Introduction à la normalisation relationnelle

La théorie de la normalisation

permet de définir une méthode de conception de "bonnes" tables, c'est-à-dire sans redondances et sans perte d'information

Exemples :

COURS	code	trimestre	libellé	professeur
	SID 130	**T2**	**Bases de données**	**Akoka**
	SID 230	**T3**	**B. D. A.**	**Akoka**
	SID 132	**T2**	**Micro-informatique**	**Wattiau**
	SID 132	**T3**	**Micro-informatique**	**Wattiau**

Redondances : on dit 2 fois que le cours SID 132 a pour libellé Micro-informatique et est donné par WATTIAU

COURS	code	trimestre	professeur
	SID 130	**T2**	**Akoka**
	SID 230	**T3**	**Akoka**
	SID 132	**T2**	**Wattiau**
	SID 132	**T3**	**Wattiau**

et

COURS	code	libellé
	SID 130	**Bases de données**
	SID 230	**B. D. A.**
	SID 132	**Micro-informatique**
	SID 132	**Micro-informatique**

Pourquoi la normalisation ?

- pour éliminer les redondances
- pour mieux comprendre les relations sémantiques entre les données
- pour éviter les incohérences de mise à jour
- pour éviter, autant que possible, les valeurs nulles

Exemple :

Relation COURS

Nomprof	Ville	Département	Nometud	Age	Nomcours	Note
Dupont	**Lille**	**59**	**Alfred**	**22**	**Math**	**12**
Dupont	**Lille**	**59**	**Arthur**	**25**	**Math**	**05**
Martin	**Arras**	**62**	**Alfred**	**22**	**Anglais**	**18**
Martin	**Arras**	**62**	**Pierre**	**23**	**Anglais**	**11**
Dupont	**Lille**	**59**	**Pierre**	**23**	**Anglais**	**13**
Charles	**Lille**	**59**	**Pierre**	**23**	**Anglais**	**12**

- *des données redondantes :* Dupont à Lille (59)
- *des risques d'incohérence :* déménagement de Dupont à Marseille
- *des valeurs nulles :* représenter un prof qui n'a pas d'étudiant

↙ *entraînent des anomalies à l'interrogation*

⇓ *Problème du choix des relations*

Comment normaliser un schéma relationnel ?

- Approche par décomposition :
  - on part d’une table contenant tous les attributs
  - et on décompose jusqu’à ce qu’il n’y ait plus de redondances
- Approche par synthèse :
  - à partir de l’ensemble des attributs
  - et des dépendances fonctionnelles
  - on constitue les tables

Jointure de deux relations

Soient R (A1, A2, …, An) et S (B1, B2, …, Bp) deux relations

La jointure de R et S est la relation T qui a pour attributs l'union des attributs de R et S et pour tuples l'ensemble des tuples construits à partir de R et S sur les valeurs identiques des attributs communs

0n note T = R $\bowtie$ S

Exemple :

R	Nom	Salaire
	Dupond	10000
	Durand	5400
	Martin	12000

S	Nom	Adresse
	Dupond	Issy
	Durand	Sète
	Martin	Sète

T	Nom	Salaire	Adresse
	Dupond	10000	Issy
	Durand	5400	Sète
	Martin	12000	Sète

Décomposition sans perte

La décomposition de R en R1, R2, …, Rn est sans perte si, pour toute extension de R, on a :

$$R_1 \bowtie R_2 \bowtie \ldots R_n = R$$

Dépendance fonctionnelle

- Il y a une dépendance fonctionnelle (df), notée A1, A2, …, An -> B1, B2, …, Bp dans la table T si, à tout moment, lorsque deux tuples ont les mêmes valeurs pour les colonnes A1, A2, …, An, ils ont alors nécessairement les mêmes valeurs pour les colonnes B1, B2, …, Bp

Axiomes d’Armstrong

Propriétés des dépendances fonctionnelles

1. Réflexivité   X -> Y
2. Augmentation   X ->Y => XZ -> Y
3. Transitivité   X -> Y et Y -> Z => X -> Z

Conséquences

4. Union   X -> Y et X -> Z => X -> YZ
5. Pseudo-transitivité   X -> Y et WY -> Z => WX -> Z
6. Décomposition   X -> Y et Z $\subseteq$ Y => X -> Z

Dépendance fonctionnelle élémentaire

- Dépendance fonctionnelle élémentaire

X -> A telle que

1) A est un attribut unique
2) A $\not\subseteq$ X
3) Il n'existe pas X' $\subseteq$ X tel que X' -> A

- Dans la recherche des dfs, on peut se limiter sans restriction aux dfs élémentaires

Exemple :

Table COURS (NOMPROF, VILLE, DEPARTEMENT, NOMETUDIANT, AGE, COURS, NOTE)

Dépendances fonctionnelles valides :

NOMPROF ➲ VILLE
NOMPROF NOMETUDIANT ➲ VILLE
NOMETUDIANT ➲ AGE

Dépendances fonctionnelles invalides :

AGE ➲ NOMETUD

Dépendances fonctionnelles élémentaires

NOMPROF ➲ VILLE
VILLE ➲ DEPARTEMENT
NOMPROF ➲ DEPARTEMENT
NOMETUD ➲ AGE
NOMETUD NOMCOURS ➲ NOTE
NOMCOURS ➲ NOMPROF

Fermeture transitive d'un ensemble de dépendances fonctionnelles

= ensemble des dfs enrichi de toutes les dfs déduites par transitivité ou pseudo-transitivité

(on se limite, sans perte de généralité, aux dfs élémentaires)

On note F^+ la fermeture transitive de F

Exemple : NOMPROF VILLE DEPARTEMENT NOMETUD AGE NOMCOURS NOTE

La fermeture transitive de :

NOMPROF ➲ VILLE
VILLE ➲ DEPARTEMENT
NOMETUDIANT ➲ AGE
NOMETUDIANT COURS ➲ NOTE
COURS ➲ NOMPROF

contient en plus :

NOMPROF ➲ DEPARTEMENT
NOMCOURS ➲ VILLE
NOMCOURS ➲ DEPARTEMENT

Équivalence de deux ensembles de dépendances fonctionnelles élémentaires :

Deux ensembles de DF élémentaires sont équivalents s'ils ont la même fermeture transitive.

Couverture minimale d'un ensemble de DFs :

Ensemble C de DF élémentaires tels que :

1. C^+ contient toutes les DF élémentaires de l'ensemble d'attributs considéré
2. $\forall \; f \in C$, $C - \{f\}$ n'est pas équivalent à C

Exemple : Relation COURS
(NOMPROF, VILLE, DEPARTEMENT, NOMETUDIANT, AGE, COURS, NOTE)

$$\left.\begin{array}{l}
\textbf{1. NOMPROF} \rightarrow \textbf{VILLE} \\
\textbf{2. VILLE} \rightarrow \textbf{DEPARTEMENT} \\
\textbf{3. NOMPROF} \rightarrow \textbf{DEPARTEMENT} \\
\textbf{4. NOMETUDIANT} \rightarrow \textbf{AGE} \\
\textbf{5. NOMETUDIANT, COURS} \rightarrow \textbf{NOTE} \\
\textbf{6. COURS} \rightarrow \textbf{NOMPROF}
\end{array}\right\} \textbf{n'est pas minimal car 3 est redondante}$$

$\{1, 2, 4, 5, 6\}$ est une couverture minimale

$\{1, 3, 4, 5\}$ n'est pas une couverture

Clé de relation
Nouvelle définition

Une clé est un sous-ensemble minimum d'attributs qui détermine tous les autres.

X clé de R(A1, A2, …, An) tel que :

1. X ➲ A1A2 …An
2. il n'existe pas Y ⊆ X tel que Y ➲ A1A2 …An

Exemple : Relation COURS
(NOMETUD, NOMCOURS)

Remarques : il peut y avoir plusieurs clés

Première forme normale (1FN)

Une relation est en 1ère Forme Normale (1FN) si tout attribut contient une valeur atomique

Contre-exemples :

1. PERSONNE (NOM, PRENOMS)
   Mise en 1FN : PERSONNE (NOM, PRENOM1, PRENOM2)

2. PERSONNE (NOM, PRENOM, ADRESSE)
   Mise en 1FN : PERSONNE (NOM, PRENOM, N°RUE, RUE, CODEPOSTAL, VILLE)

Une relation est en deuxième forme normale (2FN) si :

1. elle est en 1 FN
2. tout attribut n'appartenant pas à une clé ne dépend pas que d'une partie de cette clé

Exemple : COURS (NOMPROF, VILLE, DEPARTEMENT, NOMETUD, AGE, NOMCOURS, NOTE)

1 seule clé (NOMETUD, NOMCOURS)

Les DF : *1.* NOMPROF ➲ VILLE

2. VILLE ➲ DEPARTEMENT

3. NOMETUD ➲ AGE

4. NOMETUD, NOMCOURS ➲ NOTE

5. NOMCOURS ➲ NOMPROF

Problème pour les attributs NOMPROF, VILLE, DEPARTEMENT et AGE

COURS (<u>NOMETUD NOMCOURS</u> NOTE)
R1 (<u>NOMCOURS</u> NOMPROF VILLE DEPARTEMENT)
R2 (<u>NOMETUD</u> AGE)
} **sont en 2 FN**

Une relation est en troisième forme normale (3FN) si :

1. elle est en 2 FN
2. tout attribut n'appartenant pas à une clé ne dépend pas d'un attribut non clé

Exemple : COURS (NOMETUD NOMCOURS NOTE)
R1 (NOMCOURS NOMPROF VILLE DEPARTEMENT)
R2 (NOMETUD AGE)

R1 n'est pas en 3 FN
car NOMPROF ➲ VILLE et VILLE ➲ DEPARTEMENT
R1 (NOMCOURS NOMPROF)
R3 (NOMPROF VILLE DEPARTEMENT)

R3 n'est pas en 3 FN car VILLE ➲ DEPARTEMENT
R3 (NOMPROF VILLE)
R4 (VILLE DEPARTEMENT)

Algorithme de mise sous 3 FN

0FN ➲ 1FN : mise sous forme atomique des attributs

1FN ➲ 2FN : pour chaque partie X de clé déterminant des attributs non clés Y1, …, Yn

1. on crée une relation supplémentaire avec X pour clé et Y1, …, Yn comme attributs non clés
2. on retire Y1, …, Yn de la relation initiale

2FN ➲ 3FN : pour chaque attribut non clé Y déterminant des attributs non clés Z1, …, Zn

1. on crée une relation R' supplémentaire avec Y comme clé et Z1, …, Zn comme attributs non clés
2. on retire Z1, …, Zn de la relation initiale

R' n'est pas nécessairement en 3 FN. Si c'est le cas, réitérer le processus sur R'.

- Dans une décomposition d'une relation en plusieurs autres, on dit que la décomposition préserve une dépendance fonctionnelle s'il reste, après décomposition, une relation contenant tous les attributs de la DF

- *Propriété :*

  Toute relation a au moins une décomposition en 3 FN qui :

  1. préserve une couverture minimale de DF
  2. est sans perte

CONCLUSION

- La normalisation permet de :
  - Construire des tables sans redondance
  - Vérifier la bonne conception des tables issues de la modélisation conceptuelle
  - Restructurer une base existante