

I./ Introduction :

Nous avons vu en Première qu'on pouvait utiliser certaines structures (comme les listes et les dictionnaires) pour stocker et utiliser des données. Lorsque la quantité de données à traiter devient trop importante, ces structures se révèlent insuffisantes. On utilise alors des *bases de données*, qui vont permettre d'organiser les données de façon à faciliter leur exploitation. Afin de pouvoir manipuler ces bases de données, on utilise des logiciels appelés « *Système de Gestion de Bases de Données* », ou *SGBD*.

Un SGBD a plusieurs rôles :

- Écrire, lire, et modifier les informations contenues dans la base de données.
- Gérer les autorisations d'accès à la base de données, en fonction du statut des utilisateurs.
- Gérer les problèmes liés aux accès concurrents, quand plusieurs personnes travaillent simultanément sur une même base de données.
- Assurer la sécurité des données en cas de panne grâce à la redondance des données.

II./ Les bases de données relationnelles :

Il existe plusieurs manières d'organiser de grands ensembles de données, ce qui donne donc naissance à plusieurs types différents de bases de données. Cette année, nous traiterons uniquement des *bases de données relationnelles*.

Une base de données relationnelle est composée d'une ou de plusieurs *relation(s)*. Une relation peut être vue comme un tableau à 2 dimensions (C'est pourquoi on utilise aussi le terme « *table* » pour désigner une relation). Voici un exemple de relation :

Relation « FILMS »

id	titre	réalisateur	année	note critique
1	Terminator	Cameron	1974	****
2	Titanic	Cameron	1997	***
3	Avatar	Cameron	2009	****
4	Avatar 2	Cameron	2022	***
5	Avatar 3	Cameron	2024	?
6	E.T.	Spielberg	1982	**
7	Jurassic Park	Spielberg	1993	****
8	Docteur Folamour	Kubrick	1964	**
9	2001	Kubrick	1968	****
10	Eyes Wide Shut	Kubrick	1999	*
11	Memento	Nolan	2000	**
12	Interstellar	Nolan	2014	****
13	Je vous salue, Marie	Godard	1985	*
14	Nouvelle vague	Godard	1990	***

En-tête

Enregistrement (ou t-uplet)

Attribut (ici « titre »)

La première ligne est l'*en-tête*, et donne le nom de chacune des colonnes. Chacune des colonnes est un *attribut*. Chaque ligne constitue un *enregistrement*, ou un *t-uplet*.

La relation précédente serait par exemple un extrait de la base de données d'un site web sur le cinéma. Elle comporte cinq attributs. Chacun de ces attributs va avoir un certain *domaine*, c'est-à-dire un ensemble de valeurs admissibles.

Dans notre exemple, le premier attribut « id », correspond au numéro du film. Il s’agit donc d’un nombre entier (noté **INT**). De même pour le quatrième attribut « année ». Les deuxième, troisième et cinquième attributs, « titre », « réalisateur » et « note critique » sont des chaînes de caractères (notées **TEXT**).

Il existe d’autres types de données utilisables, comme **FLOAT** pour les nombres flottants, **DATE** pour des dates, ou **TIME** pour des horaires.

Chaque relation composant une base de données doit respecter une contrainte importante d’unicité : aucune relation ne peut comporter de lignes identiques. Pour que cette condition soit toujours respectée, on utilise un attribut particulier, que l’on appelle la *clef primaire*. Dans notre exemple, seul le premier attribut « id » peut servir de clef primaire. En effet deux films différents peuvent avoir le même titre, le même réalisateur, la même année de sortie ou la même note critique.

Le gestionnaire du site web sur le cinéma souhaite alors rajouter des informations dans sa base de données, notamment l’année de naissance et la nationalité du réalisateur. Une première idée serait de rajouter autant d’attributs (de colonnes) que d’informations supplémentaires. Notre relation « FILMS » deviendrait alors :

id	titre	réalisateur	année	note critique	année naissance réalisateur	pays réalisateur
1	Terminator	Cameron	1974	****	1954	Canada
2	Titanic	Cameron	1997	***	1954	Canada
3	Avatar	Cameron	2009	****	1954	Canada
4	Avatar 2	Cameron	2022	***	1954	Canada
5	Avatar 3	Cameron	2024	?	1954	Canada
6	E.T.	Spielberg	1982	**	1946	USA
7	Jurassic Park	Spielberg	1993	****	1946	USA
8	Docteur Folamour	Kubrick	1964	**	1928	USA
9	2001	Kubrick	1968	****	1928	USA
10	Eyes Wide Shut	Kubrick	1999	*	1928	USA
11	Memento	Nolan	2000	**	1970	UK/USA
12	Interstellar	Nolan	2014	****	1970	UK/USA
13	Je vous salue, Marie	Godard	1985	*	1930	France/Suisse
14	Nouvelle vague	Godard	1990	***	1930	France/Suisse

Cette solution n’est pas optimale, car on s’aperçoit que de nombreuses données identiques sont dupliquées, ce qui va entraîner une consommation importante de mémoire. Une solution plus intéressante serait de scinder cette grande table en deux relations différentes :

Relation « FILMS »

id	titre	réalisateur	année	note critique
1	Terminator	Cameron	1974	****
2	Titanic	Cameron	1997	***
3	Avatar	Cameron	2009	****
4	Avatar 2	Cameron	2022	***
5	Avatar 3	Cameron	2024	?
6	E.T.	Spielberg	1982	**
7	Jurassic Park	Spielberg	1993	****
8	Docteur Folamour	Kubrick	1964	**
9	2001	Kubrick	1968	****
10	Eyes Wide Shut	Kubrick	1999	*
11	Memento	Nolan	2000	**
12	Interstellar	Nolan	2014	****
13	Je vous salue, Marie	Godard	1985	*
14	Nouvelle vague	Godard	1990	***

Relation « REALISATEURS »

id	réalisateur	année naissance réalisateur	pays réalisateur
1	Cameron	1954	Canada
2	Spielberg	1946	USA
3	Kubrick	1928	USA
4	Nolan	1970	UK/USA
5	Godard	1930	France/Suisse

Il y a maintenant deux relations, chacune ayant sa clef primaire. Un problème subsiste cependant : il y a duplication du nom du réalisateur entre les deux tables. Cela peut poser un problème : si jamais le gestionnaire du site se rend compte qu'il a mal orthographié le nom d'un réalisateur, il devra le modifier dans la table « REALISATEURS » et dans toutes les lignes concernées de la table « FILMS ». Pour résoudre ce nouveau problème, on introduit le concept de *clef étrangère* (ou *clef secondaire*) :

Relation « FILMS »

id	titre	id_réalisateur	année	note critique
1	Terminator	1	1974	****
2	Titanic	1	1997	***
3	Avatar	1	2009	****
4	Avatar 2	1	2022	***
5	Avatar 3	1	2024	?
6	E.T.	2	1982	**
7	Jurassic Park	2	1993	****
8	Docteur Folamour	3	1964	**
9	2001	3	1968	****
10	Eyes Wide Shut	3	1999	*
11	Memento	4	2000	**
12	Interstellar	4	2014	****
13	Je vous salue, Marie	5	1985	*
14	Nouvelle vague	5	1990	***

Relation « REALISATEURS »

id	réalisateur	année naissance réalisateur	pays réalisateur
1	Cameron	1954	Canada
2	Spielberg	1946	USA
3	Kubrick	1928	USA
4	Nolan	1970	UK/USA
5	Godard	1930	France/Suisse

Dans cette nouvelle relation « FILMS », l'attribut « id_réalisateur » permet de faire le lien avec la relation « REALISATEURS », en faisant une référence avec la clef primaire « id » de la table « REALISATEURS ».

Une base de données peut donc comporter plusieurs tables. Afin de pouvoir en décrire facilement la structure, on définit le *schéma relationnel de la base de données*, en fournissant :

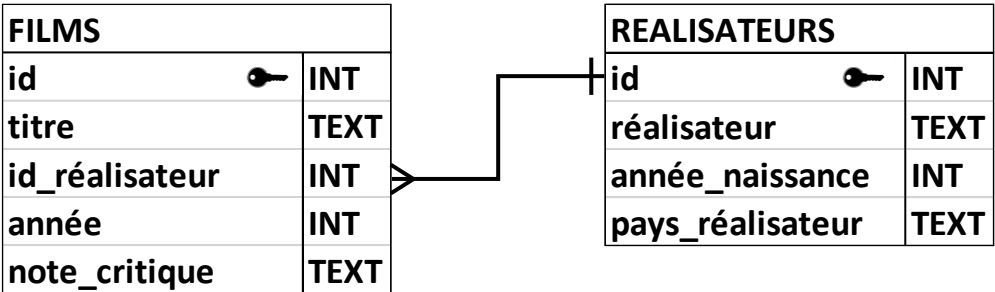
- Les noms des différentes relations,
- Pour chaque relation, la liste des attributs avec leur domaine respectif,
- Pour chaque relation, la clef primaire et la clef étrangère éventuelle.

Pour notre dernier exemple, le schéma relationnel serait :

FILMS(id : INT, titre : TEXT, #id_réalisateur : INT, année : INT, note_critique : TEXT)
REALISATEURS(id : INT, réalisateur : TEXT, année_naissance_réalisateur : INT, pays_réalisateur : TEXT)

Les attributs soulignés sont des clefs primaires, le # signifie que l'on a une clef étrangère.

On peut aussi représenter un schéma relationnel sous forme graphique, à l'aide d'un diagramme.



Les petits dessins de clefs correspondent aux clefs primaires. Le trait reliant l'attribut « id_réalisateur » de la relation « FILMS » et l'attribut « id » de la relation « REALISATEURS » indique que ces deux attributs doivent partager les mêmes valeurs. Ces valeurs sont prises de manière unique dans la relation du côté « + » et zéro, une ou plusieurs fois du côté « > ».