

# 데이터모델링의 성능

## 1. 성능 데이터 모델링의 개요

- 성능 데이터 모델링 고려사항
  - 데이터 모델링을 할 때 정규화를 정확하게 수행한다.
  - 데이터베이스 용량산정을 수행한다.
  - 데이터베이스에 발생하는 트랜잭션의 유형을 파악한다.
  - 용량과 트랜잭션의 유형에 따라 반정규화를 수행한다.
  - 이력모델의 조정, PK/FK조정, 슈퍼타입/서브타입 조정 등을 수행한다.
  - 성능관점에서 데이터 모델을 검증한다.

## 2. 정규화와 성능

### a. 정규화 이론

- 1차 정규형

- 복수의 속성값을 갖는 속성을 분리함 이로써 속성의 원자성확보

- 



COURSE\_ID에 대해 ORDER BY를 할 수 없으며 WHERE 조건절에서도 여러 속성값들중 어떤 것을 검색해야 하는지 알 수 없음

- 



- 2차 정규형

- 부분종속성을 분리

- 



- 학번 -> 학부 -> 등록금
- 학번이 같으면 학부, 등록금이 같다

- 



- 3차 정규형

- 제 2정규형에 속하면서 기본키가 아닌 모든 속성이 기본키에 이행적 함수 종속이 되지 않으면 제 3정규화 해주어야 한다.

- 



- 



### 3. 반정규화의 정의

- 시스템의 성능을 향상시키기 위해 데이터 모델을 통합하는 프로세스이다.
- 정규화로 인해서 테이블의 수가 증가하게 되면, 다수의 JOIN이 발생함에 따라 성능 저하가 발생할 수 있다. 이 문제를 해결하고자 정규화 과정이 모두 마무리 한 다음 반정규화를 실시한다.
- 단, 테이블을 합치는 것만이 반정규화는 아니다.

### 4. 정규화와 반정규화

- 정규화와 반정규화는 Trade off 관계에 있다.
- 정규화의 특징
  - 정합성과 데이터무결성이 보장
  - 테이블이 복잡해지고 성능이 떨어질 수 있음
  - 입력(Create), 수정(Update), 삭제>Delete)의 성능은 향상되고 조회(Read)의 경우 나빠질수도 있고 좋아질 수도 있다.
- 반정규화의 특징
  - 반 정규화를 하면 테이블이 단순화되며 성능이 향상
  - 정합성과 데이터무결성이 보장되지 않을 수있음
  - 반 정규화는 의도적으로 중복을 만들어 검색(Read) 성능을 향상시킨다.
  - 하지만 속성이 각기 다른 테이블에 중복되어 나타나기 때문에 입력(Create), 수정(Update), 삭제>Delete)의 성능은 낮아진다.
  - 애플리케이션 레벨에서 꼭 해당 속성에 포함된 모든 테이블에 트랜잭션을 보내야 한다.



### 5. 반정규화 적용 절차

- 반정규화 대상조사

- 범위처리빈도수 조사
- 대량의 범위 처리 조사
- 통계성 프로세스 조사
- 테이블 조인 개수
- 다른 방법유도 검토
  - 뷰(VIEW) 테이블
  - 클러스터링 적용
  - 인덱스의 조정
  - 응용어플리케이션
- 반정규화 적용

## 6. 테이블 반정규화

- 테이블 병합
  - 1:1 관계, 1:M 관계를 통합하여 조인 횟수를 줄여 성능을 향상
    - 컬럼을 조작해서 join을 줄이기
    - - 수평분할(행을 기준으로 분할) : 테이블 분할에 레코드를 기준
      - 데이터모델링의 성능

- author\_id가 1~1000 이라면 topic\_1000테이블에 author\_id가 1001~1500이라면 topic\_1500테이블에 넣음
- 서로다른 테이블에서 조회하기 때문에 속도가 빨라짐
- 수직 분할(컬럼을 기준으로 분할) : 하나의 테이블이 가지는 컬럼의 개수가 증가하는 경우 활용, 갱신 위주의 컬럼, 수직 분할, 조회 빈도가 높은 컬럼을 분할
- 
- 
- topic 테이블에 조회가 많이 일어나고 description을 매번 조회하지 않아도 되는 경우 description컬럼과 title컬럼을 따로해서 테이블을 형성함
  - description컬럼 데이터들이 용량이 커서 시간이 오래걸리는 경우 유용함
- 중복 테이블 생성
  - 대량의 데이터들에 대한 집계함수(GROUP BY, SUM 등)를 사용하여 실시간 통계정보를 계산하는 경우에 효과적인 수행을 위해 별도의 통계 테이블을 두거나 중복 테이블을 추가
- 관계의 반정규화
  - 중복관계 추가
    - 데이터를 처리하기 위한 여러 경로를 거쳐 조인이 가능하지만 이 때 발생할 수 있는 성능저하를 예방하기 위해 추가적인 관계를 맺는 방법이 관계의 반정규화임

## 7. 대량 데이터에 따른 성능

- 한 테이블에 많은 수의 컬럼을 가지고 있으면 성능저하
- 반정규화된 테이블의 사례
- 테이블에 대한 수평분할/수직분할의 절차
  - 데이터 모델링을 완성한다.
  - 데이터베이스 용량산정을 한다.
  - 대량 데이터가 처리되는 테이블에 대해서 트랜잭션 처리 패턴을 분석한다.
  - 칼럼 단위로 집중화된 처리가 발생하는지, 로우단위로 집중화된 처리가 발생하는지 분석하여 집중화된 단위로 테이블을 분리하는 것을 검토한다.

