



# 정규화



컴퓨터과학과 정재화

## 학습목차

- ① 좋은 릴레이션과 나쁜 릴레이션
- ② 함수적 종속성의 이해
- ③ 정규화



01

# 좋은 릴레이션과 나쁜 릴레이션

- 잘못된 데이터베이스 모델링
- 갱신 이상
- 좋은 릴레이션의 개념

## ☆ 나쁜 릴레이션의 예

.....

고객번호	고객명	전화번호	등급	할인율
C1	유관순	010-9999-0001	GOLD	8%
C2	손병희	010-9999-0002	VIP	10%
C3	지청천	010-9999-0003	VIP	10%
C4	안창호	010-9999-0004	GOLD	5%
C5	안중근	010-9999-0005	VIP	10%



# 잘못된 데이터베이스 모델링

---

## ▷ 데이터의 중복

- + 일관성 유지의 어려움
- + 저장 공간 낭비

## ▷ 갱신 이상

- + 삽입 이상: 레코드 추가 시 불필요한 컬럼값 없이는 추가하지 못하는 경우
- + 삭제 이상: 삭제 시 의도하지 않았던 다른 데이터가 삭제되는 경우
- + 수정 이상: 일부 레코드 수정 시 데이터의 일관성이 훼손되는 경우

## 갱신 이상 – 삽입 이상

고객번호	고객명	전화번호	등급	할인율
C1	유관순	010-9999-0001	GOLD	5%
C2	손병희	010-9999-0002	VIP	10%
C3	지청천	010-9999-0003	VIP	10%
C4	안창호	010-9999-0004	GOLD	5%
C5	안중근	010-9999-0005	VIP	10%

고객번호	고객명	전화번호
????	???	???????

불필요한 정보

등급	할인율
NEW	3%

추가하고자 하는 정보

## 갱신 이상 - 삭제 이상

고객번호	고객명	전화번호	등급	할인율
C1	유관순	010-9999-0001	GOLD	5%
C2	손병희	010-9999-0002	VIP	10%
C3	지청천	010-9999-0003	VIP	10%
C4	안창호	010-9999-0004	GOLD	5%
C5	안중근	010-9999-0005	VIP	10%
C6	윤봉길	010-9999-0006	NEW	3%

삭제되면 안 되는 정보

삭제하고자 하는 정보

## 갱신 이상 - 수정 이상

고객번호	고객명	전화번호	등급	할인율	
C1	유관순	010-9999-0001	GOLD	5%	
C2	손병희	010-9999-0002	VIP	10%	비일관성 → 8%
C3	지청천	010-9999-0003	VIP	10%	
C4	안창호	010-9999-0004	GOLD	5%	
C5	안중근	010-9999-0005	VIP	10%	→ 10% 수정 실패
C6	윤봉길	010-9999-0006	NEW	3%	



## 좋은 릴레이선의 개념

### ▷ 컴퓨터 프로그래머적 관점에서의 모델

어떻게 데이터를 저장해야 하는가?

### ▷ 릴레이선의 스키마가 얼마나 효율적으로 실세계를 반영하고 있는지 평가

### ▷ 고려사항

- + 릴레이선 내의 컬럼 간의 관계 분석
- + 불필요한 데이터의 종속과 중복 제거
- + 새로운 컬럼들이 데이터베이스에 추가될 때, 기존 컬럼과의 관계 수정을 최소화

## 좋은 릴레이션

.....

고객번호	고객명	전화번호	등급	할인율
C1	유관순	010-9999-0001	GOLD	5%
C2	손병희	010-9999-0002	VIP	10%
C3	지청천	010-9999-0003	VIP	10%
C4	안창호	010-9999-0004	GOLD	5%
C5	안중근	010-9999-0005	VIP	10%
C6	윤봉길	010-9999-0006	NEW	3%

## 좋은 릴레이션

고객번호	고객명	전화번호	등급(FK)	등급(PK)	할인율
C1	유관순	010-9999-0001	GOLD	GOLD	5%
C2	손병희	010-9999-0002	VIP	VIP	10%
C3	지청천	010-9999-0003	VIP	NEW	3%
C4	안창호	010-9999-0004	GOLD		
C5	안중근	010-9999-0005	VIP		
C6	윤봉길	010-9999-0006	NEW		

02

# 함수적 종속성의 이해

- 함수적 종속성의 정의
- 함수적 종속성의 확장
- 커버와 카노니컬 커버

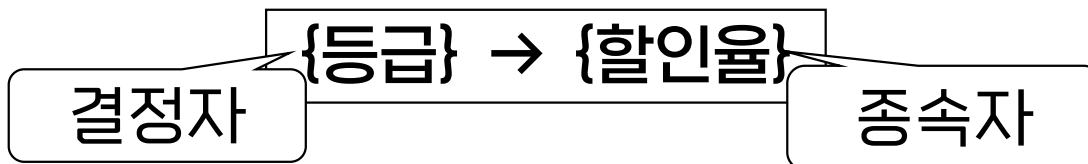
## ☆ 함수적 종속성의 개념

- ◇ FD(Functional Dependency)
- ◇ 릴레이션 인스턴스를 분석하여 속성들 간의 연관관계를 표현한 것
- ◇ 릴레이션의 효율성을 향상시켜 좋은 릴레이션으로 변환하는데 이용되는 중요한 개념
- ◇ 형식적 정의

임의의 릴레이션 스키마  $R$ 의 인스턴스  $r(R)$ 에 포함되는 서로 다른 두 레코드  $r_1, r_2$ 가 컬럼 집합  $X$ 와  $Y$ 에 대해,  $r_1[X] = r_2[X]$ 일 때,  $r_1[Y] = r_2[Y]$  이면 함수적 종속성  $X \rightarrow Y$ 가 성립한다.

# 함수적 종속성의 판별

고객번호	고객명	전화번호	등급	할인율
C1	유관순	010-9999-0001	GOLD	5%
C2	손병희	010-9999-0002	VIP	10%
C3	지청천	010-9999-0003	VIP	10%
C4	안창호	010-9999-0004	GOLD	5%
C5	안중근	010-9999-0005	VIP	10%
C6	윤봉길	010-9999-0006	NEW	3%



## 함수적 종속성의 확장

- ▷ 함수적 종속성은 릴레이션의 효율성 여부에 중요한 판단기준
  - + 그러나 릴레이션의 인스턴스만으로  
내재된 모든 함수적 종속성을 찾아내기 어려움
- ▷ 판별되지 않은 모든 함수적 종속성을 도출하기 위해  
추론 규칙 등을 사용하여 함수적 종속성을 확장
- ▷ 클로저(closure)
  - + 판별된 함수적 종속성 집합으로부터 유추할 수 있는  
모든 함수적 종속성 집합

$$F^+$$

## ☆ 함수적 종속성 추론 규칙

### ▷ 암스트롱 공리(Armstrong's axiom)

1. 재귀성 규칙 :  $X \supseteq Y$ 이면,  $X \rightarrow Y$ 이다
2. 부가성 규칙:  $X \rightarrow Y$ 이면,  $XZ \rightarrow YZ$ 이다
3. 이행성 규칙:  $X \rightarrow Y$ 이고,  $Y \rightarrow Z$ 이면,  $X \rightarrow Z$ 이다.
4. 분해 규칙:  $X \rightarrow YZ$ 이면,  $X \rightarrow Y$ 이다.
5. 합집합 규칙:  $X \rightarrow Y$ 이고,  $X \rightarrow Z$ 이면,  $X \rightarrow YZ$ 이다.
6. 의사 이행성 규칙 :  $X \rightarrow Y$ 이고,  $WY \rightarrow Z$ 이면,  
 $WX \rightarrow Z$ 이다.



## **함수적 종속성의 판별**

고객번호 → 고객명

고객명 → 등급

{고객번호, 고객명} → 할인율

고객번호	고객명	전화번호	등급	할인율
C1	유관순	010-9999-0001	GOLD	5%
C2	손병희	010-9999-0002	VIP	10%
C3	지청천	010-9999-0003	VIP	10%
C4	안창호	010-9999-0004	GOLD	5%
C5	안중근	010-9999-0005	VIP	10%
C6	윤봉길	010-9999-0006	NEW	3%

고객번호 → {고객명, 등급, 할인율}



## 카노니컬 커버

- ▷ 함수적 종속성 추론 규칙으로 확장된 클로저에는 자명한 종속성과 중복된 종속성을 포함
- ▷ 불필요한 함수적 종속성을 제거한 표준형으로 변환 후 정규화를 수행
- ▷ 표준형 조건

- +  $F$ 의 모든 함수적 종속성의 종속자의 속성은 반드시 1개
- +  $F$ 에서  $X \rightarrow A$ 를  $X$ 의 진부분집합  $Y$ 에 대하여  $Y \rightarrow A$ 로 교체했을 때, 그 집합이  $F$ 와 동등한 집합이 불가능
- +  $F$ 에서 어떤 함수적 종속성을 제거했을 때, 그 집합이  $F$ 와 동등한 집합이 불가능

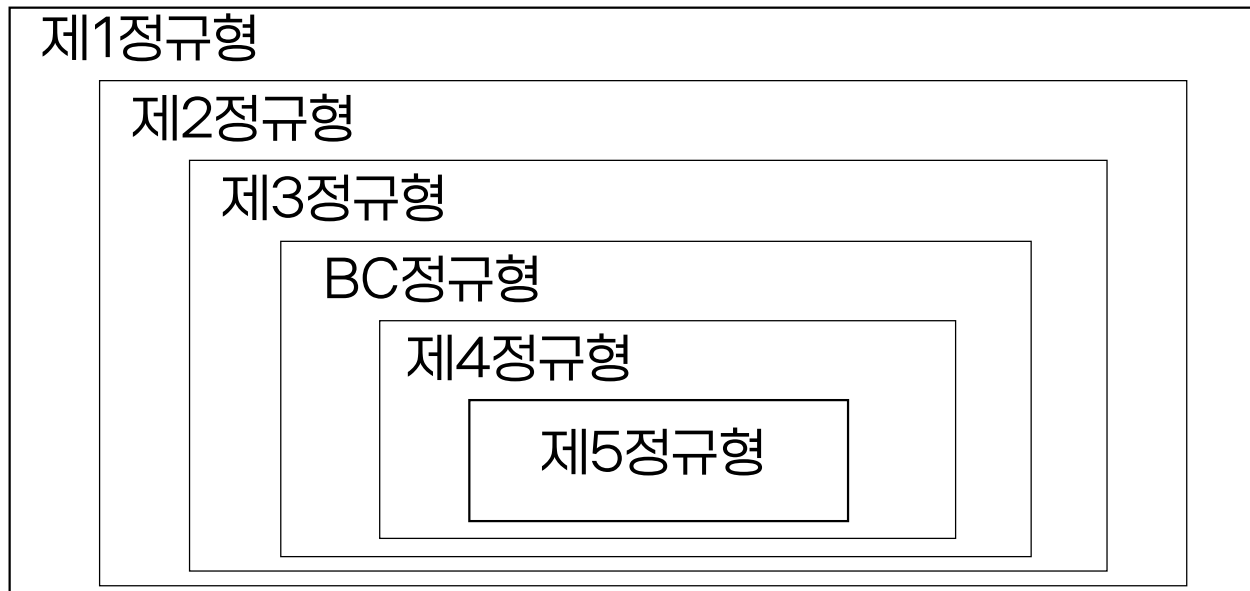
03

# 정규화

- 정규형과 정규화의 개념
- 제2정규형
- 제3정규형
- BC정규형

# 정규형

- ▷ 이상 현상을 최소화하도록 특정 조건을 갖춘 릴레이선의 형식
- ▷ 정규형의 분류



# 정규화의 목적

## ▷ 정규화의 정의

특정 정규형의 조건을 만족하도록  
릴레이션과 속성을 재구성하는 과정

## ▷ 정규화의 기능

- + 데이터베이스 내에 모든 릴레이션을 효율적으로 표현
- + 바람직하지 않은 삽입, 수정, 삭제 등의 이상 발생 방지
- + 보다 간단한 관계 연산에 기초하여 검색 알고리즘을 효과적으로 작성할 수 있도록 지원
- + 새로운 형태의 데이터가 삽입될 때, 릴레이션 재구성의 필요성을 축소

## 제1정규형

---

- ▷ 가장 조건이 단순한 정규형
- ▷ 형식적 정의

릴레이션 스키마에서 정의된  
모든 속성의 도메인이 원자값을 갖는 상태

- ▷ 관계형 모델의 특성을 만족하는 경우, 자동 적용되는 정규형



# 제1정규형을 위배한 릴레이션

도크 릴레이션

도크번호	입항시간	출항시간	목적	도크관리자	담당도선사
D1	09:00, 11:00	10:15, 11:45	선적	김규식	김순애
D1	11:50	12:45	하역	김규식	김구
D2	09:00, 12:00	10:00, 12:45	관광, 주유	한용운	이동휘
D2	13:00	15:00	정비	한용운	윤봉길

원자값을 갖지 않는 속성



## 예제 릴레이선에 대한 부연설명

---

- ◇ 선박이 항구에 정박하기 위해 정밀한 작업이 요구
  - ⊕ 풍향과 풍속, 파도와 안개 상황 등을 고려
  - ⊕ 단순히 수신호뿐만 아니라 소형 배들로 정박하려는 대형 배를 밀어 안전하게 위치시키는 작업을 고려
- ◇ 도선사: 배를 도크에 안전하게 접안시키고  
항로로 인도하는 일을 하는 사람
- ◇ 도크: 선박의 건조, 수리, 계선, 하역 작업 등을 위해  
축조된 설비 및 시설의 총칭



# 제1정규화를 수행한 릴레이션

도크 릴레이션

도크번호	입항시간	출항시간	목적	도크관리자	담당도선사
D1	09:00	10:15	선적	김규식	김순애
D1	11:00	11:45	선적	김규식	김순애
D1	11:50	12:45	하역	김규식	김구
D2	09:00	10:00	관광	한용운	이동휘
D2	12:00	12:45	주유	한용운	이동휘
D2	13:00	15:00	정비	한용운	윤봉길

원자적 상태의 속성값으로만 구성(제1정규형)

## 함수적 종속성 판별

---

### ▷ 함수적 종속성의 형식적 정의

임의의 릴레이션 스키마  $R$ 의 인스턴스  $r(R)$ 에 포함되는 서로 다른 두 레코드  $r_1, r_2$ 가 컬럼 집합  $X$ 와  $Y$ 에 대해,  $r_1[X] = r_2[X]$ 일 때,  $r_1[Y] = r_2[Y]$  이면 함수적 종속성  $X \rightarrow Y$ 가 성립한다.

## 함수적 종속성 판별

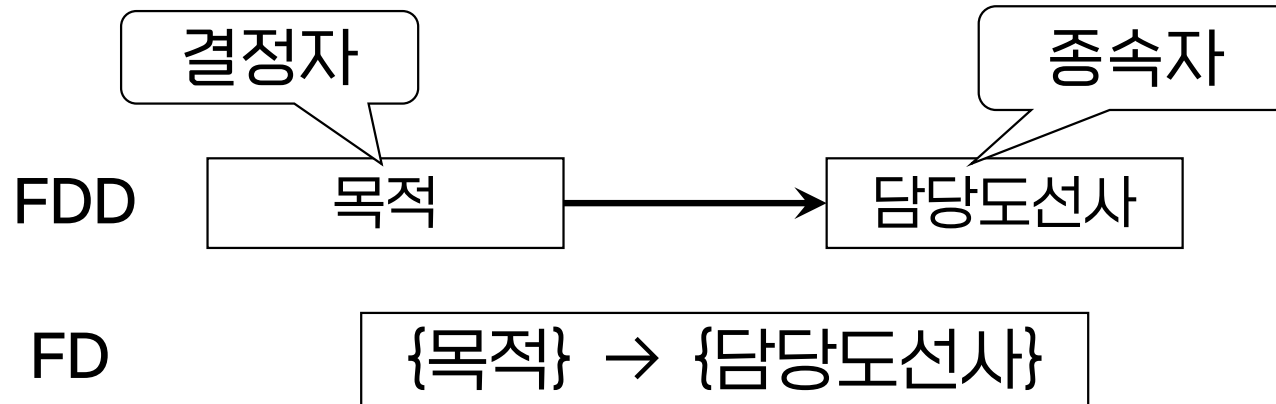


도크번호	입항시간	출항시간	목적	도크관리자	담당도선사
D1	09:00	10:15	선적	김규식	김순애
D1	11:00	11:45	선적	김규식	김순애
D1	11:50	12:45	하역	김규식	김구
D2	09:00	10:00	관광	한용운	이동휘
D2	12:00	12:45	주유	한용운	이동휘
D2	13:00	15:00	정비	한용운	윤봉길

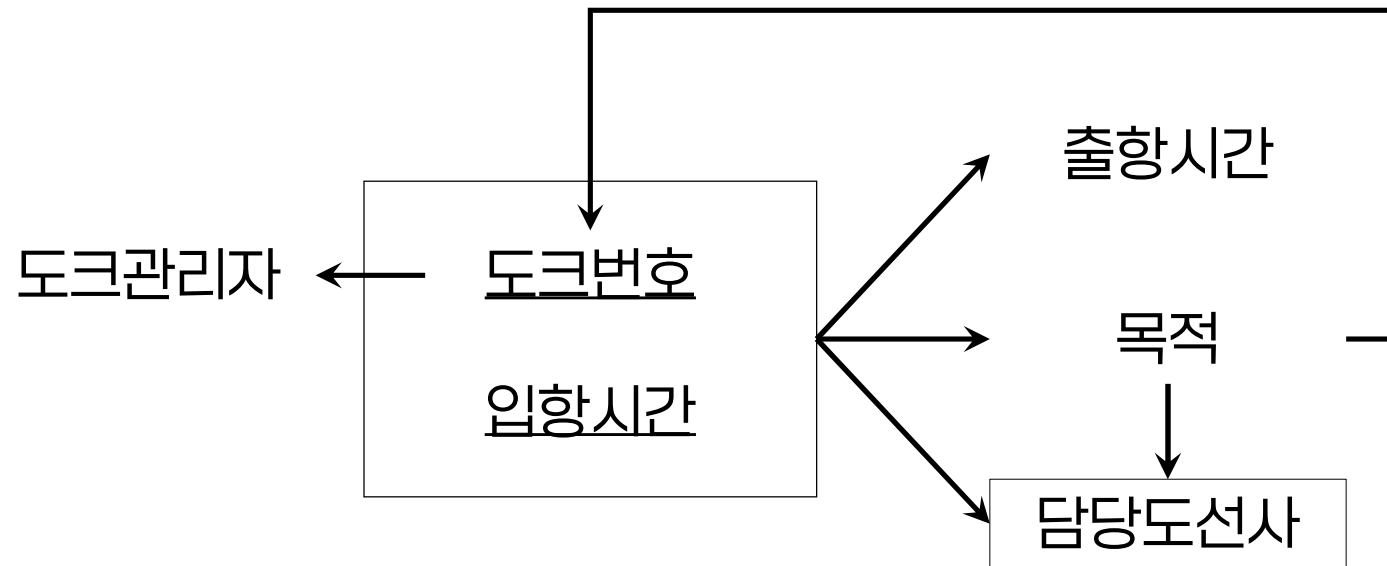
## 함수적 종속성 다이어그램

- ◇ FDD(Functional Dependency Diagram)
- ◇ 릴레이션 내의 속성간의 종속 관계를 직관적이고 이해하기 쉽게 도식화한 표현 방식

- + 직사각형: 속성 또는 속성 집합
- + 화살표: 함수적 종속성



# 도크 릴레이션의 함수적 종속성 다이어그램





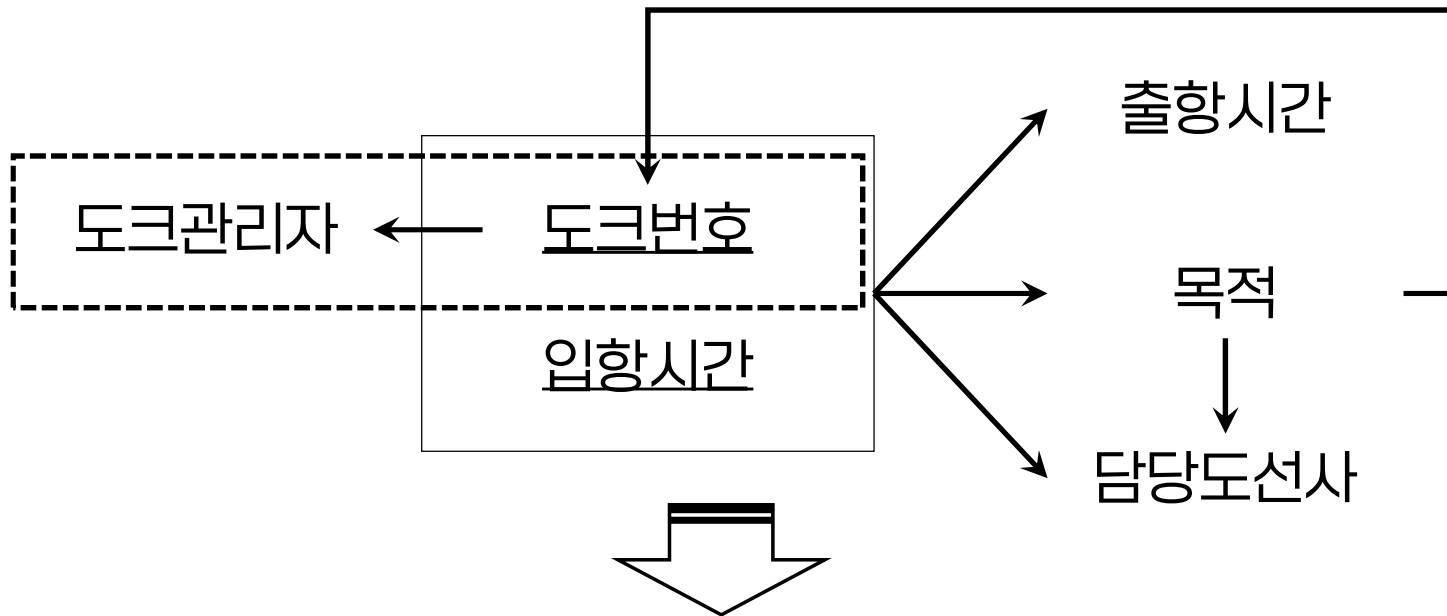
## 제2정규형의 정의

- ▷ 릴레이션이 제1정규형을 만족하고 기본키의 부분집합이 특정 속성을 종속하고 있지 않은 상태
- ▷ 형식적 정의

주어진 릴레이션의 인스턴스가 기본키가 아닌 속성(비주속성)이 기본키(주속성)에 완전히 종속되어 있는 상태

- ▷  $X \rightarrow Y$ 에 대한 함수 종속 관계
  - ⊕ 부분 함수 종속:  $Y$ 가  $X$ 의 전체가 아닌 일부분에 종속
  - ⊕ 완전 함수 종속:  $Y$ 가  $X$ 의 전체에 종속

## 제2정규형의 적용



기본키에 완전 종속되도록 릴레이션을 분해



## 릴레이션의 무손실 분해

### ▷ 무손실 분해(lossless decomposition)

#### + 정의

스키마  $R$ 에 함수적 종속성  $X \rightarrow Y$ 가 존재하고  
 $X \cap Y = \emptyset$  이면,  $R$ 을  $R - Y$ 와  $XY$ 로 분해

### ▷ 도크관리 릴레이션 무손실 분해

+ {도크번호}  $\rightarrow$  {도크관리자}

+ {도크번호}  $\cap$  {도크관리자} =  $\emptyset$

도크관리 - {도크관리자}, {도크번호, 도크관리자}



## 제2 정규화의 적용

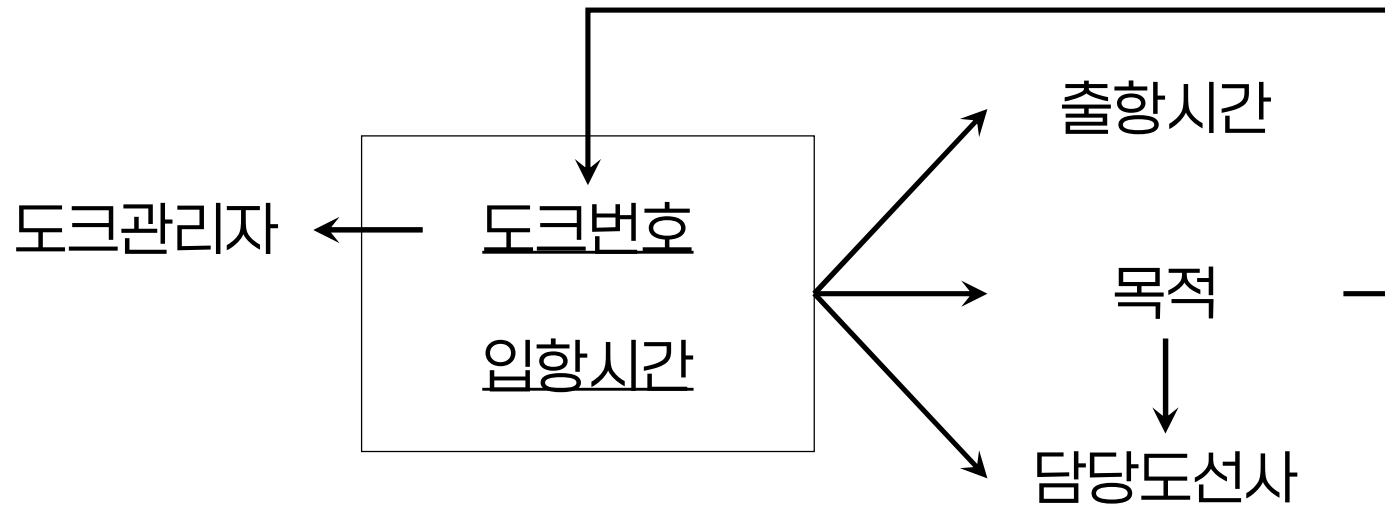
도크관리 릴레이션

도크번호	입항시간	출항시간	목적	담당도선사
D1	09:00	10:15	선적	김순애
D1	11:00	11:45	선적	김순애
D1	11:50	12:45	하역	김구
D2	09:00	10:00	관광	이동휘
D2	12:00	12:45	주유	이동휘
D2	13:00	15:00	정비	윤봉길

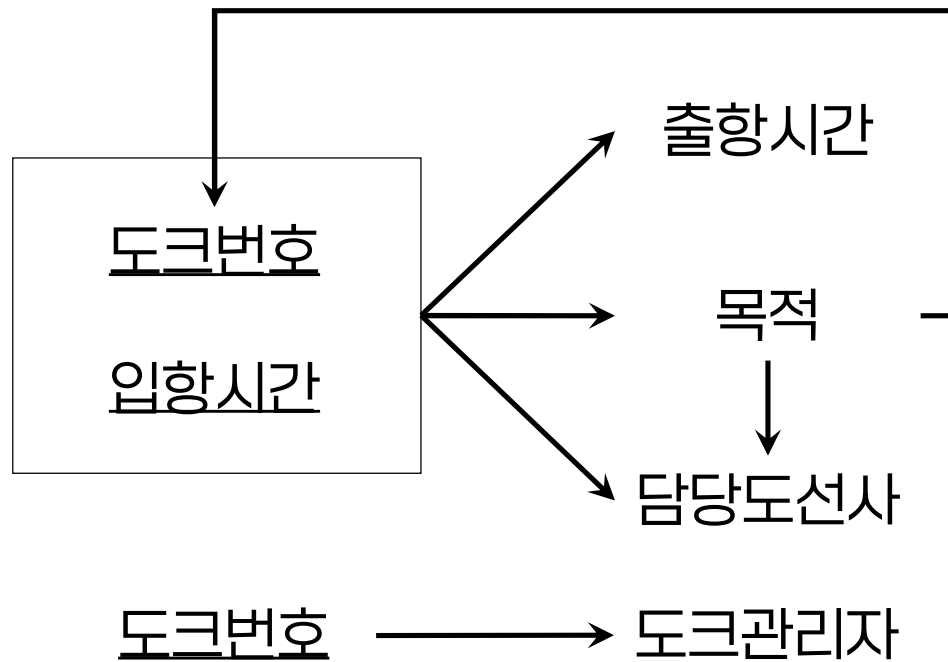
관리자 릴레이션

도크번호	도크관리자
D1	김규식
D2	한용운

## 제2정규형의 FDD



## 제2정규형의 FDD



## ☆ 제3정규형의 정의

### ▷ 형식적 정의

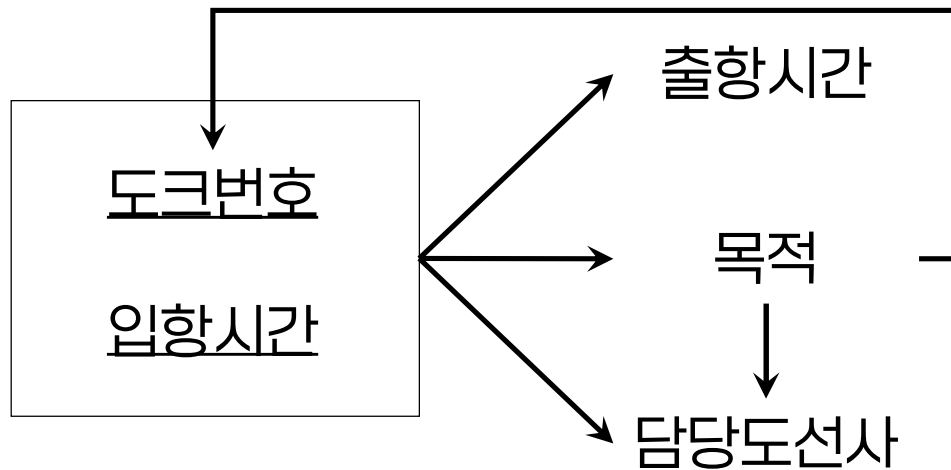
릴레이션이 제2정규형을 만족하고, 기본키가 아닌 속성들이 어떤 키에도 이행적으로 종속되지 않은 상태

### ▷ 이행적 종속성



$X \rightarrow Y$ 이고  $Y \rightarrow Z$ 이면  $X \rightarrow Z$ 이다

## 제3정규화의 적용



{도크번호, 입항시간} → {목적}

{목적} → {담당도선사}

⇒ {도크번호, 입항시간} → {담당도선사}

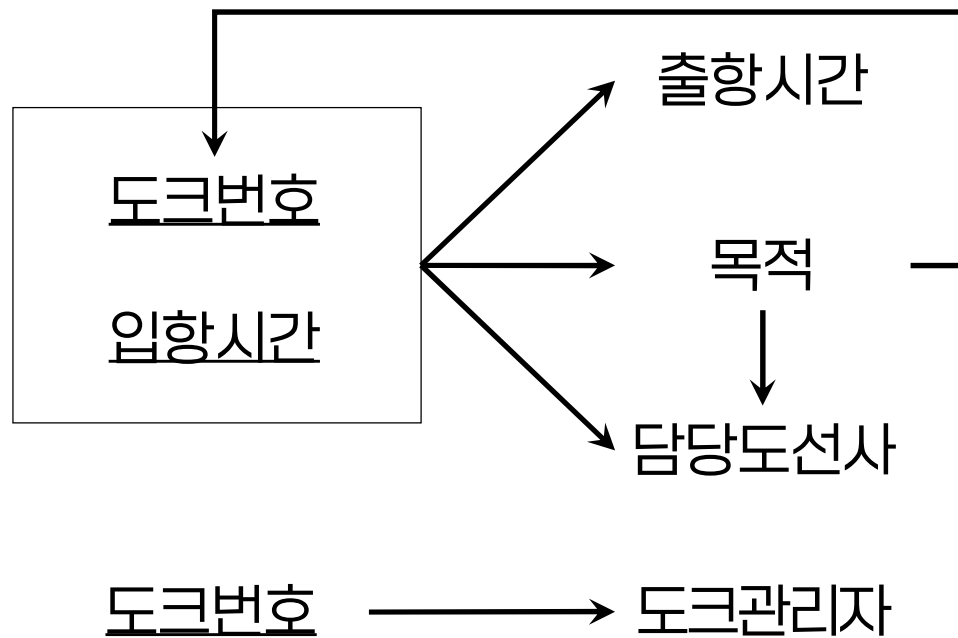
# 이행적 종속성을 포함한 릴레이션

.....

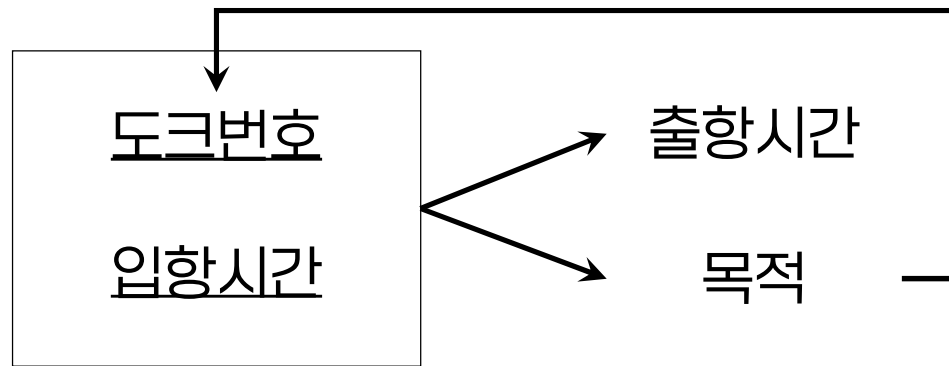
## 도크관리 릴레이션

도크번호	입항시간	출항시간	목적	담당도선사
D1	09:00	10:15	선적	김순애
D1	11:00	11:45	선적	김순애
D1	11:50	12:45	하역	김구
D2	09:00	10:00	관광	이동휘
D2	12:00	12:45	주유	이동휘
D2	13:00	15:00	정비	윤봉길

## 제3정규화의 적용



## 제3정규화의 적용



도크번호 → 도크관리자

목적 → 담당도선사



## 제3정규화의 적용

.....

도크관리 릴레이션

도크번호	입항시간	출항시간	목적	담당도선사
D1	09:00	10:15	선적	김순애
D1	11:00	11:45	선적	김순애
D1	11:50	12:45	하역	김구
D2	09:00	10:00	관광	이동휘
D2	12:00	12:45	주유	이동휘
D2	13:00	15:00	정비	윤봉길

## 제3정규화의 적용

입출항관리 릴레이션

도크번호	입항시간	출항시간	목적
D1	09:00	10:15	선적
D1	11:00	11:45	선적
D1	11:50	12:45	하역
D2	09:00	10:00	관광
D2	12:00	12:45	주유
D2	13:00	15:00	정비

도선 릴레이션

목적	담당도선사
선적	김순애
하역	김구
관광	이동휘
주유	이동휘
정비	윤봉길

# ☆ BC정규형의 정의

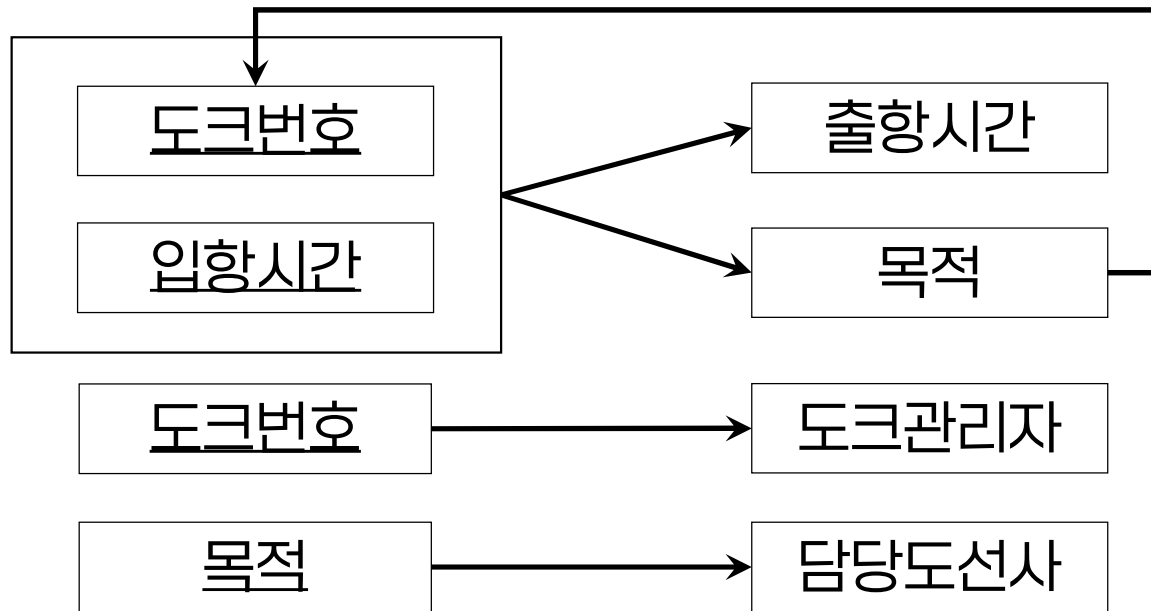
## ▷ 형식적 정의

릴레이션이 제3정규형을 만족하고 릴레이션에 성립하는  $X \rightarrow Y$  형태의 모든 함수적 종속성에 대하여  $X$ 가 수퍼키인 상태

## ▷ 입출항관리 릴레이션의 함수적 종속성

- ⊕ {도크번호, 입항시간}  $\rightarrow$  {목적}
- ⊕ {도크번호, 입항시간}  $\rightarrow$  {출항시간}
- ⊕ {목적}  $\rightarrow$  {도크번호}

# BC정규화의 적용



# BC정규화의 적용



# BC정규화의 적용

---

입출항관리 릴레이션

도크번호	입항시간	출항시간	목적
D1	09:00	10:15	선적
D1	11:00	11:45	선적
D1	11:50	12:45	하역
D2	09:00	10:00	관광
D2	12:00	12:45	주유
D2	13:00	15:00	정비

# BC정규화의 적용

.....

도크스케줄 릴레이션

목적	입항시간	출항시간
선적	09:00	10:15
선적	11:00	11:45
하역	11:50	12:45
관광	09:00	10:00
주유	12:00	12:45
정비	13:00	15:00

도크 릴레이션

목적	도크번호
선적	D1
하역	D1
관광	D2
주유	D2
정비	D2

## 정규화 결과

### 도크 릴레이션

도크번호	입항시간	출항시간	목적	도크관리자	담당도선사
D1	09:00	10:15	선적	김규식	김순애
D1	11:00	11:45	선적	김규식	김순애
D1	11:50	12:45	하역	김규식	김구
D2	09:00	10:00	관광	한용운	이동휘
D2	12:00	12:45	주유	한용운	이동휘
D2	13:00	15:00	정비	한용운	윤봉길



# **정규화 결과**

도크스케줄 릴레이션

목적	입항시간	출항시간
선적	09:00	10:15
선적	11:00	11:45
하역	11:50	12:45
관광	09:00	10:00
주유	12:00	12:45
정비	13:00	15:00

도크 릴레이션

목적	도크번호
선적	D1
하역	D1
관광	D2
주유	D2
정비	D2

관리자 릴레이션

도크번호	도크관리자
D1	김규식
D2	한용운

도선 릴레이션

목적	담당도선사
선적	김순애
하역	김구
관광	이동휘
주유	이동휘
정비	윤봉길

# 역정규화의 개념

---

## ▷ 정규화

- + 릴레이션 분할을 통해 데이터의 중복성을 최소화하는 과정
- + 사용 과정에서 많은 조인 연산을 유발

## ▷ 역정규화

- + 정규화의 반대 과정
- + 정규화를 통해 분리되었던 릴레이션을 통합하는 재조정을 수행하고 정보의 부분적 중복을 허용하는 기법
- + 데이터 접근 성능 개선 목적

## ▷ 정규화 되지 않은 스키마와 역정규화 스키마는 구별

다음 시간



# 저장 객체