

Bases de Données (USSIOG) Normalisation

Thibault Bernard

Thibault.bernard@univ-reims.fr

Conception des BD

Emprunt_Amis (N°DVD, Nom, Prénom, Contact, Date_emp)

N°DVD	Nom	Prénom	Contact	Date_emp
120	Valonne	Michelle	0609111134	12/09/17
430	Rivette	Jean	0614567565	13/03/18
532	Gaillot	Gael	0326123454	15/4/17
634	Léponge	Bob	0615653575	25/2/18
253	Guerald	Felix	0493457584	26/1/17

Pourquoi cette relation a t-elle des anomalies?

Exemples

- Anomalie de mise à jour
- Anomalie d'insertion
- Anomalie de suppression

Normalisation

- **Objectif** : supprimer les anomalies précédentes et élimination des redondances (éviter les incohérences et minimiser l'espace de stockage)
- Normaliser une relation consiste à décomposer une relation ayant des anomalies en plusieurs sous-relations sans anomalies
- La normalisation s'inscrit dans la partie conception d'une base de données. C'est un outil théorique de vérification
- La normalisation repose sur les liens sémantiques entre attributs : les dépendances fonctionnelles

Dépendances fonctionnelles

- **Définition** : Soit R une relation, X et Y 2 ss-ens d'attributs, Y dépend fonctionnellement de X ou X détermine fonctionnellement Y

ssi $\forall r$ extension de R , $\forall t_1, t_2$ tuples de r , on a

$$\Pi_X(t_1) = \Pi_X(t_2) \Rightarrow \Pi_Y(t_1) = \Pi_Y(t_2)$$

- A chaque valeur de X dans R correspond exactement 1 seule valeur de Y dans R

- **Notation**

X détermine Y : $X \rightarrow Y$

X, Y ou XY correspond à $X \cup Y$

Notion de Dépendance Fonctionnelle

- Le salaire dépend du diplôme
- Le montant du contrat d'assurance automobile dépend du malus
- Le numéro Insee d'une personne correspond exactement à son nom, son prénom, à son âge ...
- **Notation**
- N°DVD → Titre
- N°DVD → Réalisateur
- Titre, Réalisateur → Genre

Propriétés de Dépendances Fonctionnelles

- Pour une relation R donnée, l'ensemble des DF est très grand.
Objectif : réduire cet ensemble à une taille manipulable.

Règles d'Armstrong

Soient X, Y et Z des groupes d'attributs d'une relation R

- **Réflexivité** Si $Y \subseteq X$ alors $X \rightarrow Y$
- **Augmentation** Si $X \rightarrow Y$ alors $X, Z \rightarrow Y, Z$
- **Transitivité** Si $X \rightarrow Y$ et $Y \rightarrow Z$ alors $X \rightarrow Z$

Exercice

Soit $R(A,B,C,D,E)$ avec comme Dépendances Fonctionnelles

- $A \rightarrow B$
- $D \rightarrow E$

Montrer que $(A,D) \rightarrow (B,E)$.

Propriétés dérivées

Soient X, Y, Z, W 4 groupes d'attributs d'une relation R

- **Union** Si $X \rightarrow Y$ et $X \rightarrow Z$ alors $X \rightarrow (Y, Z)$
- **Décomposition** Si $X \rightarrow (Y, Z)$ alors $X \rightarrow Y$ et $X \rightarrow Z$
- **Pseudo-transitivité** Si $X \rightarrow Y$ et $W, Y \rightarrow Z$ alors $W, X \rightarrow Z$

Montrer ces trois propriétés !

Exercice

Soit $R(A,B,C,D,E,F)$ avec comme dépendances fonctionnelles

- $A \rightarrow BC$
- $B \rightarrow E$
- $CD \rightarrow EF$

Montrer que $AD \rightarrow E$ peut être déduite à l'aide des propriétés.

Fermeture d'un ensemble de Dépendances Fonctionnelles

- **Définition** L'ensemble de toutes les DF qui peuvent être déduites à partir d'un ensemble de DF de départ S d'une relation R est appelé fermeture de S et est notée S^+ .
- Comment calculer S^+ ? Pourquoi calculer S^+ ?
- Avec les règles d'Armstrong : Représente toutes les contraintes que le SGBD doit tester.

Fermeture d'un ensemble de DF

- On a vu qu'à partir de certaines DF, à l'aide des règles d'Armstrong, on peut en déduire d'autres.
- **Définition** L'ensemble des DF, noté S^+ , obtenues par application des règles d'Armstrong s'appelle la fermeture de S
 - DF initiales enrichies de toutes les DF obtenues par transitivité et augmentation
 - Le calcul est fastidieux.
- **Définition** Deux ensembles de DF S_1 et S_2 sont dits équivalents ssi ils ont même fermeture i.e. $S_1^+ = S_2^+$

Fermeture d'un attribut

- Savoir si une DF $X \rightarrow Y$ appartient à la fermeture de S , ens. de DF de départ. On calcule la fermeture de X , noté X^+ , i.e. l'ensemble des attributs qui dépendent de X .
- Algorithme de saturation

Données: X groupe d'attributs, S ens. de DF

Sortie: X^+ fermeture de X

$Res \leftarrow X$

Tant que Res évolue faire

Pour chaque DF $Y \rightarrow Z$ de S

Si Y est dans Res alors $Res \leftarrow Res \cup Z$

finfaire

Exemple

Soit $R(A,B,C,D,E,F)$ avec $S=\{A \rightarrow BC, E \rightarrow CF, B \rightarrow E, CD \rightarrow EF\}$

Calculer AB^+

On trouve $AB^+ = \{A,B,C,E,F\}$

Corollaire Pour un ensemble S de DF, une DF $X \rightarrow Y$ appartient à S^+ ssi Y est un sous ensemble de X^+ .

Calcul de S+

- Il est facile d'en déduire une méthode pour calculer S+. Soit R une relation de schéma $R(A_1, \dots, A_n)$ avec S un ensemble de DF, on calcule la fermeture de tout groupe d'attributs X de l'ensemble des parties de $\{A_1, \dots, A_n\}$

Fastidieux !

- Résumé : pour calculer la fermeture d'un groupe attributs voir la fermeture d'un ensemble de DF, on a deux méthodes:
 - 1) les règles d'Armstrong
 - 2) l'algorithme de saturation

Retour sur la notion de clé

Définition Clé candidate d'une relation $R(A_1, \dots, A_n)$ est un groupe d'attributs X tel que

- Unicité $X \rightarrow A_1, \dots, A_n$
- Irréductibilité $\forall X' \subset X$ et NON $(X' \rightarrow A_1, \dots, A_n)$

Remarque Une super clé = propriété d'unicité uniquement (par ex l'ensemble des tuples d'une relation est une super clé)

Couverture irrédondante

Définition 2 ensembles de DF sont dits équivalents ssi ils ont la même fermeture

Parmi l'ensemble de toutes les DF d'un ensemble de DF initial, certaines DF sont redondantes, d'autres découlent de celles qui sont exprimées.

Objectif : construire un ensemble de DF équivalent à l'ensemble initial mais uniquement avec les DF les plus « pertinentes ».

A partir de cet ensemble, on pourra générer toutes les autres DF.

Couverture irrédondante (minimale, irréductible)

Soit R une relation, S un ens de DF. But: isoler les DF les plus pertinentes. En effet, certaines DF sont redondantes, certaines se déduisent des autres.

Définition Une DF est dite élémentaire si elle est de la forme $X \rightarrow A$, où X est un groupe d'attributs et A est un attribut unique n'appartenant pas à X et où il n'existe pas $X' \subset X$ tel que $X' \rightarrow A$.

- Le dépendant (A) ne contient qu'un seul attribut
- Le déterminant (X) est irréductible

DF non redondante

A partir d'un ensemble SE de DF élémentaires initial, l'ensemble de toutes les DF élémentaires sont obtenues uniquement par la règle de transitivité.

Soit SE un ensemble de DF élémentaires. Une DF élémentaire est dite redondante si elle se déduit des autres par la règle de transitivité.

Dans $S = \{A \rightarrow B, A \rightarrow C, B \rightarrow C\}$, la DF $A \rightarrow C$ est redondante car elle peut être obtenue par transitivité des 2 autres.

Deux méthodes pour éliminer les DF redondantes: méthode directe ou méthode suivante

Éliminer les DF redondantes

La DF élémentaire $X \rightarrow A$ de l'ensemble de DF S est elle redondante?

On calcule X^+ sur $S \setminus \{X \rightarrow A\}$.

Si on obtient encore A alors $X \rightarrow A$ est redondante.

Définition Une couverture irrédondante d'un ensemble S de DF est un ensemble noté $Irr(S)$ de dépendances élémentaires vérifiant que

1. Toute DF élémentaire est dans S^+
2. Aucune DF de $Irr(S)$ n'est redondante

Remarque 2. signifie que $\forall s \in Irr(S), \{Irr(S) \setminus s\}^+ \neq \{Irr(S)\}^+$

Construction de la couverture irrédondante

Soit S un ensemble de DF initial. Pour construire $Irr(S)$, on simplifie cet ensemble: $\forall s : X \rightarrow Y$ de S ,

[R1] rendre le membre droit de s atomique

[R2] rendre le membre gauche de s irréductible

[R3] éliminer les DF redondantes de l'ensemble obtenu par R1 et R2.

Pour 1. On applique la règle de décomposition

Pour 2. On ne considère que les membres gauche X de s ayant plus d'un attribut. On applique l'algorithme de saturation sur $X \setminus \{\text{un attribut composant } X\}$. Si on retrouve le membre droit de s alors le membre gauche n'était pas irréductible.

Pour 3. On applique la méthode décrite page précédente ou par observation directe.

Calcul de $\text{Irr}(S)$ pour la relation $R(A,B,C,D)$ avec $S=\{A \rightarrow BC, B \rightarrow C, A \rightarrow B, AB \rightarrow C, AC \rightarrow D\}$

[R1] $A \rightarrow BC$ donne $A \rightarrow B$ et $A \rightarrow C$

S devient donc $SR1 = \{A \rightarrow C, B \rightarrow C, A \rightarrow B, AB \rightarrow C, AC \rightarrow D\}$

[R2] Les membres gauche de $AB \rightarrow C, AC \rightarrow D$ sont ils irréductibles?

Pour $AB \rightarrow C$, on calcule

$\{AB \setminus \{A\}\} +$ sur $SR1$ donne $B + = \{B, C\}$, on retrouve C , donc AB n'est pas irréductible et A peut être retiré de AB

Pour $AB \rightarrow C$

$\{AB \setminus \{B\}\} + \text{sur } S \text{ R1}$ qui donne $A + = \{A, B, C, D\}$, on retrouve C , donc AB n'est pas irréductible et B peut être retiré de AB . On obtient $A \rightarrow C$ et $B \rightarrow C$ qui existent déjà.

Pour $AC \rightarrow D$, on calcule

$\{AC \setminus \{A\}\} + \text{sur } S \text{ R1}$ donne $C + = \{C\}$, A ne peut pas être retiré de AD

$\{AC \setminus \{C\}\} + \text{sur } S \text{ R1}$ donne $A + = \{A, B, C, D\}$, C peut être retiré de AC

On obtient $A \rightarrow D$

Après la règle R2. on obtient

$SR2 = \{A \rightarrow B, A \rightarrow C, A \rightarrow D, B \rightarrow C\}$

[R3] élimination des DF redondantes

1. Observation: $A \rightarrow C$ est la seule DF qui peut être obtenu par transitivité à partir des autres donc elle est redondante

2. Méthode: calculer respectivement

Pour $A \rightarrow B$, $A +$ sur $S \ R2 \setminus \{A \rightarrow B\}$ non redondant

Pour $A \rightarrow C$, $A +$ sur $S \ R2 \setminus \{A \rightarrow C\}$ redondant

Pour $A \rightarrow D$, $A +$ sur $S \ R2 \setminus \{A \rightarrow D\}$ non redondant

Pour $B \rightarrow C$, $B +$ sur $S \ R2 \setminus \{B \rightarrow C\}$ non redondant

Résultat:

• $\text{Irr}(S) = \{A \rightarrow B, A \rightarrow D, B \rightarrow C\}$

Le calcul de l'ensemble de toutes les DF élémentaires à partir d'un ensemble initial de DF élémentaires se fait par application de la règle de transitivité.

Normalisation

Objectif : éliminer les anomalies de conception de la BD

- Relation est dans une forme normale particulire 1NF, 2NF, 3NF ... si elle satisfait un ensemble de conditions prédéfinie. $1NF \supseteq 2NF \supseteq 3NF$
- Procédure: Transformer une relation qui se trouve dans une forme normale particulière en un sous ensemble de relations sans une forme normale plus satisfaisante.
- Procédure réversible (aucune perte d'informations)

Décomposition sans perte d'informations

Soit R une relation. La décomposition de $R = R_1 \bowtie R_2 \bowtie \dots \bowtie R_n$ est une décomposition sans perte d'information.

Théorème de décomposition

Toute relation $R(A_1, A_2, \dots, A_n)$ est décomposable sans perte d'informations en R_1 et R_2 si il existe une DF $X \rightarrow Y$ et

- $R_1 = \Pi_{XY}(R)$ et
- $R_2 = \Pi_{\Delta \setminus \{Y\}}(R)$ avec $\Delta = A_1, A_2, \dots, A_n$

Préservation des DF

Soit une relation R et S l'ensemble des DF définies sur R . La décomposition de R en $R_1, \dots, R_i, \dots, R_k$ préserve l'ensemble des DF S définies sur R si $(\bigcup S_i)$ et S sont équivalents (i.e. $\{(\bigcup S_i)\}^+ = S^+$) où S_i est l'ensemble des DF de S qui ne concernent que R_i .

Il existe des algorithmes permettant de vérifier si une décomposition préserve un ensemble de DF et de vérifier si une décomposition est sans perte.

Démarche globale

- Calcul de la couverture irrédondante
- Calcul des clés candidates
- Normalisation en 2NF
- Normalisation en 3NF
- Normalisation en 3BCNF (si préservation des DF)

1NF - 2NF

Définition Une relation est 1NF si chacun des attributs qui la compose ne peut contenir qu'une valeur atomique.

Définition Une relation est 2NF ssi R est 1NF et tout attribut n'appartenant pas à une clé candidate ne dépend pas d'une partie de cette clé

$R(A,B,C,D,E)$ avec $S = \{AB \rightarrow CDE, B \rightarrow C\}$

$R(A B C D E)$ se décompose en $R_1(A B D E)$ et $R_2(B C)$

3NF

Définition Une relation R est 3NF ssi R est 2NF et tout attribut n'appartenant pas à une clé ne dépend pas d'un autre attribut n'appartenant pas à cette clé.

$R(A,B,C,D,E)$ avec $S = \{AB \rightarrow CDE, D \rightarrow E\}$

$R(A B C D E)$ se décompose en $R_1(A B C D)$ et $R_2(D E)$

Théorème Toute relation admet une décomposition en 3NF sans perte d'information et avec préservation des dépendances fonctionnelles.

Algorithme de normalisation en 3NF

1. $R(D)$ et S un ensemble initiale de DF
2. Calculer $Irr(S)$
3. Partition de $Irr(S)$ en S_1, S_2, \dots, S_p tq toutes les DF d'un même groupe aient même membre gauche
4. Construction des $R_j(D_j)$ avec D_j constitué de l'union des attributs de S_j
5. Si aucune des clés de la relation initiale n'est contenue dans une des relations, ajouter une relation constituée d'une des clés candidates.

Forme normale 3BCNF

- DF entre parties de clés et attributs non clés
- DF entre attributs non clés
- D'autres anomalies: DF entre attributs non clés vers une partie de clés

Définition Une relation R est en 3BCNF ssi R est 1NF et que les seules DF élémentaires soient celles dans lesquelles une clé détermine un attribut.

Théorème Toute relation admet une décomposition en 3BCNF sans perte d'information mais parfois avec perte de DF.

Algorithme de mise en 3BCNF

Relation $R(D)$ et S ensemble de DF associés

1. $T \leftarrow \emptyset$
2. $G \leftarrow \emptyset$
3. Calculer $\text{Irr}(S)$; $\text{Clé}(S)$;
4. Pour chaque $X \rightarrow A$ de $\text{Irr}(S)$ faire
 - Si X est non clé de R alors
 - $T \leftarrow T \cup R_i(X,A)$
 - $G \leftarrow G \cup \{\text{DF définies sur } R_i(X,A)\}$
5. Enlever de R tous les attributs déterminés (membre droit) par G
6. $T \leftarrow T \cup \{R\}$
7. $G \leftarrow G \cup \{\text{DF définies sur } R(\text{réduit})\}$
8. Supprimer de T toutes les relations $R_i(D_i)$ tq il existe $R_j(D_j)$ avec $D_j \subseteq D_i$