

1. Introduction – rappels.

Exemple : Météo France publie et stocke chaque jour des informations météorologiques issues de ses stations (réseau RADOME).

Ces données d'observations peuvent par exemple être destinées à des messages internationaux d'observation en surface (SYNOP) qui circulent sur le système mondial de télécommunication (SMT) de l'Organisation Météorologique Mondiale (OMM). Elles contiennent des paramètres atmosphériques mesurés (température, humidité, direction et force du vent, pression atmosphérique, hauteur de précipitations) ou observés (temps sensible, description des nuages, visibilité) depuis la surface terrestre. Selon l'instrumentation et des spécificités locales, d'autres paramètres peuvent être disponibles (hauteur de neige, état du sol, etc.).



Réseau Météo France RADOME

Problème : Comment stocker, accéder et exploiter ces quantités gigantesques de données ?

Exemple : Schéma de la relation « RADOME »

Les attributs sont disposés en colonnes : « num_sta », « Nom usuel », « Département », etc.

RADOME =

Chaque ligne est un tuple (ou n-uplet), instance de la relation « RADOME ».

Par exemple : (7020, PTE DE LA HAGUE, 50, etc.).

num_sta	nom_usuel	département	commune	altitude	latitude	longitude
7005	ABBEVILLE	80	ABBEVILLE	70	50.135833	1.834667
7015	LILLE-LESQUIN	59	LESQUIN	47	50.570000	3.097500
7020	PTE DE LA HAGUE	50	AUDERVILLE	6	49.725167	-1.939833
...

L'ordre des lignes n'est pas important. Par contre, chaque tuple doit être unique. Pour les identifier, on définit une **clé primaire** qui peut être constituée d'un ou plusieurs attributs.

Le **schéma d'une base de données relationnelle**, est constitué d'un ensemble de schémas relationnels.

Exemple : Schéma de la base de donnée relationnelle « BDD Météo France »

Cette base de données comporte de nombreux schémas de relation. On ne s'intéresse qu'aux deux suivants. En plus du schéma de relation « RADOME », il y a le schéma « SYNOP ».

On peut écrire :

BDD Météo France =

RADOME =

SYNOP =

Attributs inclus dans les données SYNOP essentielles OMM :

Descriptif	Attribut	Type	Unité
Indicatif OMM station	num_sta	car	
Date	Date	car	
Pression au niveau mer	pmer	int	Pa
Variation de pression en 3 heures	tend	int	Pa
Type de tendance barométrique	cod_tend	int	code
Direction du vent moyen 10 mn	dd	int	degré
Vitesse du vent moyen 10 mn	ff	réel	m/s
...

NOTA :

2. L'algèbre relationnelle

Problème : On souhaite effectuer des requêtes sur la base de données « BDD Météo France ».

Par exemples :

- Quelles sont les stations situées dans le var ?
- Quelles sont les stations où la pression au niveau de la mer est passée en dessous de 100000 Pa le 13/06/2013 ?
- Quelles sont les villes où la vitesse du vent moyen a dépassé les 20 m/s en 2013 ?

On entend par **algèbre relationnelle**, une **collection d'opérations** formelles qui agissent sur des relations et produisent une relation en résultat : $R_3 \leftarrow R_1 \text{ Op } R_2$.

Dans la plupart des Systèmes de Gestion de Base de Données relationnels, la réponse à une requête s'obtient par l'utilisation d'un ou plusieurs opérateurs relationnels.

Parfois, il est possible d'exprimer de plusieurs manières une même requête. Même si le résultat est identique pour chacune d'entre-elles, il faut privilégier celle qui est la plus efficace en terme de rapidité de réponse. Certaines requêtes s'optimisent.

On précisera dans ce qui suit les opérateurs de l'algèbre relationnelle, ainsi que les **requêtes SQL (Structured Query Language)** correspondantes, puisqu'elles sont largement utilisées dans les Systèmes de Gestion de Base de Données.

2.1. Opérateurs relationnels

- **Projection :** Sélection de certaines colonnes, élimination d'attributs.

$$R_2 \leftarrow \pi_{\text{attribut}_i, \text{attribut}_j, \text{etc.}} (R_1)$$

SQL : $R_2 \leftarrow \text{SELECT attribut}_i, \text{attribut}_j, \text{etc. FROM } R_1$

Exemple : « Quels sont les numéros de station et les villes correspondantes ? »

Algèbre relationnelle :

SQL :

- **Sélection (restriction)** : Sélection de certaines lignes (sous-ensemble de tuples) qui vérifient une condition).

Le schéma du résultat est identique à celui sur lequel se porte la sélection.

Opérateurs pour exprimer la condition : =, ≠, ≤, <, >, ≥, ∧, ∨, ¬

$$R_2 \leftarrow \sigma_{\text{attribut=condition}}(R_1)$$

SQL : $R_2 \leftarrow \text{SELECT } * \text{ FROM } R_1 \text{ WHERE attribut = condition}$

Exemple : « Quels sont les numéros de station et les villes correspondantes situées dans le var ? »

Algèbre relationnelle :

SQL :

- **Renommage** : Un attribut « a » est renommé en « b ».

Le schéma du résultat est identique à celui sur lequel se porte le renommage.

$$R_2 \leftarrow \rho_{a \rightarrow b}(R_1)$$

SQL : $R_2 \leftarrow \text{SELECT } a \text{ AS } b \text{ FROM } R_1$

- **Jointure** : Combiner une paire de tuples de deux relations en un seul tuple. On ne s'intéresse ici qu'à la **jointure symétrique simple** qui permet de recoller deux relations ayant un attribut en commun.

$$R_3 \leftarrow \sigma_{R_1.\text{attribut}=R_2.\text{attribut}}(R_1 \times R_2) \quad (R_3 \leftarrow R_1 \bowtie_{R_1.\text{attribut}=R_2.\text{attribut}} R_2)$$

SQL :

Exemple : « Quelles sont les villes où la vitesse du vent moyen a dépassé les 20 m/s ? »

Algèbre relationnelle :

SQL :

- **Fonctions d'agrégation** : Ces fonctions qui ne sont pas couvertes par l'algèbre relationnelle classique, permettent d'effectuer des calculs statistiques basiques sur les valeurs : **MIN**, **MAX**, **SUM** (somme), **AVG**(moyenne), **COUNT**(comptage du nombre de lignes).

Exemple : « Quelle est la ville et la valeur de la pression où la plus haute valeur a été mesurée ? »

Algèbre relationnelle :

SQL :

NOTA : Formalisme de générale l'instruction SQL « **SELECT** »

SELECT <liste d'expressions>

FROM <liste de tables>

WHERE <conditions>

GROUP BY <liste d'attributs>

HAVING <conditions>

ORDER BY <liste d'attributs>

2.2. Opérateurs ensemblistes.

Les opérateurs ensemblistes sont dédiés à des relations de même schéma relationnel.

- **Union** : $R_3 \leftarrow R_1 \cup R_2$: Relation contenant les tuples appartenant à R_1 et à R_2 .

Il y a suppression des lignes identiques.

Syntaxe SQL : $R_3 \leftarrow R_1 \text{ UNION } R_2$

Exemple : « Quels sont les numéros de station situés dans le var et celles situées dans les bouches du Rhône ? »

Algèbre relationnelle :

SQL :

- **Intersection** : $R_3 \leftarrow R_1 \cap R_2$: Relation contenant les tuples appartenant à R_1 et à R_2 .

On garde les lignes identiques.

Syntaxe SQL : $R_3 \leftarrow R_1 \text{ INTERSECT } R_2$

Exemple : « Quels sont les numéros de station situés dans le var et à une altitude supérieure à 70 m ? »

Algèbre relationnelle :

SQL :

- **Différence** : $R_3 \leftarrow R_1 - R_2$: Relation contenant les tuples de R_1 n'appartenant pas à R_2 .

Attention : relation non commutative et non associative.

Syntaxe SQL :

Exemple : « Quels sont les numéros de station situés dans le var et qui ne sont pas à une altitude supérieure à 70 m ? »

Algèbre relationnelle :

SQL :

- **Produit cartésien** : $R_3 \leftarrow R_1 \times R_2$: Relation contenant l'ensemble des possibilités d'association entre une valeur de R_1 et une valeur de R_2 .

On l'utilise notamment pour définir l'opérateur relationnel de type « **jointure** » (voir ci-avant).

NOTA : Le langage SQL utilise encore d'autres instructions. On pourra s'y reporter.