

정규화

정규화의 개념과 이상 현상

- 이상(anomaly)현상 : 불필요한 데이터 중복으로 인해 릴레이션에 대한 데이터 삽입, 수정, 삭제 연산을 수행할 때 발생할 수 있는 부작용

▼ 이상 현상의 종류

- 삽입 이상 : 새 데이터를 삽입하기 위해 불필요한 데이터도 함께 삽입해야 하는 문제
- 갱신 이상 : 중복 튜플 중 일부만 변경하여 데이터가 불일치하게 되는 모순의 문제
- 튜플을 삭제하면 꼭 필요한 데이터까지 함께 삭제되는 데이터 손실의 문제

-
- 삽입 이상 (Insertion anomaly):

릴레이션에 새 데이터를 삽입하려면 불필요한 데이터도 함께 삽입해야 하는 문제

- 갱신 이상 (update anomaly):

릴레이션의 중복된 튜플들 중 일부만 수정하여 데이터에 모순이 발생하는 문제

- 삭제 이상(Deletion anomaly) :

릴레이션에 튜플을 삭제하면 꼭 필요한 데이터까지 연쇄 삭제 현상이 발생하는 문제

-
- 함수적 종속성(FD; Functional Dependency) : 속성들 간의 관련성

정규화 : 함수적 종속성을 이용하여, 릴레이션을 연관성 이상 현상이 발생하지 않는 올바른 릴레이션으로 만들어 나가는 과정 ,

- 기본 목표는 관련이 없는 종속성은 별개의 릴레이션으로 표현
- 주의 사항 : 정규화를 통해 릴레이션은 무손실 분해가 되어야함
- 릴레이션이 의미상 동등한 릴레이션들로 분해되어야 하고 분해로 인한 정보 손실이 발생하지 않아야 함
- 분해된 릴레이션들을 자연조인 하면 분해 전의 릴레이션으로 복원 가능해야 함

함수 종속

▼ 함수 종속 : X가 Y를 함수적으로 결정한다

- 릴레이션 내의 모든 튜플에서 하나의 X값에 대한 Y값이 항상 하나임, X와 Y는 하나의 릴레이션을 구성하는 속성들의 부분 집합, Y가 X에 함수적으로 종속되어 있다와 같은 의미

▼ 함수 종속 관계 판단 시 유의 사항

- 속성 자체의 특성과 의미를 기반으로 함수 종속성을 판단해야 함, 속성 값은 계속 변할 수 있으므로 현재 릴레이션에 포함된 속성 값만으로 판단하면 안됨
- 일반적으로 기본키와 후보키는 릴레이션의 다른 모든 속성들을 함수적으로 결정함, 기본키나 후보키가 아니어도 다른 속성 값을 유일하게 결정하는 속성은 함수종속 관계에서 결정자가 될 수 있음

▼ 완전 함수 종속

- 릴레이션에서 속성 집합 Y가 속성 집합 X에 함수적으로 종속되어 있지만 속성 집합 X의 전체가 아닌 일부분에는 종속되지 않음을 의미, 일반적으로 함수 종속은 완전 종속을 의미함 / 예) 당첨여부는 {고객아이디, 이벤트번호}에 완전 함수 종속됨

▼ 부분 함수 종속 (PFD)

- 릴레이션에서 속성 집합 Y가 속성 집합 X의 전체가 아닌 일부분에도 함수적으로 종속됨을 의미 / 예) 고객이름은 {고객 아이디, 이벤트 번호}에 부분 함수 종속됨
-

기본 정규형과 정규화 과정

▼ 제 1 정규형 (1NF)

- 릴레이션의 모든 속성이 더는 분해되지 않는 원자 값만 가지면 제 1 정규형을 만족
- 제 1 정규형을 만족해야 관계 데이터 베이스의 릴레이션이 될 자격이 있음
- 제 1 정규형은 릴레이션에 속상 모든 속성의 도메인이 원자값으로만 구성되어 있으면 제 1 정규형에 속한다.
- 제 1 정규형은 만족하지만 이상 현상이 발생하는 릴레이션 예시

기본키인 {고객아이디, 이벤트 번호}에 완전 종속되지 못하고 일부분인 고객 아이디에 종속되는 등급과 할인을 속성이 존재하기 때문

▼ 제 2 정규형 (2NF):

- 릴레이션이 제 1 정규형에 속하고 기본키가 아닌 모든 속성이 기본키에 완전 함수 종속되면 제 2 정규형을 만족함
- 제 1 정규형에 속하는 릴레이션이 제 2 정규형을 만족하게 하려면 부분 함수 종속을 제거하고 모든 속성이 기본키에 완전 함수 종속되도록 분해함

▼ 제 2 정규형은 만족하지만 이상 현상이 발생하는 릴레이션의 예

- 이행적 함수 종속이 가능하기 때문
- 해결방법 : 이행적 함수 종속이 제거되도록 고객 릴레이션을 분해 → 분해된 릴레이션들은 제 3정규형에 속하게 됨
- 이행적 함수 종속 : 릴레이션을 구성하는 3개의 속성 집합 X, Y, Z
- 함수 종속 관계 $X \rightarrow Y$ 와 $Y \rightarrow Z$ 가 존재하면 논리적으로 $X \rightarrow Z$ 가 성립되는데 , 이때 Z 가 X 에 이행적으로 함수 종속되었다고 함

▼ 제 3 정규형(3NF : Third Normal Form)

- 릴레이션이 제 2 정규형에 속하고, 기본키가 아닌 모든 속성이 기본키에 이행적 함수 종속이 되지 않으면 제 3 정규형을 만족함
- 보이스 / 코드 정규형 (BCNF)
필요성 : 하나의 릴레이션에 여러개의 후보키가 존재하는 경우 , 제 3 정규형까지 모두 만족해도 이상 현상이 발생할 수 있음

릴레이션의 함수 종속 관계에서 모든 결정자가 후보키이면 보이스/코드 정규형에 속함

- 의미
강한 제 3 정규형 : 후보키를 여러 개 가지고 있는 릴레이션에 발생할 수 있는 이상 현상을 해결하기 위해 제 3 정규형보다 좀 더 엄격한 제약조건을 제시

▼ 제 4 정규형

- 릴레이션이 보이스/코드 정규형을 만족하면서 함수 종속이 아닌 다치 종속을 제거하면 제 4 정규형에 속함

▼ 제 5 정규형

- 릴레이션이 제 4 정규형을 만족하면, 후보키를 통하지 않는 조인 종속 (JD) 을 제거하면 제 5 정규형에 속함

▼ 정규화 시 주의사항

- 모든 릴레이션이 제 5정규형에 속해야만 바람직한 것은 아님
- 일반적으로 제 3 정규형이나 보이스/코드 정규형에 속하도록 릴레이션을 분해하여 데이터 중복을 줄이고 이상 현상을 해결하는 경우가 많음