

Bases de données avancées (cours n°1)

Table des matières

1	Normalisation des relations	1
1.1	Exemple : Cartes grises	1
1.2	Rappels	2
1.3	Dépendances fonctionnelles	3
1.3.1	Dépendance fonctionnelle (DF) – rappel	3
1.3.2	Dépendances fonctionnelles élémentaires	3
1.3.3	Dépendances fonctionnelles directes	4
1.4	Formes normales	4
1.5	Étapes d’une normalisation	6
2	Transformation d’un schéma SR au schéma EA en passant par le GDF (Rappel)	6
2.1	SR vers GDF :	6
2.2	GDF vers schéma EA :	6
3	Normalisation : exemple	7

1 Normalisation des relations

1.1 Exemple

Hypothèse : une voiture ne peut pas changer de propriétaire deux fois dans la même journée.

VOITURE

no_vehic	date_mc	no_type	nom_mod	classe	couleur	marque	puissance	pays_fab
V01	12/02/2009	M01	Clio	berline	rouge	Renault	5	France
V02	01/03/2009	M01	Clio	berline	noir	Renault	5	Espagne
V03	01/02/2019	M02	308	berline	vert	Peugeot	4	France
...

APPARTIENT

no_cond	nom_cond	adresse	no_vehic	date_achat	immat	nom_pref	lieu_pref
C01	Lagaffe	Epinal	V01	12/02/2009	123AB88	Vosges	Epinal
C02	Talon	La Bresse	V02	01/03/2009	456AD88	Vosges	Epinal
C03	Lefranc	Metz	V03	01/02/2019	FD777EF	Moselle	Metz
C02	Talon	La Bresse	V01	06/04/2020	FP649BP	Vosges	Epinal
...

Schéma relationnel :

VOITURE (no_vehic, date_mc, no_type, nom_mod, classe, couleur, marque, puissance, pays_fab)
 APPARTIENT (no_cond, nom_cond, adresse, no_vehic#, date_achat, immat, nom_pref, lieu_pref)

Schéma relationnel normalisé :

VOITURE (no_vehic, date_mc, no_type#, couleur, pays_fab)

MODELE (no_type, nom_mod, classe, marque, puissance)

CONDUCTEUR (no_cond, nom_cond, adresse)

APPARTIENT (immat#, date_achat, no_cond#)

DEPARTEMENT (no_dept, nom_pref, lieu_pref)

IMMATRICULATION (immat, no_dept#, no_vehic#)

PAYS (pays_fab)

1.2 Rappels

Définition (Relation)

Soient n ensembles D_1, D_2, \dots, D_n appelés domaines. Une *relation* R est définie comme un sous-ensemble du produit cartésien $D_1 \times D_2 \times \dots \times D_n$. Une relation est donc un ensemble de n -uplets $\langle d_1, d_2, \dots, d_n \rangle$ avec $d_1 \in D_1, d_2 \in D_2, \dots, d_n \in D_n$:

$$R = \{ \langle d_1, d_2, \dots, d_n \rangle \mid d_i \in D_i, \forall i \in \{1, \dots, n\} \}.$$

Définition (Schéma de relation)

Un *schéma de relation* définit une classe de relations de même nature par la donnée

- d'un nom de schéma de relation
- de n ensembles D_1, D_2, \dots, D_n (les domaines)
- de n nom distincts A_1, A_2, \dots, A_n (les attributs)
- des règles de cohérence de schéma de relation.

Définition (Schéma relationnel)

Un *schéma de relationnel* est la représentation du schéma d'une BdD sous forme d'un ensemble de schémas de relation.

Notion de clé :

- clé candidate (unicité, minimalité)
- clé primaire, clé alternative
- clé étrangère

1.3 Dépendances fonctionnelles

1.3.1 Dépendance fonctionnelle (DF) – rappel

Définition. Soit $R(X, Y, Z)$ un schéma de relation dans lequel X, Y, Z désignent des ensembles d'attributs, Z pouvant être vide. On dit qu'il existe une *dépendance fonctionnelle* de Y par rapport à X , notée $X \rightarrow Y$, si et seulement si

$$\forall x, x', y, y', z, z' :$$

$$(< x, y, z > \in R \text{ et } < x', y', z' > \in R \text{ et } x = x') \Rightarrow (y = y').$$

La connaissance d'une valeur de X entraîne la connaissance de la valeur de Y .

1.3.2 Dépendances fonctionnelles élémentaires

Définition. Un attribut B est en dépendance fonctionnelle **élémentaire** par rapport à un attribut ou ensemble d'attributs A si et seulement si

- $A \rightarrow B$
- $\forall C \subsetneq A : C \not\rightarrow B$

Exemple :

LIGNE_COM(no_com, no_prod, date_com, qte_com)

no_com	no_prod	date_com	qte_com
C1	P1	D1	10
C2	P1	D2	20
C2	P2	D2	30
...

- La dépendance fonctionnelle

$$\text{no_com, no_prod} \rightarrow \text{qte_com}$$

n'est pas élémentaire car

$$\{\text{no_com}\} \subset \{\text{no_com, no_prod}\} \text{ et } \text{no_com} \rightarrow \text{date_com}$$

- La dépendance fonctionnelle

$$\text{no_com, no_prod} \rightarrow \text{qte_com}$$

est élémentaire car

$$\text{no_com} \not\rightarrow \text{qte_com} \text{ et } \text{no_prod} \not\rightarrow \text{qte_com}$$

1.3.3 Dépendances fonctionnelles directes

Définition. Un attribut B est en dépendance fonctionnelle **directe** par rapport à un attribut ou ensemble d'attributs A si et seulement si

- $A \rightarrow B$
- $\nexists C : A \rightarrow C \text{ et } C \rightarrow B$

La DF $A \rightarrow B$ ne résulte pas d'une transitivité.

Exemple :

VEHICULE(no_immat, marque, type, couleur)

no_immat	marque	type	couleur
FD123CD	Renault	Clio	rouge
EF456GH	Peugeot	306	noir
EZ444BB	Peugeot	306	rouge
...

La dépendance fonctionnelle $\text{no_immat} \rightarrow \text{marque}$ n'est pas directe car

$\text{no_immat} \rightarrow \text{type}$ et $\text{type} \rightarrow \text{marque}$ (transitivité)

Les dépendances fonctionnelles

$\text{no_immat} \rightarrow \text{type}$ et $\text{no_immat} \rightarrow \text{couleur}$

sont **directes**.

1.4 Formes normales

Première forme normale (1FN)

Une relation R est en 1FN, si et seulement si chacun des attributs qui ne fait pas partie de la clé primaire est en DF par rapport à la clé (valeurs atomiques pour chaque attribut).

Deuxième forme normale (2FN)

Une relation R est en 2FN, si et seulement si chacun des attributs qui ne fait pas partie de la clé primaire est en *DF élémentaire* par rapport à la clé.

Troisième forme normale (3FN)

Une relation R est en 3FN, si et seulement si chacun des attributs qui ne fait pas partie de la clé primaire est en *DF élémentaire et directe* par rapport à la clé.

Forme normale de Boyce-Codd (BCFN)

Une relation $R(\underline{C_1, \dots, C_k}, A_1, \dots, A_n)$ est en BCFN, si et seulement si les seules DF que cette relation inclut sont celles de la clé :

$$\forall i \in \{1, \dots, n\} \quad \forall j \in \{1, \dots, k\} : A_i \not\rightarrow C_j.$$

Aucun attribut de la clé primaire n'est en dépendance fonctionnelle par rapport à un attribut ne faisant pas partie de la clé.

Remarque qu'une relation peut être en 3FN mais pas en BCFN.

1.5 Étapes d'une normalisation

- Ajouter des identifiants si nécessaire
- Définir les clés primaires
- Identifier les redondances
- Pour chaque relation :
 - établir la liste des DF issues de la clé primaire ;

- si la clé est composée de plusieurs attributs :
 - chercher et **noter** les DF issues d'une partie de la clé ;
 - dans la liste des DF issues de la clé, identifier les DF NE (non-élémentaire) ;
 - rendre la relation en 2FN : retirer les attributs qui posent problème et exprimer les DF retirées par ailleurs ;
- si la clé est composée d'un seul attribut/la relation est en 2FN :
 - chercher et **noter** les DF existantes entre les attributs ne faisant pas partie de la clé primaire ;
 - dans la liste des DF issues de la clé, identifier les DF ND (non-directe) ;
 - rendre la relation en 3FN : retirer les attributs qui posent problème et exprimer les DF retirées par ailleurs.

2 Normalisation : exemple

Gestion de stocks.

Soient les schémas de relation suivants, décrivant de manière simplifiée la base de données utilisée pour une gestion de stocks, et pour lesquels une description en extension est donnée ci-dessous :

COMMANDE (num_com, mat_client, date_com, nom_cl, ad_cl)

LIGNE_COM (num_com, num_prod, nom_fourn, ad_fourn, nom_prod, prix_unit, qte_com)

Travail demandé

1. Déterminer une clé pour chacun des ces 2 schémas de relation.
2. Ces données représentent de nombreuses redondances qui impliquent que les 2 relations COMMANDE et LIGNE_COM ne sont pas en 3FN. Il vous faut donc :
 - identifier ces redondances ;
 - indiquer quelles formes normales ne sont pas respectées et pourquoi ;
 - transformer ces 2 schémas de relation de manière à obtenir une représentation du schéma conceptuel de la base de données sous la forme de schémas de relation en 3FN ;
 - donner la définition en extension des relations ainsi obtenues.
3. Donner le graphe des dépendances fonctionnelles existantes entre les différents attributs de cette base et en déduire le schéma conceptuel de cette BdD à l'aide du modèle Entité-Association.

COMMANDE :

num_com	mat_client	date_com	nom_cl	ad_cl
C1	1	11-09-2022	Linford Christie	Londres
C2	2	11-09-2022	Frankie Fredericks	Windhoek
C3	1	12-09-2022	Linford Christie	Londres
C4	3	13-09-2022	Mickaël Johnson	Washington
C5	3	13-09-2022	Mickaël Johnson	Washington

LIGNE_COM :

num_com	num_prod	nom_fourn	ad_fourn	nom_prod	prix_unit	qte_com
C1	P1	Gail Devers	Washington	chaussures	120	2
C1	P5	Merlene Ottey	Kingston	survêtement	59	1
C2	P2	Merlene Ottey	Kingston	lacets	5	2
C2	P3	Gail Devers	Washington	short	18	3
C2	P4	Sarah Miles	Washington	maillot	20	3
C2	P5	Merlene Ottey	Kingston	survêtement	59	2
C3	P2	Merlene Ottey	Kingston	lacets	5	2
C4	P1	Gail Devers	Washington	chaussures	120	1
C4	P3	Gail Devers	Washington	short	18	5
C4	P5	Merlene Ottey	Kingston	survêtement	59	2
C5	P2	Merlene Ottey	Kingston	lacets	5	3
C5	P4	Sarah Miles	Washington	maillot	20	6
C5	P5	Merlene Ottey	Kingston	survêtement	59	1

Rappel : Transformation d'un schéma SR au schéma EA en passant par le GDF

SR vers GDF :

Rappel : tous les attributs qui ne font pas partie de la clé primaire sont en DF par rapport à elle. Si la clé primaire est composée d'un seul attribut, on utilise des flèches directes, sinon on utilise un point (attention aux synonymes).

GDF vers schéma EA :

Entité :

- une source de DF et tous les attributs qui sont au bout des flèches directes et qui ne sont pas eux-mêmes des sources de DF ;
- point de départ isolé d'une flèche vers un point ;
- point de départ isolé d'une flèche simple ;

Association :

- flèche directe entre deux sources de DF (association avec cardinalités 11 ou 01) ;
- ensemble de flèches arrivant et partant d'un point (association avec des cardinalités maximum n).

Exemple : Cours.

A_COURS (no_groupe, no_ens, no_salle, dh, duree)

+ contrainte : "un groupe a toujours cours dans la même salle"

DF : no_groupe \rightarrow no_salle

Remarques :

Un cours a lieu pour un groupe dans une salle avec un enseignant.

Un groupe peut avoir plusieurs cours avec des enseignants différents.

Un enseignant peut dispenser des cours à des groupes différents, dans des salles différentes.

Dans une salle, il peut y avoir des cours pour des groupes différents à des horaires différents.

Dans la suite, la première étape sera rédigée en **mauve**, la deuxième en **orange** et la troisième en **vert**.

— **1ère clé** : A_COURS (no_groupe, dh, no_ens, no_salle, duree)

Un groupe à une date et heure données est dans une salle avec un enseignant pour une durée.

no_groupe	dh	no_ens	no_salle	duree
gr1	dh1	e1	s1	1
gr2	dh1	e2	s2	2
gr1	dh2	e2	s1	2

Liste des dépendances fonctionnelles :

DF issues de la clé primaire :

no_groupe, dh \rightarrow no_ens **ED**

no_groupe, dh \rightarrow no_salle **NE**

no_groupe, dh \rightarrow duree **ED**

DF issue d'une partie de la clé primaire :

no_groupe \rightarrow no_salle **ED**

Normalisation :

Il existe des DF non-élémentaires issues de la clé primaire, A_COURS n'est pas en 2FN.

Passage de A_COURS en 2FN :

— on retire l'attribut qui pose problème (DF NE) :

A_COURS2 (no_groupe, dh, no_ens, duree) **2FN et 3FN**

— on exprime par ailleurs la DF retirée :

A_LIEU(no_groupe, no_salle) **2FN et 3FN**

On examine maintenant la nouvelle relation A_LIEU : la clé primaire est composée d'un seul attribut : la DF $\text{no_groupe} \rightarrow \text{no_salle}$ est donc élémentaire (on la note à côté de la DF), la relation est aussi en 2FN.

Pour vérifier si le nouveau schéma relationnel est en 3FN, on cherche la transitivité dans les relations : on cherche et on note les DF qui existent entre les attributs qui ne font pas partie de la clé primaire.

A_COURS2 : un enseignant peut avoir des cours de durées différentes et la même durée peut être associée à des cours de différents enseignants, il n'y a donc pas de DF entre les attributs no_ens et $\text{duree} \Rightarrow$ toutes les DF issues de la clé primaire sont aussi directes (on note), A_COURS2 est donc en 3FN (on note).

A_LIEU : il y a un seul attribut (no_salle) en dehors de la clé primaire \Rightarrow toutes les DF issues de la clé primaire sont aussi directes (on note), A_LIEU est donc en 3FN (on note).

Schéma relationnel normalisé 3FN :

A_COURS2 (no_groupe, dh, no_ens, duree)

A_LIEU(no_groupe, no_salle)

— **2e clé** : A_COURS (no_ens, dh, no_groupe, no_salle, duree)

Un enseignant à une date et heure données est dans une salle avec un groupe pour une durée.

no_ens	dh	no_groupe	no_salle	duree
e1	dh1	gr1	s1	1
e2	dh1	gr2	s2	2
e2	dh2	gr1	s1	2

Liste des dépendances fonctionnelles :

DF issues de la clé primaire :

- (1) no_ens, dh \rightarrow no_groupe ED
- no_ens, dh \rightarrow no_salle END (1+2)
- no_ens, dh \rightarrow duree ED

DF entre attributs ne faisant pas partie de la clé primaire :

- (2) no_groupe \rightarrow no_salle ED

Normalisation :

Un enseignant peut enseigner à plusieurs groupes, dans plusieurs salles et pendant des durées différentes. Même raisonnement pour les créneaux (dh). Il n'existe donc pas de DF d'une partie de la clé primaire vers les attributs ne faisant pas partie de la clé : toutes les DF issues de la clé primaire sont élémentaires (on note), A_COURS est donc en 2FN.

Par contre, il existe des DF non-directes : la DF no_ens, dh \rightarrow no_salle n'est pas directe car elle peut être obtenue par transitivité

no_ens, dh \rightarrow no_groupe et
no_groupe \rightarrow no_salle.

Les deux autres DF issues de la clé primaire sont directes (on note).

Passage de A_COURS en 3FN :

— on retire l'attribut qui pose problème (DF ND) :

A_COURS2 (no_ens, dh, no_groupe, duree) 2FN et 3FN

— on exprime par ailleurs la DF retirée :

A_LIEU(no_groupe, no_salle) 2FN et 3FN

On examine maintenant la nouvelle relation A_LIEU : pour les mêmes raisons que pour la 1e clé, cette relation est en 3FN.

Schéma relationnel normalisé 3FN :

A_COURS2 (no_ens, dh, no_groupe, duree)
A_LIEU (no_groupe, no_salle)

— **3e clé** : A_COURS (no_salle, dh, no_groupe, no_ens, duree)

Dans une salle, à une date et heure données il peut y avoir une seul groupe avec un enseignant pour une durée.

no_salle	dh	no_groupe	no_ens	duree
s1	dh1	gr1	e1	1
s2	dh1	gr2	e2	2
s1	dh2	gr1	e1	2

Liste des dépendances fonctionnelles :

DF issues de la clé primaire :

no_salle, dh \rightarrow no_groupe ED
no_salle, dh \rightarrow no_ens ED
no_salle, dh \rightarrow duree ED

DF d'une partie de la clé primaire par rapport à un attribut ne faisant pas partie de la clé :

no_groupe \rightarrow no_salle ED

Normalisation :

On cherche d'abord les DF NE pour savoir si A_COURS est en 2FN. Dans une même salle, il peut y avoir des groupes, enseignants et durées différents pour des créneaux différents. Même raisonnement pour les créneaux (dh).

Il n'existe donc pas de DF d'une partie de la clé primaire vers les attributs ne faisant pas partie de la clé \Rightarrow toutes les DF issues de la clé primaire sont élémentaires (on note), A_COURS est donc en 2FN.

On cherche les DF ND pour savoir si A_COURS est en 3FN. Il n'existe pas de DF entre les attributs ne faisant pas partie de la clé primaire (un groupe peut avoir plusieurs enseignants et inversement, etc.). Toutes les DF issues de la clé primaire sont donc directes (on note), A_COURS est donc aussi en 3FN.

Par contre, elle n'est pas en BCFN car il y existe une DF d'une partie de la clé (no_salle) par rapport à un attribut ne faisant pas partie de la clé (no_groupe).

Passage de A_COURS en BCFN :

— on remplace l'attribut no_salle par no_groupe dans la clé primaire de A_COURS :

A_COURS2 (no_groupe, dh, no_ens, duree) 2FN, 3FN et BCFN

— La DF no_groupe \rightarrow no_salle doit être exprimée dans la Bdd : on crée le schéma de relation

A_LIEU(no_groupe, no_salle) 2FN, 3FN et BCFN.

Schéma relationnel normalisé 3FN :

A_COURS2 (no_groupe, dh, no_ens, duree)

A_LIEU(no_groupe, no_salle)



C'est à vous...!

1. Quelles sont les formes normales qui sont respectées ?
 - (a) les DF issues de la clé primaire sont toutes élémentaires :
 - (b) les DF issues de la clé primaire sont toutes directes :
 - (c) les DF issues de la clé primaire sont élémentaires et directes :
 - (d) les valeurs d'un attribut sont des ensembles de valeurs :
 - (e) la CP (clé primaire) est composée d'un seul attribut :
 - (f) la CP est composée de plusieurs attributs et il y a un seul attribut qui ne fait pas partie de la clé primaire :
 - (g) Lors de la normalisation d'une relation
 - ☐ on élimine d'abord les DF NE et ensuite les DF ND
 - ☐ on élimine d'abord les DF ND et ensuite les DF NE
 - ☐ l'ordre n'a pas d'importance, le résultat est le même
2. Lors de la transformation d'un schéma EA vers le SR,
 - (a) quelle est la propriété de la clé candidate qui n'est pas toujours respectée par l'identifiant d'une association ?
 - (b) en considérant 01 comme 11, à quoi faut-il faire attention dans le schéma relationnel ?
.....
 - (c) en considérant 01 comme 0n, à quoi faut-il faire attention dans le schéma relationnel ?
.....
3. Lors de la transformation d'un schéma SR vers le schéma EA,
 - (a) un attribut qui est une source de DF correspondra à
 - (b) une "feuille" du graphe correspondra à
 - (c) une fois les entités identifiées, il reste dans le graphe :
—
—
—