬럼 명	사번	이름	주소	입사일자	퇴사일자	봉급	감독자사번
키 형태	PK						FK

테이블 명 : 사원

컬럼 명	사번	이름	주소	...	감독자사번
키 형태	PK				FK
건본 데이터	1	홍길동	서울		NULL
	2	박찬호	대구		1
	3	박세리	부산		2
	4	선동렬	광주		1

3-2 슈퍼 타입과 서브 타입형 데이터 모델

☐ 상위실체와 하위실체 (Super Type과 Sub Type Entity)

- 실체 (entity)는 상호 배타적인 2개 이상의 실체로 분할 가능
- 두 실체는 상호 공통의 속성이나 관계를 가질 수 있음
- 하위실체는 상위실체의 속성, 관계, 기능을 암시적으로 포함

사원

