

Cours de bases de données, ,
<http://sql.bdpedia.fr>

La normalisation relationnelle

La normalisation

La normalisation, c'est l'art de créer des schémas relationnels où toutes les relations sont en troisième forme normale, et sans perte d'information.

Dans cette session :

- Exemple de décomposition d'un schéma
- Algorithme de normalisation
- Approche globale

Ces diapositives correspondent au support en ligne disponible sur le site <http://sql.bdpedia.fr/>

Point de départ : relation globale et dépendances

On part d'un schéma contenant tous les attributs connus. Commençons par le suivant :

Appart(idAppart, surface, idImmeuble, nbEtages, dateConstruction)

On identifie les dépendances fonctionnelles

idAppart → *surface, idImmeuble, nbEtages, dateConstruction*

et

idImmeuble → *nbEtages, dateConstruction*

Remarque: En troisième forme normale ?

Non, car la seconde DF montre une dépendance dont la partie gauche n'est pas la clé, *idAppart*.

La décomposition

On identifie les dépendances fonctionnelles **minimales** et **directes**.

idAppart → *surface*, *idImmeuble*

et

idImmeuble → *nbEtages*, *dateConstruction*

On crée une relation pour chacune :

- Appart(*idAppart*, *surface*, *idImmeuble*)
- Immeuble (*idImmeuble*, *nbEtages*, *dateConstruction*)

On obtient des relations en 3FN, sans perte d'information.

Regardons les occupants

Le schéma global de départ est le suivant :

Occupant(idPersonne, nom, prnom, idAppart, surface)

La clé est le couple (*idPersonne, idAppart*). On a les dépendances suivantes :

idAppart → *surface*

et

idImmeuble → *nbEtages, dateConstruction*

Remarque: En troisième forme normale ?

Non, car pour les **deux** DF, la partie gauche n'est pas la clé,

La décomposition

Comme avant, à partir des DF minimales et directes.

- Personne(idPersonne, prénom, nom)
- Appart (idAppart, surface)

Pas suffisant, car on a perdu le lien entre les appartements et les personnes.

On ajoute une relation avec la clé.

- Occupant (idPersonne, idAppart)

On est en 3FN, sans perte d'information (jointures pour reconstituer)

Algorithme de normalisation

On part d'un schéma de relation R **global** et d'un ensemble de dépendances fonctionnelles minimales et directes.

On détermine alors les clés de R

- Pour chaque DF minimale et directe $X \rightarrow A_1 \cdots, A_n$, on crée une relation $(X, A_1 \cdots, A_n)$ de clé X
- Pour chaque clé C non représentée dans une des relations précédentes, on crée une relation (C) de clé C .

On obtient un schéma normalisé

Et en pratique ?

Pas tout à fait suffisant : les identifiants n'existent pas naturellement dans la vraie vie...

Exemple pour les films :

(titre, anne, prnomMes, nomMES, anneNaiss)

Pas de DF... Il faut les **ajouter** et décider des **entités** et de leur identifiant.

Ici, décidons que nous avons des entités Film et Réalisateur, avec idFilm et idRéalisateur. Soit :

idFilm → titre, anne, idRéalisateur

et

idRéalisateur → prnomMes, nomMES, anneNaiss

Maintenant on normalise et on obtient un schéma normalisé.

Illustration : table de départ

À partir de cette table pleine d'anomalies.

titre	année	nomMES	prénomMES	annéeNaiss
Alien	1979	Scott	Ridley	1943
Vertigo	1958	Hitchcock	Alfred	1899
Psychose	1960	Hitchcock	Alfred	1899
Kagemusha	1980	Kurosawa	Akira	1910
Volte-face	1997	Woo	John	1946
Pulp Fiction	1995	Tarantino	Quentin	1963
Titanic	1997	Cameron	James	1954
Sacrifice	1986	Tarkovski	Andrei	1932

On obtient après normalisation

idFilm	titre	année	idR
1	Alien	1979	101
2	Vertigo	1958	102
3	Psychose	1960	102
4	Kagemusha	1980	103
5	Volte-face	1997	104
6	Pulp Fiction	1995	105
7	Titanic	1997	106
8	Sacrifice	1986	107

La table des films

idR	nom	prénom	annéeNaiss
101	Scott	Ridley	1943
102	Hitchcock	Alfred	1899
103	Kurosawa	Akira	1910
104	Woo	John	1946
105	Tarantino	Quentin	1963
106	Cameron	James	1954
107	Tarkovski	Andrei	1932

La table des réalisateurs

Schéma normalisé et sans perte d'information

À retenir

Il est **toujours** possible de se ramener à un schéma normalisé.

- En déterminant les DF et les clés
- En appliquant l'algorithme de normalisation

Problème : les DF ne sont p s donn es naturellement. Il faut les "injecter" par un processus de conception bas  sur les entit s et leurs identifiants.

D marche globale : m thode entit  / association.