

# 데이터베이스 정규화

SSAFY 9기 구미 6반

김도연

이 테이블에서 문제점을 찾아봅시다..

상품 번호	상품명	판매자	판매자 전화번호
1	마우스	김철수	111-1111-1111
2	건전지	황동길	222-2222-2222
3	키보드	황동길	222-2222-2222
4	휴대폰	이재용	333-3333-3333
5	연필	김명철	444-4444-4444
6	볼펜	김명철	444-4444-4444
7	빗자루	최민철	555-5555-5555

# 앞 예시의 문제점은...

## ◆ 삽입 이상

- 새로운 판매자를 등록하려는데 등록할 수 없다
- 새로운 판매자는 판매하는 물건이 없기때문에 등록이 불가능

## ◆ 삭제 이상

- 마우스를 더이상 판매하지 않아서 삭제하는 경우
- 김철수 판매자의 정보까지 함께 사라짐, 데이터 손실

## ◆ 갱신 이상

- 길명철 판매자의 이름을 변경하는 경우
- 하나의 값만 변경하려는데 길명철이 판매하고있는 5번 6번 상품을 모두 수정해야 한다
- 일부만 수정하면 데이터가 불일치하는 문제가 생김

➤ 문제 발생의 원인: 상품 정보와 판매자 정보를 억지로 한 테이블로 만들어서!

# 데이터베이스 정규화

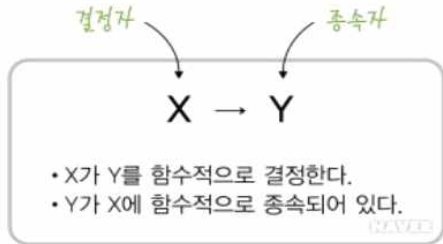
- ◆ 문제가 발생할만한 테이블을 분할하는 과정을 의미함
- ◆ 관계형 데이터베이스 설계에서 중복을 최소화 하여 데이터를 구조화 하는 프로세스
- ◆ 서로 관련이 있는 속성만으로 테이블을 구성하도록 테이블을 분할하여 이상 현상을 줄이는 과정을 의미함

# 속성간의 관련이란?

◆ 속성간 관련이 있다=함수적 종속관계가 있다

◆ 함수적 종속관계란

- X 속성값에 대한 Y 속성값이 하나로 결정되면 X가 Y를 함수적으로 결정한다고 함
- 혹은 Y가 X에 함수적으로 종속되어있다 라고함
- 기호로는  $X \rightarrow Y$  화살표로 표기한다



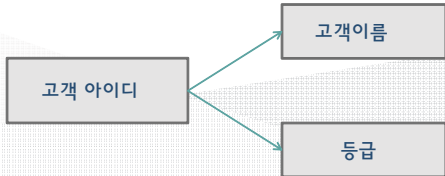
# 함수 종속 관계 예시

◆ 고객 아이디 속성값에 대응되는 고객 이름 속성과 등급 속성의 값이 단 하나이다

- 고객 아이디가 고객 이름과 등급을 결정
- 고객아이디 → (고객이름, 등급)

◆ 함수 종속중에 완전함수 종속, 부분함수 종속이 있고 이 두가지 개념을 알아둘 필요가 있다

고객 아이디	고객 이름	등급
apple	정수만	gold
banana	김명수	vip
carrot	고창석	gold
orange	김도연	silver

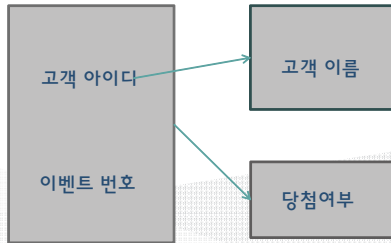


<함수 종속 다이어그램>

# 완전 함수 종속

- ◆ 테이블에서 속성집합 Y가 속성집합 X에 함수적으로 종속되어 있지만 속성집합 X의 전체가 아닌 일부분에는 종속되지 않음을 의미함
- 아래 예시에서 고객 이름은 고객 아이디에 완전 함수 종속이다  
(고객아이디, 이벤트번호)에는 당첨여부가 완전함수 종속이다

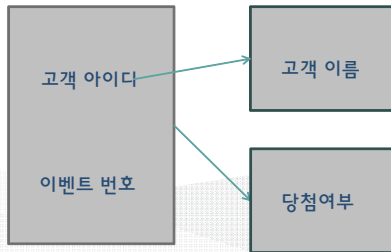
고객 아이디	이벤트 번호	당첨 여부	고객 이름
apple	e001	y	정수만
apple	e005	n	정수만
apple	e010	y	정수만
banana	e002	n	김선우
banana	e005	y	김선우
carrot	e003	y	고창석
carrot	e007	y	고창석
orange	e004	n	김도연



# 부분 함수 종속

- ◆ 속성집합 Y가 속성집합 X의 전체가 아닌 일부분에도 함수적으로 종속됨을 의미
- ◆ (고객아이디,이벤트번호)에 고객이름이 부분함수 종속되어있다

고객 아이디	이벤트 번호	당첨 여부	고객 이름
apple	e001	y	정수만
apple	e005	n	정수만
apple	e010	y	정수만
banana	e002	n	김선우
banana	e005	y	김선우
carrot	e003	y	고창석
carrot	e007	y	고창석
orange	e004	n	김도연





# 정규화의 장점

- ◆ 이상현상 방지
- ◆ 데이터 중복을 최소화 하여 저장공간 줄임
- ◆ 자료의 불일치 최소화
- ◆ 구조의 안정화

## 정규화의 단점

- ◆ 테이블을 나누기때문에 Join 연산 증가-> 연산 시간이 증가됨
- ◆ 이러한 단점들 때문에 쪼개놓은 테이블을 합치는 경우 존재: 역 정규화

# 정규형 단계

- ◆ 제 1 정규형 (1NF)
- ◆ 제 2 정규형 (2NF)
- ◆ 제 3 정규형 (3NF)
- ◆ 보이스-코드 정규형(BCNF)
- ◆ 제 5 정규형 (5NF)
  - 실제 데이터베이스 설계시 5 정규형까지 분해하는것은 바람직하지 않은 경우가 많음
  - 따라서 일반적으로는 보이스-코드 정규형까지만 분리함.
  - 보이스-코드 까지만 해도 이상현상을 해결하고 데이터 중복을 줄일 수 있음.

# 제 1 정규형

- ◆ 모든 속성들이 원자값 만을 갖는 경우 제 1 정규형을 만족한다.
- ◆ 릴레이션(=테이블) 이 모두 단일 속성으로만 구성
  - 단일속성?: 속성값이 배열, 혹은 구조체 처럼 여러 값이 한번에 들어가있지 않은 상태

학번	과목	교수	학점
100	파이선, 자바	김교수, 이교수	A,B
102	자바	이교수	A
103	파이선, C	김교수, 최교수	B, A

학번	과목	교수	학점
100	파이선	김교수	A
100	자바	이교수	B
102	자바	이교수	A
103	파이선	김교수	B
103	C	최교수	A

# 제 1 정규형 문제점

## ◆ 갱신 이상

- 파이선 교수가 변경되면 모든 파이선 과목 교수를 변경해야 함, 갱신 이상 발생

## ◆ 삽입 이상

- 자바스크립트라는 과목을 신설하는데 학생이 없어서 추가 불가능

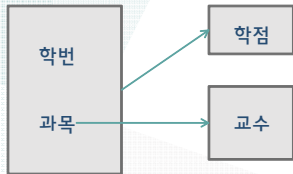
## ◆ 삭제 이상

- 103번 학생이 자퇴하면 C 언어 과목이 사라진다

학번	과목	교수	학점
100	파이선	김교수	A
100	자바	이교수	B
102	자바	이교수	A
103	파이선	김교수	B
103	C	최교수	A

## 제 2 정규형

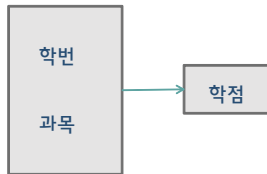
◆ 제 1 정규형을 만족하며 주 키에 속하지 않는 모든 속성이 주 키에 완전 종속인 테이블



학번	과목	교수	학점
100	파이선	김교수	A
100	자바	이교수	B
102	자바	이교수	A
103	파이선	김교수	B
103	C	최교수	A

학번	과목	학점
100	파이선	A
100	자바	B
102	자바	A
103	파이선	B
103	C	A

과목	교수
파이선	김교수
자바	이교수
C	최교수



## 제 2 정규형 문제

### ◆ 이행 종속성이 존재함

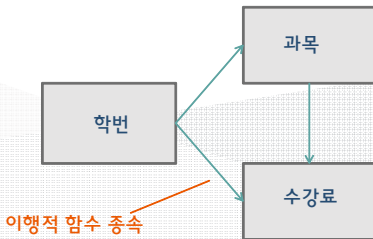
- 수강료가 학번에게 이행적으로 함수 종속됨

### ◆ 갱신이상: 파이썬을 수정하면 모든 파이썬을 수정해야 함

### ◆ 삭제이상: 200번 학생이 자퇴하면 C언어 과목 삭제됨

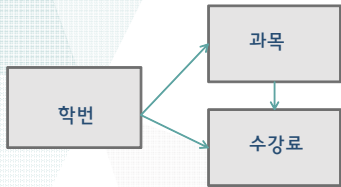
### ◆ 삽입이상: 자바스크립트 과목 신설, 하지만 학생이 없어서 추가 불가능

학번	과목	수강료
100	파이썬	20000
150	자바	15000
200	C	10000
250	파이썬	20000
300	파이썬	20000



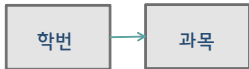
# 제 3 정규형

◆ 제 2 정규형을 만족하면서 모든 속성이 주 키에 이행적 함수 종속이 아닌 테이블

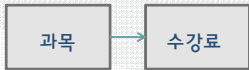


학번	과목	수강료
100	파이썬	20000
150	자바	15000
200	C	10000
250	파이썬	20000
300	파이썬	20000

학번	과목
100	파이썬
150	자바
200	C
150	파이썬
300	파이썬



과목	수강료
파이썬	20000
자바	15000
C	10000



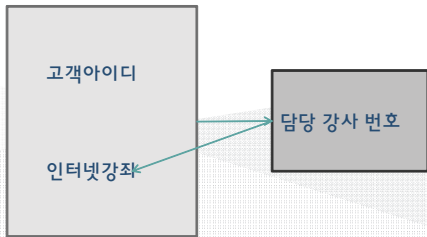


# 제 3 정규형 문제점

## ◆ 주 키에 포함되지 않는 속성과 주 키의 일부 속성이 종속 관계 있을때

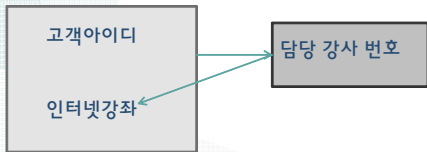
- 담당 강사번호가 주 키에 포함되지 않는데 인터넷 강좌의 속성을 결정함
- 삽입이상: 새로운 강사와 강의 추가 불가능-> 듣는 학생이 없기때문에 고객아이디 null 인데 주 키이므로 불가
- 갱신이상: p004 강사의 강좌를 중급토익으로 변경할시 모든 기초토익을 수정해야함
- 삭제이상: banana가 강의를 취소하면 p002 강사의 정보도 함께 손실

고객 아이디	인터넷 강좌	담당 강사 번호
apple	영어회화	p001
banana	기초토익	p002
carrot	영어회화	p001
carrot	기초토익	p004
orange	영어회화	p003
orange	기초토익	p004



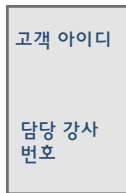
# 보이스-코드 정규형(BCNF)

- ◆ 모든 함수 종속성의 결정자(화살표의 왼쪽 속성, 다른 속성을 결정하는 속성)는 후보키가 되어야 함

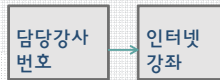


고객 아이디	인터넷 강좌	담당 강사 번호
apple	영어회화	p001
banana	기초토익	p002
carrot	영어회화	p001
carrot	기초토익	p004
orange	영어회화	p003
orange	기초토익	p004

고객 아이디	담당 강사 번호
apple	p001
banana	p002
carrot	p001
carrot	p004
orange	p003
orange	p004



담당강사번호	인터넷강좌
p001	영어회화
p002	기초토익
p003	영어회화
p004	기초토익



# 참고자료

- ◆ 영남대 컴퓨터공학과 데이터베이스 수업자료
- ◆ 네이버 지식백과 데이터베이스 개론  
(<https://terms.naver.com/list.naver?cid=58430&categoryId=58430&so=st4.asc>)
- ◆ 정규화 장점, 단점
  - (<https://worldt.tistory.com/99>)
  - (<https://pythontoomuchinformation.tistory.com/529>)
  - (<https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=star005&logNo=40113598512>)