Chapter 8-4:

관계 데이터베이스 설계 (Relational Database Design)

Database System Concepts, 6th Ed.

Chapter 8: Relational Database Design

- 좋은 릴레이션 설계를 위한 고려 사항
- 함수 종속 이론
- 원자형 도메인과 제 1 정규형
- 함수 종속을 이용한 분해
- 분해(decomposition)를 위한 알고리즘 (BCNF, 3NF)
- 다중값 종속을 이용한 분해 (4NF)
- 다른 정규형들
- 데이터베이스 설계 절차

BCNF는 충분한가?

충분히 정규화되지 않은 BCNF 데이터베이스 스키마가 있다

다음 릴레이션을 고려해 보자

inst_info (ID, child_name, phone) --- BCNF이다

instructor는 하나 이상의 전화와 다수의 자녀를 가질 수 있다.

ID	*child_name*	phone
99999	David	512-555-1234
99999	David	512-555-4321
99999	William	512-555-1234
99999	William	512-555-4321

inst_info

BCNF는 충분한가? (계속)

따라서, *inst_info* 를 아래와 같이 분해하는 것이 더 낫다:

inst_child

ID	*child_name*
99999	David
99999	Williams

inst_phone

ID	*phone*
99999	512-555-1234
99999	512-555-4321

고차원 정규화의 필요성을 나타내며, 이들 두 릴레이션은 4NF이다

다중값 종속(Multivalued Dependencies , MVD)

R을 릴레이션 스키마라 하고 $\alpha \subseteq R$ 이고 $\beta \subseteq R$이라 하자.
아래와 같은 다중값 종속이

$$\alpha \rightarrow\rightarrow \beta$$

어떤 적법한 릴레이션 r(R)에 있어 $t_1[\alpha] = t_2[\alpha]$인 r내의 모든 튜플 t_1과 t_2의 쌍에 대해 r내에 다음과 같은 튜플 t_3와 t_4가 존재하면 R상에 존재한다

$$t_1[\alpha] = t_2[\alpha] = t_3[\alpha] = t_4[\alpha]$$

$$t_3[\beta]=t_1[\beta]$$

$$t_3[R - \beta] = t_2[R - \beta]$$

$$t_4[\beta] = t_2[\beta]$$

$$t_4[R - \beta] = t_1[R - \beta]$$

MVD (Cont.)

$\alpha \rightarrow\rightarrow \beta$ 의 테이블 표현

$t_1[\alpha] = t_2[\alpha] = t_3[\alpha] = t_4[\alpha]$

$t_3[\beta]=t_1[\beta]$

$t_3[R - \beta] = t_2[R - \beta]$

$t_4[\beta] = t_2[\beta]$

$t_4[R - \beta] = t_1[R - \beta]$

	α	β	$R - \alpha - \beta$
t_1	$a_1 \ldots a_i$	$a_{i+1} \ldots a_j$	$a_{j+1} \ldots a_n$
t_2	$a_1 \ldots a_i$	$b_{i+1} \ldots b_j$	$b_{j+1} \ldots b_n$
t_3	$a_1 \ldots a_i$	$a_{i+1} \ldots a_j$	$b_{j+1} \ldots b_n$
t_4	$a_1 \ldots a_i$	$b_{i+1} \ldots b_j$	$a_{j+1} \ldots a_n$

ID $\rightarrow\rightarrow$ *child_name*

ID $\rightarrow\rightarrow$ *phone*

ID	***child_name***	**phone**
99999	David	512-555-1234
99999	William	512-555-4321
99999	David	512-555-4321
99999	William	512-555-1234

예제

R을 아래와 같은 3개의 공집합이 아닌 부분 집합으로 분할된 애트리뷰트의 집합을 가진 릴레이션 스키마라 하자.

Y, Z, W

Y →→ Z(Y가 Z를 다중 결정한다)를 만족하는 필요 충분 조건은 모든 가능한 릴레이션 r(R)에 대해 아래와 같을 때이다

$< y_1, z_1, w_1 > \in r$ 이고 $< y_1, z_2, w_2 > \in r$ 이면

$< y_1, z_1, w_2 > \in r$ 이고 $< y_1, z_2, w_1 > \in r$ 이다.

Y	Z	W
y1	z1	w1
y1	z2	w2
y1	z1	w2
y1	z2	w1

Z와 W의 행위는 동일하기 때문에 Y →→ Z를 만족하는 필요 충분조건은 Y →→ W를 따른다는 사실을 유의하라.

예제(계속)

앞의 예에서:

$ID \rightarrow\rightarrow child\_name$
$ID \rightarrow\rightarrow phone\_number$

위의 형식적 정의는 특정 Y값(ID)이 주어지면 그것에 Z의 값(*child_name*) 집합과 W의 값(*phone_number*) 집합을 관련짓도록 하고 이들 두 집합 Z와 W는 어떤 의미에서 서로 독립적이라는 개념을 정형화하도록 한다

유의

- $Y \rightarrow Z$ 이면 $Y \rightarrow\rightarrow Z$ 이다

- 실제로 (위의 표기에서) $Z_1 = Z_2$를 가진다

다중값 종속의 사용

두 가지 방식으로 다중값 종속을 사용한다.

1. 릴레이션이 주어진 함수 종속과 다중값 종속 집합하에서 적법한가의 여부를 결정하는데 사용.
2. 적법한 릴레이션의 집합에 제약 조건을 지정하는데 사용.

릴레이션 r이 주어진 다중값 종속을 만족하지 않으면, r에 튜플들을 추가해 다중값 종속을 만족하는 릴레이션 r′를 구축할 수 있다.

다중값 종속 이론

D를 함수 및 다중값 종속의 집합이라 하자. D의 폐포 D^+는 D에 논리적으로 내포된 모든 함수적 및 다중값 종속의 집합이다.

함수적 및 다중값 종속에 대한 건전하고 완전한 추론 규칙들

1. **재귀 규칙.** α가 애트리뷰트의 집합이고 $\beta \subseteq \alpha$이면, $\alpha \rightarrow \beta$가 성립한다
2. **증가 규칙.** $\alpha \rightarrow \beta$이고 γ가 애트리뷰트의 집합이면, $\gamma\alpha \rightarrow \gamma\beta$가 성립한다.
3. **이행 규칙.** $\alpha \rightarrow \beta$이고 $\beta \rightarrow \gamma$이면, $\alpha \rightarrow \gamma$가 성립한다.

암스트롱의 공리(**Armstrong's Axioms**)

다중값 종속 이론(계속)

4. 보충 규칙. $\alpha \rightarrow\rightarrow \beta$이면 $\alpha \rightarrow\rightarrow R - \beta - \alpha$가 성립한다.

5. 다중값 증가 규칙. $\alpha \rightarrow\rightarrow \beta$이고 $\gamma \subseteq R$ 및 $\delta \subseteq \gamma$이면

   $\gamma\alpha \rightarrow\rightarrow \delta\beta$ 가 성립한다.

6. 다중값 이행 규칙. $\alpha \rightarrow\rightarrow \beta$이고 $\beta \rightarrow\rightarrow \gamma$ 이면 $\alpha \rightarrow\rightarrow \gamma - \beta$가

   성립한다.

7. 중복 규칙. $\alpha \rightarrow \beta$이면 $\alpha \rightarrow\rightarrow \beta$가 성립한다.

8. 유착 규칙. $\alpha \rightarrow\rightarrow \beta$이고 $\gamma \subseteq \beta$ 및 $\delta \subseteq R$ 이고 $\delta \cap \beta = \emptyset$

   이며 $\delta \rightarrow \gamma$ 인 δ가 존재하면, $\alpha \rightarrow \gamma$가 성립한다.

D^+ 계산의 단순화

다음과 같은 규칙 (규칙 1-8을 사용해 증명함)을 사용해 D^+ 계산을 단순화할 수 있다.

- **다중값 합집합 규칙.** $\alpha \rightarrow\rightarrow \beta$이고 $\alpha \rightarrow\rightarrow \gamma$ 이면, $\alpha \rightarrow\rightarrow \beta\gamma$ 가 성립한다.
- **공통 집합 규칙.** $\alpha \rightarrow\rightarrow \beta$이고 $\alpha \rightarrow\rightarrow \gamma$ 이면, $\alpha \rightarrow\rightarrow \beta \cap \gamma$ 가 성립한다.
- **차집합 규칙.** $\alpha \rightarrow\rightarrow \beta$이고 $\alpha \rightarrow\rightarrow \gamma$ 이면, $\alpha \rightarrow\rightarrow \beta - \gamma$와 $\alpha \rightarrow\rightarrow \gamma - \beta$ 가 성립한다.

4th Normal Form (4NF)

4NF의 정의:

$\alpha \subseteq R$ 이고 $\beta \subseteq R$인 $\alpha \rightarrow\rightarrow \beta$ 형태의 D^+내 모든 다중값 종속에 대해 다음중 하나의 조건이 만족되면, 함수 및 다중값 종속의 집합 D에 대해 릴레이션 스키마 R은 4NF이다.

- $\alpha \rightarrow\rightarrow \beta$ 가 자명하다 (즉, $\beta \subseteq \alpha$ 또는 $\alpha \cup \beta = R$)

- α 가 스키마 R의 수퍼키이다.

BCNF 정의:

$\alpha \subseteq R$이고 $\beta \subseteq R$일 때 $\alpha \rightarrow \beta$ 형태의 F^+내의 모든 함수적 종속에 대하여 다음 조건 중 하나가 만족되면 함수적 종속 집합 F에 대하여 릴레이션 스키마 R은 BCNF이다:

- $\alpha \rightarrow \beta$ 가 자명하다 (즉, $\beta \subseteq \alpha$)

- α 가 R의 수퍼 키이다

- 릴레이션이 4NF이면, 그것은 BCNF이다.

4NF Decomposition Algorithm

result: = {*R*};
done := false;
compute D^+;
Let D_i denote the restriction of D^+ to R_i

while (**not** *done*)
  if (there is a schema $\mathbf{R}_i$ in *result* that is not in 4NF) **then**
    begin
      let $\alpha \rightarrow\rightarrow \beta$ be a nontrivial multivalued dependency that holds
      on R_i such that $\alpha \rightarrow R_i$ is not in D_i, and $\alpha \cap \beta = \phi$;
      result := $(result - R_i) \cup (R_i - \beta) \cup (\alpha, \beta)$;
    end
  else *done*:= true;

유의: 각 R_i는 4NF이며, 분해는 무손실 조인이다.

Example

$R = (A, B, C, G, H, I)$

$F = \{ A \rightarrow\rightarrow B$

$\quad B \rightarrow\rightarrow HI$

$\quad CG \rightarrow\rightarrow H \}$

R is not in 4NF since $A \rightarrow\rightarrow B$ and A is not a superkey for R

Decomposition

a) $R_1 = (A, B)$ (R_1 is in 4NF)

b) $R_2 = (A, C, G, H, I)$ (R_2 is not in 4NF, decompose into R_3 and R_4)

c) $R_3 = (C, G, H)$ (R_3 is in 4NF)

d) $R_4 = (A, C, G, I)$ (R_4 is not in 4NF, decompose into R_5 and R_6)

$A \rightarrow\rightarrow B$ and $B \rightarrow\rightarrow HI$ ➔ $A \rightarrow\rightarrow HI$, (MVD transitivity), and

and hence $A \rightarrow\rightarrow I$ *(MVD restriction to R_4)*

e) $R_5 = (A, I)$ (R_5 is in 4NF)

f) $R_6 = (A, C, G)$ (R_6 is in 4NF)

다른 정규형들

조인 종속 **(Join dependencies)**는 다중값 종속을 일반화 시킨 것이다.

- **project-join normal form (PJNF)** 이라고 부른다 (혹은 5th NF 이라고 부른다)

더 일반적인 제약조건을 사용하여 **domain-key normal form**을 유도해 낼 수 있다.

이러한 일반화된 제약조건을 사용하는 경우 문제점: 이해하기 어려우며, 적절한 추론 규칙이 없음.

따라서 극히 드물게 사용됨.

데이터베이스 설계 절차

스키마 *R* 이 주어졌다고 가정하면,

- *R*은 E-R 다이어그램을 릴레이션 스키마 집합으로 변환할 때 생성될 수 있다.
- *R*은 관련된 모든 속성을 포함하는 하나의 릴레이션이 될 수 있다 (**universal relation** 이라 부른다). 다음, 정규화 과정은 *R* 을 작은 스키마 릴레이션으로 분해한다.
- *R*은 임시적인 설계의 결과 생성될 수 있는데, 이후 정규형을 만족하는지 검사/분해하여 정규화 스키마 릴레이션를 얻는다.

End of Chapter