Dépendances (fonctionnelles) dans les bases de données

Les dépendances entre données

Les dépendances sont des propriétés inhérentes du système de données. Elles expriment les différentes façons dont les données sont associées les unes aux autres.

Dépendance Fonctionnelle

Définition (**dépendance fonctionnelle**) : Etant donnée une relation R(X, Y, Z) (où X, Y, Z sont des ensembles de constituants, Z pouvant être vide), on dit qu'il existe une dépendance fonctionnelle entre X et Y notée, **X→Y**, si et seulement si, quelles que soient les X, Y, Z-valeurs (x, y, z) et (x, y', z')

$$||R(x, y, z)|| \text{ et } ||R(x, y', z')|| \Rightarrow y = y'$$

Un schéma d'une relation noté $\underline{R}$ = <U, {X->Y}> définit l'ensemble des relations R construites sur l'ensemble de constituants U qui vérifient toutes la dépendance fonctionnelle X->Y

Exemple :

P (professeur), **H** (heure), **N** (salle), **Y** (classe), **T** (matière)

"le professeur **p** qui enseigne la matière **t** fait cours à l'heure **h** en salle **n** à la classe **y**"

P -> T

H, Y -> N

P, H -> Y

H, N -> P

Ces dépendances fonctionnelles permettent de définir le schéma :

P (professeur), **H** (heure), **N** (salle), **Y** (classe), **T** (matière)

<u>R</u> = <{P, H, N, Y, T}, {P->T; P, H->Y; H, N->P; H, Y-> N}>

Un autre observateur pourrait fournir une liste de dépendances différente, à savoir :

P->Y

H, Y->P

H, N->Y

La question serait alors de savoir si ces deux listes définissent le <u>**même**</u> schéma ou ***deux*** schémas différents?

Décomposition d’une relation

L’opération de décomposition consiste à examiner dans quelles conditions une relation R peut être remplacée par deux relations R1 et R2 de telle sorte que l’espace des constituants de R1 et R2 soit plus petit que celui de R, et que R1 et R2 contiennent les **mêmes informations** que R.

Définition (**décomposition binaire d'une relation**) : Etant donnée une relation R(X, Y, Z), R est décomposable suivant La décomposition ({X, Y}, {X, Z}) si il existe deux relations R1 et R2 telles que :

1. R1 et R2 sont des projections de R :

   R1 = R[X, Y], R2 = R[X, Z]

2. Le produit de R1 et R2 est R : R = R1*R2

La notion de décomposition utilise deux opérations de l'algèbre des relations : la **projection** et le **produit**.

Conditions de décomposition binaire d'une relation

Proposition 1 : pour qu'une relation $R(X, Y, Z)$ soit décomposable suivant la décomposition $(\{X, Y\}, \{X, Z\})$, il faut et il suffit que pour toute X-valeur x de R :

$$R[x, Y, Z] = R[x, Y] \times R[x, Z]$$

$$R[x, Y, Z] = \{(y, z) \mid (x, y, z) \in R\}$$

Proposition 2 : pour qu'une relation $R(X, Y, Z)$ soit décomposable, il faut et il suffit que pour toute X-valeur x
Et toute Y-valeur y de $R[X, Y]$:

$$R[x, y, Z] = R[x, Z]$$

R(LIVRE, COURS, ETUDIANT, PROFESSEUR)

"l'étudiant **e** est en cours **c** qui utilise le livre **l** avec le professeur **p**"

LIVRE	COURS	ETUDIANT	PROFESSEUR
b1	algèbre	pierre	michel
b1	algèbre	jean	michel
b3	analyse	jeanne	michel
b2	algèbre	pierre	michel
b2	algèbre	jean	michel
b3	analyse	jacques	michel

Cette relation R est décomposable par application de la proposition 1 avec R1=R[LIVRE, COURS], R2=R[COURS, ETUDIANT, PROFESSEUR]

R1

LIVRE	COURS
b1	algèbre
b2	algèbre
b3	analyse

R2

COURS	ETUDIANT	PROFESSEUR
algèbre	pierre	michel
algèbre	jean	michel
analyse	jeanne	michel
analyse	jacques	michel

Une autre décomposition possible serait obtenue avec
R3=R[COURS, ETUDIANT] et
R4=R[COURS, LIVRE, PROFESSEUR]

R3

LIVRE	ETUDIANT
algèbre	pierre
algèbre	jean
analyse	jeanne
analyse	jacques

R4

COURS	LIVRE	PROFESSEUR
algèbre	b1	michel
algèbre	b2	michel
analyse	b3	michel

D'après les propositions 1 et 2 les conditions pour qu'une relation soit décomposable exige que l'on vérifie les valeurs contenues dans la relation R. Cette propriété semble inutilisable dans un environnement de base de données car on ne peut pas vérifier cette condition à chaque fois, cela serait trop coûteux en temps.

Proposition 3 : Si une relation R(X, Y, Z) possède une dépendance fonctionnelle X->Y, alors R est décomposable et on peut écrire :

$$R = R[X, Y]*R[X, Z]$$

Définition (décomposition n-aire d'une relation) Etant donnée Une relation R(U), et un ensemble X de parties de U, {X1, …, Xn} tel que l'union des Xi soit égale à U, R est décomposable en n parties s'il existe n relations R1, R2, … Rn Telles que :

(1) les Ri sont des projections de R, c'est à dire

$Ri=R[Xi]$ $i=1, …, n$

(2) le produit des Ri est R, c'est à dire $R = * Ri$ $i=1, …, n$

Problèmes posés par un schéma relationnel

$\underline{R} = \langle\{PR, P, F, AF, N\rangle, \{PR, P \rightarrow F, N; F \rightarrow AF\}\rangle$

PR : numéro de projet
P : numéro de produit
F : numéro de fournisseur
AF : adresse du fournisseur
N : nombre de pièces approvisionnées par un fournisseur pour un projet et pour un produit.

La relation R signifie que le fournisseur **f** d'adresse **af** a approvisionné **n** pièces du produit **p** pour le projet **pr**.

R

PR	P	F	AF	N
pr1	écrou	paul	grenoble	10
pr1	boulon	paul	grenoble	10
pr2	vis	pierre	paris	50
pr3	rondelle	jean	grenoble	100
pr3	écrou	pierre	paris	30
pr2	boulon	paul	grenoble	20

Ce schéma contient un certain nombre d'anomalies qui peuvent être classées en :

(1) anomalie de mise à jour : si on veut mettre à jour l'adresse de paul il faut le mettre dans toutes les entités qui contiennent la valeur Paul.
(2) anomalie d'insertion : on ne peut introduire un fournisseur que s'il approvisionne une pièce pour un projet.
(3) anomalie de suppression : la suppression des entités où se trouve le projet pr3 nous fait perdre l'information concernant le fournisseur jean et en particulier son adresse.

Conformément à la proposition 3 en utilisant la dépendance fonctionnelle F->AF on obtient ainsi deux relations :

R1(PR, P, F, N) et R2(F, AF)

R1

PR	P	F	N
pr1	écrou	paul	10
pr1	boulon	paul	10
pr2	vis	pierre	50
pr3	rondelle	jean	100
pr3	écrou	pierre	30
pr2	boulon	paul	20

R2

F	AF
paul	grenoble
pierre	paris
jean	grenoble

Un autre type de problème : que se passe-t-il lorsqu'on remplace un schéma relationnel par un autre schéma ?

R=<{VOL, AVION, PILOTE}, {VOL->AVION; VOL->PILOTE; AVION->PILOTE}>

"le pilote p effectue le vol v sur l'avion a"

R

VOL	AVION	PILOTE
v1	a1	p1
v2	a1	p1
v3	a2	p2
v4	a3	p2

On peut décomposer la relation R suivant la dépendance VOL->AVION.

Ceci revient à utiliser à la place du schéma initial un nouveau schéma composé de deux schémas :

R1 = <{VOL, PILOTE}, {VOL->PILOTE}>

R2 = <{VOL, AVION}, {VOL->AVION}>

R1

VOL	AVION
v1	a1
v2	a1
v3	a2
v4	a3

R2

VOL	PILOTE
V1	p1
V2	p1
V3	p2
V4	p2

Si l'utilisateur décide de remplacer dans R1 le n-uplet (v4, a3) par (v4, a1), cette modification ne détruit pas la dépendance fonctionnelle VOL->AVION, par contre si on effectue le produit R1*R2 pour régénérer l'information initiale on obtient :

R1*R2

VOL	AVION	PILOTE
v1	**a1**	**p1**
v2	**a1**	**p1**
v3	a2	p2
v4	**a1**	**p2**

Constat !

Le choix d'un schéma relationnel n'est pas un problème simple??

- Comment savoir si une transformation conduit à un meilleur schéma ?
- Comment savoir si le schéma obtenu est équivalent au premier ?

Implications Logiques des Dépendances Fonctionnelles

Définition (consequence logique) : Etant donné un schéma relationnel R = <U, **F**>, nous dirons qu'une dépendance fonctionnelle **f** est une conséquence logique de **F**, notée **F** |= **f** si f est vérifiée dans toutes les relations R de R. Ce qui revient à dire qu'il ne peut exister de relation R de R tel que **f** ne soit pas vérifiée dans R.

Exemple :

R = <{P, H, N, Y, T}, {P->T; P, H->Y; H, N->P; H, Y-> N}>

P, H->N est une conséquence logique de **P, H->Y** et **H, Y-> N**

Définition (*fermeture d'un ensemble de dépendances fonctionnelles*) : La fermeture d'un ensemble F de dépendances fonctionnelles, notée F^+, est l'ensemble des dépendances fonctionnelles qui sont des conséquences logiques de F. Si $F = F^+$ on dit que l'ensemble des dépendances est une famille complète.

Remarque : Selon ces deux définitions il est normal de considérer que les notations $\underline{R} = \langle U, F \rangle$ et $\underline{R} = \langle U, F^+ \rangle$ définissent exactement le même schéma relationnel.

Définition (dérivation) : On dira que f est **dérivée** de F, ce qui sera noté F |- f, s'il existe une séquence f1, …, fn telle que fn = f, et où chaque fi est soit un élément de F ou est dérivée des dépendances précédentes grâce à l'utilisation des règles de dérivation.

Définition (validité et complétude d'un ensemble de règles de dérivation) : Un ensemble de règles de dérivation est **valide** si lorsque f est dérivée de F alors f est une conséquence logique de F, ceci peut s'exprimer formellement par F|-F implique F|=f. Un ensemble de règles de dérivation est **complet** si lorsque f est une conséquence logique de F alors f est dérivable de F, formellement F|=f implique F|-f.

Résultat : pour les dépendances fonctionnelles, il existe un ensemble de règles de dérivation qui est **valide** et **complet**.

Soit R = <U, F> un schéma relationnel. On considère les règles de dérivation :

DF1 (Réflexivité) : Si $X \subseteq Y \subseteq U$ alors $Y \to X$

DF2 (Augmentation) : Si $X \to Y$ et $Z \subseteq W \subseteq U$ alors $X, W \to Y, Z$

DF3 (Transitivité) : Si $X \to Y$ et $Y \to Z$ alors $X \to Z$

A partir des règles DF1, DF2 et DF3 on peut déduire de nouvelles règles qui sont utiles dans les manipulations sur les dépendances fonctionnelles

DF4 (Pseudo-transitivité) : Si X->Y et Y, W->Z alors X, W->Z
DF5 (Union) : Si X->Y et X->Z alors X->Y, Z
DF6 (Décomposition) : Si X->Y et Z⊆Y alors X->Z