

Normalisation - Décomposition

La **normalisation** d'une relation consiste à la **décomposer** en un ensemble de relations telles qu'aucune des relations obtenues ne possède les anomalies de redondance, de mise à jour et de suppression.

Pour être parfaites, les relations doivent respecter certaines règles. Cet ensemble de règles se nomme : **les formes normales**.

Cette théorie a été élaborée par E.F. Codd en 1970. Son objectif est d'éviter les anomalies dans les bases de données relationnelles :

- Problèmes de mise à jour.
- Suppression des redondances d'informations.
- Simplification de certaines contraintes d'intégrité.

Pour parfaire une base de données relationnelle, il est nécessaire de connaître les trois premières formes normales et la forme normale dite Boyce-Codd ; les suivantes ne sont que des extensions peu usitées.

1. 1FN - Première forme normale

Une relation est en première forme normale si :

- Tous les attributs ne contiennent qu'une seule valeur atomique (non divisible).
- Les attributs ne contiennent pas de valeurs répétitives.

Exemple :

Clients (NumCli, Nom, Prénom, Adresse, Téléphone)

Cette relation n'est pas en première forme normale, car **Adresse n'est pas atomique**. En effet voici une représentation d'un fichier ainsi décrit :

NumCli	Nom	Prénom	Adresse	Téléphone
1	Baptiste	Jean-Luc	25, rue de la forêt 12000 Rodez	0565420000
2	Auguy	Jean	Impasse des lys 15000 Aurillac	0471670000
3	Rascalou	André	2, rue droite 12000 Rodez	0565450000

Cette représentation si elle était mise en pratique générerait un accès aux données plus lent. Le simple fait de vouloir extraire les habitants d'une ville précise devra mettre en œuvre des procédures d'extraction de sous-chaînes sans fournir de garantie quant au résultat retourné.

Voici une représentation 1FN correcte :

Clients (NumCli, Nom, Prénom, Adresse, CodeP, Ville, Téléphone)

NumCli	Nom	Prénom	Adresse	Code Postal	Ville	Téléphone
1	Baptiste	Jean-Luc	25, rue de la forêt	12000	Rodez	0565420000
2	Auguy	Jean	Impasse des lys	15000	Aurillac	0471670000
3	Rascalou	André	2, rue droite	12000	Rodez	0565450000

Maintenant, récupérer les habitants d'une ville précise ne pose plus aucun problème, une simple requête SQL y parviendra de façon rapide et fiable.

2. 2FN - Deuxième forme normale

Une relation est en deuxième forme normale si :

- Elle est en première forme normale.
- Si tous les attributs non-clés ne dépendent pas d'une partie de la clé primaire.

Autrement dit, toute propriété de la relation doit dépendre intégralement de toute la clé.

Par exemple :

Commande (Numcli, CodeArticle, Date, Qté commandée, Désignation)

Cette relation est-elle en première forme normale ? Oui.

Est-elle en deuxième forme normale ? Non, car la propriété Désignation ne dépend pas intégralement de la clé (Numcli, CodeArticle, Date).

Commandes :

NumCli	CodeArticle	Date	Qté commandée	Désignation
1	Art1	28/02/2018	5	Bocal d'un kilo de Tripoux
3	Art2	28/02/2018	9	Aligot congelé
5	Art3	28/02/2018	10	Cèpes séchés
6	Art41	28/02/2018	15	Bouteille de Marcillac rouge

Connaissant {1,Art1,28/02/2018} pouvons-nous connaître de façon sûre et unique « **Bocal d'un kilo de Tripoux** » ? La réponse est évidemment non ! « **Bocal d'un kilo de Tripoux** » ne dépend pas intégralement de la clé {1,Art1,28/02/2018}.

Voici comment corriger :

Commandes(Numcli, CodeArticle, Date, Qté commandée)

Articles(CodeArticle, Désignation)

NumCli	CodeArticle	Date	Qté commandée
1	Art1	28/02/2018	5
3	Art2	28/02/2018	9
5	Art3	28/02/2018	10
6	Art41	28/02/2018	15

Articles :

CodeArticle	Désignation
Art1	Bocal d'un kilo de Tripoux
Art2	Aligot congelé
Art3	Cèpes séchés
Art41	Bouteille de Marcillac rouge

3. 3FN - Troisième forme normale

Une relation est en troisième forme normale si :

- Elle est en deuxième forme normale.
- Si toutes les dépendances fonctionnelles par rapport à la clé sont directes (s'il n'y a pas de DF transitives entre les attributs non clé).

Autrement dit, tous les attributs n'appartenant pas à la clé ne dépendent pas d'un attribut non-clé.

Exemple :

La relation Commande(NuméroCommande, #CodeClient, Nom client, #RefArticle) est-elle en troisième forme normale ?

Est-elle en première forme normale ? Oui

Est-elle en deuxième forme normale ? Oui

Est-elle en troisième forme normale ? Non !

En effet Nom client dépend d'une propriété non clé : CodeClient

Commandes :

NuméroCommande	CodeClient	Nom client	RefArticle
1	C1	Baptiste	Art25
3	C5	Savary	Art20
5	C2	Martinez	Art10
6	C1	Baptiste	Art15

Voici comment corriger :

Commande(NuméroCommande, #CodeClient, #RefArticle) Clients(CodeClient, Nom client)

Commandes :

NuméroCommande	CodeClient	RefArticle
1	C1	Art25
3	C5	Art20
5	C2	Art10
6	C1	Art15

Clients :

CodeClient	Nom client
C1	Baptiste
C2	Martinez
C5	Savary

4. BCNF - 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.
- Les seules dépendances fonctionnelles élémentaires qu'elle comporte sont celles dans lesquelles une clé détermine un attribut.

Considérons la relation Vaches(Race, Pays, Région) avec les dépendances fonctionnelles supposées :

Région ---> Pays

(Race, Pays) ---> Région

Race	Pays	Région
Aubrac	France	Auvergne
Salers	France	Auvergne
Limousine	France	Limousin
Highland	Royaume-Uni	Highlands

Cette relation est bien en troisième forme normale car aucun attribut non clé ne dépend d'une partie de la clé ou d'un attribut non clé. Cependant, on y trouve de nombreuses redondances, par exemple les deux premières lignes possèdent des pays et des régions identiques.

Afin d'éliminer ces redondances, Boyce et Codd ont introduit une forme normale qui porte leur nom (*Boyce Codd Normal Form/BCNF*) :

Une relation est en BCNF si et seulement si les seules dépendances fonctionnelles élémentaires sont celles dans lesquelles une clé détermine un attribut.

La relation Vaches pourra être décomposée en deux relations :

- Races (Race, Région)
- Régions (Région, Pays)

La dépendance fonctionnelle (Race, Pays) Région est perdue, mais elle peut être recomposée par jointure.

Cette normalisation est très importante dans la pratique si l'on veut éviter de stocker des informations redondantes. On considère, en général, que la troisième forme normale est suffisante dans les cas courants.

5. 4FN - Quatrième forme normale

Une relation est en quatrième forme normale si et seulement si :

- Elle est en BCNF.
- Lorsqu'il existe une dépendance multivaluée élémentaire, celle-ci est unique.

Une relation en BCNF peut encore comporter des redondances.

Par exemple :

L'étudiant (NumEtu) pratique une ou plusieurs langues et suit un ou plusieurs cours.

Imaginons cette relation :

Etudiants (NumEtu, Langue, Cours)

Et les postulats suivants :

Il n'y a pas de dépendances fonctionnelles.

La clé est : (NumEtu, Langue, Cours).

La relation Etudiants est en troisième forme normale et en troisième forme normale Boyce-Codd.

Cependant Etudiants contient encore des redondances :

Etudiants	NumEtu	Langue	Cours
	1	Anglais	Mathématiques
	1	Anglais	Histoire
	2	Anglais	Economie
	2	Espagnol	Economie
	3	Espagnol	Economie
	3	Espagnol	Droit

L'étudiant numéro 1 ne connaît que l'anglais, mais suit 2 cours (les mathématiques et l'histoire). Il y a donc une redondance sur la langue.

L'étudiant numéro 2 connaît deux langues (l'anglais et l'espagnol) mais ne suit qu'un cours. Il y a une redondance sur le cours.

Les dépendances fonctionnelles ne suffisent pas à définir toutes les dépendances entre les données.

Dépendances multivaluées

Pour une valeur d'étudiant, on a toutes les valeurs possibles de langue et, pour chacune de ces valeurs, toutes les valeurs possibles de cours, mais langue et cours sont indépendantes entre elles : on dit que l'on a une dépendance multivaluée entre la colonne NumEtu et la colonne Langue et une dépendance multivaluée entre la colonne NumEtu et la colonne Cours.

On voit l'inconvénient de cette forme puisque, si l'on supprime une valeur possible de la colonne Cours, par exemple l'organisme de formation retire l'économie de son catalogue, il faut supprimer toutes les lignes où économie est inscrit.

Pour éviter ce genre de problèmes, il faut passer à la quatrième forme normale, qui se peut se définir ainsi :

Une relation est en quatrième forme normale si et seulement si les seules dépendances multivaluées élémentaires sont celles dans lesquelles une clé détermine la valeur d'une colonne.

Ici, les colonnes sur lesquelles portent des dépendances multivaluées font partie de la clé, donc la table n'est pas en 4FN et il faut la décomposer en deux tables :

Etudiants1(NumEtu, Langue)

Etudiant2(NumEtu, Cours)

6. 5FN - Cinquième forme normale

Cette forme normale n'étant quasiment jamais utilisée, en voici juste la définition :

Une association est en cinquième forme normale si et seulement si :

- Elle est en quatrième forme normale.
- Elle ne possède pas de dépendance de jointure.

La cinquième forme normale est une généralisation de la quatrième forme normale qui nécessite de prendre en compte les dépendances de jointure induites par la connaissance des clés d'une relation.