

Modèle Relationnel

Département d'informatique

Présentation

- Les données sont représentées par des tables, sans préjuger de la façon dont les informations sont stockées dans la machine.
- Les tables constituent donc la *structure logique* du modèle relationnel.
- Au niveau physique, le système est libre d'utiliser n'importe quelle technique de stockage (fichiers séquentiels, indexage, adressage dispersé, séries de pointeurs, compression, ...) dès lors qu'il est possible de relier ces structures à des tables au niveau logique.
- Les tables ne représentent donc qu'une abstraction de l'enregistrement physique des données en mémoire.
- Le succès du modèle relationnel auprès des chercheurs, concepteurs et utilisateurs est dû à la puissance et à la simplicité de ses concepts.
- Contrairement à certains autres modèles, il repose sur des bases théoriques solides, notamment la théorie des ensembles et la logique des prédicats du premier ordre.

Présentation

Les objectifs du modèle relationnel sont :

- Proposer des schémas de données faciles à utiliser ;
- Améliorer l'indépendance logique et physique;
- Mettre à la disposition des utilisateurs des langages de haut niveau ;
- Optimiser les accès à la base de données ;
- Améliorer l'intégrité et la confidentialité ;
- Fournir une approche méthodologique dans la construction des schémas.

Présentation

De façon informelle, on peut définir le modèle relationnel de la manière suivante :

- Les données sont organisées sous forme de tables à deux dimensions, encore appelées relations, dont les lignes sont appelées n-uplet ou *tuple* en anglais;
- Les données sont manipulées par des opérateurs de l'algèbre relationnelle ;
- L'état cohérent de la base est défini par un ensemble de contraintes d'intégrité;
- Au modèle relationnel est associée la théorie de la normalisation des relations qui permet de se débarrasser des incohérences au moment de la conception d'une base de données relationnelle.

Éléments du modèle relationnel

1) Attribut

Un attribut est un identificateur (un nom) décrivant une information stockée dans une base.

Exemples : l'âge d'une personne, le nom d'une personne, le numéro de sécurité sociale.

2) Domaine

- Le domaine d'un attribut est l'ensemble, fini ou infini, de ses valeurs possibles.

Exemple:

- l'attribut numéro de sécurité sociale a pour domaine l'ensemble des combinaisons de quinze chiffres
- l'attribut nom a pour domaine l'ensemble des combinaisons de lettres

Éléments du modèle relationnel

3) Relation

- Une relation est un sous-ensemble du produit cartésien de n domaines d'attributs ($n > 0$).
- Une relation est représentée sous la forme d'un tableau à deux dimensions dans lequel les n attributs correspondent aux titres des n colonnes.

4) Schéma de relation

Un schéma de relation précise le nom de la relation ainsi que la liste des attributs avec leurs domaines.

Éléments du modèle relationnel

5) Degré

Le degré d'une relation est son nombre d'attributs.

6) Occurrences ou n-uplets ou tuples

- Une occurrence, ou n-uplet, ou tuple, est un élément de l'ensemble figuré par une relation.
- Une occurrence est une ligne du tableau qui représente la relation.

7) Cardinalité

La cardinalité d'une relation est son nombre d'occurrences.

Éléments du modèle relationnel

8) Clé candidate

- Une clé candidate d'une relation est un ensemble minimal des attributs de la relation dont les valeurs identifient à coup sûr une occurrence.
- La valeur d'une clé candidate est donc distincte pour toutes les tuples de la relation.
- La notion de clé candidate est essentielle dans le modèle relationnel. Ainsi, il ne peut jamais y avoir deux tuples identiques au sein d'une relation.
- Les clés candidates d'une relation n'ont pas forcément le même nombre d'attributs. Une clé candidate peut être formée d'un attribut arbitraire, utilisé à cette seule fin.

Éléments du modèle relationnel

9) Clé primaire

- La clé primaire d'une relation est une de ses clés candidates.
- Pour signaler la clé primaire, ses attributs sont généralement soulignés.

10) Clé étrangère

Une clé étrangère dans une relation est formée d'un ou plusieurs attributs qui constituent une clé primaire dans une autre relation.

11) Schéma relationnel

Un schéma relationnel est constitué par l'ensemble des schémas de relation.

12) Base de données relationnelle

Une base de données relationnelle est constituée par l'ensemble des n-uplets des différentes relations du schéma relationnel.

Éléments du modèle relationnel

Matricule	Nom	Prénom	Sexe
14/36039425	BERKANI	BADER EDDINE	M
14/36038714	BOUCHEDAK	WAHIBA	F
14/36046796	DJAHEL	BILLEL	M

Etudiant(Matricule: AN, Nom: A, Prénom: A, Sexe:C)

AN= alphanumérique

A = alphabétique

C= caractère

Passage du modèle entités-associations au modèle relationnel

- Pour traduire un schéma du modèle entités-associations vers le modèle relationnel, on peut appliquer des règles de passage.
- La normalisation devrait toujours être effectuée avant le passage au modèle relationnel. Dans les faits, elle est parfois faite *a posteriori*, ce qui impose toujours une surcharge de travail importante.

Règles de passage:

- 1) Chaque type-entité donne naissance à une relation.
- 2) Chaque attribut de ce type-entité devient un attribut de la relation.
- 3) L'identifiant est conservé en tant que clé de la relation.
- 4) Chaque type-association dont aucune patte n'a pour cardinalité maximale 1 donne naissance à une relation.

Passage du modèle entités-associations au modèle relationnel

- 5) Chaque attribut de ce type-association devient un attribut de la relation.
- 6) L'identifiant, s'il est précisé, est conservé en tant que clé de la relation, sinon cette clé est formée par la concaténation des identifiants des type-entités qui interviennent dans le type-association.
- 7) Un type-association dont au moins une patte a une cardinalité maximale à 1 ne devient pas une relation. Il décrit en effet une dépendance fonctionnelle. La relation correspondant au type-entité dont la patte vers le type-association a une cardinalité maximale valant 1, se voit simplement ajouter comme attribut l'identifiant de l'autre type-entité.

Normalisation

- Les formes normales sont différents stades de qualité qui permettent d'éviter la redondance dans les bases de données relationnelles afin d'éviter ou de limiter :
 - les pertes de données,
 - les incohérences au sein des données,
 - l'effondrement des performances des traitements
- Le processus de normalisation consiste à remplacer une relation donnée par certaines *projections* afin que la *jointure* de ces projections permette de retrouver la relation initiale
- Le processus est réversible (*i.e.* sans perte d'information).

Normalisation

- Il existe une hiérarchie dans les règles de normalisation : une relation en 5^{ème} forme normale est forcément en 4^{ème} forme normale, une relation en 4^{ème} forme normale est forcément en forme normale de Boyce-Codd, etc.
- Il existe des méthodes systématiques pour normaliser une relation dans chacune des formes normales.
- La normalisation peut être effectuée pendant la phase de conception sur le modèle entités-associations.
- Dans le cas où la normalisation est faite en amont(modèle entité-association), lors de la conception, il n'est pas nécessaire de la recommencer sur le modèle relationnel.

Normalisation

- La normalisation est effectuée en s'appuyant sur les notions de dépendance fonctionnelle, dépendance multivaluée et dépendance de jointure.
- Il est important de prendre conscience que la dépendance fonctionnelle, la dépendance multivaluée et la dépendance de jointure sont des notions sémantiques. Elles tirent leurs origines dans les contraintes du monde réel.
- Comme ces contraintes participent à la sémantique de la situation, elles doivent avoir une manifestation dans la base de données.
- Les dépendances doivent être spécifiées dans la définition de la base de données afin que le SGBD puisse les appliquer.

Normalisation

- Les concepts de normalisation fournissent en fait un moyen indirect de déclarer ces dépendances.
- La normalisation d'une base de données est une manifestation observable des dépendances observées dans le monde réel.
- La dépendance fonctionnelle permet de définir les premières formes normales jusqu'à la forme normale de Boyce-Codd (1FN, 2FN, 3FN et BCNF).
- La dépendance multivaluée permet de définir la quatrième forme normale (4FN)
- La dépendance de jointure permet de définir la cinquième forme normale (5FN).

Dépendance fonctionnelle

Dépendance fonctionnelle (DF)

Soit $R(A_1, A_2, \dots, A_n)$ un schéma de relation, et X et Y des sous-ensembles de $A_1, A_2, \dots, A_n$. On dit que X **détermine** Y ou que Y **dépend fonctionnellement** de X si, et seulement si, des valeurs identiques de X impliquent des valeurs identiques de Y .

On le note : $X \rightarrow Y$.

- Il existe une dépendance fonctionnelle entre un ensemble d'attributs X et un ensemble d'attributs Y , que l'on note $X \rightarrow Y$, si connaissant une occurrence de X on ne peut lui associer qu'une seule occurrence de Y .
- Il est essentiel de noter qu'une dépendance fonctionnelle est une assertion sur toutes les valeurs possibles et non sur les valeurs actuelles : elle caractérise une intention et non une extension de la relation.

Dépendance fonctionnelle

Dépendance fonctionnelle élémentaire

Une dépendance fonctionnelle élémentaire est une dépendance fonctionnelle de la forme $X \rightarrow A$, où A est un attribut unique n'appartenant pas à X et où il n'existe pas X' inclus au sens strict dans X (i.e. $X' \subset X$) tel que $X' \rightarrow A$.

- Une dépendance fonctionnelle est élémentaire si la cible est un attribut unique et si la source ne comporte pas d'attributs superflus. La question sur l'élémentarité d'une dépendance fonctionnelle ne doit donc se poser que lorsque la partie gauche de la dépendance fonctionnelle comporte plusieurs attributs.

Dépendance fonctionnelle

Dépendance fonctionnelle directe

Une dépendance fonctionnelle $X \rightarrow A$ est une dépendance fonctionnelle directe s'il n'existe aucun attribut B tel que l'on puisse avoir $X \rightarrow B$ et $B \rightarrow A$.

- En d'autres termes, cela signifie que la dépendance entre X et A ne peut pas être obtenue par transitivité.

Première forme normale (1FN)

Une relation est en 1FN si, et seulement si, tout attribut contient une valeur atomique (non multiples, non composées).

Exemple:

le pseudo schéma de relation Etudiant(Num, nom, prénom, adresse, module) n'est pas en première forme normale.

Il faut le décomposer en :

- Etudiant(Num-E, nom, prénom, adresse)
- Etudie(Num-E, Num-M)
- Module(Num-M, libellée)

Première forme normale (1FN)

Remarques:

- La première forme normale impose que chaque ligne d'une relation ait une seule valeur pour chaque colonne (i.e. attribut), ce qui est justement la définition d'une table. Donc, une table est nécessairement en première forme normale au sens du modèle relationnel.
- Il faut noter que le modèle relationnel peut être étendu de manière à permettre des colonnes à valeur complexe. On parle alors de modèle relationnel étendu.

Deuxième forme normale (2FN)

Une relation est en deuxième forme normale si, et seulement si, elle est en première forme normale et si toutes les dépendances fonctionnelles entre la clé et les autres attributs sont élémentaires.

Une relation est en deuxième forme normale si, et seulement si, elle est en première forme normale et si tout attribut n'appartenant pas à la clé ne dépend pas que d'une partie de la clé.

- Une relation peut être en deuxième forme normale par rapport à une de ses clés candidates et ne pas l'être par rapport à une autre. Une relation avec une clé primaire réduite à un seul attribut est, par définition, forcément en deuxième forme normale.

Deuxième forme normale (2FN)

Exemple:

Commande(Num-C, Num-Cl, Nom-P, Type, Quantité, Prix).

Cette relation indique qu'un client (identifié par Num-Cl) a passé une commande (identifiée par Num-C) de produit. Elle est bien en première forme normale. Par contre, les attributs Nom-P, Type, Quantité, Prix ne dépendent que de Num-C, et pas de Num-Cl. Cette relation n'est donc pas en deuxième forme normale.

Solution:

Une solution simple pour la normaliser est de la remplacer par :

Commande(Num-C, Num-Cl, Nom-P, Type, Quantité, Prix).

Troisième forme normale (3FN)

Une relation est en troisième forme normale si, et seulement si, elle est en deuxième forme normale et si toutes les dépendances fonctionnelles entre la clé et les autres attributs sont directes.

Une relation est en troisième forme normale si, et seulement si, elle est en deuxième forme normale et si tout attribut n'appartenant pas à la clef ne dépend pas d'un attribut non-clé.

- Une relation peut être en troisième forme normale par rapport à une de ses clés candidates et ne pas l'être par rapport à une autre. Une relation en deuxième forme normale avec au plus un attribut qui n'appartient pas à la clé primaire est, par définition, forcément en troisième forme normale.

Troisième forme normale (3FN)

Exemple:

Commande(Num-C, Num-Cl, Nom-P, Type, Quantité, Prix).

Comme nous l'avons vu plus haut, cette relation est bien en deuxième forme normale. Par contre, les attributs Type et *Prix* dépendent de l'attribut Nom-P. La relation n'est donc pas en troisième forme normale.

Solution:

Pour la normaliser, il faut la décomposer de la manière suivante :

- Commande(Num-C, Num-Cl, Num-P, Quantité).
- Produit(Num-P, Nom-P, Type, Prix)

Remarques

- Nous avons dit qu'il est souvent préférable de choisir un identifiant arbitraire de type entier. Cette pratique semble aller à l'encontre de la troisième forme normale.
- La notion de dépendance fonctionnelle est donc une question d'interprétation faisant appel à la connaissance du système modélisé et au bon sens.
- Il faut enfin noter que la normalisation n'est pas une fin en soi et qu'elle ne doit pas nécessairement être systématiquement appliquée

Forme normale de BOYCE-CODD

Une relation est en forme normale de BOYCE-CODD (BCNF) si, et seulement si, elle est en troisième forme normale et si les seules dépendances fonctionnelles élémentaires sont celles dans lesquelles une clé détermine un attribut.

Cette forme normale permet de renforcer certaines lacunes de la troisième forme normale.

Exemple:

- Module(Nom-M)
- Classe(Num-C)
- Enseignant(Nom-E)
- Cours(Nom-E, Num-C, Nom-M)

Forme normale de BOYCE-CODD

Supposons, de plus, qu'un module n'est enseignée qu'une seule fois dans une classe et que par un seul enseignant, et qu'un enseignant n'enseigne qu'une seule matière. Chacune des relations respecte bien la troisième forme normale. Cependant, dans la relation Cours, nous avons les dépendances fonctionnelles élémentaires suivantes :

- $\text{Nom-M}, \text{Num-C} \rightarrow \text{Nom-E}$
- $\text{Nom-E} \rightarrow \text{Nom-M}$

Il existe donc des dépendances fonctionnelles élémentaires dont la source n'est pas la clé de la relation.

Forme normale de BOYCE-CODD

Le non respect de la forme normale de BOYCE-CODD entraîne une redondance. pour chaque Nom-E identifiant un enseignant, il faut répéter le Nom-M identifiant le Module qu'il enseigne.

Pour normaliser la relation Cours, il faut la décomposer pour aboutir au schéma relationnel suivant :

- Module(Nom-M)
- Classe(Num-C)
- Enseignant(Nom-E, Nom-M)
- Cours(Nom-E, Num-C)

Forme normale de BOYCE-CODD

- Dans la pratique, la plupart des problèmes de conception peuvent être résolus en appliquant les concepts de troisième forme normale et de forme normale de BOYCE-CODD.
- Les quatrième et cinquième formes normales traitent encore d'autres cas de redondance, mais qui ne sont pas expliqués par des dépendances fonctionnelles.

Dépendance multivaluée (DM)

Soit $R(A_1, A_2, \dots, A_n)$ un schéma de relation contenant n propriétés, soit X , Y et Z des sous-ensembles de $A_1, A_2, \dots, A_n$ et soit X_i , Y_i et Z_i des instances de ces sous-ensembles.

Il existe une dépendance multivaluée (DM) entre les ensembles de propriétés X , Y lorsque :

$$(X_1, Y_1, Z_1) \in R \text{ et } (X_1, Y_2, Z_2) \in R \Rightarrow (X_1, Y_1, Z_2) \in R \text{ et } (X_1, Y_2, Z_1) \in R$$

- On la note $X \twoheadrightarrow Y$, ce qui se lit X multidétermine Y .
- Remarque : $X \twoheadrightarrow Y \Rightarrow X \twoheadrightarrow A_i - (X \cup Y)$.

Dépendance multivaluée (DM)

Comme illustration, supposons une situation où un Médecin dans un hôpital est qualifié pour effectuer un certain type d'interventions chirurgicales sur certains patients. Cette situation est modélisée par le schéma relationnel suivant :

- Médecin(Nom-M)
- Intervention(Type-I)
- Patients(Age)
- Intervenir(Nom-M, Type-I, Age)

Dépendance multivaluée (DM)

Supposons maintenant que le Médecin qui effectue un ensemble de types d'interventions pour un ensemble de Patients, est capable d'effectuer chacun de ces types d'interventions sur des patients de différents Age.

Dans ce cas, il existe des dépendances multivaluées dans la relation *Intervenir* : $Nom-M \twoheadrightarrow Type-Intervention$ et $Nom-M \twoheadrightarrow Age$.

Quatrième forme normale (4FN)

Une relation est en quatrième forme normale (4FN) si, et seulement si, elle est en forme normale de BOYCE-CODD et si elle ne possède pas de dépendance multivaluée ou si, $X \twoheadrightarrow Y$ étant la dépendance multivaluée, il existe une propriété A telle que $X \rightarrow A$.

Dans la section précédente, nous avons présenté un schéma relationnel qui n'était pas en quatrième forme normale en raison du schéma de relation Intervenir.

Quatrième forme normale (4FN)

Pour normaliser la relation Intervenir, il faut la décomposer pour aboutir au schéma relationnel suivant :

- Médecin(Nom-M)
- Intervention(Type-I)
- Patients(Age)
- Est-Qualifiée(Nom-M, Type-I)
- Intervenir(Nom-M, Age)

Dépendance de jointure (DJ)

Jusqu'ici, nous avons pu résoudre une redondance dans une relation en la remplaçant par deux de ses projections.

Il existe cependant des relations qui ne peuvent pas être décomposées sans perte d'information en deux projections, mais qui peuvent l'être en trois ou plus (ces cas sont assez rares en pratique).

C'est ce que permet la normalisation en cinquième forme normale.

Les dépendances de jointures font appel à des notions (*projection* et *jointure*) qui ont été définies dans le chapitre « Algèbre relationnelle ».

Dépendance de jointure (DJ)

Dépendance de jointure (DJ)

Soient $X_1, X_2, \dots, X_n$ des sous-ensembles d'un schéma de relation R .

Il y a une dépendance de jointure, notée $*\{X_1, X_2, \dots, X_n\}$ dans la relation R , si :

$$R = \Pi_{(X_1)} R \bowtie \Pi_{(X_2)} R \bowtie \dots \bowtie \Pi_{(X_n)} R$$

Dépendance de jointure triviale

Une dépendance de jointure est triviale si une des parties, X_i , est l'ensemble de toutes les attributs de R .

Cinquième forme normale (5FN)

Une relation R est en cinquième forme normale (5FN) si, pour toute dépendance de jointure non triviale $\{X_1, X_2, \dots, X_n\}$ dans R , chacun des X_i contient une clé candidate de R .

- En d'autres termes, les seules décompositions qui préservent le contenu sont celles où chacune des tables de la décomposition contient une clé candidate de la table.
- Cette forme normale est finale vis-à-vis de la projection et de la jointure : elle garantit qu'une relation en cinquième forme normale ne contient aucune anomalie pouvant être supprimée en effectuant des projections (i.e. des décompositions).

Conclusions

- Il existe d'autres formes normales:
 - La forme normale domaine-clé (FNDC)
 - La forme normale de restriction-union
 - La sixième forme normale (6NF).
- Bien que l'objectif de la normalisation soit d'amener le concepteur à obtenir des relations en forme normale *finale* (i.e. en cinquième forme normale), cet objectif ne doit pas être interprété comme une loi.
- Il peut exister, très occasionnellement, de bonnes raisons pour passer outre les principes de la normalisation. De plus, un schéma en cinquième forme normale n'est pas nécessairement un schéma pleinement satisfaisant.
- D'autres facteurs sont à considérer dans le processus de conception d'une base de données et l'expérience et l'intuition jouent un rôle important.