

정규형과 정규화 이해하기

데이터베이스 설계의 핵심

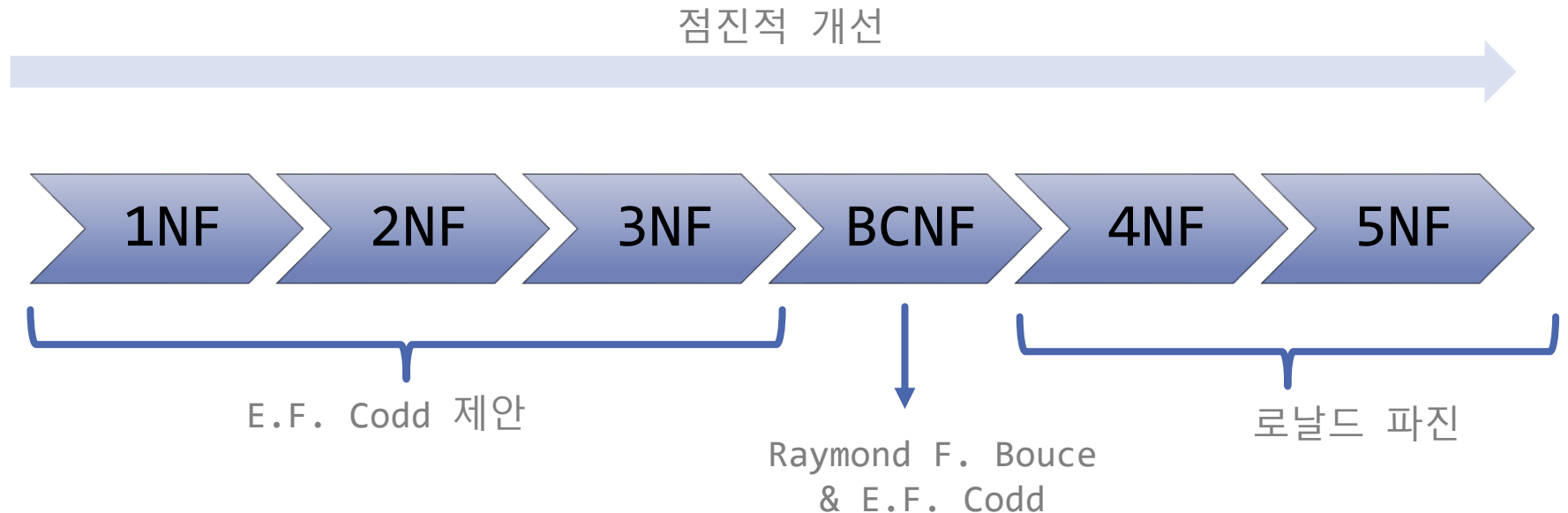
정규형과 정규화

: 정규형(Normal Form)

▶ 정규형 (Normal Form)

- ▶ 데이터 이상(anomaly) 현상을 제거하기 위한 관계형 모델의 설계 지침
- ▶ 데이터 중복을 최소화하고 무결성을 보장하기 위해 사용됨
- ▶ 삽입, 삭제, 갱신 이상을 방지하여 데이터베이스의 일관성과 효율성을 향상시킴
- ▶ 여러 단계의 정규형(1NF, 2NF, 3NF, BCNF, 4NF, 5NF)이 있으며, 각 단계는 점진적으로 데이터 구조를 개선함

비정규 릴레이션



정규형과 정규화 : 정규형(Normal Form)

▶ 정규화

- ▶ 정규형을 위배한 릴레이션을 정규형으로 만드는 과정
- ▶ 데이터베이스 테이블을 분석하고 수정하여 데이터 중복을 줄이고 무결성을 높이는 과정
- ▶ 데이터의 저장을 최적화하여 삽입, 삭제, 갱신 이상을 방지함
- ▶ 관계형 데이터베이스에서 효율적인 데이터 관리와 검색을 가능하게 함
- ▶ 단계별로 진행되며, 각 정규형을 적용하여 데이터 구조를 점진적으로 개선함
- ▶ 과도한 정규화가 성능 저하를 초래할 수 있어, 때로는 성능을 위해 의도적인 비정규화를 고려하기도 함

데이터베이스 이상 현상

: 삽입, 삭제, 갱신 이상

▶ 삽입 이상 (Insertion Anomaly)

- ▶ 개념: 필요한 데이터를 삽입할 때, 필수적인 정보가 부족하여 삽입이 어려운 상황
- ▶ 문제점: 필요한 데이터를 삽입하지 못하거나, 원치 않는 정보를 함께 입력해야 하는 경우 발생

▶ 삭제 이상 (Deletion Anomaly)

- ▶ 개념: 특정 정보를 삭제할 때 연관된 다른 유용한 정보까지 함께 삭제되는 상황
- ▶ 문제점: 중요한 정보의 손실로 이어져 데이터베이스의 유용성 감소

▶ 갱신 이상 (Update Anomaly)

- ▶ 개념: 데이터 중복으로 인해 동일한 데이터의 여러 사본을 각각 갱신해야 하는 상황
- ▶ 문제점: 일부 데이터만 갱신되어 데이터 간 불일치 발생

이러한 이상 현상들은 **데이터의 일관성과 무결성을 해치고,**
데이터관리의 복잡성을 증가시킴



정규화를 통해 **데이터베이스 설계를 개선**

함수적 종속성 (Functional Dependency)

- ▶ 하나의 속성이나 속성 그룹의 값이 다른 속성의 값을 결정하는 관계
 - ▶ = 한 속성의 값이 다른 속성의 값에 영향을 미치면
 - > 첫 번째 속성은 두 번째 속성에 **함수적으로 종속**된 것
- ▶ 예) 직원 데이터베이스에서 '직원 ID'라는 속성이 있고,
각 직원 ID에 대해 고유한 '직원 이름', '직원 주소' 등의 정보가 있다면,
'직원 ID'는 '직원 이름'과 '직원 주소'에 함수적으로 종속된 것

이 관계를 이해하는 것은 **데이터베이스의 정규화 과정에서 중요**하며
데이터 중복을 제거하고 무결성을 유지하는데 필수적이다

함수적 종속성 (Functional Dependency)

▶ 완전 함수 종속

- ▶ 정보 A가 정보 B를 식별할 때
- ▶ 예) 주민등록번호는 개인의 이름을 정확하게 결정하고 식별할 수 있음

▶ 부분 함수 종속

- ▶ 정보 A를 알기 위해 정보 B의 일부만 있어도 충분할 때
- ▶ 예) (교수 ID, 학기)로 이루어진 (복합)키에서 교수의 이름을 알기 위해서는 교수 ID만 있으면 충분함

▶ 이행적 함수 종속

- ▶ 정보 A가 정보 B를 결정하고, 정보 B가 정보 C를 결정할 때, 정보 A가 정보 C를 결정하게 되는 것
 - ▶ 이는 정보 A가 정보 C를 간접적으로 결정하는 관계가 형성
- ▶ 예) 직원 ID로 직원이 속한 부서를 알고, 그 부서로부터 부서의 위치를 알 수 있음

비정규 릴레이션

주문 테이블

주문 번호	회원 이름	주문 날짜	주문 상품명	상품 가격	수량	배송지
1001	홍길동	2024-04-30	티셔츠, 청바지	20000, 50000	1, 1	서울시 강남구
1002	홍길동	2024-05-01	운동화	80000	1	서울시 강남구
1003	김철수	2024-05-02	모자	20000	1	서울시 종로구
1004	이영희	2024-04-29	스웨터, 운동화	40000, 80000	1, 1	서울시 송파구
1005	이영희	2024-05-01	청바지	50000	2	서울시 송파구

▶ 데이터셋에서 나타나는 문제점

▶ 중복과 업데이트 이상 (Update Anomaly)

- ▶ '홍길동'이라는 이름이 여러 행에 걸쳐 중복되어 있고, 회원의 주소나 이름이 바뀌면 여러 행을 업데이트 해야 함

▶ 삽입 이상 (Insertion Anomaly)

- ▶ 새로운 상품을 주문하지 않은 새로운 회원을 추가하려면 상품 정보를 포함해야만 행을 추가할 수 있음

▶ 삭제 이상 (Deletion Anomaly)

- ▶ 특정 주문을 삭제하면 해당 회원에 대한 다른 중요한 정보도 사라질 수 있음

▶ 비원자적 속성 (Non-atomic values)

- ▶ 상품명과 상품 가격, 수량 등이 여러 값을 하나의 필드에 저장하고 있어 1NF를 위반함

제1정규형 (1NF)

: 도메인이 원자값

▶ 제1정규형 (1NF)

- ▶ **모든 컬럼이 원자값**을 가지도록 설계

- ▶ 개념: 모든 열은 원자값만을 포함해야 하며, 각 레코드는 유일해야 함

- ▶ 목적: 테이블에서 반복 그룹을 제거하고 엔트리가 간단하고 명확하도록 함

원자값(Atomic value): 더 이상 분해되지 않는 값, 더 이상 속성을 나눌 수 없는 값

제1정규형 (1NF)

: 도메인이 원자값

▶ 제1정규형 (1NF)

주문 테이블

주문 번호	회원 이름	주문 날짜	주문 상품명	상품 가격	수량	배송지
1001	홍길동	2024-04-30	티셔츠, 청바지	20000, 50000	1, 1	서울시 강남구
1001	홍길동	2024-04-30	청바지	50000	1	서울시 강남구
1002	홍길동	2024-05-01	운동화	80000	1	서울시 강남구
1003	김철수	2024-05-02	모자	20000	1	서울시 종로구
1004	이영희	2024-04-29	스웨터, 운동화	40000, 80000	1, 1	서울시 송파구
1004	이영희	2024-04-29	운동화	80000	1	서울시 송파구
1005	이영희	2024-05-01	청바지	50000	2	서울시 송파구

▶ 각 상품을 별도의 레코드로 분리하고, 각 레코드는 그 자체로 원자값만을 포함하게 됨

제2정규형 (2NF)

: 부분 함수 종속성 제거

▶ 제2정규형 (2NF)

▶ 부분 함수 종속성 제거

- ▶ 개념: 1NF를 충족하고, 모든 비주요 속성이 기본 키 전체에 대해 완전히 함수적으로 종속되어야 함
- ▶ 목적: 부분적 종속을 제거하여 각 데이터 항목이 테이블 내에서 한 번만 저장되도록 함

부분 종속(partial dependency): 일반 속성이 식별자의 일부 속성에만 종속되는 것

제2정규형 (2NF)

: 부분 함수 종속성 제거

- ▶ 제2정규형 (2NF)
 - ▶ 주문 테이블 : 주문에 대한기본 정보와 주문이 이루어진 날짜를 저장. '주문 번호'를 기본 키로 사용
 - ▶ 주문 상세 테이블 : 각 주문에 대한 상품 상세 정보를 포함.
('주문 번호', '주문 상품명')을 복합키로 사용

주문 테이블			
주문 번호	회원 이름	주문 날짜	배송지
1001	홍길동	2024-04-30	서울시 강남구
1001	홍길동	2024-04-30	서울시 강남구
1002	홍길동	2024-05-01	서울시 강남구
1003	김철수	2024-05-02	서울시 종로구
1004	이영희	2024-04-29	서울시 송파구
1004	이영희	2024-04-29	서울시 송파구
1005	이영희	2024-05-01	서울시 송파구

주문상세 테이블			
주문 번호	주문 상품명	상품 가격	수량
1001	티셔츠	20000	1
1001	청바지	50000	1
1002	운동화	80000	1
1003	모자	20000	1
1004	스웨터	40000	1
1004	운동화	80000	1
1005	청바지	50000	2

제2정규형 (2NF)

: 부분 함수 종속성 제거

- ▶ 제2정규형 (2NF)
 - ▶ 주문 테이블 : '회원 이름'과 '배송지'는 '주문 번호'에 완전 함수적으로 종속됨
 - ▶ 주문 상세 테이블 : '상품 가격'과 '상품 수량'은 복합키 ('주문 번호', '주문 상품명')에 완전 함수적으로 종속됨

주문 테이블			
주문 번호	회원 이름	주문 날짜	배송지
1001	홍길동	2024-04-30	서울시 강남구
1001	홍길동	2024-04-30	서울시 강남구
1002	홍길동	2024-05-01	서울시 강남구
1003	김철수	2024-05-02	서울시 종로구
1004	이영희	2024-04-29	서울시 송파구
1004	이영희	2024-04-29	서울시 송파구
1005	이영희	2024-05-01	서울시 송파구

주문상세 테이블			
주문 번호	주문 상품명	상품 가격	수량
1001	티셔츠	20000	1
1001	청바지	50000	1
1002	운동화	80000	1
1003	모자	20000	1
1004	스웨터	40000	1
1004	운동화	80000	1
1005	청바지	50000	2

제3정규형 (3NF)

: 이행적 함수 종속성 제거

▶ 제3정규형 (3NF)

▶ 이행적 함수 종속성 제거

- ▶ 개념: 2NF를 충족하고, 비주요 속성이 기본 키에 이행적으로 종속되지 않아야 함
- ▶ 목적: 이행적 종속성을 제거하여 데이터 중복을 더욱 줄이고 무결성을 향상시킴

이행적 함수 종속(transitive functional dependency):

A가 B에 함수적으로 종속되고 B가 C에 함수적으로 종속되면
A가 C에도 함수적으로 종속되는 것

제3정규형 (3NF)

: 이행적 함수 종속성 제거

- ▶ 제3정규형 (3NF)
 - ▶ 회원 테이블(생성) : 회원 정보를 별도로 관리하며 '회원 이름'을 기본 키로 사용
'배송지' 정보는 '회원 이름'에 직접 종속됨
 - ▶ 주문 테이블(수정) : '회원 이름' 속성을 제거하고 회원 정보는 '회원 테이블'에서 관리함

회원 테이블	
회원 이름	배송지
홍길동	서울시 강남구
홍길동	서울시 강남구
홍길동	서울시 강남구
김철수	서울시 종로구
이영희	서울시 송파구
이영희	서울시 송파구
이영희	서울시 송파구

주문 테이블		
주문 번호	회원 이름	주문 날짜
1001	홍길동	2024-04-30
1001	홍길동	2024-04-30
1002	홍길동	2024-05-01
1003	김철수	2024-05-02
1004	이영희	2024-04-29
1004	이영희	2024-04-29
1005	이영희	2024-05-01

주문상세 테이블			
주문 번호	주문 상품명	상품 가격	수량
1001	티셔츠	20000	1
1001	청바지	50000	1
1002	운동화	80000	1
1003	모자	20000	1
1004	스웨터	40000	1
1004	운동화	80000	1
1005	청바지	50000	2

제3정규형 (3NF)

: 이행적 함수 종속성 제거

- ▶ 제3정규형 (3NF)
 - ▶ 회원 테이블 : 회원의 이름과 배송지 정보를 주문 테이블로부터 분리, 이행적 종속성을 제거
 - ▶ 주문 테이블 : 주문과 관련된 핵심 정보만을 포함하도록 수정.
이제 주문 테이블은 주문 날짜와 회원 이름(회원 테이블과 연결)을 포함
 - ▶ 주문 상세 테이블 : 주문과 관련된 상품 정보를 담고 있으며 이행적 종속성은 없음

회원 테이블

회원 이름	배송지
홍길동	서울시 강남구
홍길동	서울시 강남구
홍길동	서울시 강남구
김철수	서울시 종로구
이영희	서울시 송파구
이영희	서울시 송파구
이영희	서울시 송파구

주문 테이블

주문 번호	회원 이름	주문 날짜
1001	홍길동	2024-04-30
1001	홍길동	2024-04-30
1002	홍길동	2024-05-01
1003	김철수	2024-05-02
1004	이영희	2024-04-29
1004	이영희	2024-04-29
1005	이영희	2024-05-01

주문상세 테이블

주문 번호	주문 상품명	상품 가격	수량
1001	티셔츠	20000	1
1001	청바지	50000	1
1002	운동화	80000	1
1003	모자	20000	1
1004	스웨터	40000	1
1004	운동화	80000	1
1005	청바지	50000	2