

Corrigé du R1.05-Intro BD et SQL (Séance n° 3) Dépendances Fonctionnelles et Normalisation

Rappels :

On dispose d'une relation $R(A, B, C, D, \dots)$ avec F un ensemble de DFs

1NF) Une relation R est en 1NF ssi tous ses attributs sont atomiques (simple ou monovalués).

2NF) Une relation R est en 2NF ssi elle est en 1NF et il n'existe aucune DF interdite valide sur R . Cette DF interdite est de type $X \rightarrow A$ (X est une partie d'une clef et A un attribut qui n'est pas dans une clef).

3NF) Une relation R est en 3NF ssi elle est en 2NF et il n'existe aucune DF interdite valide sur R . Cette DF interdite est de type $X \rightarrow A$ (X n'est pas une clef ou ne contient pas une clef et A un attribut qui n'est pas dans une clef)

Remarque importante

Pour simplifier la normalisation on suppose que la notion de clef dans les définitions correspond à la clef primaire (une clef candidate minimale)).

Exercice n° 1

Q1 L'un des attributs A, B, C ou D peut-il jouer le rôle de clef primaire ?

Aucun des quatre attributs ne peut jouer **seul** le rôle de clé primaire car aucun de ne détermine tous les autres, par exemple :

A : la valeur $a1$ est associée à $b1$ et $b2$
 B : la valeur $b1$ est associée à $a1$ et $a3$
 C : la valeur $c1$ est associée à $b1$ et $b2$
 D : la valeur $d2$ est associée à $c1$ et $c3$

Q2 Quelles sont les combinaisons « minimales » d'attributs qui pourraient avoir un rôle de clef primaire dans la relation (les clefs candidates minimales) telle qu'elle est donnée ?

Les trois couples d'attributs suivants sont des clefs candidates minimales pour R : (A, B) , (A, D) , (B, D) , car il n'existe aucun couple de valeurs dupliqué pour (A, B) , (A, D) , (B, D) et donc :

$A, B \rightarrow C, D$; $A, D \rightarrow B, C$; $B, D \rightarrow A, C$

Les couples d'attributs suivants sont des clefs impossibles pour R : (A, C), (B, C), (C, D), car :

- A, C : le couple de valeurs (a1, c1) est associé à b1 et b2
- B, C : le couple de valeurs (b1, c1) est associé à a1 et a3
- C, D : le couple de valeurs (c1, d2) est associé à a1 et a3

Exercice n° 2

Q3 : Dépendances fonctionnelles déduites de l'ensemble F en appliquant la propriété de transitivité.

Les DF pouvant être déduites de F en appliquant la propriété de transitivité sont les suivantes :

- NUM_PIECE \rightarrow CATEGORIE car NUM_PIECE \rightarrow LIBELLE et LIBELLE \rightarrow CATEGORIE
- NUM_PIECE \rightarrow TAUX_TVA car NUM_PIECE \rightarrow CATEGORIE et CATEGORIE \rightarrow TAUX_TVA
- LIBELLE \rightarrow TAUX_TVA car LIBELLE \rightarrow CATEGORIE et CATEGORIE \rightarrow TAUX_TVA

Q4 : Donnez un exemple de dépendance déduite de F en utilisant la propriété d'union.

Une DF pouvant être déduite de F en appliquant la propriété d'union est la suivante :
NUM_PIECE \rightarrow LIBELLE, CATEGORIE, TAUX_TVA, PRIX

Q5 : Quelle est la clef primaire de la relation PIECES ?

La clef primaire de la relation PIECES doit être choisie parmi les clefs candidates. Or il n'en existe qu'une qui soit minimale : NUM_PIECE, car c'est le seul attribut de la relation permettant de déterminer tous les autres.

Exercice n° 3

Q6 : Quelles sont les clefs candidates minimales. Justifiez votre réponse en utilisant les Axiomes d'Armstrong.

a) Il s'agit de trouver l'attribut ou l'ensemble d'attributs déterminant tous les autres.

On a : $I \rightarrow B$ et $B \rightarrow O$, donc par transitivité : $I \rightarrow O$. D'autre part : $S \rightarrow D$ et $I, S \rightarrow Q$.

Les attributs I et S suffisent à déterminer tous les autres et constituent donc une clef candidate minimale de R. Comme $D \rightarrow S$, le couple d'attributs I,D constitue la seconde clef candidate minimale. I,S est choisie comme clef primaire, d'où : R (I, S, B, O, Q, D)

b) Méthode par simplification :

B, O, I, S, Q, D --> B, O, I, S, Q, D

Simplification de la source de la DF précédente :

Supprimer B car $I \rightarrow B$

Supprimer O car $I \rightarrow O$

Supprimer Q car $I, S \rightarrow Q$

Supprimer D car $S \rightarrow D$

Q7 : Trouver une décomposition en 3NF de R. Sur le schéma obtenu, préciser les clefs primaires et étrangères.

Première forme normale :

La relation R est en 1NF car tous les attributs sont mono-valués.

Deuxième forme normale :

R n'est pas en 2NF, car B et O ne sont pas en dépendance fonctionnelle totale avec la clef ; une partie de la clef les détermine : $(I \rightarrow B, O)$.

R (I, S, B, O, Q, D) est décomposée en R1(I, S, Q, D) et R2(I, B, O)

Troisième forme normale :

- R1 n'est pas en 3NF car $I, D \rightarrow Q$. I, D n'est pas la clef primaire de R1 et Q n'est pas un attribut de la clef primaire de R1. D'où :

R1(I, S, D, Q) est décomposée en R11(I, S) et R12(I, D, Q)

- R2 n'est pas en 3NF, car on a : $B \rightarrow O$. B n'est pas la clef primaire de R2 et O n'est pas un attribut de la clef primaire de R2. D'où la décomposition de R2 en : R21(I, B) et R22(B, O)

Le schéma 3NF (où les clefs primaires sont soulignées et les clefs étrangères sont en italique gras) correspondant à R est donc le suivant :

R11(I, S)
R12(I, D, Q)
R21 (I, **B**)
R22(B, O)

Q8 : Existe-t-il d'autre décomposition en 3NF de R ?

Oui en choisissant I, D comme clef primaire.

Exercice n° 4

Q9 : Quelles sont les dépendances fonctionnelles minimales existant dans R ?

f1 : CODE_MAT \rightarrow NOM_MAT

f2 : NOM_MAT \rightarrow CODE_MAT

f3 : NUM_ETUD \rightarrow NOM_ETUD

f4 : CODE_MAT, NUM_ETUD \rightarrow MOYENNE

f5 : NOM_MAT, NUM_ETUD \rightarrow MOYENNE

Q10 : Quelles sont les clefs candidates minimales.

D'après les DF observées, il existe deux clefs candidates dans R, les couples d'attributs : (CODE_MAT, NUM_ETUD) et (NOM_MAT, NUM_ETUD).

Q11 : Choisir la meilleure clef primaire en justifiant votre choix.

La clef primaire de R est choisie parmi ces deux couples, probablement : (CODE_MAT, NUM_ETUD), car pour des clefs primaires les valeurs entières sont préférables à des chaînes.

Q12 : Construction de R' à partir de R1 et R2.

La jointure entre R1 et R2 ne peut être réalisée que sur des attributs sémantiquement compatibles, i.e. NOM_ETUD.

Construisons, à partir des tuples de R, l'extension de R1 et R2, on obtient les résultats suivants :

R1	CODE_MAT	NOM_MAT	NOM_ETUD
	1	Math	Dupont
	1	Math	Durand
	2	BD	Dupont
	2	BD	Durand
	3	Anglais	Dupont

R2	NUM_ETUD	NOM_ETUD	MOYENNE
	100	Dupont	10
	200	Durand	15
	300	Dupont	5
	200	Durand	10
	300	Dupont	12
	100	Dupont	16
	300	Dupont	10

L'opérateur de projection éliminant les duplicats, il y a, pour R1, perte d'un étudiant de nom Dupont (à cause de l'homonymie) et pour R2, perte d'une des notes de l'étudiant n° 100 (car il a la même note pour deux matières différentes).

Le résultat de la jointure entre R1 et R2 sur NOM_ETUD est illustré par la figure suivante (extension de la relation R').

La jointure effectuée ne permet pas de retrouver l'extension initiale de la relation R. Il y a en fait génération de tuples erronés, i.e. n'existant pas dans l'extension initiale de R. Cela provient du fait que la décomposition proposée ne respecte pas le théorème de décomposition.

R	CODE_MAT	NOM_MAT	NOM_ETUD	NUM_ETUD	NOM_ETUD	MOYENNE
	1	Math	Dupont	100	Dupont	10
	1	Math	Dupont	300	Dupont	5
	1	Math	Dupont	300	Dupont	12
	1	Math	Dupont	100	Dupont	16
	1	Math	Dupont	300	Dupont	10
	1	Math	Durand	200	Durand	15
	1	Math	Durand	200	Durand	10
	2	BD	Dupont	100	Dupont	10
	2	BD	Dupont	300	Dupont	5

2	BD	Dupont	300	Dupont	12
2	BD	Dupont	100	Dupont	16
2	BD	Dupont	300	Dupont	10
2	BD	Durand	200	Durand	15
3	Anglais	Dupont	100	Dupont	10
3	Anglais	Dupont	300	Dupont	5
3	Anglais	Dupont	300	Dupont	12
3	Anglais	Dupont	100	Dupont	16
3	Anglais	Dupont	300	Dupont	10
3	Anglais	Durand	200	Durand	15

Q13 : En quelle forme normale est R ? Justifiez votre réponse.

Première forme normale :

La relation R est en 1NF car tous les attributs sont mono-valués.

Deuxième forme normale :

R n'est pas en 2NF car NOM_MAT et NOM_ETUD ne dépendent pas totalement de la clef, à cause respectivement des DF f1 et f3.

Q14 : Trouvez une décomposition 3NF de R.

Deuxième forme normale :

La décomposition en 2NF de R conduit à isoler la dépendance f1 et f3 dans deux nouvelles relations tout en éliminant leur cible de la relation initiale. Le schéma en 2NF est donc le suivant :

NOTE (CODE_MAT, NUM_ETUD, MOYENNE)

ETUDIANT (NUM_ETUD, NOM_ETUD)

MATIERE(CODE_MAT, NOM_MAT)

Troisième forme normale :

Les relations NOTE, MATIERE et ETUDIANT sont en 3NF : il est évident qu'elles ne comportent aucune dépendance problématique-3NF (un seul attribut non clef dans chacune).

Exercice n° 5

Q15 : Quel est le degré de normalité des deux relations R1 et R2 ?

Première forme normale :

R1 et R2 sont en 1NF car tous leurs attributs sont atomiques.

Remarquons que la clef primaire de chacune des relation R1 et R2 est l'unique clef candidate minimale.

Deuxième forme normale :

- R1 n'est pas en 2NF car C est en dépendance partielle avec la clef à cause de : $B \rightarrow C$.
- R2 est en 2NF car aucune partie de la clef ne détermine un autre attribut.

Troisième forme normale :

R2 n'est pas en 3NF à cause de la dépendance transitive : $M \rightarrow N$ et aussi $I, J \rightarrow K$.

Q16 : Degré de normalité des relations R11, R12, R21 et R22.

Première forme normale :

les quatre relations sont en 1NF car tous leurs attributs sont atomiques.

Deuxième forme normale :

- R11 est en 2NF car ni A ni B ne détermine les autres attributs.
- R12 est en 2NF car la clef est mono-attribut.
- R21 est en 2NF car ni G ni H ne sont sources de dépendance.
- R22 n'est pas en 2NF car N est déterminé par une partie de la clef : M.

Troisième forme normale :

- R11 n'est pas en 3NF à cause de la dépendance transitive : $D \rightarrow E, F$.
- R12 est en 3NF (Trivial ! R12 n'a que deux attributs dont la clef).
- R21 est en 3NF

Q17 : Donner une décomposition en 3NF de R1 et R2.

Deuxième forme normale :

- R1 doit être normalisée en 2NF en isolant la dépendance partielle $B \rightarrow C$ dans une nouvelle relation et en éliminant sa cible. R1 est donc décomposée en S1 et S2 :

S1 (A, B, D, E, F) et S2 (B, C)

Troisième forme normale :

- S1 n'est pas en 3NF à cause de la dépendance transitive : $D \rightarrow E, F$, qu'il faut isoler dans une nouvelle relation. S1 est donc décomposée en T1 et T2 :

T1 (A, B, D) et T2 (D, E, F)

- S2 est en 3NF (Trivial ! S2 n'a que deux attributs dont la clef primaire).
- R2 doit être normalisée en 3NF en isolant la dépendance transitive $M \rightarrow N$ dans une nouvelle relation et en éliminant sa cible.

R2 est donc décomposée en U1 et U2 : U1 (G, H, I, J, K, L, M) et U2 (M, N). Dans le schéma relationnel précédent, U1 n'est pas en 3NF, car il existe une DF dont la source n'est pas clef primaire et déterminant un attribut n'appartenant à la clef primaire ($I, J \rightarrow K$).

Cette relation doit donc être décomposée en : V1 (G, H, (I, J), L, M) et V2 (I, J, K).

La décomposition en 3NF de R1 et R2 est donc :

R1 :	T1 (<u>A</u> , <u>B</u> , <u>D</u>)	R2 :	V1 (<u>G</u> , <u>H</u> , (<u>I</u> , J), L, <u>M</u>)
	T2 (<u>D</u> , E, F)		V2 (<u>I</u> , J, K)
	S2 (<u>B</u> , C)		U2 (<u>M</u> , N)

Exercice n° 6

Q1 : Intégration de l'attribut CATEGORIE dans la relation PERSONNE

L'attribut CATEGORIE est multivalué (une personne peut appartenir à plusieurs catégories). Deux solutions peuvent donc être envisagées pour normaliser la relation en 1NF :

- création de plusieurs attributs CATEGORIE1, CATEGORIE2, ...
- création de la relation « toute-clef » suivante :
TYPE_PERSONNE (CODE, CATEGORIE)

Le choix de la « meilleure » solution dépend bien sûr de la manière dont seront utilisées les données (aspect traitement).

Q2 : Calcul du montant des charges.

Pour pouvoir calculer le montant des charges, il faut conserver, pour les différents appartements, le coefficient donné. Nous intégrons donc le nouvel attribut dans APPART, dont le schéma devient :

APPART (NUM, TYPE, ADR, VILLE, SURFACE, LOYER, *CODEPROP*, COEF)

Cependant, on sait que COEF est défini en fonction uniquement de la ville et de la surface de l'appartement, i.e. $VILLE, SURFACE \rightarrow COEF$.

La relation ainsi modifiée est en 1NF (attributs atomiques) et en 2NF (clef primaire mono-attribut). Cependant, elle n'est pas en 3NF car :

$NUM \rightarrow VILLE, SURFACE$; $VILLE, SURFACE \rightarrow COEF$; $NUM \rightarrow COEF$.

En appliquant les principes de normalisation en 3NF, on isole la DF posant problème $VILLE, SURFACE \rightarrow COEF$ dans une nouvelle relation et on obtient le schéma suivant :

APPART (NUM, TYPE, ADR, (*VILLE*, *SURFACE*), LOYER, *CODEPROP*)
BAREME (VILLE, SURFACE, COEF)

Le couple d'attributs VILLE, SURFACE est une clef étrangère dans APPART faisant référence à la clef primaire de BAREME.