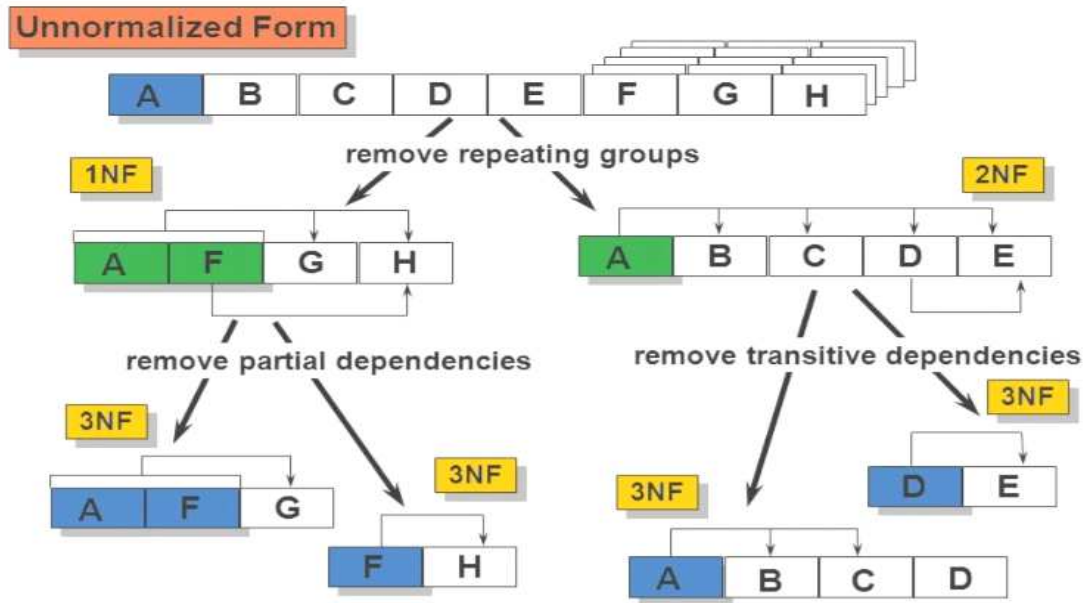


정규화 (Normalization)



이미지 출처 : <https://estuary.dev/data-normalization/>

1. 개념

정규화는 데이터베이스 설계에서 중요한 개념으로, 데이터의 중복을 최소화하고 데이터 무결성을 유지하기 위해 사용된다. 테이블을 작은 단위로 분할하여 데이터 중복을 방지하고 관련성 있는 데이터를 연결한다. 그러므로 데이터의 일관성을 유지하고 데이터 수정시 발생하는 이상현상을 방지한다.

각 단계를 거칠 때마다 테이블은 더욱 정제되어 효율적인 데이터 관리가 가능해진다. 그러나 과도한 정규화는 쿼리의 복잡성을 증가시킬 수 있으므로 실제 응용 시에는 데이터베이스의 성능과 요구사항을 고려하여 적절한 수준의 정규화를 결정해야 한다.

정규화 과정에서는 데이터의 관계를 명확하게 파악하는 것이 중요하며 각 정규형의 기준을 충족시키면서도 실제 응용에 적합한 구조를 설계하는 것이 핵심이다. 데이터베이스 설계 및 관리에 있어 정규화는 필수적인 과정으로 이를 통해 데이터의 일관성, 무결성, 그리고 효율적인 접근을 보장할 수 있다.

2. 정규형의 종류와 특징

제1정규형(1NF) : 모든 컬럼의 값이 원자값(분할 불가능한 값)을 가져야 한다.

제2정규형(2NF) : 1NF를 만족하며 모든 비주요 속성이 기본키에 완전 함수적 종속을 만족해야 한다. 즉, 부분적 종속을 제거한다.

제3정규형(3NF): 2NF를 만족하고 비주요 속성이 기본키에 이행적 종속을 만족하지 않아야 한다.

보이스-코드 정규형(BCNF) : 모든 결정자가 후보키가 되는 정규형이다. 기본키가 아닌 속성에 의해 기본키가 함수적 종속되는 상황을 제거한다.

제4정규형 (4NF) : BCNF를 만족하며, 다치 종속(multi-valued dependency)을 제거한다.

[정규형 예시]

비정규화된 테이블 (1NF 미만)

학생 ID	학생 이름	수강 과목 및 성적
001	김철수	수학:90, 화학:95
002	이영희	수학:85, 물리:90

- 위 테이블은 '수강 과목 및 성적' 컬럼에서 여러 값을 하나의 필드에 저장하고 있어 1NF를 만족하지 않는다. 이러한 구조는 데이터 중복 및 업데이트 이상의 위험을 증가시킨다.

1) 제1정규형(1NF)

- 모든 컬럼의 값이 원자값(분할 불가능한 값)을 가져야 한다.

학생 ID	학생 이름	과목명	성적
001	김철수	수학	90
001	김철수	화학	95
002	이영희	수학	85
002	이영희	물리	90

변경사항 : '수강 과목 및 성적'을 분리하여 각 과목과 성적을 별도의 레코드로 수정하였음

2) 제2정규형(2NF)

[학생 테이블]

학생 ID	학생 이름
001	김철수
002	이영희

[성적 테이블]

학생 ID	과목명	성적
001	수학	90
001	화학	95
002	수학	85
002	물리	90

변경사항: 학생ID와 과목명을 별도의 테이블로 분리하여 수정하였음.

3) 제3정규형(3NF)

[학생 테이블]

학생 ID	학생 이름
001	김철수
002	이영희

[성적 테이블]

학생 ID	과목 ID	성적
001	M101	90
001	C101	95
002	M101	85
002	P101	90

[과목 테이블]

과목 ID	과목명
M101	수학
C101	화학
P101	물리

'성적 테이블'은 학생ID와 과목ID에 의해 성적이 결정되며, '과목 테이블'은 과목ID에 의해 과목명이 결정된다.

변경사항: 과목명과 성적 사이의 이행적 종속성을 제거하기 위해 과목명을 별도의 테이블로 분리하여 수정하였음

4) 보이스-코드 정규형(BCNF), 모든 결정자가 후보키가 되는 정규형이다. 기본키가 아닌 속성에 의해 기본키가 함수적 종속되는 상황을 제거한다.

5) 제4정규형 (4NF)

특징: BCNF를 만족하며, 다치 종속(multi-valued dependency)을 제거한다.