

BASES DE DONNÉES : INTRODUCTION ET VOCABULAIRE

I. Bases de données relationnelles

1) Introduction

Une base de données (BDD ou DB, database en anglais) est une collection (généralement importante) de données cohérentes entre elles. Elle modélise des *entités* et des *associations* entre ces entités.

On utilise un Système de Gestion de Bases de Données (SGBD) pour stocker et manipuler des bases de données : l'utilisateur n'a pas à s'occuper de la façon dont sont stockées les informations, ni des algorithmes employés pour récupérer efficacement les informations demandées.

Les données sont représentées dans des *tables*, aussi appelées *relations*. Chaque nom de colonne est appelé un *attribut*. Un attribut prend ses valeurs dans un ensemble appelé *domaine*. Le *schéma de relation* d'une table est le n -uplet des attributs de la table ; il donne donc la liste des colonnes d'une relation. On précise parfois dans le schéma le domaine de chaque attribut.

Définir une base de données, c'est donc se donner un ensemble de schémas de relation,

- chaque *schéma de relation* $S = ((A_1, \text{dom}(A_1)), \dots, (A_n, \text{dom}(A_n)))$ étant un n -uplet d'attributs A_i ,
- chaque attribut A_i possédant un domaine $\text{dom}(A_i)$,
- et pour chaque schéma S , une relation $R(S)$, ensemble fini de n -uplet de $\text{dom}(A_1) \times \dots \times \text{dom}(A_n)$.

Exemple

On représente deux classes de CPGE à l'aide des deux tables suivantes.

classes

nom	profPrincipal
MPSI	DUPONT
PCSI	DUPOND

elevés

nom	prenom	dateNaissance	classe	etabOrigine
ALCAZAR	Peggy	1997-02-01	MPSI	Ottokar
ALCAZAR	Ramon	1998-07-06	PCSI	Ottokar
CASTAFIORE	Bianca	1999-05-04	PCSI	Moulinsart
HADDOCK	Archibald	1998-03-02	MPSI	Moulinsart
LAMPION	Séraphin	1999-07-14	PCSI	Ottokar
RASTAPOPOULOS	Roberto	1997-12-25	PCSI	Rackham le Rouge
TOURNESOL	Tryphon	1999-04-03	MPSI	Moulinsart

- L'attribut *classe* de la relation *elevés* a pour *domaine* {MPSI, PCSI}
- Le *schéma* de la relation *classes* est (nom, profPrincipal).

2) Clé

Dans une base de données, il faut pouvoir accéder facilement à un élément précis d'une relation ou en tester l'existence. Pour cela, on utilise une *clé*, c'est-à-dire un ou plusieurs attributs de la relation qui détermine de manière unique chaque n -uplet de la relation.

Autrement dit, si $R(S)$ est une relation de schéma S et si K est une partie de S , alors K est une clé pour R si et seulement si pour tous $t_1, t_2 \in R$ tels que t_1 et t_2 coïncident pour les attributs de K , on a $t_1 = t_2$.

La *clé primaire* d'une relation est une clé choisie par le concepteur de la base de données au moment de la création de la relation. Il est alors impossible d'ajouter des éléments qui créeraient des doublons pour la clé. On préfère choisir une clé primaire aussi simple que possible (et qui ne comporte pas trop d'attributs). Il est fréquent d'ajouter un attribut supplémentaire pour servir de clé (identifiant).

Exemple

Dans l'exemple précédent, {Nom} ne peut pas constituer une clé (deux ALCAZAR). Le couple {Nom, Prénom} pourrait servir de clé, mais le risque d'homonymie n'est pas nul. On pourra donc ajouter un numéro d'étudiant pour servir de clé.

II. Un deuxième exemple

On souhaite créer une base de données sur des films : titre, réalisateur, année de sortie, casting. On voudrait de plus stocker des informations sur les artistes (année de naissance, rôles).

Pour cela, on va utiliser trois tables : une table *films*, une table *artistes* et une table *roles* avec les schémas suivants :

Table <i>films</i>	
Attribut	Type
<i>id</i>	int(11)
titre	varchar(50)
annee	int(11)
idRealisateur	int(11)
genre	varchar(50)
resume	text
codePays	varchar(4)

Table <i>artistes</i>	
Attribut	Type
<i>id</i>	int(11)
nom	varchar(30)
prenom	varchar(30)
anneeNaiss	int(11)
anneeDeces	int(11)

Table <i>roles</i>	
Attribut	Type
<i>idFilm</i>	int(11)
<i>idActeur</i>	int(11)
nomRole	varchar(30)

Des clés primaires ont été définies pour chacune des relations : l'attribut *id* pour la relation *artistes*, l'attribut *id* pour la relation *films*, le couple (*idFilm*, *idActeur*) pour la relation *roles*.

Le réalisateur d'un film est décrit dans la table *films* par leur identifiant dans la table *artistes*.

id	titre	annee	idRealisateur	genre	resume	codePays
1	Sueurs froides	1958	3	Drame	...	USA
2	Alien	1979	4	Science-fiction	...	USA
3	Titanic	1997	6	Drame	...	USA
4	Sacrifice	1986	9	Drame	...	FR

id	nom	prenom	anneeNaiss	anneeDeces
1	Gavin	John	1931	NULL
2	Milland	Ray	1905	1986
3	Hitchcock	Alfred	1899	1980
4	Scott	Ridley	1937	NULL

Par exemple, Alfred Hitchcock a réalisé *Sueurs froides*.

La table *roles* contient des couples constitués de l'identifiant d'un film et de l'identifiant d'un artiste (ainsi que le nom du rôle).

idFilm	idActeur	nomRole
7	19	Carolyn Burnham
7	18	Lester Burnham
2	5	Ripley

id	titre	annee	idRealisateur	genre	resume	codePays
6	Sleepy Hollow	1999	13	Fantastique	...	USA
7	American Beauty	1999	17	Comédie	...	USA
8	Impitoyable	1992	20	Western	...	USA

id	nom	prenom	anneeNaiss	anneeDeces
17	Mendes	Sam	1965	NULL
18	Spacey	Kevin	1959	NULL
19	Bening	Anette	1958	NULL
20	Eastwood	Clint	1930	NULL

Par exemple, Kevin Spacey a joué le rôle de Lester Burnham dans *American Beauty*.