

# LIFBDW2 – BASES DE DONNÉES AVANCÉES

## TD6 – Formes normales, couvertures

Licence informatique – Automne 2019–2020

Les questions marquées du symbole (†) sont à préparer pour la séance

### Exercice 1 : normalisation (†)

Donner la forme normale la plus avancée des schémas de relation suivants munis de l'ensemble  $\Sigma$  de contraintes. Si ce n'est pas fait, normaliser la relation en la décomposant.

1.  $Employe(NomEmp, NomDept, NomChef)$   
avec  $\Sigma = \{NomEmp \rightarrow NomDept; NomDept \rightarrow NomChef\}$ ;
2.  $Employe(NomEmp, NomEnfant, Salaire)$   
avec  $\Sigma = \{NomEmp \rightarrow Salaire; NomEmp \twoheadrightarrow NomEnfant\}$ ;
3.  $Adresse(Rue, Ville, CodePostal)$   
avec  $\Sigma = \{Rue, Ville \rightarrow CodePostal; CodePostal \rightarrow Ville\}$ ;

### Exercice 2 : définitions de la troisième forme normale

L'objectif de cette exercice est de montrer que les deux définitions **3FN-1** et **3FN-2** ci-après de la troisième forme normale sont équivalentes. On considère les paires  $\langle R, F \rangle$  formées d'un schéma de relation  $R$  et d'un ensemble de  $F$  de dépendances fonctionnelles sur  $R$ . On donne les définitions suivantes :

**Directe** Une DF  $X \rightarrow Y$  est *directe* ssi  $\nexists Z. X \rightarrow Z \wedge Z \twoheadrightarrow X \wedge Z \rightarrow Y$ .

**Premier** Un attribut  $A \in R$  est *premier* s'il appartient à *au moins* une clé minimale de  $R$ .

**2FN**  $\langle R, F \rangle$  est en **2FN** ssi il n'existe pas de DF non-triviale  $X \rightarrow A \in F^+$  avec  $A$  non-premier et  $X$  sous-ensemble *propre*<sup>1</sup> d'une clé minimale.

**3FN-1**  $\langle R, F \rangle$  est en **3FN** ssi  $\langle R, F \rangle$  est en **2FN** et toutes les DFs non-triviales  $X \rightarrow A \in F^+$  avec  $A$  non-premier sont *directes*.

**3FN-2**  $\langle R, F \rangle$  est en **3FN** ssi pour toute DF non-triviale  $X \rightarrow A \in F^+$ , soit  $X$  est (super)clé, soit  $A$  est premier.

1. Montrer que si la condition **2FN** n'est pas respectée, alors la **3FN-2** ne l'est pas non plus.
2. Montrer que s'il existe une DF non-triviale non-directe  $X \rightarrow A \in F^+$  avec  $A$  non-premier, alors la condition **3FN-2** n'est pas respectée.
3. Conclure des deux questions précédentes que la définition **3FN-2** implique la définition **3FN-1**.
4. Montrer que s'il existe une DF  $X \rightarrow A$  avec  $A$  non-premier et  $X$  qui n'est pas (super)clé, alors la condition **3FN-1** n'est pas satisfaite.
5. Conclure sur l'équivalence des définitions **3FN-1** et **3FN-2**.

---

1.  $X$  est un sous-ensemble *propre* de  $Y$  noté  $X \subsetneq Y$  ssi  $X \subseteq Y \wedge X \neq Y$

### Exercice 3 : modélisation de recettes

On souhaite créer une base de données de recettes de cuisine, décrites comme suit :

*Une recette est identifiée par son numéro. Elle a un nom et un type particulier (soupe, entrée, dessert, ...). Elle utilise un ou plusieurs ingrédients (carottes, viande de boeuf, poivre, ...). Un ingrédient est identifié par son numéro et a un nom. Pour chaque ingrédient dans une recette, on précisera sa quantité.*

1. Dresser l'inventaire des DF valides sur la relation universelle  $R = \{NumR, NomR, TypeR, NumI, NomI, Qte\}$
2. Exhiber les problèmes de redondance à partir de cet exemple. La DF  $NumR, NumI \rightarrow Qte$  engendre-t-elle un problème de redondance ?
3. Proposer de façon intuitive une décomposition sans pertes et sans redondance de  $R$ .
4. On veut maintenant ajouter à cette modélisation qu'à chaque recette correspond un ensemble d'ustensiles (attribut *Ustensile*). Cette information se traduit-elle par une nouvelle DF ?
5. Que penser de la décomposition  $R_1 = (NumR, NomR, TypeR)$ ,  $R_2 = (NumR, NumI, Qte)$ ,  $R_3 = (NumI, NomI)$ ,  $R_4 = (NumR, Ustensile)$ .

### Exercice 4 : modélisation d'une base de films

On souhaite modéliser des données sur des films. On analyse les données de la façon suivante.

*Chaque film est identifié par un unique numéro d'exploitation. Un film est également déterminé par son titre et sa date de sortie. Un film peut être une reprise d'un ou plusieurs autres films avec lesquels il possède un taux de ressemblance. Chaque film a un réalisateur et un producteur, et implique un ou plusieurs studios et un ensemble d'acteurs. Pour chaque acteur, on souhaite stocker le nombre de minutes qu'il joue dans un film ainsi que son cachet pour le film.*

*Notons qu'un réalisateur peut-être producteur ou acteur dans un autre film, et inversement. Un réalisateur/producteur/acteur possède un identifiant unique, un nom, un prénom et une adresse. La connaissance du nom, du prénom et de l'adresse permet d'identifier uniquement une personne. Un studio est identifié par un numéro et possède un nom unique et une adresse.*

1. Faire l'inventaire des attributs (relation universelle) et des contraintes (DF, DI) du problème.
2. Donner les clés (minimales) de la relation universelle
3. Proposer une décomposition sans pertes de la relation universelle en précisant les clés primaires et étrangères.

## Corrections

### Solution de l'exercice 1

1. en 2FN : on peut normaliser en  $R_1 = (NomEmp, NomDept)$  et  $R_2 = (NomDept, NomChef)$  qui est en FNBC
2. en 1FN : on peut normaliser en  $R_1 = (NomEmp, salaire)$  et  $R_2 = (NomEmp, NomEnfant)$  qui est en 4FN et donc en FNBC.
3. en 3FN : on ne peut pas faire mieux sans perte de dépendances.

### Solution de l'exercice 2

1.  $\langle ABC, \{AB \rightarrow C, B \rightarrow C\} \rangle$  n'est pas en 2FN.  $\langle ABC, \{A \rightarrow B, B \rightarrow C\} \rangle$  est en 2FN mais pas en 3FN.
2. Supposons  $\langle R, F \rangle$  viole la 2FN, on a donc une dépendance  $X \rightarrow A$  avec  $X \subsetneq K$  où  $K$  est une clé minimale et l'attribut  $A$  non-premier. Comme  $K$  est minimale,  $X$  ne peut pas être clé et donc super clé et la condition 3FN-2 n'est pas respectée.
3. Supposons que  $\langle R, F \rangle$  satisfait une dépendance non-directe  $X \rightarrow Z \rightarrow A$  avec  $Z \not\rightarrow X$  avec  $A$  non-premier. Comme  $Z \not\rightarrow X$ ,  $Z$  n'est pas clé et viole la condition 3FN-2.
4. On a montré par la contraposée que la 3FN-2 implique la 2FN et la 3FN-2 implique l'absence de dépendances non directe, c'est-à-dire la définition de la 3FN-1.
5. Supposons qu'il existe une dépendance  $X \rightarrow A$  avec  $A$  non-premier et  $X$  qui n'est pas (super)clé. Soit  $K$  une clé minimale :
  - $X \subseteq K$  et donc  $X \subsetneq K$  car  $X$  n'est pas super clé. On a une violation de la 2FN sur la dépendance  $X \rightarrow A$
  - $X \not\subseteq K$  et  $X \not\rightarrow K$  car sinon  $X$  serait clé. On a donc une dépendance transitive  $K \rightarrow X \rightarrow A$  qui viole la condition 3FN-1.
6. La dernière question a montré par la contraposée que la 3FN-1 implique la 3FN-2 et d'après la question 4 on a l'équivalence recherchée.

### Solution de l'exercice 3

1. On a les DFs suivantes  $F = \{NumR \rightarrow NomR, TypeR; NumI \rightarrow NomI; NumR, NumI \rightarrow Qte\}$  :
  - Une recette est identifiée par son numéro. Elle a un nom et un type particulier donne  $NumR \rightarrow NomR, TypeR$ ,
  - Un ingrédient est identifié par son numéro et a un nom donne  $NumI \rightarrow NomI$ ,
  - Pour chaque ingrédient dans une recette, on précisera sa quantité donne  $NumI \rightarrow Qte$ .
2. Un problème de redondance a lieu lorsqu'une DF prends deux fois la même valeur à gauche et à droite, autrement dit, si on a une DF  $X \rightarrow Y$  et deux tuples  $t_1$  et  $t_2$  différents tels que  $t_1[XY] = t_2[XY]$ . La dernière DF ne pose pas de problème, puisque la partie gauche est une clé.
3. On va décomposer selon les parties gauches des DFs obtenues  $R_1 = (NumR, NomR, TypeR)$ ,  $R_2 = (NumR, NumI, Qte)$ ,  $R_3 = (NumI, NomI)$ .
4. Non, il n'y a pas de nouvelle DF non-triviale.
5. Cette décomposition engendre une perte de jointure... Il suffit pour cela d'exhiber la relation suivante sur  $R$  :

No. Tuple	NumR	NomR	TypeR	NumI	NomI	Qte	Ustensile
1	1	Couscous	Plat principal	1	Pois Chiches	0,5kg	Couscoussière
2	1	Couscous	Plat principal	2	Agneau	1kg	Plat

On peut vérifier que cette relation ne viole aucune DF, elle est donc valide. On construit alors les relations  $r_1, r_4$  en prenant les projections correspondantes. Ensuite, on fait la jointure naturelle entre ces relations. On obtient :

No. Tuple	NumR	NomR	TypeR	NumI	NomI	Qte	Ustensile
1	1	Couscous	Plat principal	1	Pois Chiches	0,5kg	Couscoussière
3	1	Couscous	Plat principal	2	Agneau	1kg	Couscoussière
4	1	Couscous	Plat principal	1	Pois Chiches	0,5kg	Plat
2	1	Couscous	Plat principal	2	Agneau	1kg	Plat

Cette relation est différente de la première, il y a donc perte de jointure et la décomposition n'est pas correcte. On devrait préciser que soit l'ustensile dépend de l'ingrédient et de la recette ou alors introduire une MVD pour préciser l'indépendance entre ustensiles et ingrédients dans une recette.

## Solution de l'exercice 4

- On obtient l'ensemble de DFs et DIIs suivant :
  - $IDFilm \rightarrow Titre, Date$
  - $Titre, Date \rightarrow IDFilm$
  - $IDFilm, IDReprise \rightarrow Ressemblance$
  - $IDFilm \rightarrow IDProd, IDReal$
  - $IDFilm \twoheadrightarrow IDStudio$  ou  $IDFilm \twoheadrightarrow IDActeur$
  - $IDFilm, IDActeur \rightarrow Duree, Cachet$
  - $IDPersonne \rightarrow NomPersonne, Prenom, AdressePersonne$
  - $IDStudio \rightarrow NomStudio, AdresseStudio$
  - $IDProd \subseteq IDPersonne$
  - $IDReal \subseteq IDPersonne$
  - $IDActeur \subseteq IDPersonne$
  - $IDReprise \subseteq IDFilm$
- $IDFilm, IDReprise, IDActeur, IDStudio$  et  $Titre, Date, IDReprise, IDActeur, IDStudio$ . Comme on a indépendance via les DMV entre  $IDFilm, IDActeur, IDStudio$  on va pouvoir décomposer cette clef.
- On va décomposer la relation universelle ainsi :
  - $R_1 = \{ \underline{IDFilm}, Titre, Date, \#IDProd, \#IDReal \}$ , on choisit par exemple  $IDFilm$  en clé primaire et on ajoute une contrainte d'unicité sur  $Titre, Date$  comme alternative.
  - $R_2 = \{ \underline{IDFilm}, \#IDActeur, Duree, Cachet \}$
  - $R_3 = \{ \underline{IDFilm}, \#IDStudio \}$
  - $R_4 = \{ \underline{IDFilm}, \#IDReprise, Ressemblance \}$
  - $R_5 = \{ \underline{IDStudio}, NomStudio, AdresseStudio \}$
  - $R_6 = \{ \underline{IDPersonne}, NomPersonne, Prenom, AdressePersonne \}$