("데이터베이스 설계 및 구축", 오세종, 생능출판사)

("SQL Server 2005 를 위한 데이터베이스 모델링", 김연홍, dbBADA) p40, p334~336

("사례로 배우는 데이터베이스 설계와 구축 - Database Zoo", 이종만, 이한출판사, 2006.6)

<정규화> : 데이터의 중복성을 최소화하기 위해 테이블을 분할하는 작업이다

* **함수적 종속성(FD : Functional Dependency)** : 한 릴레이션 내에 존재하는 속성들 사이의 관계에 대한 특별한 형태의 제약(속성들의 두 집합 사이의 제약조건)

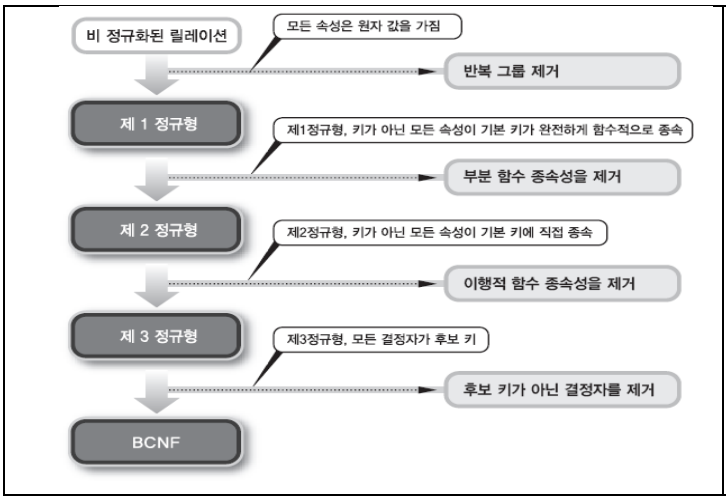
$$X \rightarrow Y$$

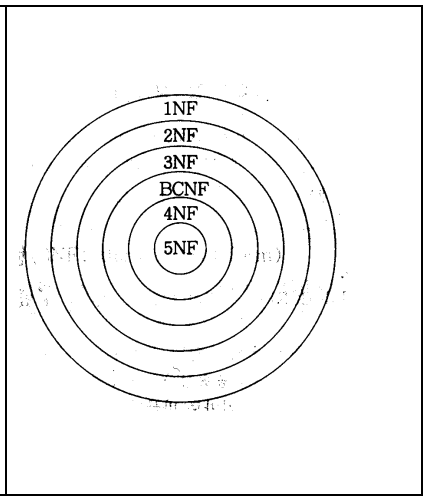
X는 Y를 함수적으로 결정한다(Functionally Determines) : 결정자

Y는 X에 함수적으로 종속된다(Functionally Dependent) : 종속자

(단, X → Y 가 성립한다고 해서 Y → X 가 성립하는 것은 아니다)

* **정규형** : 릴레이션의 속성 사이에 존재하는 함수 종속성과 키들을 기반으로 하여 그 릴레이션의 스키마를 분석할 수 있는 정형적 구조





<정규형 과정>

비정규형 테이블 → 1차정규형 → 2차정규형 → 3차정규형

↑ 반복그룹제거 ↑ 부분함수종속성 제거 ↑ 이행적함수종속성 제거

* **제1정규형(1NF)** : 비정규형 → 1차정규형(1NF)
모든 속성은 원자값을 가지도록 하고, 반복그룹을 제거한다. 1차정규형 테이블이라고 한다.

* **제2정규형(2NF)** : 1차정규형 → 2차정규형(2NF)
복합키의 일부분에 의해 종속되는 필드들이 존재할 경우 이를 제거 :(부분함수종속 제거)

* **제3정규형(3NF)** : 2차정규형 →3차정규형(3NF)
키가 아닌 일반 필드에 종속적인 필드가 존재할 경우 이를 제거 :(이행적함수종속 제거)

* **완전함수종속성(FD)** : 함수종속성 X → Y 에서 X로부터 어떤 속성 A를 제거하면 함수종속성 X → Y가 더 이상 성립하지 않는 경우의 종속성
* **부분함수종속성(PD)** : 함수종속성 X → Y 에서 X로부터 어떤 속성 A를 제거해도 함수종속성 X → Y가 성립하는 경우
* **이행함수종속성(TD)** : 릴레이션 스키마 R에서 어떤 키의 부분집합이 아닌 속성의 집합 Z가 존재하고 함수종속성 X → Y 와 Y → Z가 만족할 때, 이 두 함수종속성에 의해 추론될 수 있는 함수종속성 X → Z 가 성립하는 경우

[정규화 예제] 주문

비정규형

주문

주문ID	주문일	회원ID	회원명	회원등급	상품ID	상품명	단위	수량	단가
1	20020101	HONG	홍길동	일반	A101	숯팩	1-팩	2	22,000
					B102	연필	타스	3	11,000
					C202	전화기	1대	4	15,000
2	20020101	PARK	박찬호	일반	A101	숯팩	1-팩	2	20,000
3	20020103	PSR	박세리	우수	C202	전화기	1대	1	15,000
					D203	오이팩	2-팩	3	10,000
4	20020105	HONG	홍길동	일반	B102	연필	타스	2	11,000

다음을
제1정규형
제2정규형
제3정규형

으로 작성하시오

주문 (반복그룹 존재)

주문ID	주문일	회원ID	회원명	회원등급	상품ID	상품명	단위	수량	단가
1	20020101	HONG	홍길동	일반	A101	숯팩	1-팩	2	20,000
1	20020101	HONG	홍길동	일반	B102	연필	타스	3	11,000
1	20020101	HONG	홍길동	일반	C202	전화기	1대	4	15,000
2	20020101	PARK	박찬호	일반	A101	숯팩	1-팩	2	20,000
3	20020103	PSR	박세리	우수	C202	전화기	1대	1	15,000
3	20020103	PSR	박세리	우수	D203	오이팩	2-팩	3	10,000
4	20020105	HONG	홍길동	일반	B102	연필	타스	2	11,000

(풀이과정)

1) 제 1 정규형 (1NF)

1NF
주문

주문ID	주문일	회원ID	회원명	회원등급
1	20020101	HONG	홍길동	일반
2	20020101	PARK	박찬호	일반
3	20020103	PSR	박세리	우수
4	20020105	HONG	홍길동	일반

1NF
주문상세

주문ID	상품ID	상품명	단위	수량	단가
1	A101	숯팩	1-팩	2	20,000
1	B102	연필	타스	3	11,000
1	C202	전화기	1대	4	15,000
2	A101	숯팩	1-팩	2	20,000
3	C202	전화기	1대	1	15,000
3	D203	오이팩	2-팩	3	10,000
4	B102	연필	타스	2	11,000

2) 제 2 정규형 (2NF)

2NF
주문

주문ID	주문일	회원ID	회원명	회원등급
1	20020101	HONG	홍길동	일반
2	20020101	PARK	박찬호	일반
3	20020103	PSR	박세리	우수
4	20020105	HONG	홍길동	일반

2NF 아니다 (부분함수종속 존재)
주문상세

주문ID	상품ID	상품명	단위	수량	단가
1	A101	숯팩	1-팩	2	20,000
1	B102	연필	타스	3	11,000
1	C202	전화기	1대	4	15,000
2	A101	숯팩	1-팩	2	20,000
3	C202	전화기	1대	1	15,000
3	D203	오이팩	2-팩	3	10,000
4	B102	연필	타스	2	11,000

(반복그룹 표시)

2NF
주문

주문ID	주문일	회원ID	회원명	회원등급
1	20020101	HONG	홍길동	일반
2	20020101	PARK	박찬호	일반
3	20020103	PSR	박세리	우수
4	20020105	HONG	홍길동	일반

2NF
주문상세

주문ID	상품ID	수량
1	A101	2
1	B102	3
1	C202	4
2	A101	2
3	C202	1
3	D203	3
4	B102	2

2NF
상품

상품ID	상품명	단위	단가
A101	숯팩	1-팩	20,000
B102	연필	타스	11,000
C202	전화기	1대	15,000
D203	오이팩	2-팩	10,000

3) 제 3 정규형 (3NF)

3NF 아니다 (이행적함수종속 존재)
주문

주문ID	주문일	회원ID	회원명	회원등급
1	20020101	HONG	홍길동	일반
2	20020101	PARK	박찬호	일반
3	20020103	PSR	박세리	우수
4	20020105	HONG	홍길동	일반

(반복그룹 표시)

3NF
주문상세

주문ID	상품ID	수량
1	A101	2
1	B102	3
1	C202	4
2	A101	2
3	C202	1
3	D203	3
4	B102	2

3NF
상품

상품ID	상품명	단위	단가
A101	숯팩	1-팩	20,000
B102	연필	타스	11,000
C202	전화기	1대	15,000
D203	오이팩	2-팩	10,000

<최종 답안>

3NF
주문

주문ID	주문일	회원ID
1	20020101	HONG
2	20020101	PARK
3	20020103	PSR
4	20020105	HONG

3NF
회원ID

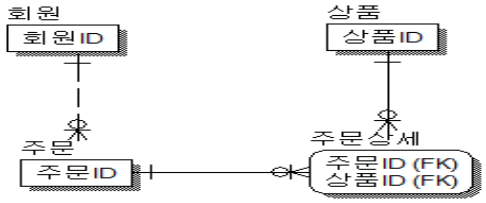
회원ID	회원명	회원등급
HONG	홍길동	일반
PARK	박찬호	일반
PSR	박세리	우수

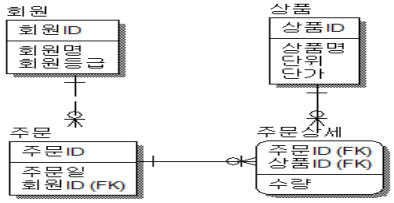
3NF
주문상세

주문ID	상품ID	수량
1	A101	2
1	B102	3
1	C202	4
2	A101	2
3	C202	1
3	D203	3
4	B102	2

3NF
상품

상품ID	상품명	단위	단가
A101	숯팩	1-팩	20,000
B102	연필	타스	11,000
C202	전화기	1대	15,000
D203	오이팩	2-팩	10,000





<최종 ERwin>

* 정규화(Normalization) : 데이터의 중복성을 최소화하기 위해 테이블을 분할하는 작업

* **데이터의 중복**은 데이터베이스를 저장하는데 필요한 저장공간을 낭비하고, 데이터의 수정, 삭제 등에 의해 일관성 문제를 일으킬 수 있다.
* 관계형 데이터베이스 설계의 **목표**는 데이터베이스 조작에 대하여 중복성을 배제하고 일관성을 유지할 수 있도록 데이터베이스의 관계구조를 구축하는 것이다.

*** 이상현상(Anomaly)**

이상이란 사용자가 어떤 테이블의 자료를 **삽입, 수정, 삭제 등의 변경 작업을 하고자 할 때 자료의 비일관성이 생기는 경우**를 말한다.
이러한 이상현상을 해결하기 위해 한 릴레이션에 존재하는 속성관계를 분석하고 속성들간의 종속성(dependency)을 분해하는 과정을 **정규화(Normalization)**라 한다.

수강과목

학번	과목번호	과목명
201219111	S111	DBMS
201219222	S222	Java
201219222	S333	알고리즘
201219333	S111	DBMS
201219333	S333	알고리즘

1) 삽입이상 : (NULL, S444, 데이터베이스) -> 엔티티무결성 제약조건 위배
2) 수정이상 : 과목번호 S111 의 과목명을 DB프로그래밍 으로 수정 -> 데이터불일치
3) 삭제이상 : (20129222, S222, Java) 삭제 -> 데이터 손실 (과목정보 삭제됨)

=>해결 : 수강(학번, 과목번호)
과목(과목번호, 과목명) 2개의 릴레이션으로 분해

*** 구조화된 관계의 문제점**

장점) 삽입, 삭제, 갱신이상 등의 문제점 해결 / 단점) 데이터베이스 실행속도저하, 참조무결성문제

<구조화되기 이전>	<정규화된 관계>
SELECT 학번, 과목명 FROM 수강과목;	SELECT 학번, 과목명 FROM 수강, 과목 WHERE 수강.과목번호 = 과목.과목번호;

* 정규화된 관계에서는 학번과 과목명이 한 테이블에 존재하지 않으므로 두 테이블을 결합시켜야

정규화된 관계의 문제점	정규화의 장점이 단점보다 크다
1) 명령어 복잡해지고 한 테이블에만 관련된 명령의 실행시간보다 더 많이 소요 2) 분리된 두 관계의 참조무결성 유지필요	→ 1) 성능우수 컴퓨터, 정교한 DBMS를 도입하여 해결 → 2) 참조무결성은 DBMS에서 자동적 통제

[기본예제 1-1] 다음 테이블에 대하여 답하시오.

사원취미 테이블을 제1정규화 하라.

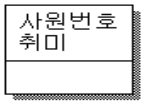
사원번호(PK)	취미
1001	등산 낚시
1002	테니스 등산
1003	볼링

(원자값이 아닌 경우) :
제1정규화하여 **테이블**로 작성하시오

<1NF>

사원번호(PK)	취미(PK)
1001	등산
1001	낚시
1002	테니스
1002	등산
1003	볼링

사원취미



[기본예제 1-2] 다음 테이블에 대하여 답하시오.

사원취미 테이블을 제1정규화 하라.

사원번호(PK)	취미1	취미2	취미3
1001	등산	낚시	
1002	테니스	등산	
1003	볼링		

(동일 속성의 반복) : 제1정규화하여 ERwin으로!

[기본예제 2] 다음 회원 테이블을 정규화하라

회원

회원번호	이름	주소	휴대폰	주민등록번호	자격증	취득일
1	홍길동	서울시 면목 8 동	(010)-111-1111	111111-1111111	운전면허 1 종	19980203
					한문 3 급	19990603
					정보처리 1 급	20000502
2	이준호	부산 연산구		222222-2222222	정보처리 1 급	20000502

(풀이과정)

비정규형 (원자값이 아니기 때문)

회원

회원번호	이름	주소	휴대폰번호	주민등록번호	자격증	취득일
1	홍길동	서울시 면목 8동	(010)-111-1111	111111-1111111	운전면허1종, 한문3급, 정보처리1급	19980203, 19990603, 20000502
2	이준호	부산 연산구		222222-2222222	정보처리1급	20000502

(원자값으로 작성)

회원

회원번호	이름	주소	휴대폰번호	주민등록번호	자격증	취득일
1	홍길동	서울시 면목 8동	(010)-111-1111	111111-1111111	운전면허1종	19980203
					한문3급	19990603
					정보처리1급	20000502
2	이준호	부산 연산구		222222-2222222	정보처리1급	20000502

1NF 아니다 (반복그룹 존재하기 때문)

회원

회원번호	이름	주소	휴대폰번호	주민등록번호	자격증	취득일
1	홍길동	서울시 면목 8동	(010)-111-1111	111111-1111111	운전면허1종	19980203
1	홍길동	서울시 면목 8동	(010)-111-1111	111111-1111111	한문3급	19990603
1	홍길동	서울시 면목 8동	(010)-111-1111	111111-1111111	정보처리1급	20000502
2	이준호	부산 연산구		222222-2222222	정보처리1급	20000502

1NF 만족 (원자값, 반복그룹 제거한 결과)

회원

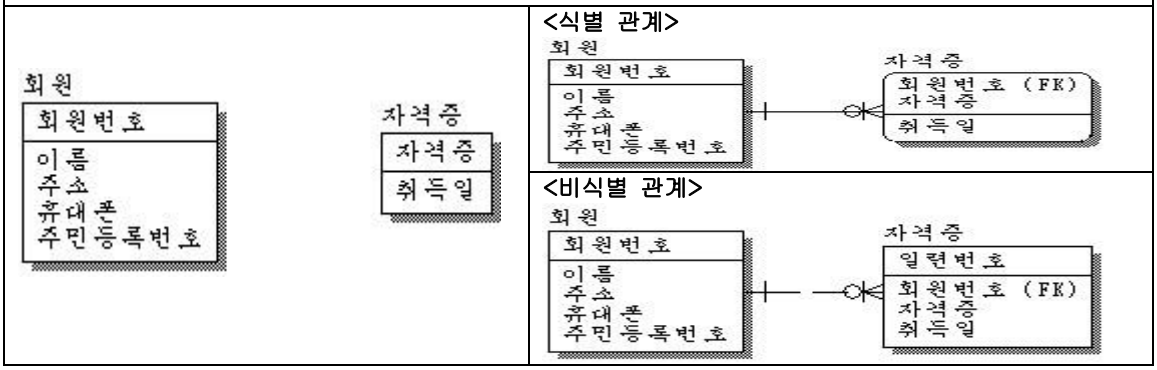
회원번호	이름	주소	휴대폰번호	주민등록번호
1	홍길동	서울시 면목 8동	(010)-111-1111	111111-1111111
2	이준호	부산 연산구		222222-2222222

회원자격증

회원번호	자격증	취득일
1	운전면허1종	19980203
1	한문3급	19990603
1	정보처리1급	20000502
2	정보처리1급	20000502

즉, 1NF => 회원(회원번호, 이름, 주소, 휴대폰번호, 주민등록번호)
회원자격증(회원번호, 자격증, 취득일) 2개의 테이블로 분해된다.

* 표시 기본키(PK)
외래키(FK)
반복그룹 영역



[기본예제 3-1] 다음 사원 테이블을 정규화하라

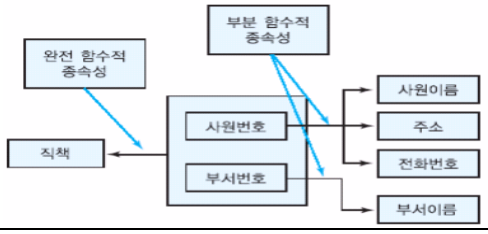
사원

사원번호	사원이름	주소	전화번호	직책	부서번호	부서이름
4257	정미림	홍제동	731-3497	팀장	1	홍보
1324	이범수	양재동	653-7412	프로그래머	2	개발
1324	이범수	양재동	653-7412	웹 디자이너	1	홍보
3609	안명석	양재동	425-8520	팀장	3	총무

함수적종속성(FD) :

사원번호 → 사원이름
사원번호 → 주소
사원번호 → 전화번호
부서번호 → 부서이름
{사원번호, 부서번호} → 직책

함수적종속다이어그램(FDD)



<최종 관계스키마>

[기본예제 3-2] 다음 학과등록 테이블을 정규화하라

테이블 명 : 학과 등록

컬럼 명	학번	과정코드	평가코드	과정명	기간
키 형태	PK				
건본 데 이 터	100	A01	A	JAVA 프로그래밍	4개월
	101	B01	D+	웹 마스터	6개월
	101	B03	A	DBMS 전문가	5개월
	100	B01	B	웹 마스터	6개월

함수적종속성(FD) :

{학번, 과정코드} → 평가코드
과정코드 → 과정명, 기간

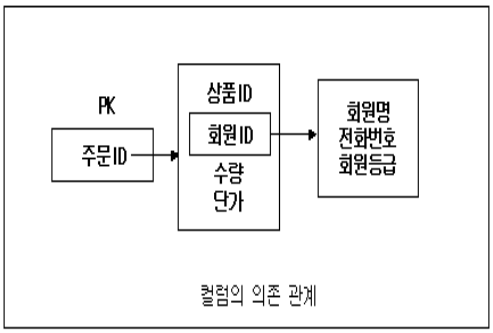
함수적종속다이어그램(FDD)

<최종 관계스키마>

[기본예제 4] 다음 주문 테이블을 정규화하라

테이블 명 : 주문

컬럼 명	주문ID	상품ID	회원ID	회원명	전화번호	회원등급	수량	단가
키 형태	PK							
건본 데이터	1	A01	HONG	홍길동	02)222-2222	우수	2	10,000
	2	B03	PARK	박찬호	043)111-1111	일반	1	20,000
	3	A02	PSR	박세리	031)222-2222	일반	3	15,000
	4	B03	HONG	홍길동	02)222-2222	우수	4	16,000



테이블 명 : 주문

컬럼 명	주문ID	상품ID	회원ID	수량	단가
키 형태	PK		FK		
건본 데 이 터	1	A01	HONG	2	10,000
	2	B03	PARK	1	20,000
	3	A02	PSR	3	15,000
	4	B03	HONG	4	20,000

테이블 명 : 회원

컬럼 명	회원ID	회원명	전화번호	회원등급
키 형태	PK			
건본 데 이 터	HONG	홍길동	02)222-2222	우수
	PARK	박찬호	043)111-1111	일반
	PSR	박세리	031)222-2222	일반

3 정규화에 의한 테이블 분리

<최종 관계스키마>

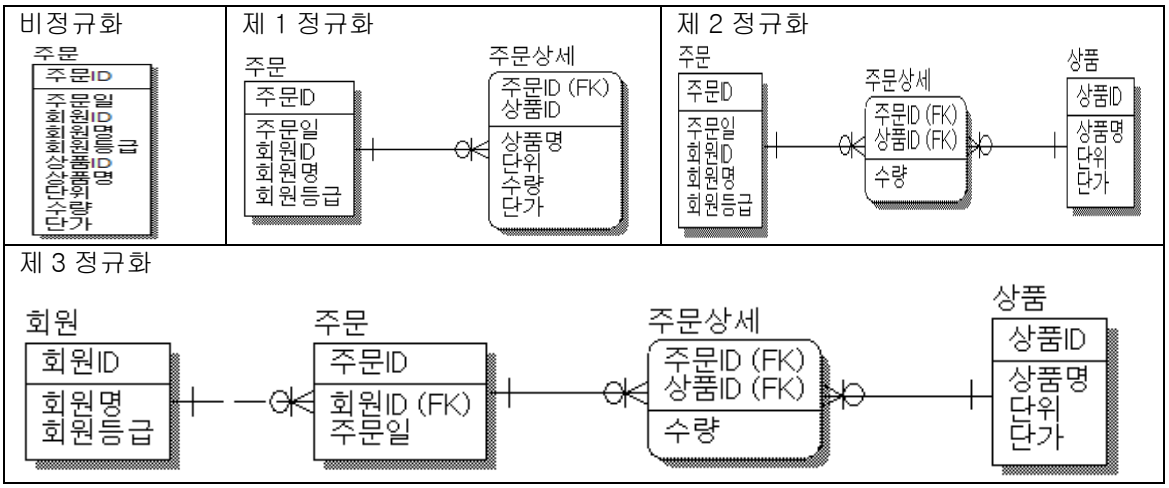
[Quiz 1] 다음을 1, 2, 3 차정규형으로 변환하시오

테이블 명 : 주문

컬럼 명	주문ID	주문일	회원ID	회원 명	회원등급	상품ID	상품 명	단위	수량	단가
키 형태	PK									
견본 데이터	1	20020101	HONG	홍 길 동	일반	A101	숲 팩	1-팩	2	20,000
						B102	연필	타 스	3	11,000
						C202	전화기	1대	4	15,000
	2	20020101	PARK	박 찬 호	일반	A101	숲 팩	1-팩	2	20,000
	3	20020103	PSR	박 세 리	우수	C202	전화기	1대	1	15,000
						D203	오이 팩	2-팩	3	10,000
	4	20020105	HONG	홍 길 동	일반	B102	연필	타 스	2	11,000

주문 테이블

<답안>



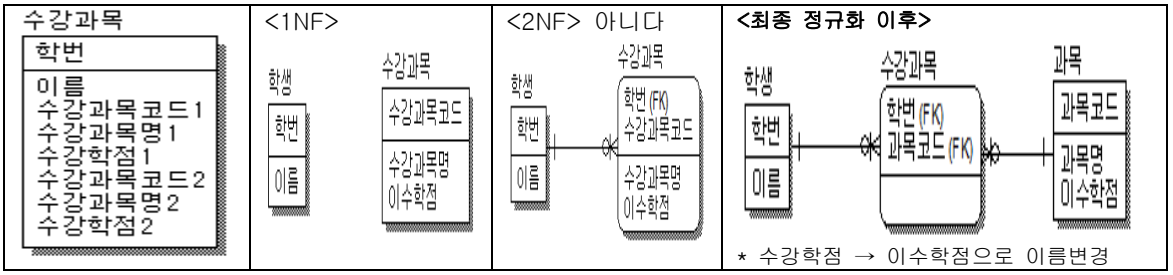
[실습예제 1] 다음은 학생의 수강과목 정보를 관리하는 엔티티에 대한 것이다.

수강과목

학번 : 20080111 이름 : 이준호

수강과목코드	수강과목명	수강학점
CS310	자료구조	3
CS313	DB프로그래밍	3

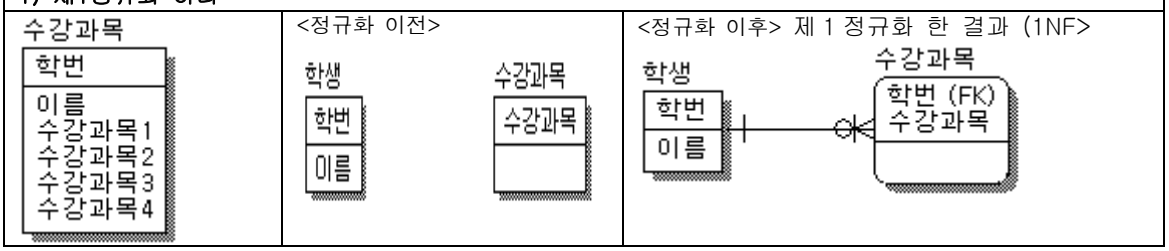
<답안>



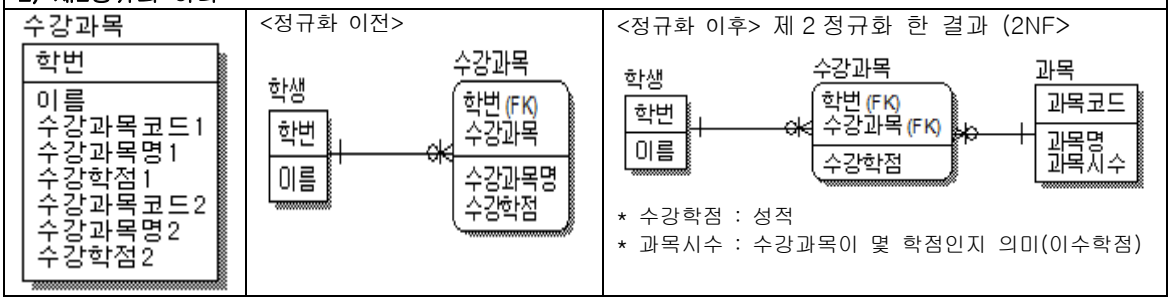
[Quiz 2] **<8장 연습문제 2~4번>** (오세종 p242~243)

다음은 학생의 수강과목 정보를 관리하는 엔티티에 대한 것이다.

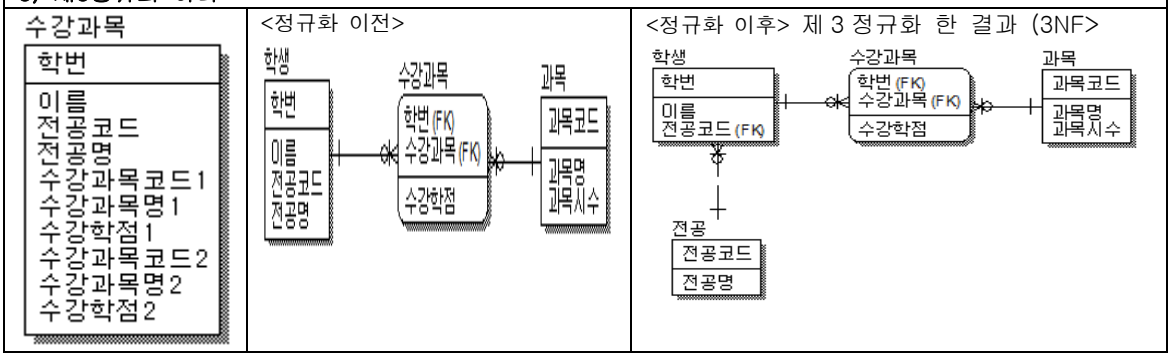
1) 제1정규화 하라



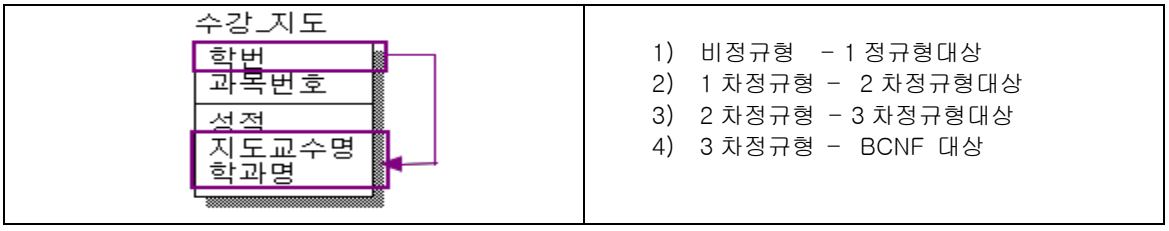
2) 제2정규화 하라



3) 제3정규화 하라



[Quiz 3] {학번,과목번호} 가 결정자이면서 PK 이고 {성적,지도교수명,학과명}이 종속자이다. 속성 중 학번이 결정자이고 {지도교수명,학과명}이 학번에만 함수종속성을 가진다면 이는 몇 차 정규형에 속하고 몇 차 정규화의 대상인가?



1) 비정규형 - 1 정규형대상
2) 1 차정규형 - 2 차정규형대상
3) 2 차정규형 - 3 차정규형대상
4) 3 차정규형 - BCNF 대상

[양식 예제] 판매전표 (오세종 8 장 p226, p228~239) → 1, 2, 3 차정규형으로 변환

판매 전표

판매일자 : *1997.03.16*
판매부서 : *DH01* **부서명** : *영업 1 부*
판매번호 : *16-12*
판매사원번호 : *12437* **사원명** : *김철수*

고객번호 : *YS02* **고객명** : *영진전자*
고객주소 : *서울시 영등포구 여의도동 137-27*

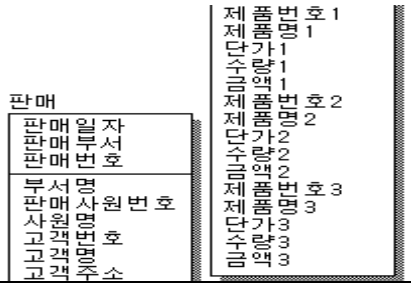
← 머리부분

제품번호	제품명	단가	수량	금 액
DW01	TV 12"	500	10	5,000
DW09	선풍기	200	20	4,000
DW20	라디오	100	30	3,000

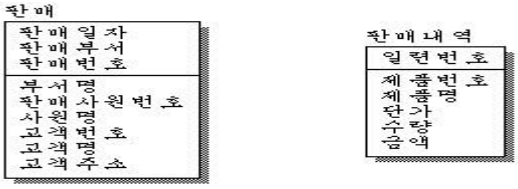
← 서술부분

❑ Tip) 장부나 전표에서 머리부(header)와 서술부(detail)는 별도의 엔티티에 저장한다

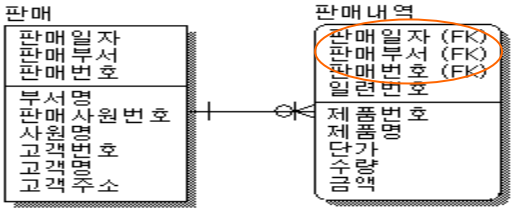
1) 판매전표를 논리적모델링하라



서술부에 있는 속성들이 반복적으로 나타나야 하므로, 반복속성을 피하려면 머리부와 서술부를 별도의 엔티티로 저장한다.

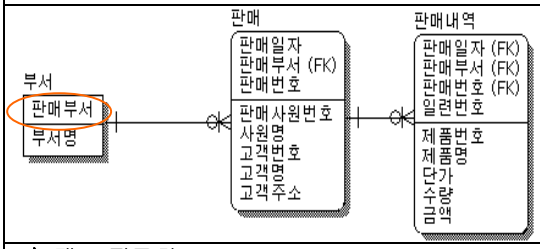


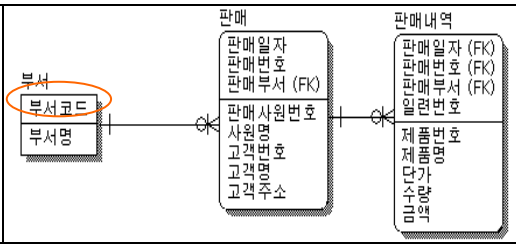
2) 제 1 정규화



서술부에 해당하는 엔티티의 식별자= {머리부 엔티티의 주식별자} + {서술부에 있는 여러 건의 데이터들을 구별하기 위한 속성}

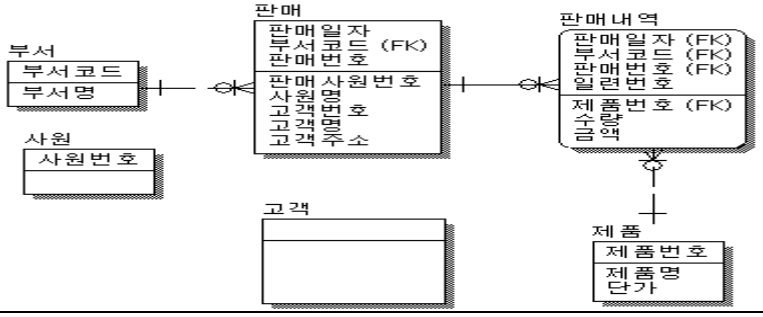
3) 제 2 정규화





4) 제 3 정규화

다음을 수정하여 제 3 정규형에 맞게 완성하시오



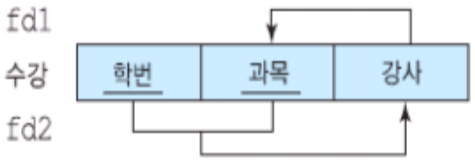
*** BCNF : 제 3 정규형을 만족하고, 모든 결정자가 후보키이어야 한다.**

수강 릴레이션에서 각 학생은 여러 과목을 수강할 수 있고, 각 강사는 한 과목만 가르친다. 이 릴레이션의 기본키는 (학번, 과목)이다. 키가 아닌 강사 애트리뷰트가 기본키에 완전함수종속하므로 제 2 정규형을 만족하고, 키가 아닌 강사 애트리뷰트가 기본키에 직접 종속하므로 제 3 정규형도 만족한다.

(학번, 과목) → 강사
강사 → 과목

수강

학번	과목	강사
11002	데이터베이스	이영준
11002	운영체제	고성현
24036	자료구조	엄영지
24036	데이터베이스	조민형
11048	데이터베이스	이영준



갱신이상 발생) 수강 릴레이션에서 키가 아닌 애트리뷰트가 다른 애트리뷰트 결정

수강 릴레이션의 후보키 : (학번, 과목)과 (학번, 강사)

즉, 수강 릴레이션은 강사 애트리뷰트가 후보키가 아닌데도 과목 애트리뷰트를 결정하므로 BCNF 가 아니다.

수강

학번	과목	강사
11002	데이터베이스	이영준
11002	운영 체제	고성현
24036	자료 구조	엄영지
24036	데이터베이스	조민형
11048	데이터베이스	이영준

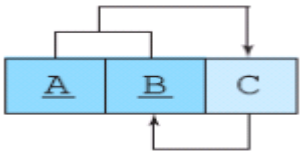
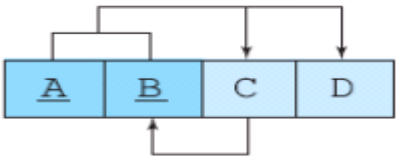
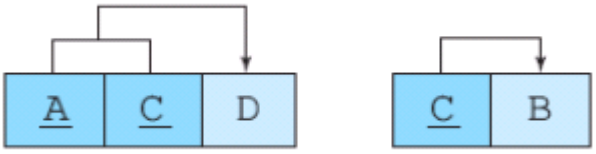
수강1

학번	강사
11002	이영준
11002	고성현
24036	엄영지
24036	조민형
11048	이영준

수강2

강사	과목
이영준	데이터베이스
고성현	운영 체제
엄영지	자료 구조
조민형	데이터베이스

* 제 3 정규형의 애트리뷰트가 a) 3 개인 경우, b) 4 개인 경우의 BCNF 정규화 방법

제 3 정규형은 만족하나 BCNF 아님	제 3 정규형이면서 BCNF 만족하는 릴레이션
 (a) 애트리뷰트가 세 개	(a) 애트리뷰트가 세 개
 (b) 애트리뷰트가 네 개	 (b) 애트리뷰트가 네 개