Usage des bases de données : opérations et requêtes

Les bases de données contenant de très nombreuses informations, la consultation ou la modification passe par certains opérateurs décrit mathématiquement et mise en œuvre informatiquement en langage SQL. La consultation ou la modification de bases de données se fera en langage SQL, par des requête, sur la base d'opérateur mathématique.

L'usage de ces operateurs est appelé une requête.

1 Exemple introductif : les cinémas

Films						
id	Nom	Realisateur	Pays	Annee	NoteCritique	NotePublic
1	Star Wars	Lucas	USA	1977	7.0	8.3
2	Asterix : Mission Cleopatre	Chabat	France	2012	4.1	8.0
3	La mort aux trousses	Hitchcock	USA	1959	9.0	8.1
4	Les oiseaux	Hitchcock	USA	1963	9.4	8.1
5	La dolce vita	Fellini	Italie	1960	8.8	8.8
6	Bienvenu chez les chtis	Boon	France	2008	4.2	9.1
7	La cité de Dieu	Meirelles	Bresil	2002	8.8	8.2
8	Le loup de Wall Street	Scorsese	USA	2013	8.2	7.6

Cine1	
Titre	Seance
Asterix : Mission Cleopatre	20 :00
Star Wars	20 :00
La cité de Dieu	20 :30
La dolce vita	20 :30
Le loup de Wall Street	21 :00
Les oiseaux	22 :00
Bienvenu chez les chtis	22 :00

Cine2	
Titre	Seance
La mort aux trousses	20 :00
Star Wars	20 :00
La cité de Dieu	20 :15
La dolce vita	20 :30
Le loup de Wall Street	20 :40
Asterix : Mission Cleopatre	22 :00
Bienvenu chez les chtis	22 :00
Le loup de Wall Street	22 :15

2 Opérateurs simples de l'algèbre relationnelle

2.1 Projection

2.1.1 Defnition

Soit $R(S)$ une relation de schéma S, et $X \subset S$.
On appelle projection de R selon X la relation :

$$\Pi_X(R) = \{e(x)/e \in R\}$$

2.1.2 Ecriture SQL

```sql
SELECT Ai,..., An
FROM R
;
```

2.1.3 Application 1

Détailler le contenu de $\Pi_{(Nom,Annee)}(Films)$.
Quelle requête permet d'obtenir ce résultat ?

2.1.4 Application 2

Soit $E = (Pays)$. Quel est le contenu de $\Pi_E(Films)$? Ecrire la requête correspondante
Que vaut $\#\Pi_E(Films)$?

2.2 Selection (ou restriction)

2.2.1 définition

Soit $R(S)$ une relation de schéma S ; $A \in S$ et $a \in dom(A)$.
On appelle $\sigma_{A=a}(R)$ la sélection de R selon $A = a$ une nouvelle relation formée des enregistrements de R respectant la condition $A = a$.

2.2.2 Ecriture SQL

```sql
SELECT *
FROM R
WHERE condition
;
```

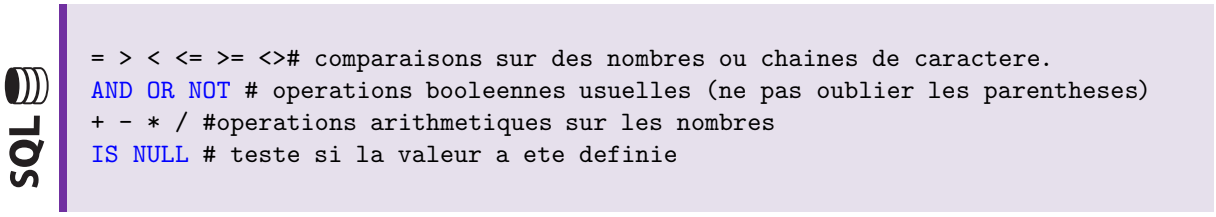
2.2.3 Application 3

Détailler le contenu de :

$$\sigma_{Realisateur="Hitchcock"}(Films)$$

Ecrire la requête permettant d'obtenir ce résultat.

2.2.4 Tests et opérations possibles pour la condition WHERE



```
= > < <= >= <># comparaisons sur des nombres ou chaines de caractere.
AND OR NOT # operations booleennes usuelles (ne pas oublier les parentheses)
+ - * / #operations arithmetiques sur les nombres
IS NULL # teste si la valeur a ete definie
```

2.3 Remarque 1 : projection et selection

2.3.1 Application 4

Proposer une requête pour obtenir une relation contenant le nom des films et leur année, uniquement les films postérieurs à 2000.

2.4 Remarque 2 : opérations sur les données

En langage SQL, il est aussi possible de réaliser les opérations précédentes dans le champ `SELECT`. L'opération peut alors s'effectuer entre une donnée et une constante, ou entre deux données d'un même enregistrement.

2.4.1 Application 5

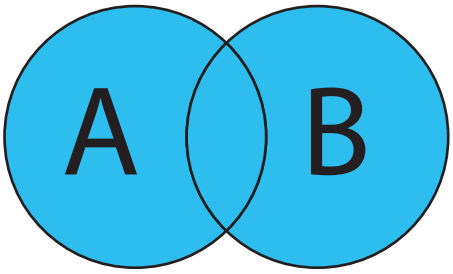
Proposer une requête pour obtenir une relation contenant le nom des films et leur âge.

Proposer une requête pour obtenir la liste des noms des films dont la note du public dépasse de 1 point la note de la critique.

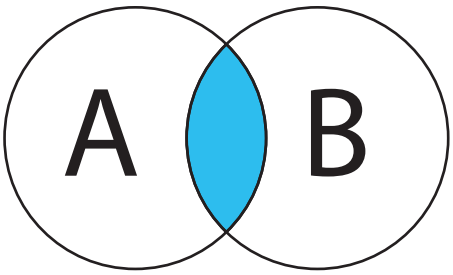
3 Opérateurs sur les ensembles

3.1 Principe

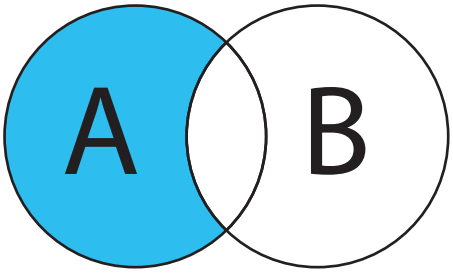
3.1.1 Union



3.1.2 Intersection



3.1.3 Différence



3.1.4 Ecriture SQL

Union	$A \bigcup B$	`SELECT * FROM A UNION SELECT * FROM B ;`
Intersection	$A \bigcap B$	`SELECT * FROM A INTERSECT SELECT * FROM B ;`
Différence	$A \setminus B$	`SELECT * FROM A EXCEPT SELECT * FROM B ;`

3.1.5 Application 6

Q1 : Détailler la relation $(cine1 \bigcup cine2)$?
Q2 : Détailler la relation $cine1 \bigcap cine2$
Q3 : Détailler la relation $(cine1 \setminus cine2)$?

4 Opérateurs plus évolués

4.1 Renommage

Pour changer le titre de l'attribut lors d'une projection :

```
SELECT Nom AS titre, Realisateur FROM Films ;
```

4.1.1 Application 7

Extraire la table des "cartons" cinématographiques (note public strictement supérieure à 8.5) et leur réalisateur.

4.2 Elimination des doublons

Pour éléminer des redondances dans les résultats, il faut simplement ajouter `DISTINCT` après le mot `SELECT`.

```
SELECT DISTINCT Realistaeur FROM Films ;
```

4.3 Ordonnancement et limitation des résultats

La commande ORDER BY permet de trier les lignes dans un résultat d'une requête SQL. La commande LIMIT est utilisée pour spécifier le nombre maximum de résultats que l'ont souhaite obtenir. Cette commande est souvent associé à un OFFSET, c'est-à-dire un décalage sur le jeu de résultat.

```
SELECT colonne1,colonne2 FROM table
ORDER BY colonne1 LIMIT 10 OFFSET 5 ;
```

4.3.1 Application 8

Ecrire une requête qui affiche le nom des 3 meilleurs films selon le public. Puis une autre pour afficher ceux au rang 4 à 6.

5 Fonctions d'agrégation

La notion d'agrégation fait ici référence au traitement de plusieurs données, agrégées en une seule, par exemple la moyenne des valeurs d'une colonne.

Les mots clés sont les suivants :

Comptage du nombre de valeurs prises par un attribut	COUNT()
Maximum des valeurs prises par un attribut	MAX()
Minimum des valeurs prises par un attribut	MIN()
Somme des valeurs prises par un attribut	SUM()
Moyenne des valeurs prises par un attribut	AVG()

5.1 Regroupement et filtrage

Les fonctions précédentes traitent un groupe de données, dans le cas le plus simple toute la colonne.

Pour affiner ce type de traitement, il est possible de regrouper les enregistrements par une même valeur d'attribut. Le mot clé à utiliser est alors `GROUP BY`.

```sql
SELECT AVG(Annee) FROM Films GROUP BY Realisateur ;
```

En complément, il est possible d'éliminer certaines données. On ne garde que les données selectionnées avec le mot clé `HAVING`

```sql
SELECT Pays, AVG(NotePublic) FROM Films GROUP BY Pays HAVING AVG(NotePublic)>8.5 ;
```

5.1.1 Application 9

Q1 : donner la moyenne des notes du public ;
Q2 : donner la meilleur note de la critique de chaque réalisateur ;
Q3 : donner le nombre de film par pays.

6 Sous-requêtes ou requêtes imbriquées

Une requête peut avoir besoin de données ayant déjà subi un prétraitement. Le prétraitement nécessaire peut être mise en place par une requête imbriquée, qui apparait entre parenthèse dans la requête principale.

```
SELECT Nom FROM Films WHERE NoteCritique>=(SELECT AVG(NoteCritique) FROM Films);
```

La requête imbriquée peut renvoyer :

- un résultat simple,
- une colonne,
- une table.

6.1 Application 10

1. Afficher la moyenne des notes de la critique.
2. Afficher le nom des films et la différence entre sa note (de la critique) et la moyenne.
3. Afficher le nom des films qui ont une meilleure note que la moyenne de la critique.
4. Afficher le nom du films issus des 3 meilleurs pays (pays avec la meilleure moyenne de note de la critique).

7 Jointure

7.1 Produit

Le produit entre deux tables A et B est noté :

$$P = A \times B$$

Le schéma de P est $(A_1, \ldots, A_n, B_1, \ldots, B_m)$

Ses enregistrements sont toutes les combinaisons des enregistrements de A et des enregistrements de B.

7.1.1 Exemple

Viande
type viande
Steak
Poulet
Magret

Accompagnement
type accompagnement
Riz
Frites

Viande×Accompagnement	
type viande	*type accompagnement*
Steak	Riz
Steak	Frites
Poulet	Riz
Poulet	Frites
Magret	Riz
Magret	Frites

7.1.2 Ecriture SQL

```sql
SELECT * FROM Viande,Accompagnement;
```

7.1.3 Application 11 : tournoi sportif

On dispose de la table `Equipes`. On souhaite organiser un tournoi par poule : chaque équipe doit réncontrer toutes les autres équipes. Proposer une requête qui permette d'arriver à la table souhaitée.

Equipes	
Nom	**Entraineur**
Bordeaux	X
Marseille	Y
Toulouse	Z

domicile	**visiteur**
Bordeaux	Marseille
Bordeaux	Toulouse
Marseille	Bordeaux
Marseille	Toulouse
Toulouse	Bordeaux
Toulouse	Marseille

7.2 Jointure

Une jointure permet d'extraire des données de tables différentes, mais qui restent liées entre elles.

Une jointure de deux tables consiste à considérer le produit cartésien de ces deux tables (ou d'extractions), en identifiant deux colonnes (une de chaque table) de sorte à ne garder dans le produit cartésien que les n-uplets tels que les valeurs soient les mêmes pour ces deux attributs.

Les données des deux colonnes choisies doivent donc être cohérentes entre elles. Pour cette raison, on choisit typiquement d'un côté une clé étrangère vers l'autre table, et de l'autre la clé primaire.

7.2.1 Exemple

etudiants		
Nom	Année de Naissance	Ville
Alice	1997	Strasbourg
Bob	1998	Paris
Eve	1995	Strasbourg

villes	
NomV	Département
Strasbourg	67
Paris	75

etudiants×villes				
Nom	Année de Naissance	Ville	NomV	Département
Alice	1997	Strasbourg	Strasbourg	67
Alice	1997	Strasbourg	Paris	75
Bob	1998	Paris	Strasbourg	67
Bob	1998	Paris	Paris	75
Eve	1995	Strasbourg	Strasbourg	67
Eve	1995	Strasbourg	Paris	75

Suppression des lignes				
Nom	Année de Naissance	Ville	NomV	Département
Alice	1997	Strasbourg	Strasbourg	67
Alice	1997	Strasbourg	Paris	75
Bob	1998	Paris	Strasbourg	67
Bob	1998	Paris	Paris	75
Eve	1995	Strasbourg	Strasbourg	67
Eve	1995	Strasbourg	Paris	75

Jointure				
Nom	Année de Naissance	Ville	NomV	Département
Alice	1997	Strasbourg	Strasbourg	67
Bob	1998	Paris	Paris	75
Eve	1995	Strasbourg	Strasbourg	67

7.2.2 Ecriture SQL

Soient deux tables $T1$ et $T2$ contenant trois attributs chacune.

```sql
SELECT T1.Att1, T1.Att2, T2.Att3 FROM T1 JOIN T2 ON T1.Att1 = T2.Att3;
```

7.2.3 Application 12

Donner la liste des films français et de leurs horaires dans le cinema 1

7.3 Autojointure

Imaginons un régime autoritaire où, pour assurer le calme, les individus se surveillent entre eux.

Liste		
Id	Nom	est_surveillé_par
1	Alice	2
2	Alain	1
3	Bob	1
4	Bobby	1
5	Chico	1
6	Dimitri	2

1. Tracer la représentation graphique associée au schéma relationnel (clé primaire et lien 1-*) ;
2. Ecrire la requête permettant d'afficher la table des noms des individus et de leur surveillants respectifs.