

## 제 3 장 정규화

3.1 함수 종속

3.2 기본 정규형

3.3 고급 정규형

# 정규화

## ■ 이상(anomaly) 현상

- ◆ 데이터베이스의 설계가 잘못되어 데이터베이스에 데이터를 삽입하거나 삭제 또는 갱신할 때의 문제
- ◆ 삽입 이상
  - 파일에 레코드를 추가할 때, 불필요한 항목의 값이 들어오는 경우, 두 개 이상의 릴레이션의 조인시 속성으로 결과가 나타남
- ◆ 삭제 이상
  - 키값의 삭제시 의도와는 달리 다른 데이터가 삭제될 때 발생
- ◆ 갱신 이상
  - 하나의 파일이 불필요한 중복을 포함하면서 수정될 때, 몇 개의 다른 논리적인 항목들이 수정될 필요성이 있을 때 일어남

## ■ 이러한 이상 현상을 제거하는 알고리즘이 정규화 이론

## 함수 종속(1)

- 어떤 릴레이션  $R$ 에서  $X$ 와  $Y$ 를 각각  $R$ 의 애트리뷰트 집합의 부분 집합이라 하자. 애트리뷰트  $X$ 의 값 각각에 대해 항상 애트리뷰트  $Y$ 의 값이 오직 하나만 연관되어 있을 때  $Y$ 는  $X$ 에 함수 종속
- $X \rightarrow Y$ 로 표기
- $X$ 를 결정자(determinant)라 하고  $Y$ 를 종속자(dependent)

## 함수 종속(2)

학생

학번	이름	학년	학과
99023	황현미	4	전산과
99006	김수동	3	전자과
00153	박수근	2	전산과
01013	김수동	1	전자과
98111	홍순길	4	전기과

- 학생 릴레이션을 기호로 표시하면  
학번 → 이름  
학번 → 학년  
학번 → 학과
- 학번 → (이름, 학년, 학과)

## 함수종속(3)

### ■ 함수 종속도

- ◆ 릴레이션 내의 애트리뷰트들 간의 복잡한 함수 종속 관계를 쉽게 이해하기 위해 도식으로 표현

수강

학번	과목코드	성적	학년
98023	C413	A	4
98023	E412	A	4
99006	C123	B	3
01013	C312	A	1
01013	C324	C	1
01013	C413	A	1
98111	C312	A	4
98111	C324	A	4
98111	C413	B	4
98111	E412	C	4
00153	C312	B	2



학번, 과목코드 → 성적  
학번 → 학년

## 함수 종속(4)

### ■ 암스트롱의 공리

- ◆ 재귀 규칙(reflexivity rule) :  $X \supseteq Y$ 이면  $X \rightarrow Y$ 이다.
- ◆ 첨가 규칙(augmentation rule) :  $X \rightarrow Y$ 이면  $XZ \rightarrow YZ$ 이다.
- ◆ 이행 규칙(transitivity rule) :  $X \rightarrow Y$ 이고  $Y \rightarrow Z$ 이면  $X \rightarrow Z$ 이다.
- ◆ 연합 규칙(union rule) :  $X \rightarrow Y$ 이고  $X \rightarrow Z$ 이면  $X \rightarrow YZ$ 이다.
- ◆ 분해 규칙(decomposition rule) :  $X \rightarrow YZ$ 이면  $X \rightarrow Y$ 이고  $X \rightarrow Z$ 이다.
- ◆ 가이행 규칙(pseudotransitivity rule) :  $X \rightarrow Y$ 이고  $WY \rightarrow Z$ 이면  $WX \rightarrow Z$ 이다.

# 정규형

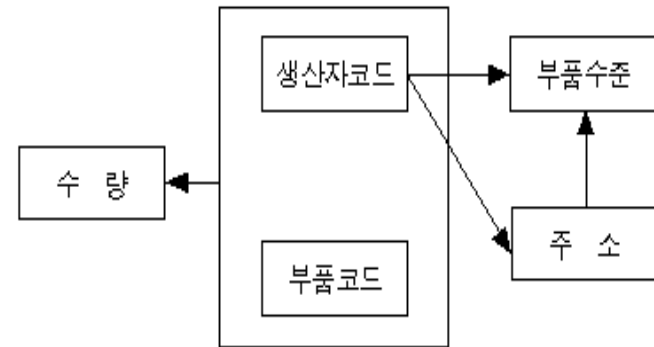
- 서로 관련이 없는 내용들은 별개의 릴레이션으로 표현해야 함
- 이렇게 표현된 릴레이션이 어떤 특정 제약 조건을 만족할 때 그 제약 조건을 요건으로 하는 정규형에 속한다고 말함
- 제 1 정규형에서부터 제 5 정규형까지 있음
- 이 정규형들은 차수가 높아질수록 만족시켜야 할 제약 조건이 가중
- 기본 정규형
  - ◆ 제 1 정규형부터 보이스-코드 정규형

# 제1정규형(1NF: First Normal Form)(1)

- 어떤 릴레이션 R에 속한 모든 도메인이 원자 값(atomic value)으로만 이루어져 있다면 제 1 정규형(1NF)에 속한다.

재고정보

생산자코드	부품수준	주소	부품코드	수량
S1	20	대구	P1	300
S1	20	대구	P2	200
S1	20	대구	P3	400
S1	20	대구	P4	200
S1	20	대구	P5	100
S1	20	대구	P6	100
S2	10	부산	P1	300
S2	10	부산	P2	400
S3	10	부산	P2	200
S4	20	대구	P2	200
S4	20	대구	P4	300
S4	20	대구	P5	400



1. 삽입이상
2. 삭제이상
3. 갱신이상



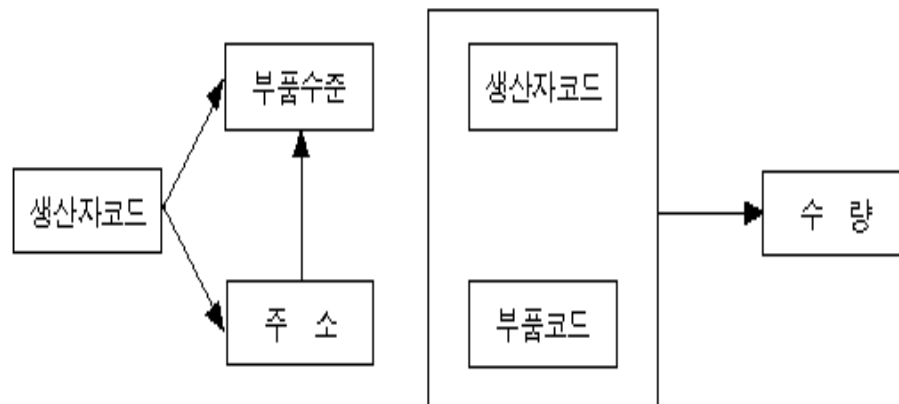
# 제1정규형(1NF: First Normal Form)(2)

생산자

생산자코드	부품수준	주소
S1	20	대구
S2	10	부산
S3	10	부산
S4	20	대구
S5	30	제주

재고

생산자코드	부품코드	수량
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400



생산자(생산자코드, 부품수준, 주소)

주 키 : {생산자코드}

재고(생산자코드, 부품코드, 수량)

주 키 : {생산자코드, 부품코드}

외래 키 : {생산자코드} 참조 : 생산자

## 제2정규형(2NF : Second Normal Form)(1)

- 어떤 릴레이션 R이 1NF이고 주 키에 속하지 않은 애트리뷰트 모두가 주 키에 완전 함수 종속이면, 제 2 정규형(2NF)에 속한다.
- 종속성 보존이란 분해되기 전의 릴레이션에 가지고 있던 모든 함수 종속이 분해된 릴레이션들의 함수 종속으로부터 유도되어야 한다
- 재고정보 릴레이션에 가지고 있는 함수 종속  
    {생산자코드, 부품코드} → 수량  
    생산자코드 → 주소  
    주소 → 부품수준  
    생산자코드 → 부품수준
- 생산자 릴레이션에 가지고 있는 함수 종속  
    생산자코드 → 주소      1) 삽입이상  
    주소 → 부품수준      2) 삭제이상  
    생산자코드 → 부품수준      3) 갱신이상

## 제2정규형(2NF : Second Normal Form)(2)

- ‘생산자’ 릴레이션은 추출연산을 통해서 ‘생산자 주소’와 ‘지역부품정보’ 릴레이션으로 나눌 수 있다.

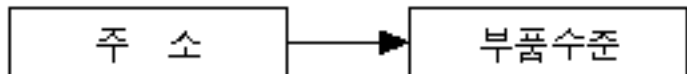
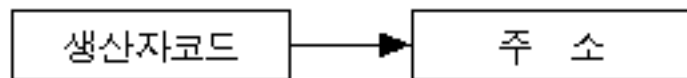
생산자주소(생산자코드, 주소)

주 키 : {생산자코드}

외래 키 : {주소} 참조 : 지역부품정보

지역부품정보(주소, 부품수준)

주 키 : {주소}



생산자주소

생산자코드	주소
S1	대구
S2	부산
S3	부산
S4	대구
S5	제주

지역부품정보

주소	부품수준
제주	30
대구	20
부산	10

## 제3정규형(3NF : Third Normal Form)(1)

- 어떤 릴레이션 R이 2NF이고 주 키에 속하지 않은 모든 애트리뷰트들이 주 키에 이행적 함수 종속이 아닐 때 제 3 정규형(3NF)에 속한다.

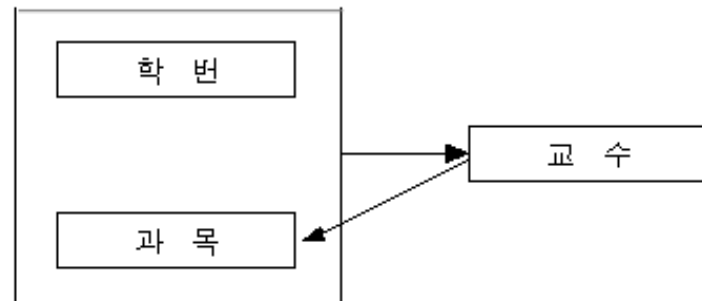
## 제3정규형(3NF : Third Normal Form)(2)

### ■ 조건

- ◆ 각 과목에 대해 한 학생은 한 교수의 강의만 수강한다.
- ◆ 교수는 한 과목만 강의한다.
- ◆ 어떤 과목은 여러 교수가 강의할 수 있다.
- ◆ (단, ‘수강현황’ 릴레이션의 교수명 애트리뷰트의 값에 동명이인은 없는 것으로 가정)

수강현황

학번	과목	교수
100	데이터통신	정기동
100	객체지향언어	이한필
150	객체지향언어	이한필
175	C++	조항구
175	데이터통신	황선영
200	C++	조항구
200	객체지향언어	김순동



- 1) 삽입 이상
- 2) 삭제 이상
- 3) 갱신 이상

## 제3정규형(3NF : Third Normal Form)(3)

- 학생과목(학번, 과목)  
주키(학번, 과목)
- 교수과목(교수, 과목)  
주키(교수)  
외래키(과목) 참조 : 학생과목
- 학생교수(학번, 교수)  
주키(학번, 교수)  
외래키(교수) 참조: 교수과목
- 교수과목(교수, 과목)  
주키(교수)

# 보이스/코드정규형(BCNF)(1)

- 릴레이션  $R$ 의 모든 결정자가 후보 키이면 릴레이션  $R$ 은 보이스/코드 정규형(BCNF)에 속한다.

## 보이스/코드정규형(BCNF)(2)

- 후보키(생산자코드, 부품코드)와 (생산자성명, 부품코드)

제고현황

생산자코드	생산자 성명	부품코드	수량
S1	홍길동	P1	300
S1	홍길동	P2	200
S1	홍길동	P3	400
S1	홍길동	P4	200
S1	홍길동	P5	100
S1	홍길동	P6	100
S2	박수근	P3	500
S3	김신영	P2	400
S3	김신영	P4	300
S3	김신영	P6	100

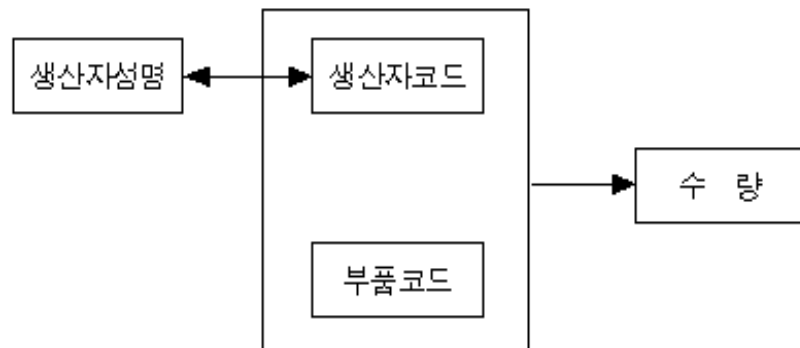
(단, 생산자성명은  
동명이인이 없다고 가정)

생산자코드 또는 생산자성명을 갱신하고자 할 때  
데이터의 중복으로 인한 갱신 이상(update anomaly)발생



## 보이스/코드정규형(BCNF)(3)

### ■ 재고현황 릴레이션의 함수 종속도



- 생산자코드성명(생산자코드, 생산자성명)

주 키 : {생산자코드} 또는 {생산자성명}

외래 키 : {생산자성명} 참조 : 생산자부품(주 키가 {생산자코드}일 때)

또는

{생산자코드} 참조 : 생산자부품(주 키가 {생산자성명}일 때)

- 생산자부품(생산자코드, 부품코드, 수량) 또는 생산자부품(생산자성명, 부품코드, 수량)

주 키 : {생산자코드, 부품코드} 또는 {생산자성명, 부품코드}

## 제 4 정규형(4NF)(1)

과목	교수	교재
포트란	{ 정기동 황수연 }	{ 전산개론 포트란77 }
파스칼	정기동	{ 전산개론 자료구조론 파스칼프로그램 }

(비정규화)

과목교수교재

과목	교수	교재
포트란	정기동	전산개론
포트란	정기동	포트란77
포트란	황수연	전산개론
포트란	황수연	포트란77
파스칼	정기동	전산개론
파스칼	정기동	자료구조론
파스칼	정기동	파스칼 프로그램

(정규화)

과목교재

과목	교재
포트란	전산개론
포트란	포트란77
파스칼	전산개론
파스칼	자료구조론
파스칼	파스칼프로그램

과목교수

과목	교수
포트란	정기동
포트란	황수연
파스칼	정기동

만일 이 릴레이션에 두 개의 튜플  $\langle c1, p1, t1 \rangle$ 과  $\langle c1, p2, t2 \rangle$ 가 포함되어 있다면 또 다른 두 개의 튜플  $\langle c1, p1, t2 \rangle$ 과  $\langle c1, p2, t1 \rangle$ 도 포함되어야 한다.

## 제 4 정규형(4NF)(2)

### ■ 다중값 종속(MVD)

- ◆ A, B, C 세 개의 애트리뷰트(부분 집합)를 가진 릴레이션 R에서 애트리뷰트 쌍 (A, C)-값에 대응하는 B-값의 집합이 A-값에만 종속되고 C-값에는 독립이면 B는 A에 다중값 종속된다고 하고  $A \twoheadrightarrow B$  로 표기
- ◆ 릴레이션 R(A,B,C)에서 MVD  $A \twoheadrightarrow B$ 가 성립하면  $A \twoheadrightarrow C$ 도 동시에 성립
- ◆  $A \twoheadrightarrow B|C$

## 제 4 정규형(4NF)(3)

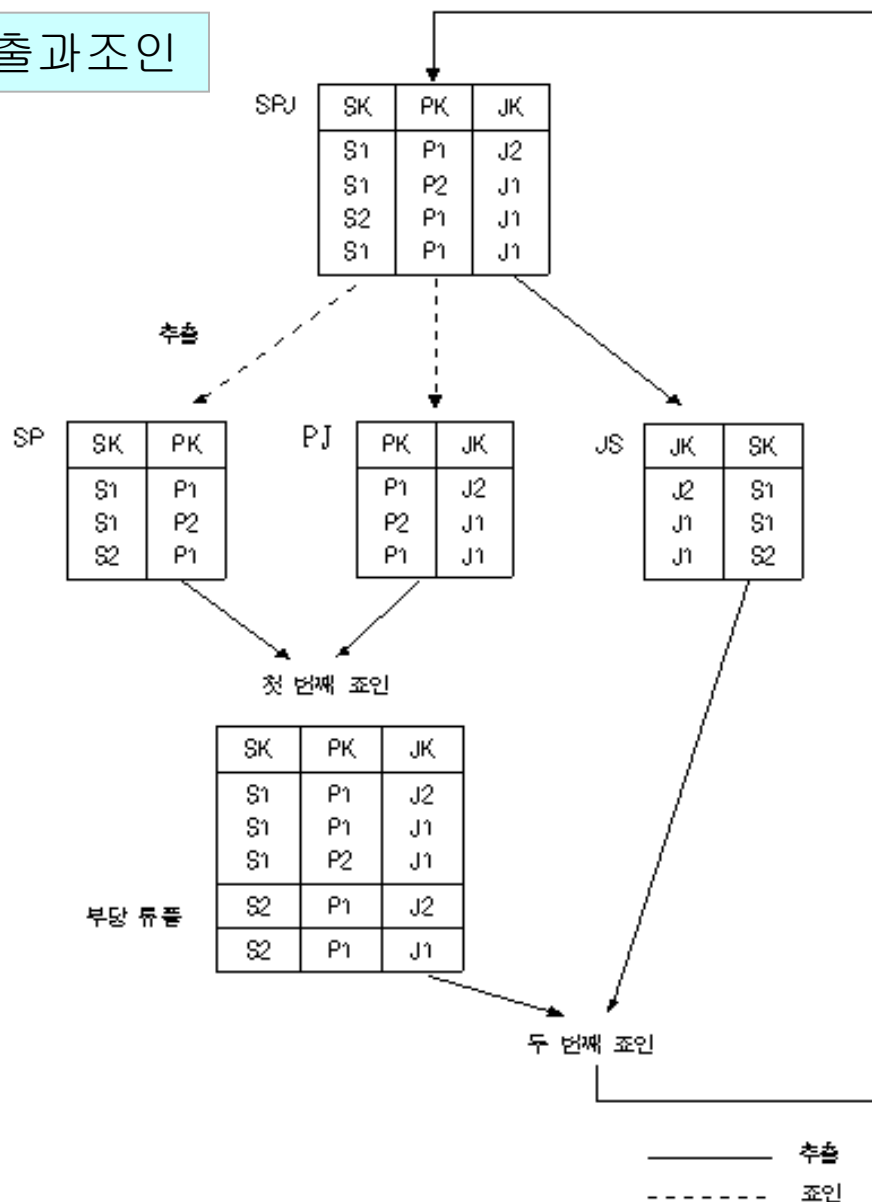
- $X \twoheadrightarrow Y$ 가 당연한 MVD가 되는 조건
  - ◆  $X \supseteq Y$  이거나
  - ◆  $X \cup Y$ 가 곧 릴레이션 R이 될 때이다.
- 릴레이션 R에 비당연 MVD  $A \twoheadrightarrow B$ 가 존재할 때 R의 모든 애트리뷰트들도 또한 A에 함수 종속(즉, R의 모든 애트리뷰트 X에 대해  $A \rightarrow X$ 이고 A가 후보 키)이면 릴레이션 R은 4NF에 속한다.
  - ◆ 어떤 릴레이션이 4NF이라면 MVD가 없거나 MVD  $A \twoheadrightarrow B|C$ 가 있을 경우 A에 대응되는 B와 C의 값은 각각 하나씩이고 이때 A는 후보 키라는 것을 의미

# 제5정규형(5NF : Fifth Normal Form)(1)

- N개 이상의 릴레이션으로 분해하여야만 정보 무손실 분해가 되는 릴레이션을 N-분해 릴레이션이라 한다.

# 제5정규형(5NF : Fifth Normal Form)(2)

추출과조인



## 제5정규형(5NF : Fifth Normal Form)(3)

### ■ 제약 조건

- ◆ 만일  $\langle S1, P1, J2 \rangle, \langle S2, P1, J1 \rangle, \langle S1, P2, J1 \rangle$ 이 SPJ에 포함되어 있으며,  $\langle S1, P1, J1 \rangle$ 도 또한 SPJ에 포함되어 있어야 한다.
- ◆ 릴레이션 SPJ가 세 개의 추출 SP, PJ, JS를 조인한 결과와 같다는 것은 만일 SP가  $\langle S1, P1 \rangle$ , PJ가  $\langle P1, J1 \rangle$ , JS가  $\langle J1, S1 \rangle$ 을 포함한다면 SPJ는  $\langle S1, P1, J1 \rangle$ 을 포함해야 한다는 것과 동등
- ◆ 임의의 J2, S2, P2에 대해  $\langle S1, P1, J2 \rangle, \langle S2, P1, J1 \rangle, \langle S1, P2, J1 \rangle$ 이 SPJ에 있다면  $\langle S1, P1 \rangle, \langle P1, J1 \rangle, \langle J1, S1 \rangle$ 도 추출 SP, PJ, JS에 각각 있어야 함

## 제5정규형(5NF : Fifth Normal Form)(4)

- 조인 종속(JD) : 어떤 릴레이션 R의 애트리뷰트의 부분 집합을 A, B, ..., Z라 하자. 이때 만일 릴레이션 R이 그의 추출 A, B, ..., Z를 조인한 것과 똑같다면 R은 조인 종속 \*(A, B, ..., Z)를 만족시킨다고 한다.
  - ◆ SPJ 릴레이션은 조인 종속 \*(SP, PJ, JS)를 만족하고 있고 또 3-분해 릴레이션임



## 제5정규형(5NF : Fifth Normal Form)(5)

- 릴레이션  $R(A,B,C)$ 가 조인 종속  $*(AB, AC)$ 를 만족하기만 하면 릴레이션  $R$ 은 두 개의 MVD  $A \twoheadrightarrow B|C$ 를 만족한다.
- 릴레이션  $R$ 에 존재하는 모든 조인 종속(JD)이 릴레이션  $R$ 의 후보 키를 통해서만 성립된다면 릴레이션  $R$ 은 제 5 정규형(5NF) 또는 PJ/NF(projection-join normal form)에 속한다.
- 릴레이션 학생(학번, 이름, 학과, 학년)
  - ◆  $*((\text{학번}, \text{이름}, \text{학과}), (\text{학번}, \text{학년}))$
  - ◆  $*((\text{학번}, \text{이름}), (\text{학번}, \text{학년}), (\text{이름}, \text{학과}))$