

Chapitre II : Optimisation de schémas relationnels

Introduction :

L'exploitation d'une base de données porte essentiellement sur les mises à jour et recherches des informations qui y sont sauvegardées. Son optimisation dépend en grande partie de la bonne conception de son schéma. Ainsi, il ne suffit pas de concevoir un schéma relationnel pour une base de données mais d'en concevoir un qui soit le plus optimisé possible. Dans ce chapitre, nous parlons de l'optimisation des schémas relationnels.

L'optimisation de schémas relationnels est une étape très importante dans la conception d'une base de données. Une base de données doit être optimisée avant d'être implémentée sur un SGBD. L'optimisation d'un schéma relationnel passe par la normalisation de ses relations. Pour normaliser une relation on construit toutes les dépendances fonctionnelles entre ses attributs, puis on fait de sorte que la relation respecte des formes normales telles que la 1FN, la 2FN, la 3FN, la FNBC. Cependant, pour éviter les redondances, chaque relation doit également respecter la 4FN et la 5FN.

I. Les dépendances entre attributs

I. 1. Dépendance fonctionnelle

Il y'a dépendance fonctionnelle entre deux attributs d'une entité si à chaque valeur de l'un correspond une et une seule valeur de l'autre. L'ensemble contenant la/les attribut(s) de gauche est appelé source et celui contenant la/les attribut(s) de droite est appelé but de la DF. Une DF dont la source est composée d'un seul attribut est appelée DF simple. Si la source est composée de plusieurs attributs on parle de DF composée.

Remarque : Si pour une relation R (X, Y, Z) :

- ✓ $X \rightarrow Y$ et $Y \rightarrow Z$ alors on a $X \rightarrow Z$.
- ✓ $X \rightarrow Y$ et $X \rightarrow Z$ Alors on a $X \rightarrow Y, Z$

Graphe minimale d'une relation : Si F est l'ensemble des DF entre les attributs d'une relation R, on appelle **graphe minimum** de R le sous-ensemble de F contenant les DF élémentaires non déductibles dont le but contient un seul attribut et tels que toutes les DF appartenant à F peuvent en être déduites.

I. 2. Dépendance multivaluée

Soient $R (R_1, R_2, R_3, \dots, R_p)$ une relation, X et Y des sous-ensembles de $\{R_1, R_2, R_3, \dots, R_p\}$, on dit que X multi-détermine Y , noté $X \twoheadrightarrow Y$ si à chaque valeur de X correspond plusieurs valeurs de Y indépendamment des autres attributs. Pour chaque valeur prise par X , les valeurs prises par Y sont indépendantes des valeurs prises par Z .

Alors X et Y sont liés par une dépendance multi-valuée si pour tous enregistrements (x, y_1, z_1) et (x, y_2, z_2) dans l'instance de R les enregistrements (x, y_1, z_2) et (x, y_2, z_1) appartiennent aussi à l'instance de R .

Exemple

Soit la relation **Voiture** (Marque, Type, Energie)

Si l'existence des enregistrements :

- t_1 (Peugeot, Berline, Essence) ;
- t_2 (Peugeot, SUV, Diesel)

Implique celle des enregistrements

- c_1 (Peugeot, Berline, Diesel) ;
- c_2 (Peugeot, SUV, Essence)

Alors on a la dépendance multi-valuée $\text{Marque} \twoheadrightarrow \text{Type}$ (Marque multi-détermine Type ou Type est multi-dépendant de Marque)

Remarque :

Si dans la relation $R (X, Y, Z)$, $X \twoheadrightarrow Y$ alors $X \twoheadrightarrow Z$. On note $X \twoheadrightarrow Y \mid Z$.

I. 3. Dépendance de jointure

Soient $R (R_1, R_2, R_3, \dots, R_p)$ une relation, A_1, A_2, \dots, A_m des sous-ensembles de $\{R_1, R_2, R_3, \dots, R_p\}$, il existe une dépendance de jointure $*\{A_1, A_2, \dots, A_m\}$ si :

$$R = \Pi_{A_1}(R) \bowtie \Pi_{A_2}(R) \bowtie \Pi_{A_3}(R) \bowtie \dots \bowtie \Pi_{A_p}(R)$$

Les dépendances multi-valuées sont des cas particuliers de dépendances de jointure. Si une relation $R (X, Y, Z)$ vérifie $X \twoheadrightarrow Y$ et donc $X \twoheadrightarrow Z$, elle satisfait la dépendance de jointure $*\{XY, XZ\}$.

$$R = \Pi_{X,Y}(R) \bowtie \Pi_{X,Z}(R)$$

Exemple

Location (NumL, NomL, PrenomL, NumA, NumP, NomP, PrenomP, Nb_Chambre)

On considère que chaque appartement a un seul propriétaire et qu'on ne garde dans la base que le dernier locataire de chaque appartement.

$A_1 = \{\text{NumL}, \text{NomL}, \text{PrenomL}\}$

$A_2 = \{\text{NumA}, \text{Nb_Chambre}, \text{NumP}, \text{NumL}\}$

$A_3 = \{\text{NumP}, \text{NomP}, \text{PrenomP}\}$

On a une dépendance de jointure $*\{A_1, A_2, A_3\}$

Location = $\Pi_{\text{NumL}, \text{NomL}, \text{PrenomL}}(\text{Location}) \bowtie \Pi_{\text{NumA}, \text{Nb_Chambre}, \text{NumP}, \text{NumL}}(\text{Location}) \bowtie \Pi_{\text{NumP}, \text{NomP}, \text{PrenomP}}(\text{Location})$

II. Formes normales

Il existe plusieurs formes normales permettant d'éviter les redondances, les risques d'incohérence, les anomalies de modification (une occurrence d'une information est modifiée et pas les autres), de suppression (une information importante est perdue lors de la suppression d'un enregistrement), etc.

II. 1. La troisième forme normale (3FN)

Une relation est en troisième forme normale s'il n'y a pas de dépendance fonctionnelle :

- d'une partie d'une clé vers un attribut non clé ;
- entre attributs non clé.

Cependant une relation en 3FN peut comporter des redondances. Pour éviter ces redondances d'autres formes normales sont ajoutées : la Forme Normale de Boyce-Codd (FNBC), la quatrième forme normale, la cinquième forme normale, ...

II. 2. La forme normale de Boyce-Codd (FNBC)

Une relation est en forme normale de Boyce-Codd (FNBC) si et seulement si ses clés candidates sont les uniques sources de DF, c'est-à-dire pour toute DF $A \rightarrow B$ telle que B n'est pas inclus dans A, A est une clé candidate.

Remarque : Toute dépendance fonctionnelle en FNBC est en 3FN.

II. 3. La quatrième forme normale (4FN)

Une relation est en 4FN si les seules dépendances sont celles dans lesquelles une clé multi-détermine un attribut. Toute dépendance en 4FN est également en FNBC.

On dit également qu'une relation est en 4FN, si elle est en FNBC et si elle ne contient pas de Dépendance Multi-valuée.

Les dépendances fonctionnelles sont des cas particuliers de dépendance multi-valuée.

Exemple

Soient **Personne** (CNI, NumPP, Nom, Prenom, Adresse) une relation et F l'ensemble des dépendances fonctionnelles entre ses attributs.

$F = \{ \text{CNI} \rightarrow \text{NumPP}, \text{Nom}, \text{Prenom}, \text{Adresse} ; \text{NumPP} \rightarrow \text{CNI}, \text{Nom}, \text{Prenom}, \text{Adresse} \}$

En plus de ces DF toutes les DF ayant pour source une des sources de ces 3 DF et pour but un sous-ensemble des attributs du but correspondant appartient à F.

$\text{CNI} \rightarrow \text{NumPP} ; \text{CNI} \rightarrow \text{Nom} ; \text{CNI} \rightarrow \text{Prenom} ; \text{CNI} \rightarrow \text{Nom}, \text{Prenom} ; \text{etc.}$

Cette relation est en FNBC car toutes les sources de DF sont des clés candidates

Elle est également en 4FN car il n'y a aucune dépendance multi-valuée entre ces attributs.

II. 4. La cinquième forme normale (5FN)

Une relation est en 5FN si toutes ses dépendances de jointure sont impliquées par les clés.

Exemple

La relation **Location** de la section I.1.3 n'est pas en 5FN. Par contre après l'avoir décomposée on obtient 3 relations en 5FN.

L₁ (NumLoc, NomLoc, PrenomLoc)

L₂ (NumAppart, Nb_Chambre, NumProp, NumLoc)

L₃ (NumProp, NomProp, PrenomProp)

III. Les axiomes d'Armstrong

Soient A, B, C des sous-ensembles quelconques d'attributs d'une relation R et AB l'union de A et B.

Les règles permettant de trouver de nouvelles DF, à partir d'un ensemble de DF, connues sous le nom d'axiomes d'Armstrong sont les suivantes :

- **Réflexivité** : si B est inclus dans A, alors $A \rightarrow B$;
- **Augmentation** : si $A \rightarrow B$, alors $AC \rightarrow BC$;
- **Transitivité** : si $A \rightarrow B$ et $B \rightarrow C$ alors $A \rightarrow C$.

A partir des axiomes d'Armstrong, nous pouvons déduire :

- **Union** : $X \rightarrow Y$ et $X \rightarrow Z$, alors $X \rightarrow YZ$
- **Pseudo-Transitivité** : $X \rightarrow Y$ et $YW \rightarrow Z$, alors $XW \rightarrow Z$
- **Décomposition** : $X \rightarrow Y$ et Z inclus dans Y, alors $X \rightarrow Z$

IV. La Fermeture d'un ensemble d'attributs

La fermeture d'un ensemble d'attributs X (notée X^+) d'une relation représente l'ensemble des attributs de cette relation qui peuvent être déduits de X à partir d'une famille de Dépendances Fonctionnelles F en appliquant les axiomes d'Armstrong.

Y est inclus dans X^+ si et seulement si $X \rightarrow Y$.

La fermeture d'un ensemble d'attributs X sous F contient les attributs de X et tous les attributs dans le but d'une DF de F dont la source contient des attributs inclus dans X^+ .

V. La Couverture minimale d'un ensemble de DF

Si on élimine les dépendances fonctionnelles redondantes de F, c'est-à-dire celles qui peuvent être déduites d'un ensemble minimale F' inclus dans F, on obtient la couverture minimale de F.

VI. La décomposition d'un schéma relationnel

La décomposition d'un schéma relationnel a pour objectif de remplacer un schéma R par un ensemble de schéma $T = \{R_1, R_2, R_3, \dots, R_n\}$ tels que $R = R_1 \cup R_2 \cup R_3 \cup \dots \cup R_n$. Les Schémas R_i ($1 \leq i \leq n$) n'étant pas obligatoirement disjoints. Elle permet d'éliminer les redondances et les anomalies de mise à jour. Elle permet donc d'optimiser le schéma d'une relation et de faciliter les tâches d'administration.

VII. Cas pratique

Examen (Numero_Cand, Nom_Cand, Prenom_Cand, Age_Cand, Type_Cand, Nom_Etab, Adresse_Etab, Telephone_Etab, Nom_Mat, Coefficient_Mat, Type_Mat, Nom_Ville, NbEtab_Ville, NbCandidat_Ville, Code_Correct, Nom_Correct, Prenom_Correct, Grade_Correct, Age_Correct, Numero_Jury, Annee_Jury, Serie_Jury, Session_Jury, Note)

On considère que :

- ✓ Un candidat est issu d'un établissement et un seul et compose dans un jury et un seul ;
- ✓ Un correcteur est issu d'un établissement et un seul et corrige dans un jury et seul ;
- ✓ Un correcteur est spécialiste dans une et une seule matière ;
- ✓ Un établissement loge 0 ou plusieurs jurys ;
- ✓ Un jury est logé dans et un seul établissement ;
- ✓ Deux établissements peuvent porter le même nom dans deux villes différentes ;
- ✓ Deux candidats ne peuvent avoir le même numéro de téléphone qui est obligatoire.

Travail à faire

1. Donnez l'ensemble F des DF entre les attributs de cette relation ;
2. Donnez le graphe minimal des DF de la relation Location ;
3. Donnez la couverture minimale de l'ensemble F ;
4. Décomposez la relation en relations respectant la 5FN ;
5. Démontrez $\{\text{Code_Correct} \rightarrow \text{Serie_Jury}\}$ à partir des axiomes d'Armstrong.