

BASES DE DONNÉES : INTRODUCTION ET VOCABULAIRE

1 Bases de données relationnelles

1.1 Introduction

Une base de données (BDD ou DB, database en anglais) est une collection (généralement importante) de données cohérentes entre elles. Dans une base de données relationnelle, on 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* 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 tables,

- chaque *schéma* $S = ((A_1, \text{dom}(A_1)), \dots, (A_n, \text{dom}(A_n)))$ 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.

nom	profPrincipal
MPSI	DUPONT
PCSI	DUPOND

nom	prenom	dateNaissance	classe	etabOrigine
ALCAZAR	Peggy	2002-02-01	MPSI	Ottokar
ALCAZAR	Ramon	2003-07-06	PCSI	Ottokar
CASTAFIORE	Bianca	2003-05-04	PCSI	Moulinsart
HADDOCK	Archibald	2003-03-02	MPSI	Moulinsart
LAMPION	Séraphin	2004-07-14	PCSI	Ottokar
RASTAPOPOULOS	Roberto	2002-12-25	PCSI	Rackham le Rouge
TOURNESOL	Tryphon	2004-04-03	MPSI	Moulinsart

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

Un étudiant fait partie d'une et une seule classe, alors qu'une classe comporte un nombre quelconque d'étudiants : on dira que l'association classe-étudiant est une association 1-*

1.2 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 primaire*, 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* est en général 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 de la table `eleves` (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.

Lorsqu'un attribut fait référence à une clé primaire d'une autre table, on parle de *clé étrangère*. Par exemple, si on considère l'attribut `Nom` de la table `classes`, alors l'attribut `classe` de la table `eleves` en est une clé étrangère.

2 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).

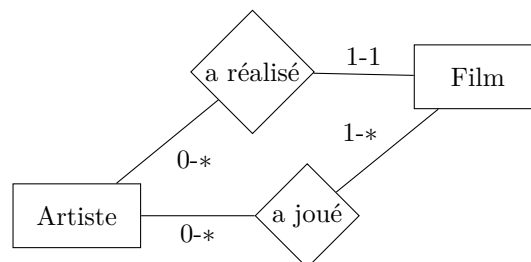
On fait l'hypothèse (erronée) qu'un film n'est réalisé que par une seule personne.

Notre modèle comportera deux *entités* : les films et les artistes. Les informations concernant celles-ci seront stockées dans deux tables de notre base de données.

Les *associations* entre ces entités seront les suivantes « a joué dans », « a réalisé ».

Un acteur peut avoir joué dans plusieurs films et un film peut comporter plusieurs acteurs : l'association acteur-film est une association **-**.

Comme il existe deux associations entre les personnes et les films, il est possible de les détailler davantage, en dessinant le *diagramme entité/relation* suivant :



Ce diagramme reprend les informations suivantes :

- Un film a été réalisé par une et une seule personne.
- Une personne peut n'avoir réalisé aucun film, en avoir réalisé un seul ou plusieurs.
- Une personne peut n'avoir joué aucun rôle, un seul, ou plusieurs personnages.
- Un film peut comporter un ou plusieurs personnages joués par des acteurs (c'est une hypothèse de notre modélisation).

Les deux associations « a joué dans », « a réalisé » pourront être représentées dans deux tables supplémentaires. Néanmoins, comme l'association Film-Réalisateur est une association 1 – 1, il est possible de la représenter à l'aide d'un attribut de la table `films`.

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	2018
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
14	32	Mills
14	22	Somerset
14	18	Doe
2	5	Ripley

id	titre	annee	idRealisateur	genre	resume	codePays
13	Van Gogh	1990	29	Drame	...	FR
14	Seven	1995	31	Policier	...	USA
15	L'armée des douze singes	1995	33	Science-fiction	..	USA

id	nom	prenom	anneeNaiss	anneeDeces
20	Eastwood	Clint	1930	NULL
21	Hackman	Gene	1930	NULL
22	Freeman	Morgan	1937	NULL
23	Crowe	Russell	1964	NULL

Par exemple, Morgan Freeman a joué le rôle de Somerset dans *Seven*.