

XIV – Bases de données relationnelles

Introduction

Un des enjeux actuels de l'informatique est de regrouper, ordonner et mettre à disposition un grand nombre de données en vue d'une utilisation par des experts ou des particuliers. Cette mise en commun de données nécessite de les structurer afin d'en extraire facilement les informations recherchées.

1/ Limites des structures de données plates

1.1) Un exemple d'utilisation de tableaux

Imaginons que l'on cherche à recenser tous les élèves de CPGE d'un lycée donné, en prenant en compte les caractéristiques suivantes : prénom, nom, filière, numéro de classe, lycée d'origine et note obtenue au bac. On peut choisir de structurer les données sous forme d'une liste de classe, chaque classe étant elle-même une liste qui commence par le nom de la filière, le numéro de la classe et finit par la liste des élèves. Enfin, chaque élève sera représenté par une liste contenant prénom, nom, lycée d'origine et note au bac.

Exemple : voici un aperçu du lycée imaginaire que nous étudierons dans ce chapitre.

```
[ ['BCPST', 1, [ ['Jean', 'Bonnot', 'Lycée Maphore', 14.2], ['Bernard', 'Tichaud', 'Lycée Négal', 12.8], ... ] ], ['BCPST', 2, [ ['Robert', 'Lingot', 'Lycée Négal', 15.4], ['Elodie', 'Nosauze', 'Lycée Raynade', 17.1], ... ] ] ]
```

On peut facilement obtenir la liste des élèves d'une classe donnée. Par contre, il est plus fastidieux de trier les élèves selon leur lycée d'origine.

Exemple : faire cela en exercice.

On aurait pu également lister les élèves selon leur lycée d'origine, ce qui permet de répondre plus facilement à la dernière question. En contrepartie, il est plus compliqué d'obtenir la liste des élèves de chaque classe. Il n'y a pas de solution miracle !

Le problème

Chaque élève a des liens d'appartenance à la fois avec les classes et avec les lycées. Or, la représentation sous la forme d'un tableau nous contraint à privilégier l'un de ces liens, ce qui complique forcément certaines recherches.

Ou alors on se contente d'une liste d'élèves en rajoutant la filière et le numéro de classe à chaque élève, ce qui prend plus de place du fait de la redondance d'informations.

1.2) Utilisation du modèle relationnel

Faisons maintenant abstraction de tout lien d'appartenance. On peut représenter de manière assez similaire les classes et les élèves, en convenant par exemple qu'une classe est un triplet (id,filiere,n°) comme (01,'BCPST',1) et qu'un élève est un quintuplet (prenom,nom,id,lycee,note). Chacune de ces représentations est appelée *relation* ou *table*.

Exemple : voici ce que cela donne avec l'exemple précédent

Table Classes

id	filiere	n°
01	'BCPST'	1
02	'BCPST'	2

Table Eleves

prenom	nom	id	lycee	note
'Jean'	'Bonnot'	01	'Maphore'	14.2
'Bernard'	'Tichaud'	01	'Négal'	12.8
'Robert'	'Lingot'	02	'Négal'	15.4
'Elodie'	'Nosaure'	02	'Raynade'	17.1

C'est plus léger que le dernier modèle car on évite de stocker des 6-uplets. On pourra aussi ajouter des informations dans la table Classes sans avoir à modifier la table Eleves. Et si l'on veut ajouter des caractéristiques aux lycées d'origine, il suffit de créer une table spécifique pour ceux-ci, avec un identifiant que l'on retrouvera dans la table Eleves et toute autre information utile.

C'est ce type de représentation que l'on va utiliser dans le monde des **bases de données relationnelles** (que l'on abrègera en BDDR ou BDD).

1.3) Abstraction des bases de données relationnelles

L'**algèbre relationnelle** est une théorie qui définit les opérations pouvant être effectuées sur des relations. Pour fixer les idées, disons que ces dernières sont des ensembles de tuples, qui correspondent aux données de notre BDD. Cette théorie est au cœur des logiciels de gestion des BDD.

D'un point de vue pratique, on utilisera le langage SQL pour effectuer des requêtes dans les BDD. On l'étudiera plus en détail lors du prochain chapitre.

Les bases de données relationnelles permettent de simplifier la gestion de données et d'éviter les problèmes évoqués précédemment, en garantissant à la fois une structure de données efficace et une rapidité d'accès à celles-ci.

De plus – et c'est un point crucial – l'utilisateur n'a pas besoin de connaître la manière dont les données sont stockées pour interroger la BDD. Ainsi, nul besoin d'être un expert pour adresser des requêtes simples à la BDD, qui peut alors se retrouver à la disposition d'un grand nombre d'utilisateurs.

2/ Vocabulaire des bases de données relationnelles

2.1) Modèle utilisé

Les données sont représentées à l'aide de tableaux à deux dimensions appelés *tables*. Par convention, les données d'un même élément de la BDD se trouvent sur la même ligne. Ainsi, les lignes correspondent aux différents éléments ou *enregistrements* de la BDD, tandis que les colonnes donnent les différentes caractéristiques ou *attributs* de ceux-ci.

Exemple : voici la table Eleves2018 qui va nous servir de fil rouge dans ce chapitre.

prenom	nom	filiere	numero	lycee_origine	note_bac
Jean	Bonnot	BCPST	1	Maphore	14.2
Jean	Aymard	BCPST	1	Raynade	13.7
Robert	Lingot	BCPST	2	Négal	15.4
Elodie	Nosaure	BCPST	2	Raynade	17.1
Bernard	Tichaud	BCPST	1	Négal	12.8
Pauline	Ohm	BCPST	1	Pulture	16.8
Elodie	Plomate	BCPST	2	Maphore	11.9
Sophie	Garrot	BCPST	2	Minherr	18.0
Pascale	Feutrage	BCPST	1	Maphore	14.7
Bernard	Bonnot	BCPST	1	Dimant	12.1
Paul	Hymaire	BCPST	1	Pulture	15.2

2.2) Concepts fondamentaux

Attributs ce sont les caractéristiques dont on souhaite enregistrer les valeurs. Ils forment un ensemble : autrement dit, ils sont deux à deux distincts et l'ordre n'a pas d'importance. Ainsi, dans notre exemple, on ne parlera pas du 2^e attribut mais de l'attribut nom.

Domaine l'ensemble des valeurs que peut prendre un attribut A est son *domaine*. On le note $\text{Dom}(A)$. Il est fixé lors de la création de la BDD et n'influence pas les recherches.

Exemple : $\text{Dom}(\text{numero}) = \mathbb{N}^*$ et $\text{Dom}(\text{note_bac}) = [0; 20]$.

Schéma relationnel un p -uplet d'attributs deux à deux distincts $A_1, A_2, \dots, A_p$ est appelé *schéma relationnel*. On le notera $S = (A_1, A_2, \dots, A_p)$... même si l'ordre importe peu.

Exemple : $S = (\text{prenom}, \text{nom}, \text{filiere}, \text{numero}, \text{lycee_origine}, \text{note_bac})$ ici.

- On note $A \in S$ pour dire que A est un attribut de S .
- Par extension, si $X = (A_1, \dots, A_n)$ avec les A_i des attributs de S , on notera $X \subset S$.
Ainsi, $(\text{prenom}, \text{filiere}, \text{lycee_origine}) \subset S$.

Relation on appelle *relation* (ou *table*) associée au schéma relationnel S un ensemble fini de p -uplets de $\text{Dom}(A_1) \times \text{Dom}(A_2) \times \dots \times \text{Dom}(A_n)$. On le note $\mathcal{R}$. Pour préciser que la relation $\mathcal{R}$ est associée au schéma S , on la notera $\mathcal{R}(S)$.

Exemple : la table Eleves2018 est une relation $\mathcal{R}$ associée au schéma S .

Enregistrements on appelle *enregistrements* (ou *valeurs*) les éléments d'une relation $\mathcal{R}$. Ils correspondent aux lignes dans la représentation en tableau. Leur nombre $\# \mathcal{R}$ est le *cardinal* de la relation.

Exemple : $(\text{Jean}, \text{Aymard}, \text{BCPST}, 1, \text{Raynade}, 13.7) \in \mathcal{R}(\mathcal{S})$.

Définition Soient $\mathcal{R}(\mathcal{S})$ une relation et $e \in \mathcal{R}(\mathcal{S})$.

- On note $e \cdot A$ la composante de e associée à l'attribut $A \in \mathcal{S}$.
- Par extension, si $X = (A_1, \dots, A_n) \subset \mathcal{S}$, on notera $e(X) = (e \cdot A_1, \dots, e \cdot A_n)$.

2.3) Une contrainte essentielle

Proposition

Dans une relation $\mathcal{R}(\mathcal{S})$, les enregistrements forment un ensemble.

Ainsi, les enregistrements sont deux à deux distincts, ce qui évite la redondance des données. De plus, l'ordre des enregistrements n'est pas fixé, mais on verra plus loin que ceci n'a aucune importance.

3/ Algèbre relationnelle élémentaire

3.1) Définition

L'algèbre relationnelle permet de formaliser les requêtes que l'on adresse aux relations constituant une BDD relationnelle. C'est un ensemble d'opérateurs que l'on applique à des relations et qui fournissent des relations : on parle alors d'*opérateurs relationnels*.

On peut les combiner et former des requêtes élaborées à partir d'opérateurs élémentaires. On cherchera ainsi à exprimer toute requête à l'aide d'une expression algébrique des opérateurs élémentaires, si possible peu coûteuse en temps de calcul et en mémoire (optimisation de recherche).

Exemples : quelle est la moyenne des notes obtenues au bac par les élèves de BCPST1 ? Quels sont les élèves de BCPST venant du lycée Négal ?

3.2) Opérateurs ensemblistes usuels

On les illustrera à l'aide des tables des élèves de la classe de BCPST2 en 2017 et 2018.

BCPST2_2017

prenom	nom	lycee_origine
Robert	Lingot	Négal
Sandy	Kilos	Pulture
Sophie	Garrot	Minherr
Thomas	Nifique	Maphore

BCPST2_2018

prenom	nom	lycee_origine
Robert	Lingot	Négal
Elodie	Nosaure	Raynade
Sophie	Garrot	Minherr
Elodie	Plomate	Maphore

Considérons deux relations $\mathcal{R}_1$ et $\mathcal{R}_2$ suivant le même schéma relationnel $\mathcal{S}$.

Union l'union de deux relations $\mathcal{R}_1$ et $\mathcal{R}_2$ est l'ensemble des enregistrements appartenant à la relation $\mathcal{R}_1$ ou à la relation $\mathcal{R}_2$. On le note $\mathcal{R}_1 \cup \mathcal{R}_2$.

Exemple : dans l'exemple ci-dessus, on obtient les 6 élèves qui sont passés en BCPST2 entre 2017 et 2018.

Intersection l'intersection de deux relations $\mathcal{R}_1$ et $\mathcal{R}_2$ est l'ensemble des enregistrements communs à $\mathcal{R}_1$ et à $\mathcal{R}_2$. On le note $\mathcal{R}_1 \cap \mathcal{R}_2$.

Exemple : Sophie Garrot et Robert Lingot sont les deux 5/2 en 2018.

Différence la différence entre deux relations $\mathcal{R}_1$ et $\mathcal{R}_2$ est l'ensemble des enregistrements de la relation $\mathcal{R}_1$ que l'on ne retrouve pas dans la relation $\mathcal{R}_2$. On le note $\mathcal{R}_1 \setminus \mathcal{R}_2$.

Exemple : Sandy Kilos et Thomas Nifique ont intégré (ou abandonné) en 2017.

3.3) Opérateurs spécifiques aux BDD

Soit $\mathcal{R}$ une relation de schéma $\mathcal{S}$. Dans les exemples, on prendra la table $\mathcal{R} = \text{Eleves2018}$.

Projection la projection est une opération consistant à ne garder qu'une partie des attributs.

Définition Pour $X \subset \mathcal{S}$, on appelle projection de $\mathcal{R}$ selon X la relation $\pi_X(\mathcal{R}) = \{e(X) / e \in \mathcal{R}\}$.
Le schéma de cette nouvelle relation est alors X .

Exemple : $\pi_{\text{prenom}}(\mathcal{R})$ fournit une table contenant 8 prénoms distincts.

Sélection la sélection est une opération consistant à ne conserver que les enregistrements vérifiant une certaine propriété. Celle-ci est en fait un prédicat sur $\mathcal{R}$.

Définition On appelle sélection de $\mathcal{R}$ selon la propriété $\mathcal{P}$ la relation $\sigma_{\mathcal{P}}(\mathcal{R}) = \{e \in \mathcal{R} / \mathcal{P}(e)\}$.

Exemple : $\sigma_{\text{prenom}=\text{Elodie ou lycee}=\text{Raynade}}(\mathcal{R})$ fournit une table de 3 élèves.

On peut distinguer différents types de sélections.

- **Sélection simple** : lorsque $\mathcal{R}$ suit $\mathcal{S}$, que $A \in \mathcal{S}$ et $a \in \text{Dom}(A)$, on dit que $\sigma_{A=a}(\mathcal{R})$ est une sélection simple. On peut en construire avec les différents opérateurs de comparaison $=, \neq, <, >, \leq, \geq, \in, \notin, \subset, \supset, \not\subset$ ou $\not\supset$.
- **Sélection composée** : grâce aux opérateurs ensemblistes, il est possible d'exprimer des conditions complexes en combinant des sélections simples.

Exemple : si l'on pose $\mathcal{R}_1 = \sigma_{\mathcal{P}_1}(\mathcal{R})$ et $\mathcal{R}_2 = \sigma_{\mathcal{P}_2}(\mathcal{R})$, alors $\sigma_{\mathcal{P}_1 \text{ et } \mathcal{P}_2}(\mathcal{R}) = \mathcal{R}_1 \cap \mathcal{R}_2$ et $\sigma_{\mathcal{P}_1 \text{ ou } \mathcal{P}_2}(\mathcal{R}) = \mathcal{R}_1 \cup \mathcal{R}_2$.

Une propriété composée d'opérateurs logiques correspond ainsi à des opérations ensemblistes sur des relations. On peut alors l'écrire directement, c'est plus clair !

Exemple : on préférera $\sigma_{\mathcal{P}_1} \text{ et } \mathcal{P}_2(\mathcal{R})$ à $\sigma_{\mathcal{P}_1}(\mathcal{R}) \cap \sigma_{\mathcal{P}_2}(\mathcal{R})$.

Sélection et projection vont souvent de paire et nous serons amenés fréquemment à les composer. Ainsi, si l'on s'intéresse aux noms et prénoms des élèves de BCPST issus du lycée Négal, on prendra $\pi_{\text{nom}, \text{prenom}} \left(\sigma_{\text{filier}=\text{BCPST et lycee}=\text{Négal}} \right) (\mathcal{R})$

Renommage c'est l'opération consistant à renommer un des attributs du schéma.

Définition Soient $A \in \mathcal{S}$ et $B \notin \mathcal{S}$ un attribut tel que $\text{Dom}(A) = \text{Dom}(B)$.
On note $\rho_{A \rightarrow B}(\mathcal{R})$ la relation obtenue en renommant A en B .
Elle suit le schéma $\rho_{A \rightarrow B}(\mathcal{S}) = (\mathcal{S} \setminus \{A\}) \cup \{B\}$.

Par extension, on notera $\rho_{A_1, \dots, A_n \rightarrow B_1, \dots, B_n}$ l'opération composée $\rho_{A_1 \rightarrow B_1} \circ \dots \circ \rho_{A_n \rightarrow B_n}$.

Exemple : il est bougrement tentant d'effectuer $\rho_{\text{filier} \rightarrow \text{section}}(\mathcal{R})$.

4/ Notion de clef

4.1) Introduction

Pour garantir la non-répétition des enregistrements, les BDD utilisent une *clef* qui doit être indiquée lors de la création de la base. Typiquement, c'est un attribut entier qui est automatiquement incrémenté à chaque nouvelle entrée et qui peut ainsi servir à identifier les entrées sans ambiguïté. Ce concept de clef va aussi avoir son importance lors de l'utilisation de relations multiples.

En effet, il est souhaitable de répartir les données dans plusieurs tables pour éviter la redondance d'informations. Ainsi, la BDD sera définie par plusieurs schémas relationnels et plusieurs relations. Les clefs vont alors permettre d'utiliser simultanément plusieurs relations lors de recherches croisées.

Exemple : il est naturel de diviser la table Eleves2018 en trois tables distinctes, afin de séparer les informations concernant les classes, les élèves et leurs lycées d'origine.

Classes			Lycees		
id_C	filier	n°	id_L	nom	ville
01	BCPST	1	01	Maphore	Brest
02	BCPST	2	02	Raynade	Nantes
			03	Négal	Dakar
			04	Pulture	Poitiers
			05	Minherr	Lille
			06	Dimant	Strasbourg

Eleves

prenom	nom	id_C	id_L	note_bac
Jean	Bonnot	01	01	14.2
Jean	Aymard	01	02	13.7
Robert	Lingot	02	03	15.4
Élodie	Nosaure	02	02	17.1
Bernard	Tichaud	01	03	12.8
Pauline	Ohm	01	04	16.8
Élodie	Plomate	02	01	11.9
Sophie	Garrot	02	05	18.0
Pascale	Feutrage	01	01	14.7
Bernard	Bonnot	01	06	12.1
Paul	Hymaire	01	04	15.2

On peut alors aisément ajouter un attribut dans la table Lycees (en l'occurrence ville) sans alourdir la table Eleves : nul besoin de répéter 3 fois que le lycée Maphore est à Brest!!!

4.2) Clef, clef minimale et clef primaire

Soit $\mathcal{R}$ une relation de schéma relationnel $\mathcal{S}$.

Définition Une *clef* de la relation $\mathcal{S}$ est un ensemble d'attributs $X \subset \mathcal{S}$ tel que

$$\forall e_1, e_2 \in \mathcal{R} \quad e_1(X) = e_2(X) \implies e_1 = e_2$$

Remarque : du fait de l'unicité des enregistrements, $\mathcal{S}$ est toujours une clef.

Proposition

Soit $X \subset \mathcal{S}$. Alors X est une clef de $\mathcal{R}$ si, et seulement si, $\# \pi_X(\mathcal{R}) = \# \mathcal{R}$.

Exemples : examinons les trois tables Classes, Lycees et Eleves.

- ❶ id_C et id_L sont des clefs des relations Classes et Lycees.
- ❷ nom est une clef pour Lycees pour le moment ... mais il n'est pas exclu que l'on enregistre d'autres lycées de même nom dans la table.
- ❸ Le même problème se pose avec $(\text{prenom}, \text{nom})$ dans la table Eleves : des homonymies ne sont pas exclues lors de futurs enregistrements.

Définition Une clef est *minimale* si la suppression de l'un de ses attributs lui fait perdre son caractère de clef.

Exemple : la clef $(\text{prenom}, \text{nom})$ de la table Eleves est minimale car il y a des élèves de même nom ou de même prénom.

Une relation peut très bien posséder plusieurs clefs minimales ; évidemment, on privilégiera le cas échéant celles qui possèdent le moins d'attribut. Le cas idéal consiste évidemment en une clef composée d'un seul attribut (qui a été créé pour cela en général).

Définition Si $A \in S$ est une clef pour $\mathcal{R}$, on dit que c'est une **clef primaire**.

Exemples : les attributs id_C et id_L sont bien entendu des clefs primaires.

Proposition

Un attribut $A \in S$ est une clef primaire si, et seulement si,

$$\forall e_1, e_2 \in \mathcal{R} \quad e_1 \neq e_2 \implies e_1 \cdot A \neq e_2 \cdot A$$

Une clef primaire permet ainsi d'identifier les enregistrements de la relation, ce qui renforce l'efficacité des procédures d'interrogation. Bien entendu, le schéma seul ne permet pas de garantir qu'un attribut constitue une clef primaire car il risque de perdre son caractère de clef lorsque la BDD va grossir (ce qui est sa vocation *a priori*).

En pratique, on ajoute souvent un attribut dans la relation qui permet de numéroté artificiellement les enregistrements : c'est aussi ce qu'on appelle un **identifiant**.

Exemple : pour la table *Eleves*, il suffit de rajouter un attribut id_E et de numéroté les élèves au fur et à mesure des enregistrements.

4.3) Clef étrangère

Définition Une **clef étrangère** d'une relation $\mathcal{R}$ est un attribut de $\mathcal{R}$ prenant les mêmes valeurs que la clef primaire d'une autre relation $\mathcal{R}'$.

Ceci permettra de faire le lien entre plusieurs relations. En pratique, la clef étrangère est l'attribut (si besoin renommé) servant de clef primaire dans une autre relation.

Exemple : dans la table *Eleves*, les attributs id_C et id_L sont des clefs étrangères. Elles permettent de reconstituer la table *Eleves_2018* à partir des trois tables séparées.

5/ Opérateurs complexes de l'algèbre relationnelle

Les opérateurs relationnels croisés jusqu'ici ne s'appliquent qu'à une seule relation ou, pour les opérateurs ensemblistes, à deux relations suivant le même schéma.

Dans une base de donnée, on est amenés à manipuler des relations qui ont rarement le même schéma. Et même lorsque c'est le cas, leur union ou leur intersection n'ont pas forcément de sens. On définit donc d'autres opérateurs afin de croiser les informations présentes dans les différentes relations.

5.1) Opérations ensemblistes avancées

Soient $\mathcal{R}$ et $\mathcal{R}'$ deux relations de schémas respectifs $\mathcal{S}$ et $\mathcal{S}'$ tels que $\mathcal{S} \cap \mathcal{S}' = \emptyset$. Notons que cette condition est facilement satisfaite en renommant les attributs.

Produit cartésien le premier moyen d'assembler deux relations est d'en faire le produit cartésien. On se restreint au cas où les schémas sont disjoints pour ne pas se préoccuper de la concordance des attributs communs.

Définition On appelle *produit de $\mathcal{R}$ par $\mathcal{R}'$* la relation de schéma $\mathcal{S} \cup \mathcal{S}'$ définie par

$$\mathcal{R} \times \mathcal{R}' = \{u \mid u(\mathcal{S}) \in \mathcal{R} \text{ et } u(\mathcal{S}') \in \mathcal{R}'\}$$

Exemple : Entrees $\times$ Plats $\times$ Desserts nous donne tous les menus possibles.

Remarque : cette nouvelle relation a un cardinal important. De plus, le résultat n'a pas forcément beaucoup de sens : dans le produit Eleves $\times$ Lycee, on peut rencontrer sur une même ligne un élève et un lycée dont il n'est pas issu. Aucun intérêt !

Jointure c'est la sélection des éléments du produit de deux relations vérifiant une condition.

Définition Si $\mathcal{C}$ est une condition booléenne sur $\mathcal{S} \cup \mathcal{S}'$, on appelle *jointure de $\mathcal{R}$ et $\mathcal{R}'$ selon $\mathcal{C}$* la relation de schéma $\mathcal{S} \cup \mathcal{S}'$ définie par

$$\mathcal{R} \bowtie_{\mathcal{C}} \mathcal{R}' = \sigma_{\mathcal{C}}(\mathcal{R} \times \mathcal{R}') = \{e \in \mathcal{R} \times \mathcal{R}' \mid \mathcal{C}(e) \text{ est vraie}\}$$

Remarque : en général, la condition est l'égalité de deux attributs. On obtient alors une *jointure symétrique*, qui permet de fusionner les relations en identifiant ces attributs. En pratique, on rajoute une projection pour supprimer les attributs redondants.

Exemple : considérons Eleves $\bowtie_{\mathcal{C}}$ Lycees pour $\mathcal{C} = \text{Eleves.id}_L = \text{Lycees.id}_L$.

On a utilisé dans cet exemple une clef primaire et une clef étrangère pour effectuer une jointure symétrique. La relation obtenue a alors un sens, car on a regroupé les données compatibles des deux tables.

La jointure est plus rapide à effectuer que les compositions de produits et de sélection, on s'efforcera donc de privilégier cette opération pour optimiser les requêtes.

5.2) Agrégats et fonctions d'agrégation

L'idée consiste à regrouper les données selon les valeurs d'un attribut – les groupes obtenus sont appelés des *agrégats* – et à effectuer une opération sur chaque agrégat.

Exemple : calculons la moyenne des notes au bac des élèves issus de chaque lycée.

Pour cela, il faudra utiliser des fonctions qui peuvent prendre autant d'arguments que possible (la taille des agrégats est variable) et qui ne dépendent pas de l'ordre des arguments (il n'y a pas d'ordre dans les enregistrements d'une relation).

Définition

Une fonction d'agrégation sur un ensemble E est une fonction pouvant être appliquée à un nombre quelconque d'arguments et indépendamment de leur ordre.

Exemples : le cardinal d'une liste ; le max, le min, la somme ou la moyenne de ses éléments.

Proposition

Soient $\mathcal{R}(\mathcal{S})$ une relation, $X \subset \mathcal{S}$, $A \in \mathcal{S} \setminus X$ et f une fonction d'agrégation sur $\text{dom}(A)$. On note ${}_X\gamma_{f(A)}(\mathcal{R})$ la relation de schéma $X \cup \{f(A)\}$ obtenue :

- en regroupant les éléments de $\mathcal{R}$ selon les valeurs des attributs de X ;
- en calculant les valeurs $f(A)$ pour chacun de ces groupes.

Exemples : reprenons $\mathcal{R} = \text{Eleves2018}$

- ❶ $\text{lycee_origine} \gamma_{\text{MOYENNE}(\text{note_bac})}(\mathcal{R})$ calcule la moyenne des notes obtenues au bac lycée par lycée ;
- ❷ $\gamma_{\text{MOYENNE}(\text{note_bac})}(\mathcal{R})$ calcule la moyenne des notes au bac de tous les élèves ;

Bien entendu, on sera amenés à combiner les agrégations et des sélections, qu'elles se produisent en amont (avant) ou en aval (après).

Exemples :

- on obtient les lycées pour lesquels la moyenne des notes au bac dépasse 12 avec $\pi_{\text{lycee_origine}}(\sigma_{\text{MOYENNE}(\text{note_bac}) > 12}(\text{lycee_origine} \gamma_{\text{MOYENNE}(\text{note_bac})}(\mathcal{R})))$
- on récupère la moyenne des notes au bac des élèves de BCPST1 grâce à $\gamma_{\text{MOYENNE}(\text{note_bac})}(\sigma_{\text{filiere}=\text{BCPST et numero}=1}(\mathcal{R}))$

On privilégiera les sélections en amont : elles permettent en effet d'appliquer les fonction d'agrégations à des relations plus petites, donc la recherche est plus rapide.

Exercices

Exercice 1.

Proposer des domaines appropriés pour les attributs suivants :

- | | | |
|----------------|-----------------|--------------------|
| 1) nationalité | 2) âge | 3) adresse e-mail |
| 4) date | 5) mot de passe | 6) note de musique |

Exercice 2.

Proposer des schémas relationnels représentant :

- | | |
|----------------------------------|---------------------------------------|
| 1) l'inventaire d'un supermarché | 2) le parc informatique d'un lycée |
| 3) les animaux d'un zoo | 4) les différents éléments chimiques. |

Dans chaque cas, donner un exemple raisonnable de relation associée au schéma.

Exercice 3.

Soient $\mathcal{R}(\mathcal{S})$ une relation et $X \subset \mathcal{S}$. Montrer que $\# \pi_X(\mathcal{R}) \leq \# \mathcal{R}$.

Exercice 4.

Considérons la relation $\mathcal{R} = \text{Eleves2018}$ vue en cours. Traduire les requêtes suivantes :

- 1) Obtenir la liste des lycées d'origine
- 2) Obtenir les prénoms des élèves de BCPST1
- 3) Obtenir les noms des élèves ayant eu une note au bac inférieure à 14
- 4) Obtenir le nombre d'élèves issus de chaque lycée
- 5) Obtenir la moyenne des notes au bac pour chaque classe

Exercice 5.

On va considérer un ensemble de relations permettant de gérer un complexe hôtelier.

- Celui-ci est composé de différents bâtiments, identifiés par leur nom et leur nombre d'étoiles. Ils sont représentés dans la relation `batiment` de schéma `(nom,etoiles)`.
- Les chambres, caractérisées par leur numéro, nom du bâtiment et nombre de fenêtres, sont représentées dans la relation `chambre` de schéma `(numero,batiment,fenetres)`.
- Certaines chambres possédant plusieurs lits, on a aussi une relation `lit` de schéma `(id_lit,chambre,bat_lit)` où `id_lit` est un identifiant unique au sein du complexe.
- Enfin, les nuitées sont caractérisées par le nom du client, l'identifiant du lit et la date. Elles forment alors une relation `nuitee` de schéma `(client,lit,date)`.

Obtenir le nom des clients ayant séjourné :

- 1) dans le bâtiment Rose.
- 2) dans un bâtiment trois étoiles.
- 3) dans une chambre ayant au moins deux fenêtres.

