

5.1 Introduction

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. En d'autres termes, le processus est réversible (*i.e.* sans perte d'information).

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. Ces algorithmes de décomposition, associés à chacune des formes normales, sortent du cadre de ce cours et ne seront pas abordés.

La normalisation peut être effectuée, et c'est préférable, pendant la phase de conception sur le modèle entités-associations. Dans le cas où la normalisation est faite en amont, lors de la conception, il n'est pas nécessaire de la recommencer sur le modèle relationnel. On peut tout de même vérifier que les relations obtenues par le passage du modèle entités-associations au modèle relationnel sont toujours en forme normale, mais, sauf erreur, il ne devrait pas y avoir de problème.

5.2 Dépendance fonctionnelle (DF)

Définition 1 -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 : $X \rightarrow Y$.

Autrement dit, 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.

Définition 2 -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 $X' \rightarrow A$.

Autrement dit, 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.

Définition 3 -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.

5.3 Première et deuxième forme normale

Première forme normale

Définition 4 -première forme normale (1FN)- Une relation est en première forme normale si, et seulement si, tout attribut contient une valeur atomique (non multiples, non composées).

Par exemple, le pseudo schéma de relation *Personne* (num-personne, nom, prénom, rue-et-ville, *prénoms-enfants*) n'est pas en première forme normale. Il faut le décomposer en :

- *Personne*(num-personne, nom, prénom, rue, ville)
- *Prénoms-enfants*(num-personne, num-prénom)
- *Prénoms*(num-prénom, prénom)
- Deuxième forme normale

Définition 5 -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 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.

5.4 Troisième forme normale

Définition 6 -Troisième forme normale (3FN)- Une relation est en troisième forme normale si, et seulement si elle est en deuxième forme normale et tout attribut n'appartenant pas à la clé n'est pas en dépendance fonctionnelle directe avec un ensemble d'attributs non-clé.

5.5 Forme normale de BOYCE-CODD

Définition 7 -forme normale de BOYCE-CODD (BCNF)- Une relation est en forme normale de Boyce-Codd (BCNF) si, et seulement si, les seules dépendances fonctionnelles élémentaires sont celles dans lesquelles une clé détermine un attribut non-clé.

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

5.6 Pour aller plus loin : quatrième et cinquième forme normale

Dépendance multivaluée (DM)

Définition 8 -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 (i.e. une affectation de valeur à chacune des propriétés 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 : $X \twoheadrightarrow Y \Rightarrow X \twoheadrightarrow A_i - (X \cup Y)$.

Comme illustration, supposons une situation où un employé d'un garage est qualifié pour effectuer un certain type d'intervention sur certaines marques de voiture. Cette situation est modélisée par le schéma relationnel suivant :

- Employé(Nom-Employé)
- Intervention(Type-Intervention)
- Constructeur(Marque)
- Intervenir(Nom-Employé, Type-Intervention, Marque)

Supposons maintenant qu'un employé qui effectue un ensemble de types d'interventions pour un ensemble de marques de voiture, est capable d'effectuer chacun de ces types d'interventions sur chacune de ces marques de voitures. Dans ce cas, il existe des dépendances multivaluées dans la relation *Intervenir* : $\text{Nom-Employé} \twoheadrightarrow \text{Type-Intervention}$ et $\text{Nom-Employé} \twoheadrightarrow \text{Marque}$.

Quatrième forme normale

Définition 9 -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 $X \rightarrow A$.

Nom-Employé	Type-Intervention	Marque
Tussier	Dépannage	Peugeot
Tussier	Dépannage	Citroën
Martin	Électricité	Citroën
Martin	Électricité	Renault
Martin	Mécanique	Citroën
Martin	Mécanique	Renault
Piquard	Carrosserie	Fiat
Piquard	Carrosserie	Ford

Piquard	Alarme	Fiat
Piquard	Alarme	Ford
Piquard	Électricité	Fiat
Piquard	Électricité	Ford

Tableau 5.1: Exemple de relation n'étant pas en quatrième forme normale.

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*. La table 5.1 propose un exemple de relation correspondant à ce schéma de relation. Cette table permet d'observer le phénomène de redondance consécutif au fait que cette table n'est pas en quatrième forme normale. Dans cette table, le nombre de lignes commençant par un nom d'employé donné doit être égale au nombre d'interventions que cet employé peut faire multiplié par le nombre de marques sur lesquelles il peut travailler. Imaginons que l'employé *Piquard* puisse maintenant travailler sur des voitures de la marque *Citroën* (on désire ajouter **une** information dans la base), il faudra alors ajouter trois lignes à la table : une pour chaque type d'intervention (*Carrosserie*, *Alarme* et *Électricité*).

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

- Employé(Nom-Employé)
- Intervention(Type-Intervention)
- Constructeur(Marque)
- Etre-capable-de(Nom-Employé, Type-Intervention)
- Etre-capable-d'intervenir-sur(Nom-Employé, Marque)

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 seront définies plus loin

Définition 10 -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 :

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

Définition 11 -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)

Définition 12 -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. Il est donc superflu de décomposer de ce point de vue.

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).

Relation Fournisseur		
Num-Fournisseur	Num-Article	Num-Organisme
f1	a2	o1
f1	a1	o2
f2	a1	o1
f1	a1	o1

Tableau 5.2: Exemple de relation n'étant pas en cinquième forme normale.

Prenons, comme illustration, la relation *Fournisseur* (table 5.2) qui décrit les fournisseurs des organismes de la fonction publique.

La fonction publique a des règles très particulières concernant les fournisseurs pour réduire le potentiel de conflit d'intérêt. Un fournisseur fournit un certain nombre d'articles (par exemple *f1* fournit *a1* et *a2*). Le même article peut être fourni par plusieurs fournisseurs (par exemple *a1* est fourni par *f1* et *f2*). Un fournisseur peut être attitré à plusieurs organismes (par exemple *f1* est attitré à *o1* et *o2*). Un organisme peut avoir plusieurs fournisseurs (par exemple *o1* est servi par *f1* et *f2*). Un organisme peut utiliser plusieurs articles (c'est-à-dire que *o1* utilise *a1* et *a2*) et un article peut être utilisé par plusieurs organismes (c'est-à-dire que *a1* est utilisé par *o1* et *o2*). La règle de la fonction publique est la suivante :

- si un fournisseur fournit un certain article (comme *f1* fournit *a1*),
- le fournisseur est attitré à l'organisme (comme *f1* est attitré à *o1*), et
- l'organisme utilise un article (comme *o1* utilise *a1*),
- alors nécessairement, le fournisseur fournit l'article à l'organisme (*f1* fournit *a1* à *o1*).

Le dernier fait est déductible des trois autres.

Cette table contient de la redondance de données parce que certains faits sont répétés. Par exemple, le fait que *f1* fournit *a1* est répété à deux reprises, une fois parce qu'il fournit *a1* à *o1* et une autre fois parce qu'il fournit *a1* à *o2*. Le fait que *f1* est attitré à *o1* est aussi répété à deux reprises. Il en est de même pour *o1* qui utilise *a1*.

La relation *Fournisseur* souffre d'une dépendance de jointure : $\{(Num-Fournisseur, Num-Article), (Num-Fournisseur, Num-Organisme), (Num-Article, Num-Organisme)\}$. Pour résoudre le problème de redondance, il faut décomposer la relation en trois (cf. table 3.4).

Relation FournisseurArticle		Relation FournisseurOrganisme		Relation ArticleOrganisme	
Num-Fournisseur	Num-Article	Num-Fournisseur	Num-Organisme	Num-Article	Num-Organisme
<i>f1</i>	<i>a2</i>	<i>f1</i>	<i>o1</i>	<i>a2</i>	<i>o1</i>
<i>f1</i>	<i>a1</i>	<i>f1</i>	<i>o2</i>	<i>a1</i>	<i>o2</i>
<i>f2</i>	<i>a1</i>	<i>f2</i>	<i>o1</i>	<i>a1</i>	<i>o1</i>

Tableau 3.4: Décomposition de la relation *Fournisseur* (table 3.3) pour obtenir des relations en cinquième forme normale.

Il est important de se convaincre qu'aucune décomposition binaire de cette relation ne préserve le contenu de la relation initiale. Pour cela, il suffit de tenter de joindre deux tables parmi les trois précédentes. Aucune de ces jointures, ne produit la relation *Fournisseur*.

5.7 Remarques au sujet de la normalisation

Il existe d'autres formes normales comme la forme normale domaine-clé (FNDC), la forme normale de restriction-union ou 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.