

## Dépendances fonctionnelles et normalisations

### 1 Définitions

Une *dépendance fonctionnelle* est un énoncé du type

$$A_1, A_2, \dots, A_k \rightarrow A_{k+1}, \dots, A_n$$

où  $A_1, A_2, \dots, A_k, A_{k+1}, \dots, A_n$  sont des attributs (colonnes) d'une base de données. Elle signifie que deux tuples ayant la même valeur sur  $A_1, \dots, A_k$  doivent avoir la même valeur sur chaque colonnes  $A_{k+1}, \dots, A_n$  (en français :  $A_1, \dots, A_k$  déterminent  $A_{k+1}, \dots, A_n$ ). On dit que les attributs  $A_{k+1}, \dots, A_n$  *dépendent fonctionnellement* de  $A_1, A_2, \dots, A_k$ .

Un ensemble de dépendances fonctionnelles  $\mathcal{F}$  est *minimal* si aucune dépendance ne peut être déduite des autres en utilisant les règles suivantes :

- **trivialité** : si  $Y \subseteq X$  alors  $X \rightarrow Y$
- **augmentation** : si  $X \rightarrow Y$  alors  $X, Z \rightarrow Y, Z$  pour toute suite d'attributs  $Z$ .
- **transitivité** : si  $X \rightarrow Y$  et  $Y \rightarrow Z$  alors  $X \rightarrow Z$
- **union** : si  $X \rightarrow Y$  et  $X \rightarrow Z$  alors  $X \rightarrow Y, Z$
- **décomposition** si  $X \rightarrow Y$  et  $Z \subseteq Y$  alors  $X \rightarrow Z$

La *clôture transitive* des attributs  $A_1, \dots, A_k$  pour un ensemble de dépendances fonctionnelles  $\mathcal{F}$  est l'ensemble des attributs  $B_1, \dots, B_\ell$  qui dépendent fonctionnellement de  $A_1, \dots, A_k$ . On la note  $[A_1, \dots, A_k]_{\mathcal{F}}^+$ . Un ensemble d'attributs  $A_1, \dots, A_k$  est *une clef pour une relation*  $R(B_1, \dots, B_\ell)$  si ce sont des attributs de  $R$  et si sa clôture transitive contient  $B_1, \dots, B_\ell$ . C'est une *clef primaire* si elle est minimale, c'est-à-dire, aucun sous-ensemble strict de cette clé n'est une clé.

Un schéma est en :

- *Forme normale 1* si tout attribut est atomique.
- *Forme normale 2* si un attribut ne fait pas partie d'une clef, il ne peut pas dépendre d'une partie stricte d'une clef primaire.
- *Forme normale 3* Pour toute dépendance fonctionnelle, le membre de gauche contient une clef ou tout attribut du membre de droit appartient à une clef primaire.

Un schéma et un ensemble de dépendances fonctionnelles peut se décomposer en une collection de schémas, dans le sens où chaque relation  $R$  peut se décomposer en  $R_1, \dots, R_k$  tels que  $R_i = \pi_i(R)$  pour une certaine projection  $\pi_i$ .

On dit cette décomposition *sans perte d'information* si toute relation  $R$  du schéma d'origine peut être retrouvée à partir des relations  $R_1, \dots, R_k$  :  $R = \pi_1(R) \bowtie \dots \bowtie \pi_k(R)$ .

On dit que cette décomposition *respecte les dépendances fonctionnelles* si celles-ci sont toujours satisfaites par la nouvelle décomposition.

**Exemple/exercice.** En exemple, considérons la table suivante qui liste les tournois de tennis et leur gagnant :

Année	Tournoi	Surface	Gagnant	Sexe	Nationalités
1984	Roland Garos H	Terre battue	Yannick Noah	M	Française
1992	Wimbledone H	Herbe	Andre Agassi	M	Américaine, Arménienne
2003	Australian Open F	Hard	Serena Williams	F	Américaine

On a les dépendances fonctionnelles suivantes :

1. (Année, Tournoi)  $\rightarrow$  Gagnant
2. Tournoi  $\rightarrow$  Surface
3. Gagnant  $\rightarrow$  (Sexe, Nationalités).

Quelles sont les clefs de cette table ? Quelles sont les clefs primaires ? Pour chaque forme normale sus-citée, expliquez pourquoi la table n'est pas dans cette forme normale et réparer la table pour qu'elle respecte cette forme normale.

## 2 TD

### 2.1 Exercice 1

Soient  $R$  une relation de schéma  $\mathcal{A} = (A, B, C, D)$  (on le note  $R(A, B, C, D)$ ) et  $\mathcal{F} = \{A \rightarrow B; B \rightarrow C\}$ .

1. Quelle est la fermeture  $[A]_{\mathcal{F}}^+$  de  $A$  ?
2. Quelles sont les clés de  $R$  ? Les clés primaires ?

### 2.2 Exercice 2 : Dédution de dépendance

1. Soit le schéma  $R(A, B, C, D, E, G, H)$ . Soit  $\mathcal{F} = \{A, B \rightarrow C; B \rightarrow D; C, D \rightarrow E; C, E \rightarrow G, H; G \rightarrow A\}$ . Est-ce que les dépendances
  - $A, B \rightarrow E$ ,
  - $B, G \rightarrow C$ ,
  - $A, B \rightarrow G$
 sont déductibles de  $\mathcal{F}$  (i.e. font partie de la fermeture de  $\mathcal{F}$ ) ?
2. Soit  $\mathcal{F}_1 = \{A \rightarrow B; C, E \rightarrow H; C \rightarrow E; A \rightarrow C, H\}$ . L'ensemble  $\mathcal{F}$  est-il minimal ?
3. Soit  $\mathcal{F}_2 = \{A \rightarrow B, C; C \rightarrow E, H\}$ . Les deux ensembles de dépendances fonctionnelles  $\mathcal{F}_1$  et  $\mathcal{F}_2$  sont-ils équivalents ?

### 2.3 Exercice 3 : Décomposition et perte d'information

1. On considère le schéma de relation  $R(A, B, C)$  et la dépendance fonctionnelle suivante :

$$\mathcal{F} = \{A, B \rightarrow C\}.$$

Déterminer si la décomposition suivante est sans perte d'information :

$$R_1(A, B), R_2(B, C).$$

Étudiez le cas de la table suivante :

A	B	C
1	2	3
4	2	5

2. On considère le schéma de relation  $R(A, B, C, D, E)$  et les dépendances fonctionnelles suivantes :

$$\mathcal{F} = \{A \rightarrow C; B \rightarrow C; C \rightarrow D; D, E \rightarrow C; C, E \rightarrow A\}.$$

Déterminer si la décomposition suivante est sans perte d'information :

$$R_1(A, D), R_2(A, B), R_3(B, E), R_4(C, D, E), R_5(A, E).$$

La même question pour la décomposition :

$$R_1(A, D), R_2(A, B), R_3(B, E), R_4(C, D), R_5(D, E), R_6(A, E).$$

## 2.4 Exercice 4 : Normalisation

On considère le schéma de relation  $R(C,T,H,S,E,N)$  (Cours, Enseignant, Horaire, Salle, Étudiant, Note) et les dépendances fonctionnelles suivantes :

$$\mathcal{F} = \{C \rightarrow T; \quad H, S \rightarrow C; \quad H, T \rightarrow S; \quad C, E \rightarrow N; \quad H, E \rightarrow S\}.$$

1. Calculer une clé.
2. Mettre en Boyce-Codd Normal Form (BCNF), donner plusieurs résultats possibles.

## 2.5 Exercice 5 : Normalisation avancée

Nous voulons mettre en forme la base de données UFR dont le schéma est :

UFR(NoTD, Salle, Jour, Heure,  
NoEnseignant, Nom\_Enseignant, Prenom\_Enseignant,  
COD\_MOD, Diplome, Matiere, NoEtudiant,  
Nom\_Etudiant, PreNom\_Etudiant, Adresse, Date\_Inscription)

Sachant que :

- Un code module (COD\_MOD) précise à la fois un diplôme et une matière.
- Les TD sont annuels et il y a un et un seul TD par semaine dans chaque module.
- Un TD est assuré par un seul enseignant.
- Un numéro de TD est relatif à un module.
- Un enseignant peut assurer plusieurs TD.
- Un étudiant peut être inscrit dans plusieurs modules, mais dans un seul TD par module.
- Date\_Inscription est la date d'inscription d'un étudiant à un module.

Les dépendances fonctionnelles associées à ce schéma sont :

- (a) NoEtudiant  $\rightarrow$  Nom\_Etudiant, PreNom\_Etudiant, Adresse
- (b) NoEnseignant  $\rightarrow$  Nom\_Enseignant, Prenom\_Enseignant
- (c) COD\_MOD  $\rightarrow$  Diplome, Matiere
- (d) Diplome, matiere  $\rightarrow$  COD\_MOD
- (e) Salle, Jour, Heure  $\rightarrow$  NoTD, COD\_MOD
- (f) COD\_MOD, NoTD  $\rightarrow$  Salle, Jour, Heure, NoEnseignant
- (g) COD\_MOD, NoEtudiant  $\rightarrow$  NoTD, Date\_Inscription
- (h) NoEnseignant, Jour, Heure  $\rightarrow$  Salle
- (i) NoEtudiant, Jour, Heure  $\rightarrow$  Salle
- (j) Salle, Jour, Heure  $\rightarrow$  NoEnseignant
- (k) NoEnseignant, COD\_MOD, NoTD  $\rightarrow$  Salle, Jour, Heure

1. Ce schéma a plusieurs anomalies. Donnez des exemples de redondances et de valeurs qui doivent être **NULL** à certains endroits.
2. L'ensemble des dépendances est long. Peut-être trop?
  - (a) Calculer les clôtures transitives de NoEnseignant et du triplet Diplome, NoEtudiant, matiere.
  - (b) Les dépendances
    - Salle, Jour, Heure  $\rightarrow$  NoEnseignant,
    - NoEnseignant, COD\_MOD, NoTD  $\rightarrow$  Salle, Jour, Heure
    - COD\_MOD, NoTD  $\rightarrow$  Sallesont-elles redondantes? En déduire un ensemble minimal de dépendances fonctionnelles.

(c) Donnez au moins trois clefs différentes de cette relation.

On veut maintenant couper le schéma pour éviter ces anomalies. On veut que cette découpe vérifie les conditions suivantes :

- (i) Qu'elle se fasse sans pertes d'informations.
- (ii) Qu'elle préserve les dépendances fonctionnelles.
- (iii) Qu'elle n'ait plus besoin de `NULL`.
- (iv) Qu'elle minimise les redondances.
- (v) Et idéalement qu'elle respecte l'une des formes normales.

3. Donnez des clefs pour ces relations. Dites lesquelles des 5 conditions la décomposition suivante respecte. Dessinez les pattes de corbeaux associées.

```
TD(NoTD, COD_MOD, Jour, Heure, Salle, No_Enseignant,  
    Nom_Enseignant, Prenom_Enseignant)
```

```
INSCRIPTION(NoEtudiant, Nom_Etudiant, PreNom_Etudiant, Adresse,  
             COD_MOD, Diplome, Matiere, Date_Inscription, NoTD)
```

4. Donnez des clefs pour ces relations. Dites lesquelles des 5 conditions la décomposition suivante respecte. Dessinez les pattes de corbeaux associées.

```
ENSEIGNEMENT(NoTD, COD_MOD, Jour, Heure, Salle,  
              No_Enseignant, Nom_Enseignant, Prenom_Enseignant)  
ETUDIANT(NoEtudiant, Nom_Etudiant, PreNom_Etudiant, Adresse)  
INSCRIPTION(NoEtudiant, COD_MOD, Date_Inscription, NoTD)  
MODULE(COD_MOD, Diplome, Matiere)
```

5. Donnez des clefs pour ces relations. Dites lesquelles des 5 conditions la décomposition suivante respecte. Dessinez les pattes de corbeaux associées.

```
ENSEIGNEMENT_PLANNING(NoEnseignant, Nom_Enseignant,  
                       Prenom_Enseignant, Jour, Heure, NoTD, COD_MOD)
```

```
ENSEIGNEMENT_SALLE(NoEnseignant, Jour, Heure, Salle)
```

```
ETUDIANT_INSCRIPTION(NoEtudiant, Nom_Etudiant, PreNom_Etudiant,  
                      Adresse, NoTD, COD_MOD, Date_Inscription, Diplome, Matiere)
```

```
Etudiant_Planning(NoEtudiant, Jour, Heure, COD_MOD)
```

6. Proposez une décomposition qui satisfait les 5 conditions. Donnez les clefs et les pattes de corbeaux.