

Recherche d'un schéma relationnel

1.1 Données en table

Voici un tableau simple, comme ceux manipulés par un tableur :

titre	date	nom	prenom	annee_naissance
Hana-bi	1997	Kitano	Takeshi	1947
Big fish	2003	Burton	Tim	1958
Edward aux mains d'argent	1990	Burton	Tim	1958
Sonatine	1993	Kitano	Takeshi	1947
Pulp Fiction	1995	Tarantino	Quentin	1963
Play Time	1967	Tati	Jacques	1907
Vertigo	1958	Hitchcock	Alfred	1898
Psychose	1960	Hitchcock	Alfred	1898
Parle avec elle	2002	Almodovar	Pedro	1949
Mon oncle	1958	Tati	Jacques	1907
Volver	2006	Almodovar	Pedro	1949
Reservoir Dogs	1992	Tarantino	Quentin	1963
Alive	2003	Kitamura	Ryûhei	Null
Godzilla : Final Wars	2004	Kitamura	Ryûhei	Null

1. Précisez quels sont les attributs et les domaines de ces attributs
2. La table contient des n -uplets, comme par exemple : *Hana-bi, 1997, Kitano, Takeshi, 1947*. Combien de n -uplets contient cette table.

1.2 Vocabulaire employé dans le domaine des bases de données

Compléter le tableau avec les termes du *modèle relationnel*

terme du modèle	terme utilisés pour un tableur
	Table
	ligne
	Nom de colonne
	Cellule
	Type

1.3 Anomalies dans la table

La table donnée plus haut est utilisée pour tenir à jour les prêts de DVD au CDI du lycée.

Q1.a On voudrait insérer la ligne suivante dans la table. Quelle sera l'instruction SQL ?

```
1 (Alive, 2004, Berthe, Frederic, 1967)
```

Q1.b Une fois que vous avez ajouté mis à jour votre base de données, un élève vous rapporte le DVD-Blue Ray du film *Alive*. Quelques jours plus tard, il vous demande quelle est la filmographie du réalisateur, car il a bien aimé ce film. Vous faites une recherche. Mais vous ne pouvez pas le renseigner précisément. Pourquoi ?

Q2. Une erreur s'est glissée dans le tableau : Alfred Hitchcock est né en 1899 et non 1898. Vous le corrigez

dans la ligne *Vertigo*. Est-ce que la table est complètement mise à jour ?

Q3. Vous supprimez les lignes de *Volver* et *Parle avec elle*. Pouvez vous alors retrouver la date de naissance de Pedro Almodovar ?

Q4. Dans chaque cas, vous avez mis en évidence un problème dans la structure de la base de données. Préciser s'il s'agit d'une :

- anomalie d'insertion
- anomalie de modification
- anomalie de suppression

Q5. Expliquer en quoi cette table présente des **redondances** d'informations

Q6. Quelle est la solution pour ne plus avoir ces problèmes ?

1.4 Corriger les problèmes

On va adopter un schéma relationnel avec plusieurs tables.

La base de données doit :

1. être capable de représenter individuellement les films et les réalisateurs, de manière à ce qu'une action sur l'un n'entraîne pas systématiquement une action sur l'autre,
2. définir une méthode d'identification d'un film ou d'un réalisateur, qui permette d'assurer que la même information est représentée une seule fois. On utilisera pour chaque table une **clé primaire**.
3. préserver le lien entre les films et les réalisateurs mais sans introduire de redondance.

1.4.1 Première proposition

On sépare les données en 2 tables. On placera un index numérique à la première colonne, qui fournira un identifiant unique.

Films

id_film	titre	date	id_rea
1	Hana-bi	1997	
2	Big fish	2003	
3	Edward aux mains d'argent	1990	
4	Sonatine	1993	
5	Pulp Fiction	1995	
6	Play Time	1967	
7	Vertigo	1958	
8	Psychose	1960	
9	Parle avec elle	2002	
10	Mon oncle	1958	
11	Volver	2006	
12	Reservoir Dogs	1992	
13	Alive	2003	
14	Godzilla : Final Wars	2004	

Réalisateur :

id_rea	nom	prenom	annee_naissance
1	Kitano	Takeshi	1947
2	Burton	Tim	1958
3	Tarantino	Quentin	1963
4	Tati	Jacques	1907
5	Hitchcock	Alfred	1899
6	Almodovar	Pedro	1949
7	Kitamura	Ryûhei	1969

A faire : Compléter la première table avec les valeurs correspondantes pour id_rea.

- Q1. L'association entre les informations des 2 tables est-elle toujours aussi explicite qu'avec une seule table ?
 Q2. Vérifier enfin que la base de données produite ne présente plus aucune des anomalies citées plus haut. Donner un exemple.

1.4.2 Deuxième proposition

L'association entre ces 2 tables est assurée par une nouvelle table *FilmsRéalisés*, dont la clé primaire est constituée d'une combinaison de valeurs des 2 colonnes id_film et id_rea :

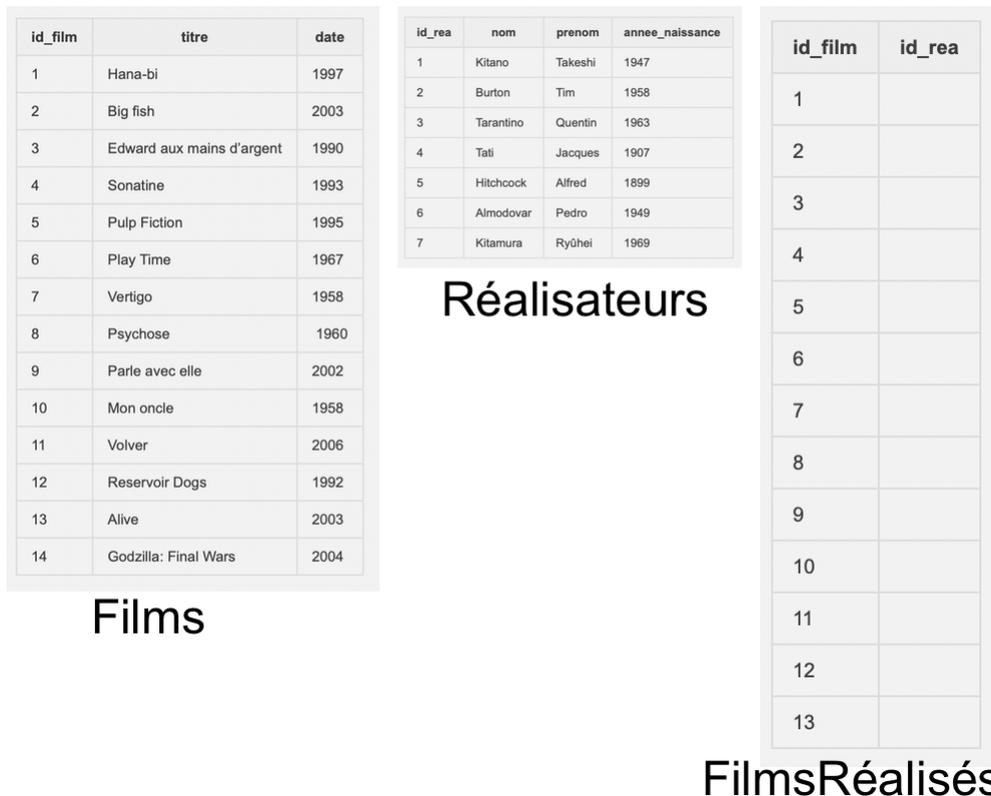


FIGURE 1 – association à 3 tables

- Q1. Compléter la table *FilmsRéalisés*.
 Q2. Ces deux descriptions suivent le modèle *entité-association*. Mais la manière avec laquelle cette association est représentée est différente.

Exercices

2.1 Ex 1 : Attributs et domaines

En base de données, les domaines (equivaut à type) sont les suivants :

- INT ou INTEGER : un entier
- FLOAT(x) : un nombre décimal avec x définissant la précision (nombre de bits de codage de la mantisse)
- REAL est un synonyme standard de FLOAT(24)
- CHAR(n) : chaîne d'au plus n caractères
- VARCHAR(n) : un type de données de longueur variable avec au maximum n caractères stockée
- TEXT
- DATE une date

Questions :

- quels sont les *types* équivalents en python.
- S'agit-il de types *simples* ou *construits*? Lesquels?

2.2 Ex 2 : Entité

Soit la Relation *FILMS* présentée ici :

id	titre	realisateur	ann_sortie	note_sur_10
1	Alien, le huitième passager	Scott	1979	10
2	Dune	Lynch	1985	5
3	2001 : l'odyssée de l'espace	Kubrick	1968	9
4	Blade Runner	Scott	1982	10

1. Proposer une instruction SQL permettant de calculer la moyenne des notes obtenues par réalisateur.
2. Y-a-t-il une ou plusieurs *entités*? Définir ce terme (entité)
3. créez une relation REALISATEURS (attributs de la relation REALISATEURS : id, nom, prenom et ann_naissance). Donner le *schéma de cette relation*.
4. Donner la définition de *clé étrangère*.
5. Donner le nouveau schéma relationnel de FILMS, afin d'établir un lien avec REALISATEURS.
6. Proposer une nouvelle instruction SQL permettant de calculer la moyenne des notes obtenues par réalisateur, mais cette fois à partir du nouveau schéma relationnel.
7. Proposez un nouveau schéma relationnel à 3 tables.

Aide : Le schéma d'une relation est donnée sous forme d'un ensemble de tuples :

$$S = ((A_1, \text{domaine}_1), (A_2, \text{domaine}_2) \dots (A_n, \text{domaine}_n))$$

les A_i sont les attributs.

2.3 Ex 3 : Un exemple de schéma relationnel avec une sandwicherie

Une sandwicherie effectuant des livraisons à domicile dispose d'une base de données dont certains extraits sont reproduits ici :

Sandwichs

id_Sandwich	Nom_sandwich	Prix
1	Cheeseburger	3,90
2	Double cheese	4,90
3	Italien	4,90
4	Foie gras	15,00

Clients

Nom	Prenom	Adresse	Numero_client
Bernard	Alain	9, rue Bienvenue, 13008 Marseille	42
Bernard	Yves	2, rue Vive la Joie, 13400 Aubagne	51

Commandes

Numero_client	id_Sandwich	Quantité	Numero_commande	Date
42	3	2	12452	2019-12-11
42	4	1	12452	2019-12-11
51	1	4	13301	2019-12-23

1. Une commande peut-elle comporter plusieurs sandwichs de types différents ?
2. Quel est le schéma relationnel de chacune des tables ? Donner, au choix, le schéma sous forme symbolique ou bien sous forme de diagramme
3. La table *Commandes* comporte-t-elle un attribut qui est clé primaire ? Un-des attribut-s qui est-sont clé-s étrangère-s ?
4. Ecrire une instruction SQL qui permet de connaître le nom, l'adresse et le contenu de la commande n°12452