

Séq. 2 – Base de données

Le modèle relationnel

Objectifs

- Identifier les concepts définissant le modèle relationnel (relation, attribut, domaine, clef primaire, clef étrangère, schéma relationnel)
- Exprimer les contraintes d'intégrité (domaine, relation et référence)
- Savoir distinguer la structure d'une base de données de son contenu
- Repérer des anomalies dans le schéma (redondances de données, anomalies d'insertion, de suppression, de mise à jour)

Ce cours est inspiré de : https://pixees.fr/informatiquelycee/n_site/nsi_term_bd_intro.html

1 Introduction

L'année dernière nous avons eu l'occasion de travailler sur des données structurées en les stockant dans des fichiers au format CSV. Même si cette méthode de stockage de l'information peut s'avérer pratique, elle montre rapidement ses limites.

Le terme base de données est apparu au début des années 60. C'est l'apparition des disques durs à la fin des années 50 qui a permis d'utiliser les ordinateurs pour stocker et manipuler des données. Avec l'apparition du Web, la quantité de données à stocker a littéralement explosé. Aujourd'hui, la plupart des sites internet (du petit site personnel au grand site d'e-commerce) utilisent au moins une base de données. Les bases de données jouent un rôle fondamental dans notre monde devenu numérique où il est extrêmement facile de dupliquer l'information.

A faire vous même 1.

Trouvez des exemples de domaines d'activité où les bases de données jouent un rôle primordial.

2 Les Bases de données relationnelles

Il existe différents types de bases de données, par exemple, les bases de données hiérarchiques, les bases de données objet, les bases de données nosql ou bien encore les bases de données relationnelles.

Les bases de données relationnelles sont le plus utilisées au monde, c'est ce type de base de données que nous allons étudier.

Les bases de données relationnelles ont été mises au point en 1970 par Edgar Franck Codd, informaticien britannique (1923-2003). Ces bases de données sont basées sur la théorie mathématique des ensembles.

2.1 Principes généraux

A faire vous même 2. S' approprier le vocabulaire

Lire livre Prepac – P. 182-183 et complétez la définition des mots suivants :

- un domaine :
C' est un ensemble de valeurs que peut prendre une donnée
Exemple : des entiers (noté \mathbb{N}), des chaînes de caractères (notés \mathcal{S}), des dates (notés \mathcal{D}), des booléens (notés \mathcal{B}), ...
- un attribut :
C' est un couple (nom, domaine) où le nom (représentant un titre de colonne) prend ses valeurs dans un domaine.
- un schéma :
C' est un ensemble ordonné d' attributs distincts
- une table ou relation :
Une relation de schéma $S=(A_1, A_2, A_3, \dots, A_n)$ est un ensemble fini de n-uplets $(x_1, x_2, x_3, \dots, x_n)$ où les x_i prennent leurs valeurs dans A_i
- Les éléments d' une table :
Ce sont les enregistrements de la relation. Leur nombre est appelé cardinal.

2.2 Illustration

La notion de *relation* est au cœur des bases de données relationnelles. Une *relation* peut être vue comme un tableau à 2 dimensions, composé d'un en-tête et d'un corps. Le corps est lui-même composé de t-uplets (lignes) et d'attributs (colonnes). L'en-tête contient les intitulés des attributs, le corps contient les données proprement dites. À noter que l'on emploie aussi le terme *table* à la place de *relation*.

Voici un exemple de relation :

id	titre	auteur	ann_publi	note
1	1984	Orwell	1949	10
2	Dune	Herbert	1965	8
3	Fondation	Asimov	1951	9
4	Le meilleur des mondes	Huxley	1931	7
5	Fahrenheit 451	Bradbury	1953	7
6	Ubik	K.Dick	1969	9
7	Chroniques martiennes	Bradbury	1950	8
8	La nuit des temps	Barjavel	1968	7
9	Blade Runner	K.Dick	1968	8
10	Les Robots	Asimov	1950	9
11	La Planète des singes	Boulle	1963	8
12	Ravage	Barjavel	1943	8
13	Le Maître du Haut Château	K.Dick	1962	8
14	Le monde des Â	Van Vogt	1945	7
15	La Fin de l'éternité	Asimov	1955	8
16	De la Terre à la Lune	Verne	1865	10

Annotations de l'image :

- en-tête : pointe vers la première ligne du tableau (les colonnes).
- corps : pointe vers l'ensemble des lignes de données.
- 1 t-uplet : pointe vers une seule ligne de données (la ligne 11).
- l'attribut "titre" : pointe vers la colonne "titre".

Le t-uplet encadré en jaune sur le schéma ci-dessus contient les éléments suivant : 11, La Planète des singes, Boulle, 1963 et 8.

L'attribut `titre` est composé des éléments suivants : 1984, Dune, Fondation, Le meilleur des mondes, ...

A faire vous même 3.

- Faites la liste des éléments appartenant à l'attribut `auteur`.

Pour chaque attribut d'une relation, il est nécessaire de définir un domaine : Le domaine d'un attribut donné correspond à un ensemble fini ou infini de valeurs admissibles.

Par exemple, le domaine de l'attribut `id` correspond à l'ensemble des entiers (noté INT) : la colonne `id` devra obligatoirement contenir des entiers.

Autre exemple, le domaine de l'attribut `titre` correspond à l'ensemble des chaînes de caractères (noté TEXT).

Dernier exemple, le domaine de l'attribut `note` correspond à l'ensemble des entiers positifs.

Comme domaines, nous trouvons aussi le décimaux, les booléens, les dates, les coordonnées GPS, ...

A faire vous même 4.

Quel est, selon vous, le domaine de l'attribut `auteur`

2.3 Le modèle relationnel

A faire vous même 5. S'approprier le vocabulaire

Lire livre Prepabac – P. 184-185 et complétez la définition des mots suivants :

- clé primaire :
C' est un attribut (ou plus rarement un ensemble d' attributs) qui permet d' identifier un unique enregistrement.
- clé étrangère:
C' est un attribut qui correspond à une clé primaire d' une autre table/relation
- contrainte d' intégrité :
C' est une règle qui définit la cohérence d' une donnée ou d' un ensemble de données d' une base de données

A faire vous même 6.

Voici un extrait d'une relation référençant des films :

Relation FILMS

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

Listez les différents attributs de cette relation. Donnez le domaine de chaque attribut. Pour chaque attribut dire si cet attribut peut jouer le rôle de clef primaire, vous n' oublierez pas de justifier vos réponses.

2.4 Contraintes d'intégrité

2.4.1 contrainte de domaine :

Au moment de la création d'une relation, il est nécessaire de renseigner le domaine de chaque attribut. Le SGBD s'assure qu'un élément ajouté à une relation respecte bien le domaine de l'attribut correspondant : si par exemple vous essayez d'ajouter une note non entière (par exemple 8.5), le SGBD signalera cette erreur et n'autorisera pas l'écriture de cette nouvelle donnée.

2.4.2 contrainte de relation :

une relation ne peut pas contenir 2 t-uplets identiques.

Par exemple, la situation ci-dessous n'est pas autorisée (ici aussi c'est le SGBD qui veille au grain) :

id	titre	auteur	ann_publi	note
1	1984	Orwell	1949	10
2	Dune	Herbert	1965	8
2	Dune	Herbert	1965	8
3	Fondation	Asimov	1951	9

Afin d'être sûr de respecter cette contrainte des t-uplets identiques, intervient la notion de `clef primaire`.

RAPPEL : Une clef primaire est un attribut dont la valeur permet d'identifier de manière unique un t-uplet de la relation. Autrement dit, si un attribut est considéré comme clef primaire, on ne doit pas trouver dans toute la relation 2 fois la même valeur pour cet attribut.

2.4.3 contrainte d'intégrité référentielle :

C'est le respect de 3 règles :

- Une clé étrangère est obligatoirement une clé primaire de la table à laquelle elle se réfère
- Un enregistrement ne peut être effacé s'il possède des enregistrements liés (i.e. sa clé primaire est une clé étrangère dans une autre table)
- La clé primaire ne peut être modifiée si l'enregistrement possède des enregistrements liés (i.e. sa clé primaire est une clé étrangère dans une autre table)

2.5 Optimisation

Nous désirons enrichir cette relation en ajoutant des informations supplémentaires sur les auteurs, nous obtenons alors :

Relation LIVRES_AUTEURS

id	titre	nom_auteur	prenom_auteur	date_nai_auteur	langue_ecriture_auteur	ann_publi	note
1	1984	Orwell	George	1903	anglais	1949	10
2	Dune	Herbert	Frank	1920	anglais	1965	8
3	Fondation	Asimov	Isaac	1920	anglais	1951	9
4	Le meilleur des mondes	Huxley	Aldous	1894	anglais	1931	7
5	Fahrenheit 451	Bradbury	Ray	1920	anglais	1953	7
6	Ubik	K.Dick	Philip	1928	anglais	1969	9
7	Chroniques martiennes	Bradbury	Ray	1920	anglais	1950	8
8	La nuit des temps	Barjavel	René	1911	français	1968	7
9	Blade Runner	K.Dick	Philip	1928	anglais	1968	8
10	Les Robots	Asimov	Isaac	1920	anglais	1950	9
11	La Planète des singes	Boulle	Pierre	1912	français	1963	8
12	Ravage	Barjavel	René	1911	français	1943	8
13	Le Maître du Haut Château	K.Dick	Philip	1928	anglais	1962	8
14	Le monde des Â	Van Vogt	Alfred Elton	1912	anglais	1945	7
15	La Fin de l'éternité	Asimov	Isaac	1920	anglais	1955	8
16	De la Terre à la Lune	Verne	Jules	1828	français	1865	10

Il y a pas mal d' informations dupliquées, par exemple, on retrouve 3 fois "K.Dick Philip 1928 anglais", même chose pour "Asimov Isaac 1920 anglais"... Cette duplication est-elle indispensable ? Non ! Est-elle souhaitable ? Non plus ! En effet, dans une base de données, on évite autant que possible de dupliquer l'information. Si nous dupliquons autant de données inutilement c'est que notre structure ne doit pas être la bonne ! Mais alors, comment faire pour avoir aussi des informations sur les auteurs des livres ?

La solution est relativement simple : travailler avec 2 relations au lieu d'une seule et créer un "lien" entre ces 2 relations :

Relation AUTEURS

id	nom	prenom	ann_naissance	langue_ecriture
1	Orwell	George	1903	anglais
2	Herbert	Frank	1920	anglais
3	Asimov	Isaac	1920	anglais
4	Huxley	Aldous	1894	anglais
5	Bradbury	Ray	1920	anglais
6	K.Dick	Philip	1928	anglais
7	Barjavel	René	1911	français
8	Boulle	Pierre	1912	français
9	Van Vogt	Alfred Elton	1912	anglais
10	Verne	Jules	1828	français

Relation LIVRES

id	titre	id_auteur	ann_publi	note
1	1984	1	1949	10
2	Dune	2	1965	8
3	Fondation	3	1951	9
4	Le meilleur des mondes	4	1931	7
5	Fahrenheit 451	5	1953	7
6	Ubik	6	1969	9
7	Chroniques martiennes	5	1950	8
8	La nuit des temps	7	1968	7
9	Blade Runner	6	1968	8
10	Les Robots	3	1950	9
11	La Planète des singes	8	1963	8
12	Ravage	7	1943	8
13	Le Maître du Haut Château	6	1962	8
14	Le monde des Â	9	1945	7
15	La Fin de l'éternité	3	1955	8
16	De la Terre à la Lune	10	1865	10

Comme vous l'avez sans doute remarqué, l'attribut `id_auteur` de la relation LIVRES permet de créer un lien avec la relation AUTEURS. `id_auteur` correspond à l'attribut `id` de la relation AUTEURS. L'introduction d'une relation AUTEURS et la mise en place de liens entre cette relation et la relation LIVRES permettent d'éviter la redondance d'informations.

Pour établir un lien entre 2 relations RA et RB, on ajoute à RA un attribut x qui prendra les valeurs de la clé primaire de RB. Cet attribut x est appelé clef étrangère (l'attribut correspond à la clé primaire d'une autre table, d'où le nom).

Dans l'exemple ci-dessus, l'attribut `id_auteur` de la relation LIVRES permet bien d'établir un lien entre la relation LIVRES et la relation AUTEURS, `id_auteur` correspond bien à la clé primaire de la relation AUTEURS, conclusion : `id_auteur` est une clef étrangère.

Pour préserver l'intégrité d'une base de données, il est important de bien vérifier que toutes les valeurs de la clef étrangère correspondent bien à des valeurs présentes dans la clé primaire (nous aurions un problème d'intégrité de la base de données si une valeur de l'attribut `id_auteur` de la relation LIVRES ne correspondait à aucune valeur de la clé primaire de la relation AUTEURS). Certains SGBD ne vérifient pas cette contrainte (ne renvoie aucune erreur en cas de problème), ce qui peut provoquer des comportements erratiques.

A faire vous même 7.

En partant de la relation FILMS ci-contre, créez une relation REALISATEURS (attributs de la relation REALISATEURS : id, nom, prenom et ann_naissance, vous trouverez toutes les informations nécessaires sur le Web).

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

Modifiez ensuite la relation FILMS afin d'établir un lien entre les relations FILMS et REALISATEURS. Vous préciserez l'attribut qui jouera le rôle de clef étrangère.

2.6 Schéma relationnel

On appelle schéma relationnel l'ensemble des relations présentes dans une base de données. Il est nécessaire de fournir les informations suivantes :

- Les noms des différentes relations
- pour chaque relation, la liste des attributs avec leur domaine respectif

- pour chaque relation, la clef primaire et éventuellement la clef étrangère

Voici un exemple pour les relations LIVRES et AUTEURS :

AUTEURS(id : INT, nom : TEXT, prenom : TEXT, ann_naissance : INT, langue_ecriture : TEXT)
LIVRES(id : INT, titre : TEXT, #id_auteur : INT, ann_publi : INT, note : INT)

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

On peut aussi l'écrire de cette façon :

AUTEURS(id : \mathcal{N} , nom : \mathcal{P} , prenom : \mathcal{P} , ann_naissance : \mathcal{N} , langue_ecriture : \mathcal{P})

LIVRES(id : \mathcal{N} , titre : \mathcal{P} , #id_auteur : \mathcal{N} , ann_publi : \mathcal{N} , note : \mathcal{N})

3 Conception de bases de données

A faire vous même 8. S'appropriier le vocabulaire

Lire livre Prepabac – P. 186-187 et faites-en un résumé :

Regardez vidéo : <https://www.youtube.com/watch?v=bhtzximjwk8>

Faire P. 190 ex 1

Faire P. 191 ex 2

Faire P. 191 ex 3

Faire P. 192 ex 4

Faire P. 192 ex 5

On considère une gestion simplifiée des voyages dans l'espace. La base de données utilisée est constituée de quatre relations nommées Astronaute, Fusee, Equipe et Vol. Voici le contenu des tables Astronaute, Fusee, Equipe et Vol.

Les clés primaires sont soulignées et les clefs étrangères sont précédées d'un # :

Astronaute				
<u>id_astronaute</u>	nom	prenom	nationalite	nb_vols
1	'PESQUET'	'Thomas'	'français'	2
2	'AMSTRONG'	'Neil'	'américain'	8
3	'MAURER'	'Mathias'	'allemand'	1
4	'MCARTHUR'	'Megan'	'américain'	5

Fusee			
<u>id_fusee</u>	modele	constructeur	nb_places
1	'Falcon 9'	'SpaceX'	6
2	'Starship'	'SpaceX'	100
3	'Soyouz'	'TsSKB Progress'	2
4	'SLS'	'Boeing'	6

Equipe	
<u>id_vol</u>	#id_astronaute
1	1
1	2
1	3
2	1
2	3
3	1
3	2
3	4
4	2
4	4

Vol		
<u>id_vol</u>	#id_fusee	Date
1	1	'12/09/2022'
2	4	'25/10/2022'
3	3	'18/11/2022'
4	2	'23/12/2022'

1. On s'intéresse ici à la notion de clés primaire et étrangère.

a. Donner la définition d'une clé primaire.

c. Le schéma relationnel de la table Astronaute est :

Astronaute (id_astronaute : INT, nom : TEXT, prenom : TEXT, nationalite : TEXT, nb_vols : INT)

Écrire le schéma relationnel de la table Fusee en précisant le domaine de chaque attribut.