Quelques Principes de Bon Database Design

Jef Wijsen

October 10, 2007

1 Dépendances fonctionnelles

Une *dépendance fonctionnelle* (DF) est une contrainte de la forme $B_1B_2\dots B_n \rightarrow C$ avec $B_1, B_2, \dots, B_n, C$ des attributs du schéma de relation en question. Une relation respecte cette DF si pour tous les tuples s, t de la relation, l'expression suivante est vraie: [1]

$$\begin{array}{rrcl} \text{si} & s(B_1) & = & t(B_1) \\ \text{et} & s(B_2) & = & t(B_2) \\ & & \vdots & \\ \text{et} & s(B_n) & = & t(B_n) \\ \text{alors} & s(C) & = & t(C)\ . \end{array}$$

C'est-à-dire, si deux tuples ont les mêmes valeurs pour tous les attributs $B_1, B_2, \dots, B_n$, alors ces deux tuples doivent avoir la même valeur pour l'attribut C. On dira aussi que C est *déterminé* par $B_1, B_2, \dots, B_n$.

Un schéma de relation peut être "enrichi" avec les DF à respecter. Pour la relation suivante, on aura:

ENom → Sal: Un employé ne gagne qu'un seul salaire.

ENom → DNom: Un employé ne peut pas travailler pour deux départements différents.

DNom → Budget: Un département ne profite que d'un seul budget.

Par contre, on ne souhaite pas imposer, par exemple, Sal → ENom, parce qu'il est possible que deux employés gagnent le même salaire.

ENom	Sal	DNom	Budget
Ed	3000	MIS	5000K
An	3500	Marketing	2000K
Tim	3600	Marketing	2000K

Exemple 1 Le schéma HORAIRE [Prof, Cours, Local, Sem, Jour, Heure] pour stocker, par exemple, le tuple ⟨J. Wijsen, Gestion de bases de données, 230, 1, vendredi, 13h15⟩. On imposera:

Cours → Prof
Prof, Sem, Jour, Heure → Cours
Local, Sem, Jour, Heure → Cours
Cours, Sem, Jour, Heure → Local

[1] Si $t = \langle a_1, a_2, \dots, a_m \rangle$ est un tuple sur le schéma $R\,[A_1, A_2, \dots, A_m]$, alors $t(A_1) = a_1$, $t(A_2) = a_2$, ...

Si on n'acceptait pas de cours à cheval sur deux semestres:

$$\mathsf{Cours} \to \mathsf{Sem}$$

Que penser de:

$$\mathsf{Cours} \to \mathsf{Jour}\ ?$$

$\triangleleft$

2 Raisonner sur les DF

Si une relation sur le schéma $\mathsf{ED}\,[\mathsf{ENom}, \mathsf{Sal}, \mathsf{DNom}, \mathsf{Budget}]$ respecte à la fois $\mathsf{ENom} \to \mathsf{DNom}$ et $\mathsf{DNom} \to \mathsf{Budget}$, elle respectera forcément $\mathsf{ENom} \to \mathsf{Budget}$ (pourquoi ?). Autrement dit, si ENom détermine DNom et DNom détermine Budget, alors ENom détermine aussi Budget.

Soit Σ un ensemble de DF. Soit X un ensemble d'attributs. Pour connaître tous les attributs déterminés par X, il suffit d'appliquer la première règle une fois, puis la deuxième règle jusqu'au "point fixe":

1. Tout attribut de X est déterminé par X.
2. *Transititvité:* Si on a la DF $B_1 B_2 \dots B_n \to C$ et si chaque B_i est déjà déterminé par X ($1 \leq i \leq n$), alors C est déterminé par X.

Exemple 2 Soit $\Sigma = \{A \to C, B \to D, BCD \to E, AEF \to G, AG \to F, DE \to A\}$. Les attributs déterminés par AB sont:

- A et B à cause de la première règle.
- C à cause de la deuxième règle, suite à $A \to C$.
- D à cause de la deuxième règle, suite à $B \to D$.
- E à cause de la deuxième règle, suite à $BCD \to E$.

$\triangleleft$

3 Boyce-Codd Normal Form

Un schéma de relation avec un ensemble Σ de DF est en *Boyce-Codd Normal Form* (BCNF) si la condition suivante est respectée:

> si $X \to C$ est une DF dans Σ avec $C \notin X$, alors X détermine <u>tous les attributs</u> du schéma.

BCNF est un "certificat de qualité": un schéma certifié BCNF ne souffre pas des anomalies vues précédemment. Par contre, ces anomalies peuvent se produire dans les schémas non BCNF. En effet, prenons le cas le plus simple d'un schéma non BCNF:

$$R\,[A, B, C] \text{ et } \Sigma = \{B \to C\}$$

Ce schéma n'est pas en BCNF: $B \to C$ est une DF dans Σ, mais B ne détermine pas A. Une relation sur ce schéma peut contenir de la redondance, tout en respectant la DF

$B \rightarrow C$: dans la relation suivante, la valeur qui se cache derrière le point d'interrogation ne peut être que c.

A	B	C
a_1	b	c
a_2	b	?

Exemple 3 Le schéma ED [ENom, Sal, DNom, Budget] n'est pas en BCNF, parce que DNom → Budget, mais DNom ne détermine ni ENom, ni Sal. ◁

4 3NF

Pour la relation suivante, les contraintes sont Cours → Prof et Prof, Heure → Cours (un prof ne peut pas donner deux cours différents au même moment). Puisque Cours ne détermine pas Heure, le schéma n'est pas en BCNF.

HORAIRE	Prof	Cours	Heure
	Ed	BD	lun, 8h15
	Ed	BD	ven, 13h15

Cependant, si on décidait d'utiliser deux schémas QUI [Cours, Prof] et QUAND [Cours, Heure], il n'est plus aisé de vérifier Prof, Heure → Cours.

QUI	Cours	Prof
	BD	Ed

QUAND	Cours	Heure
	BD	lun, 8h15
	BD	ven, 13h15

Une *clé* est un ensemble X tel que:

- X détermine tous les attributs; et
- *Minimalité:* aucun sous ensemble propre de X ne détermine tous les attributs.

Pour cet exemple, on obtient deux clés:

1. {Cours, Heure},
2. {Prof, Heure}.

Pour la table HORAIRE, il sera toujours facile à vérifier que la clé {Prof, Heure} est respectée (pas de doublons). Par contre, la décomposition en QUI [Cours, Prof] et QUAND [Cours, Heure] "scinde" cette clé en deux et rend cette vérification plus problématique. La forme normale 3NF sera plus faible que BCNF, mais aura comme avantage de ne jamais nous pousser à scinder une clé en deux. Voici sa définition:

Un schéma de relation avec un ensemble Σ de DF est en *3NF* si la condition suivante est respectée:

> si $X \rightarrow C$ est une DF dans Σ avec $C \notin X$, alors soit C fait partie d'une clé ou X détermine tous les attributs du schéma.

5 Exercices

Les exercices 88–102 et 113–114 du syllabus.