



Laboratoire d'Informatique Scientifique et Industrielle
École Nationale Supérieure de Mécanique et d'Aérotechnique
1, avenue Clément Ader - BP 40109 - 86961 Futuroscope cedex - France



From Designing Data Warehouse Applications to Tuning

Ladjel BELLATRECHE

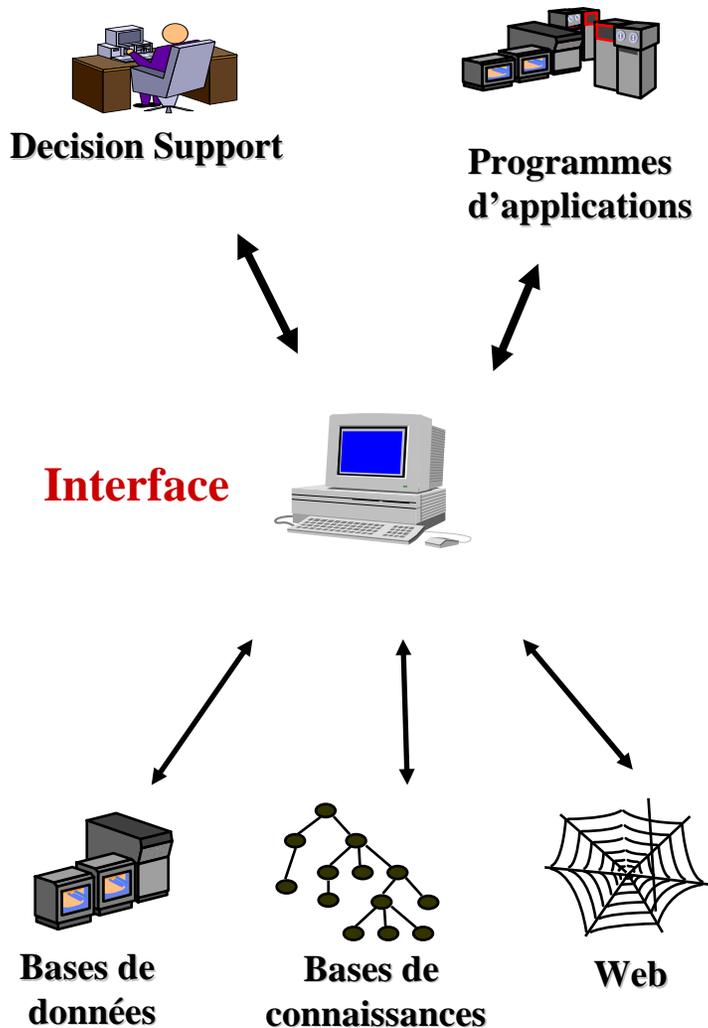
LISI/ENSMA – Université de Poitiers

bellatreche@ensma.fr

Pur Produit de l'Université de Sidi Bel Abbès

Promo : 1987 ~ 1992

- Prof. Tamer Özsu University of Waterloo - Canada
 - Optimization and tuning in Data warehouses. To Appear in Encyclopedia of Database Systems, Tamer Özsu, Springer, December 2008
- Participants :
 - Kamel Boukhalifa – Doctorant [Univ. Poitiers ~ USTHB]
 - Tahar BIDI – Magister INI
 - Soumia Benkrid – Magister INI
 - Sybille Caffiau – INRIA France
 - Prof. Pascal Richard – LISI/ENSMA Poitiers



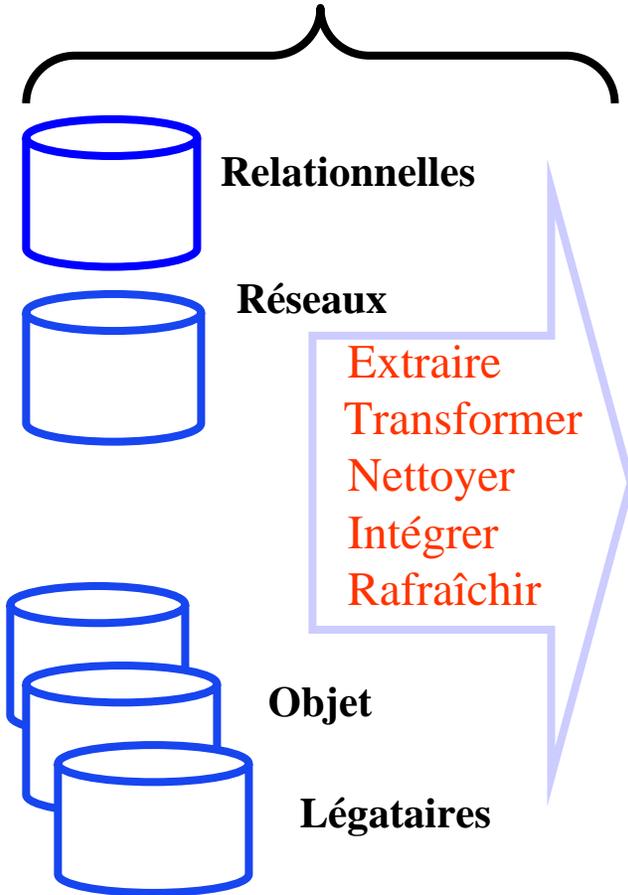
Caractéristiques des sources de données :

- ➔ Hétérogènes (schématique, sémantique),
- ➔ Autonomes
- ➔ Évolutives
- ➔ Réparties

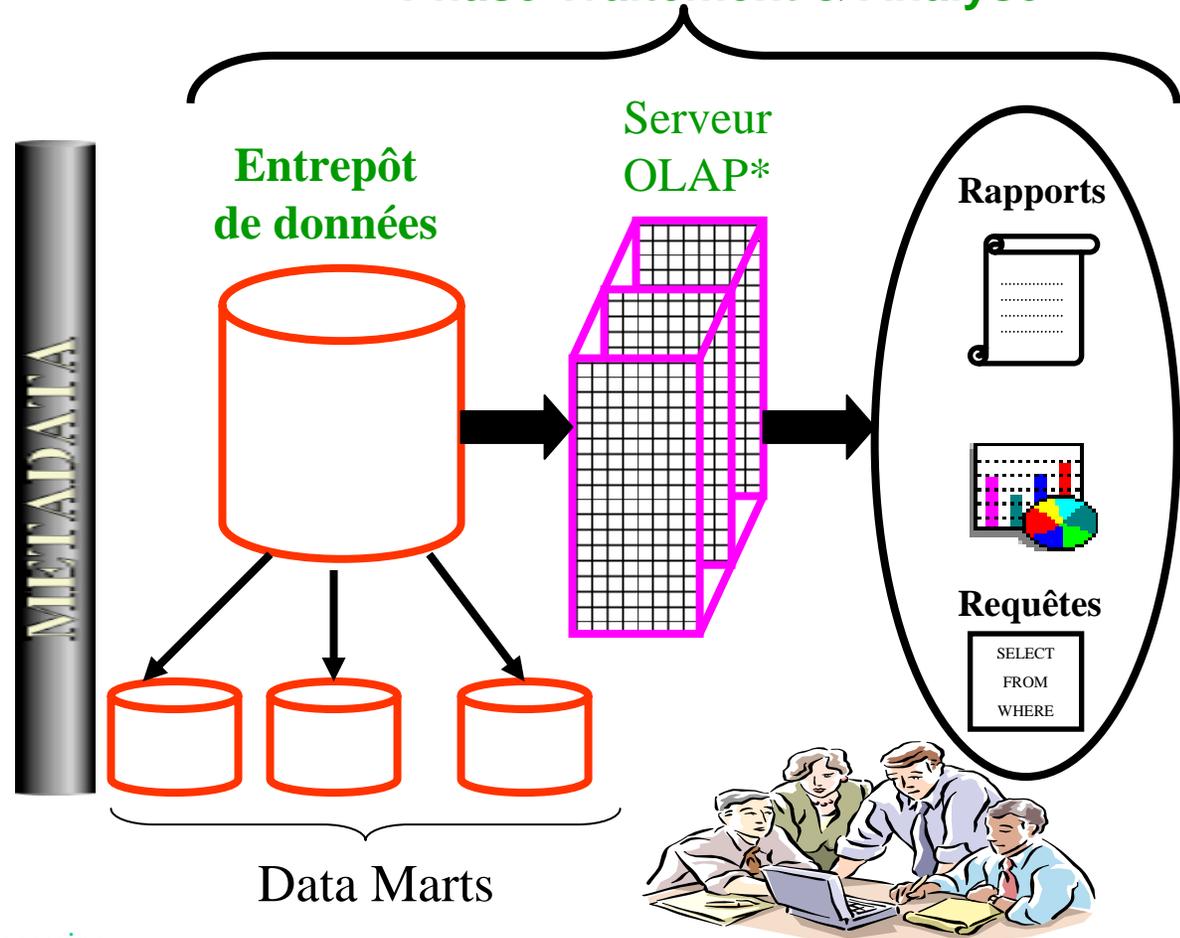
Besoins :

- ➔ Intégration données
- ➔ Gestion de l'évolution des données

Phase Intégration

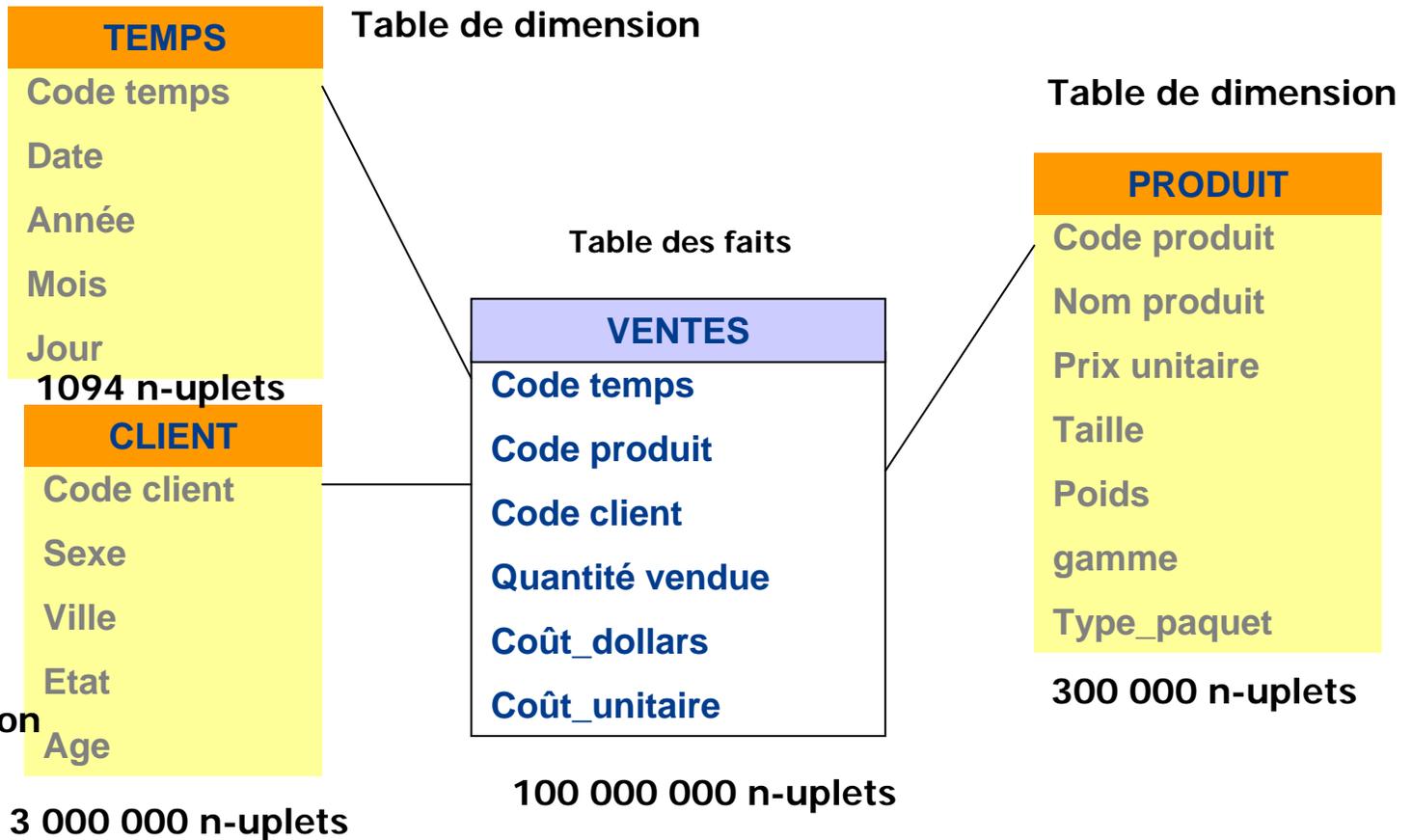


Phase Traitement & Analyse



* OLAP: On-Line Analytical Processing

schéma le plus utilisé pour modéliser les EDs



Requêtes de jointure en étoile

- Plusieurs de jointure
- Suivies par des sélection

- Volumétrie
 - des entrepôts de plusieurs téra octets
- Requêtes complexes
 - Opérations de sélection, jointure et agrégations
- Exigence des décideurs
 - Importance de la conception physique
- Evolution permanente
 - Instances, Requêtes, Schémas, Structures d'Optimisation
 - Importance du Tuning

- I. Conception physique**
- II. Techniques d'optimisation & Algorithmes de Sélection**
- III. Tuning**
- IV. Outil d'administration**
- V. Conclusion & Perspectives**

**Tables
(Conception logique)**

Profils des tables

**Profils des
requêtes**

**Conception
physique**

Techniques d'optimisation

Schéma de placement

Ressources utilisées

- 1. Techniques d'optimisation**
- 2. Contraintes de sélection**

❑ Vues matérialisées

- ❑ Une vue est une requête nommée
- ❑ Une vue matérialisée : les données résultant de sa requête sont stockées et maintenues

❑ Index avancés

- ❑ Structures permettant d'associer à une clé d'un n-uplet l'adresse relative de cet n-uplet

❑ Traitement parallèle

❑ Fragmentation

❑ Groupement

❑ Contraintes :

- ❑ Espace disque
- ❑ Coût de maintenance (rafraîchissement des données)
- ❑ Coût de calcul
- ❑ **Impossibilité de matérialiser toutes les vues**

❑ Problème de **S**élection des **V**ues (PSV):

- Sélectionner un ensemble de vues afin de maximiser ou minimiser une fonction objectif sous une contrainte
- Fonction « objectif »:

Optimiser le coût d'exécution des requêtes ou Optimiser le coût de maintenance

⇒ **Problème de Sac à Dos (NP-Complet)**

```
CREATE MATERIALIZED VIEW V1
AS
SELECT * FROM Vente V, Client C
WHERE V.noClient = C.noClient
```

❑ Contraintes :

- ❑ Espace disque, coût de maintenance, coût de calcul

❑ Index populaires dans les entrepôts:

- ❑ **Index simple**: sur une seule table ou une seule vue (arbre B, index binaire, index de projection)

- ❑ **Index de jointure**: sur plusieurs tables dans un schéma en étoile

❑ Problème de sélection d'index

❑ Algorithmes de sélection d'index

- ❑ Soit un attribut A , prenant n valeurs possibles $[v_1, \dots, v_n]$ (domaine)
- ❑ **Création d'un index bitmap sur l'attribut A :**
 - ❑ On crée n tableaux de bit, un pour chaque valeur v_i
 - ❑ Le tableau contient un bit pour chaque tuple t
 - ❑ Le bit d'un tuple t est à 1 si: $t.A = v_i$, à 0 sinon

Ouvrier(NSS, Nom, Salaire, Spécialité, N°U)

Domaine de l'attribut Spécialité est: {Soudeur, Fraiseur, Sableur, Tourneur}

SQL: CREATE INDEX BITMAP ON Ouvrier(Spécialité);

ROWID(RID)	Soudeur	Fraiseur	Sableur	Tourneur
00055 :000 :0023	0	1	0	0
00234 :020 :8922	1	0	0	0
19000 :328 :6200	0	0	0	1
21088 :120 :1002	0	0	1	0

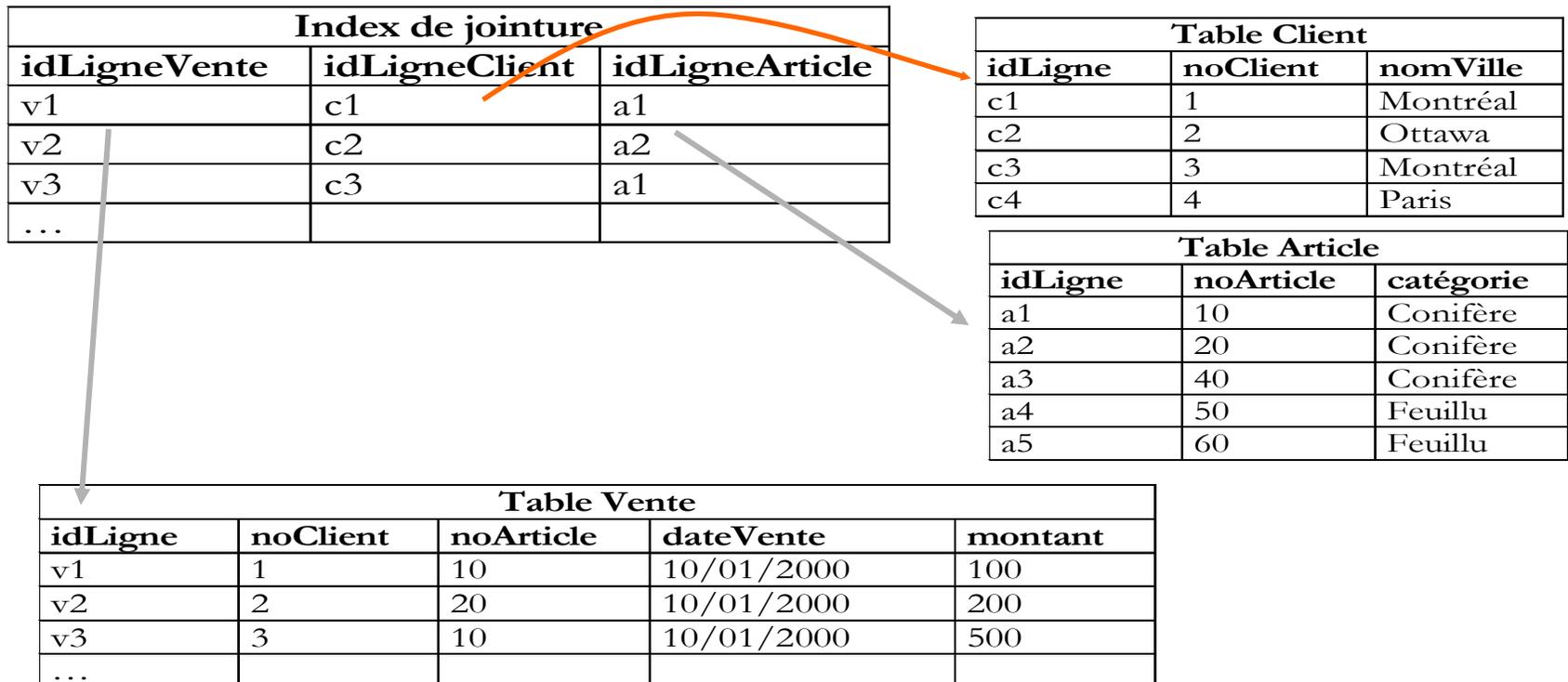
- ❑ Soit la requête suivante:

```
SELECT * FROM Ouvrier WHERE Spécialité = "Soudeur"
```

- ❑ On prend le vecteur de bit de la colonne Soudeur
- ❑ On garde toutes les cellules à 1
- ❑ On accède aux **enregistrements** par l'adresse (ROWID)
- ```
SELECT COUNT(*) FROM Ouvrier WHERE Spécialité = 'Soudeur'
```

  - On compte le nombre 1 dans la colonne Soudeur
  - On fait la somme et c'est terminé
- ❑ Très efficace si le nombre de valeurs (n) est petit
  - ❑ Exemple: Sexe

- ❑ Pré-calcule une opération de jointure
- ❑ Maintient des relations entre des clés étrangères (identifiants des tables de dimensions) et la clé primaire de la table des faits



❑ A bitmap join index is a bitmap index on one table (usually the fact table) that references columns in other tables (usually the dimension tables)

❑ Syntaxe:

```
CREATE BITMAP INDEX Ventes_Sexe
```

```
ON Ventes (Clients.Sexe)
```

```
FROM Ventes, Clients
```

```
WHERE Ventes.CID=Clients.CID;
```

**CLIENTS**

| RID <sup>c</sup> | CID | Nom     | Ville    |
|------------------|-----|---------|----------|
| 6                | 616 | Gilles  | Poitiers |
| 5                | 515 | Yves    | Paris    |
| 4                | 414 | Patrick | Nantes   |
| 3                | 313 | Didier  | Nantes   |
| 2                | 212 | Eric    | Poitiers |
| 1                | 111 | Pascal  | Poitiers |

**VENTES**

| RID <sup>s</sup> | CID | PID | TID | Montant |
|------------------|-----|-----|-----|---------|
| 1                | 616 | 106 | 11  | 25      |
| 2                | 616 | 106 | 66  | 28      |
| 3                | 616 | 104 | 33  | 50      |
| 4                | 515 | 104 | 11  | 10      |
| 5                | 414 | 105 | 66  | 14      |
| 6                | 212 | 106 | 55  | 14      |
| 7                | 111 | 101 | 44  | 20      |
| 8                | 111 | 101 | 33  | 27      |
| 9                | 212 | 101 | 11  | 100     |
| 10               | 313 | 102 | 11  | 200     |
| 11               | 414 | 102 | 11  | 102     |
| 12               | 414 | 102 | 55  | 103     |
| 13               | 515 | 102 | 66  | 100     |
| 14               | 515 | 103 | 55  | 17      |
| 15               | 212 | 103 | 44  | 45      |
| 16               | 111 | 105 | 66  | 44      |
| 17               | 212 | 104 | 66  | 40      |
| 18               | 515 | 104 | 22  | 20      |
| 19               | 616 | 104 | 22  | 20      |
| 20               | 616 | 104 | 55  | 20      |
| 21               | 212 | 105 | 11  | 10      |
| 22               | 212 | 105 | 44  | 10      |
| 23               | 212 | 105 | 55  | 18      |
| 24               | 212 | 106 | 11  | 18      |
| 25               | 313 | 105 | 66  | 19      |
| 26               | 313 | 105 | 22  | 17      |
| 27               | 313 | 106 | 11  | 15      |

| RID | Ville |    |   |
|-----|-------|----|---|
|     | P     | Pr | N |
| 1   | 1     | 0  | 0 |
| 2   | 1     | 0  | 0 |
| 3   | 1     | 0  | 0 |
| 4   | 0     | 1  | 0 |
| 5   | 0     | 0  | 1 |
| 6   | 1     | 0  | 0 |
| 7   | 1     | 0  | 0 |
| 8   | 1     | 0  | 0 |
| 9   | 1     | 0  | 0 |
| 10  | 0     | 0  | 1 |
| 11  | 0     | 0  | 1 |
| 12  | 0     | 0  | 1 |
| 13  | 0     | 1  | 0 |
| 14  | 0     | 1  | 0 |
| 15  | 1     | 0  | 0 |
| 16  | 1     | 0  | 0 |
| 17  | 1     | 0  | 0 |
| 18  | 0     | 1  | 0 |
| 19  | 1     | 0  | 0 |
| 20  | 1     | 0  | 0 |
| 21  | 1     | 0  | 0 |
| 22  | 1     | 0  | 0 |
| 23  | 1     | 0  | 0 |
| 24  | 1     | 0  | 0 |
| 25  | 0     | 0  | 1 |
| 26  | 0     | 0  | 1 |
| 27  | 0     | 0  | 1 |

```
SELECT count(*)
FROM Ventes V, Clients C
WHERE V.CID = C.CID
AND C.Ville = 'Poitiers'
```

```
CREATE BITMAP INDEX
Ventes_Pictaviens_bjix ON
Ventes(Client.Ville) FROM
Ventes, Client WHERE
Ventes.CID = Client.CID
```

Deux opérations :

1. Lire dans le bitmap la colonne P
2. Compter le nombre de 1.  
 → Pas de lecture dans la table Ventes

## □ Types

- Fragmentation horizontale : en fonction des tuples
  - Fragmentation horizontale dérivée : partitionner une table en fonction d'une autre fragmentée
  - Fragmentation verticale : en fonction des attributs
  - Fragmentation mixte
- Fragmentation horizontale est bien adaptée aux ED :
- Partitioning option [Oracle 9i]

- **Fragmentation horizontale primaire (FHP)**

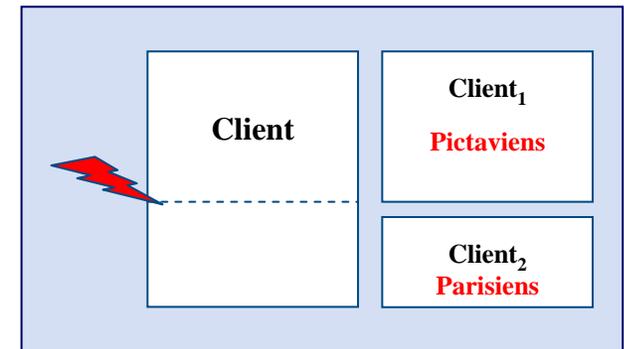
- Fragmenter une table en utilisant les prédicats de sélection définis sur cette table

**Prédicat** : Attribut  $\theta$  Valeur,  $\theta \in \{=, <, \leq, >, \geq\}$  et valeur  $\in \text{Domaine}(\text{Attribut})$ .

- Exemple: Client (Client\_id, Nom, Ville)

- Client<sub>1</sub> :  $\sigma_{\text{Ville}='Poitiers'}(\text{Client})$

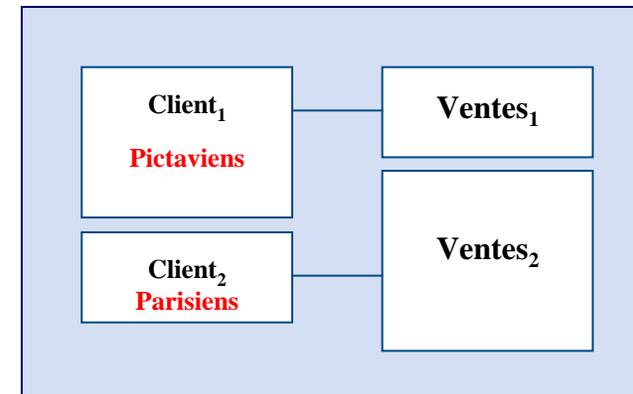
- Client<sub>2</sub> :  $\sigma_{\text{Ville}='Paris'}(\text{Client})$ ,



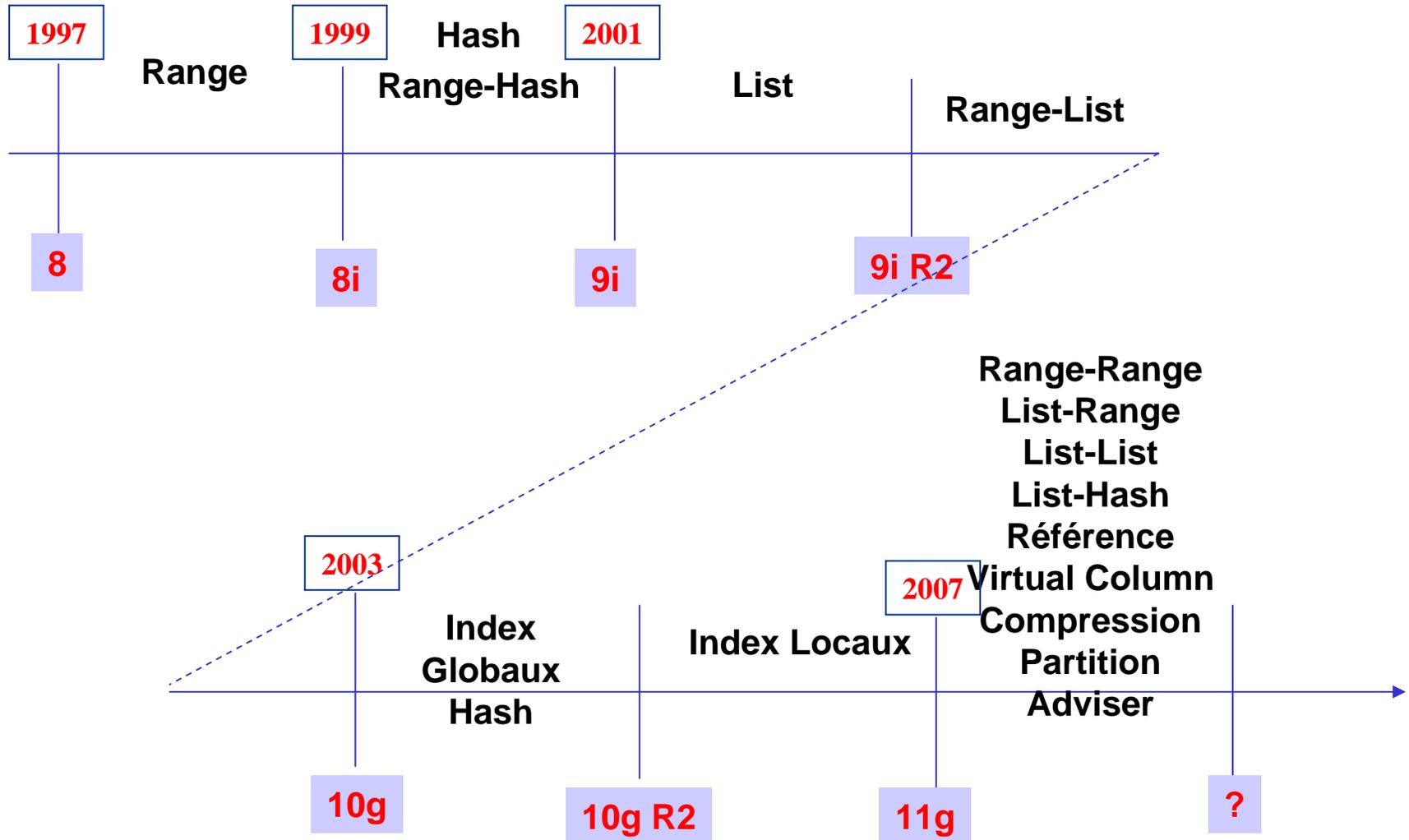
- **FHP ~ requêtes de jointure en étoile**

- Optimisation des sélections définies sur les tables de dimension

- **Fragmentation horizontale dérivée (FHD)**
  - Fragmenter une table (S) selon des attributs d'une autre table (T) :  
(existence de lien entre S et T)
  - Ventes(Client\_id, Produit\_id, Date, Quantité)
    - Ventes<sub>1</sub>=Ventes SEMI JOIN Client<sub>1</sub>
    - Ventes<sub>2</sub>=Ventes SEMI JOIN Client<sub>2</sub>



- **Impact de la FHD sur RJE**
  - **Optimisation de la jointure entre S et T**



## A. Un seul attribut

– Trois modes: Range, Hash, List

```

CREATE TABLE Client_Age_Sexe
(
 Client_id NUMBER(5),
 Nom Varchar2(20),
 Sexe Varchar2(2),
 Age Number(4),
 Département NUMBER(2))
PARTITION BY RANGE (Age)
(
 SUBPARTITION BY LIST (Sexe)
 SUBPARTITION TEMPLATE(
 SUBPARTITION Client1 VALUES ('M') TABLESPACE MasculinTBS ,
 SUBPARTITION Client2 VALUES ('F') TABLESPACE FémininTBS)
);
(
PARTITION Client_Moins_27 VALUES LESS THAN (27),
PARTITION Client_27_60 VALUES LESS THAN (61),
PARTITION Client_Plus_60 VALUES LESS THAN (MAXVALUES)
);

```

## ❑ Entrées :

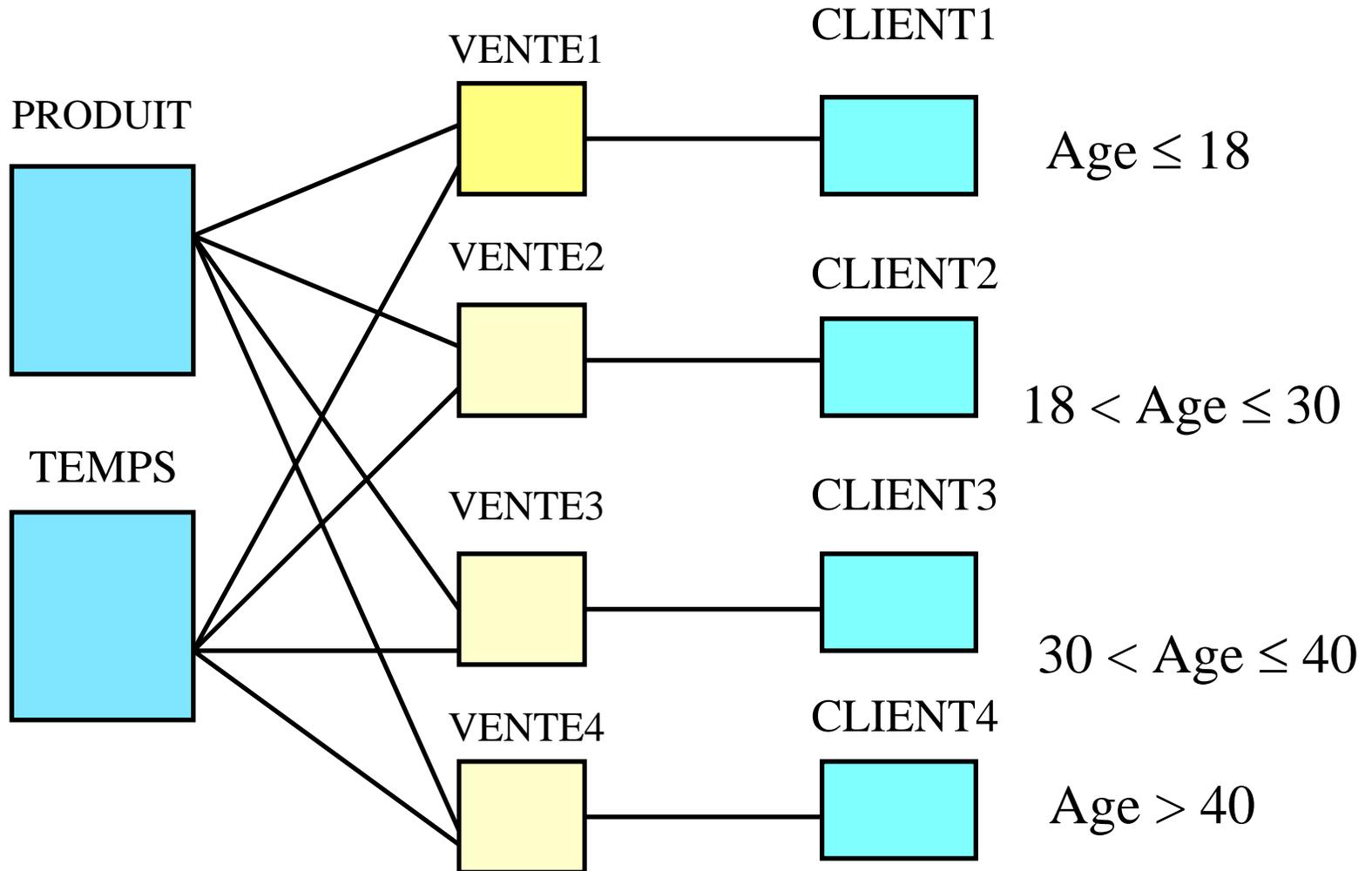
- ❑ Un schéma en étoile  $S (F, D_1, D_2, \dots, D_n)$
- ❑ Un ensemble de requêtes fréquentes  $Q$

## ❑ Sortie :

- ❑ Ensemble de sous-schémas en étoile

## ❑ Méthode :

- ❑ Décomposer une (des) table(s) de dimensions en utilisant la fragmentation horizontale
- ❑ Décomposer la table des faits en fonction des tables de dimension fragmentées



# Nombre de fragments générés

## Exemple :

Client(CID, Nom, Sexe, Age, Département).

Temps(TID, Jour, Semaine, Mois, Année).

Fragmenter Temps sur Semaine (les 52 semaines de l'année).

Fragmenter Client sur Sexe (F et M = 2 fragments) ,

Age (Moins de 26 ans, 27-40, 41-60 et +60 ans = 4 fragments) et

département (les 100 départements de France)

**Conséquence** : Nombre de fragments de la table des faits

**52x2x4x100 = 41 600 fragments**

$$N = \prod_{i=1}^k M_i$$

**M<sub>i</sub>** : nombre de fragments de la table de dimension D<sub>i</sub>  
**k** : nombre de tables de dimension fragmentées

## Entrées

- Entrepôt de données composé de
  - Un ensemble  $D$  de tables de dimension  $D = \{D_1, \dots, D_d\}$
  - Une table des faits  $F$
- Charge de requêtes les plus fréquentes  $Q = \{Q_1, \dots, Q_m\}$
- $W$  : seuil (fixé par l'administrateur)

## Sorties :

- Ensemble  $D' \subseteq D$  des tables de dimension fragmentées
- Ensemble de  $N$  fragments de faits  $F_1, F_2, \dots, F_N$

## Objectifs :

- Réduire le temps d'exécution de  $Q$
- $N \leq W$

- Problème de fragmentation horizontale à un seul domaine (PFHSD)
  - Une seule table de dimension
  - Un seul attribut A
- Réduction à partir du problème 3-Partition
  - 3-Partition NP-Complet
  - PFHSD NP-Complet
- Notre problème de fragmentation est plus compliqué
  - Plusieurs tables de dimension
  - Plusieurs attributs par table de dimension

- Décomposition des domaines des attributs de fragmentation en sous domaines
- Codage d'un schéma de fragmentation

## Exemple

Trois attributs de dimension:  
Age, Genre, Saison

|        |        |         |       |           |     |
|--------|--------|---------|-------|-----------|-----|
|        | Age<18 | 18-30   | 30-45 | 45-60     | >60 |
| Age    | 0      | 1       | 0     | 1         | 2   |
|        | M      | F       |       |           |     |
| Genre  | 0      | 0       |       |           |     |
|        | été    | automne | hiver | printemps |     |
| Saison | 0      | 0       | 0     | 1         |     |

## Schémas de fragmentation

### ❑ Client en 3 fragments sur Age (Genre n'est pas utilisé)

- Client1 : Age <18 **OU**  $30 \leq \text{Age} < 45$
- Client2 :  $18 \leq \text{Age} < 30$  **OU**  $45 \leq \text{Age} < 60$
- Client3 : Age  $\geq 60$

### ❑ Temps en 2 fragments sur Saison

- Temps1 : Saison=«été» **OU** «automne» **OU** «hiver»
- Temps2 : Saison= «printemps»



Table Ventes fragmentée  
en : **3x2=6 fragments**

Partition P0

Partition P1

# Algorithme de Sélection: Hill Climbing

## • Principe

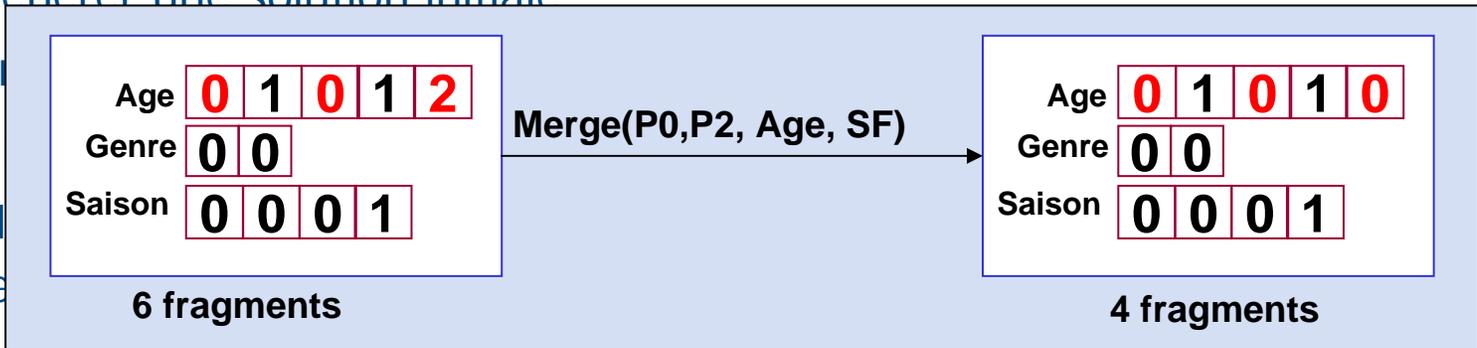
1. Générer une solution initiale

2. Améliorer la solution

• Mesure de la qualité

– Métrique

re



## • Solution initiale

– Algorithme

## • Amélioration

– Merge

- Fusionner deux partitions en une seule.

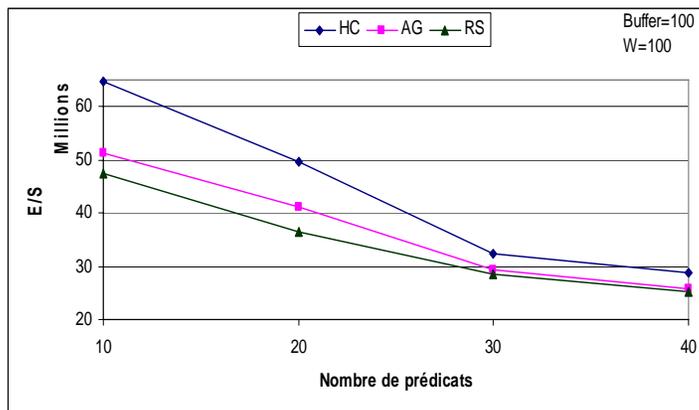
– Split

- Eclater une partition de sous domaines en deux partitions

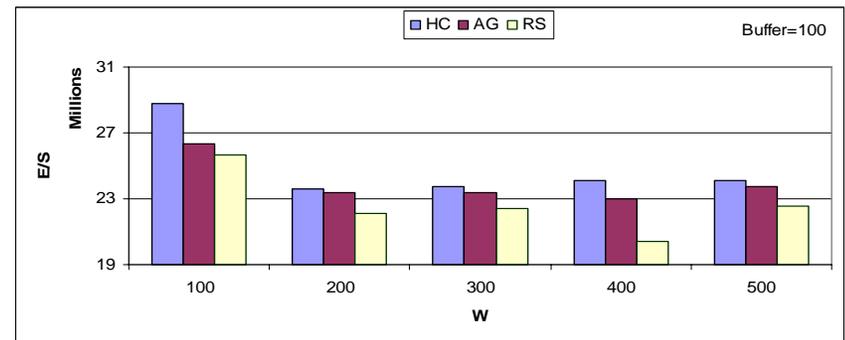


1. Théorique : Modèle de coût
2. Validation sous Oracle

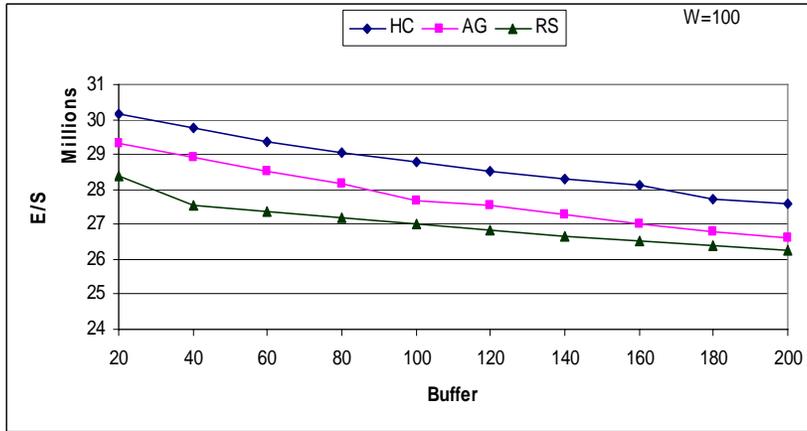
| DataSet :                                                                                                          | Requêtes :                                                                                                         | Algorithmes                                                                                                                |
|--------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| Benchmark APB-1                                                                                                    | 60 requêtes                                                                                                        |                                                                                                                            |
| <ul style="list-style-type: none"> <li>• 1 Table de faits (24786 000 )</li> <li>• 4 tables de dimension</li> </ul> | <ul style="list-style-type: none"> <li>• 40 prédicats de sélection</li> <li>• 12 attributs de sélection</li> </ul> | <ul style="list-style-type: none"> <li>• Hill Climbing</li> <li>• Algorithme génétique</li> <li>• Recuit Simulé</li> </ul> |



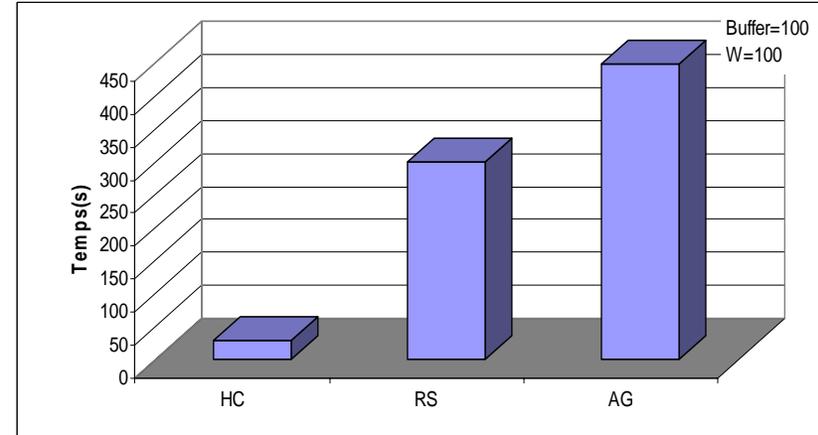
Performance vs Nombre de prédicats



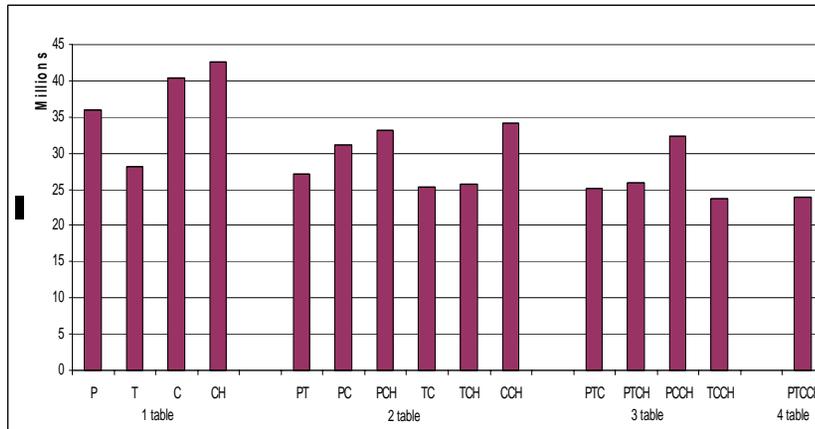
Effet du seuil W



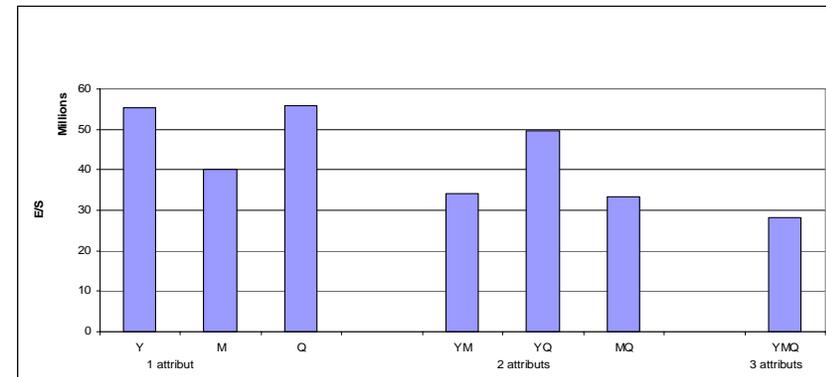
**Effet du buffer sur la performance**



**Temps d'exécution**



**Choix des tables de dimension**



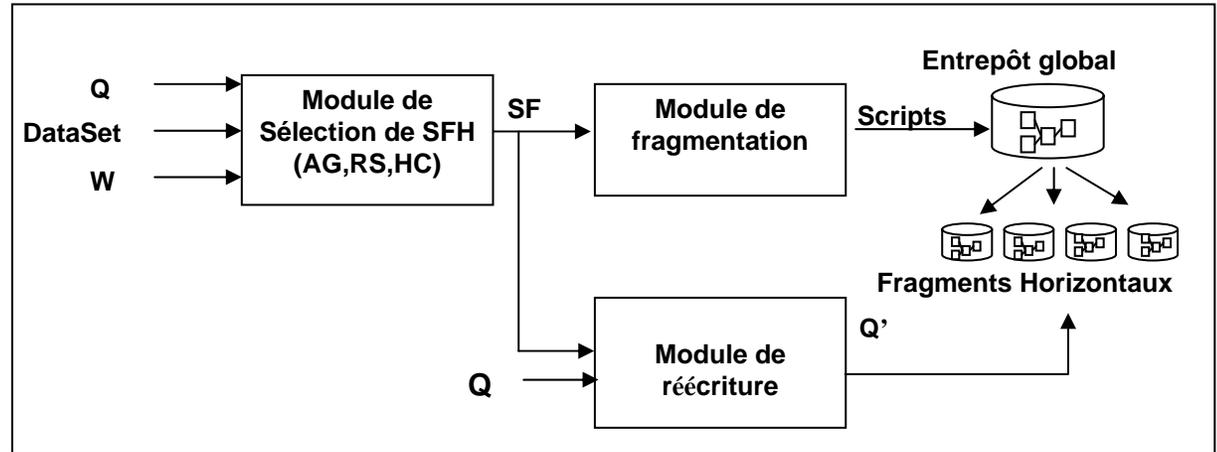
**Choix des attributs de dimension**

- Problèmes rencontrés
  - FHP sur  $n$  ( $n > 2$ ) attributs n'est pas supportée.
  - FHD sur ( $m > 1$ ) tables de dimension n'est pas supportée.
- Solutions
  1. Méthode d'implémentation de la FHP (ajout d'une nouvelle propriété)
  2. Méthode d'implémentation de la FHD (vues matérialisée)

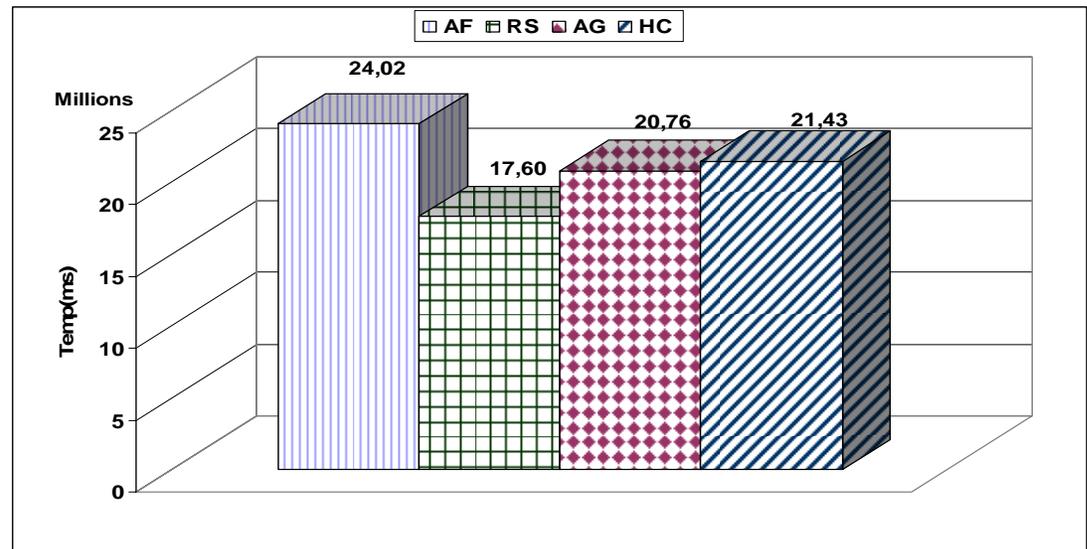
⇒ Nécessité de réécrire les requêtes sur les schémas fragmentés

  - Identifier les fragments valides pour chaque requête
  - Réécrire la requête sur ces fragments

- Architecture



- Résultats



```

SELECT count(*)
FROM Ventes V, Client C, Produit P, Temps T
WHERE V.CID = C.CID AND V.TID=T.TID AND
V.PID=P.PID
AND C.Ville = 'Poitiers'
AND T.Mois='Juin'
AND P.Catégorie='Beauté'

```

IJB

| RI<br>D | Ville |    |   |    |    | Mois |   |   |    |   | Catégorie |   |    |   |   | AND |
|---------|-------|----|---|----|----|------|---|---|----|---|-----------|---|----|---|---|-----|
|         | P     | Pr | N | Ja | Fe | M    | A | M | Ju | B | M         | J | Jr | F |   |     |
| 1       | 1     | 0  | 0 | 1  | 0  | 0    | 0 | 0 | 0  | 1 | 0         | 0 | 0  | 0 | 0 |     |
| 2       | 1     | 0  | 0 | 0  | 0  | 0    | 0 | 0 | 1  | 1 | 0         | 0 | 0  | 0 | 1 |     |
| 3       | 1     | 0  | 0 | 0  | 0  | 1    | 0 | 0 | 0  | 0 | 1         | 0 | 0  | 0 | 0 |     |
| 4       | 0     | 1  | 0 | 1  | 0  | 0    | 0 | 0 | 0  | 0 | 1         | 0 | 0  | 0 | 0 |     |
| 5       | 0     | 0  | 1 | 0  | 0  | 0    | 0 | 0 | 1  | 1 | 0         | 0 | 0  | 0 | 0 |     |
| 6       | 1     | 0  | 0 | 0  | 0  | 0    | 0 | 1 | 0  | 1 | 0         | 0 | 0  | 0 | 0 |     |
| 7       | 1     | 0  | 0 | 0  | 0  | 0    | 1 | 0 | 0  | 0 | 0         | 0 | 0  | 1 | 0 |     |
| 8       | 1     | 0  | 0 | 0  | 0  | 1    | 0 | 0 | 0  | 0 | 0         | 0 | 0  | 1 | 0 |     |
| 9       | 1     | 0  | 0 | 1  | 0  | 0    | 0 | 0 | 0  | 0 | 0         | 0 | 0  | 1 | 0 |     |
| 10      | 0     | 0  | 1 | 1  | 0  | 0    | 0 | 0 | 0  | 0 | 0         | 0 | 1  | 0 | 0 |     |
| 11      | 0     | 0  | 1 | 1  | 0  | 0    | 0 | 0 | 0  | 0 | 0         | 0 | 1  | 0 | 0 |     |
| 12      | 0     | 0  | 1 | 0  | 0  | 0    | 0 | 1 | 0  | 0 | 0         | 0 | 1  | 0 | 0 |     |
| 13      | 0     | 1  | 0 | 0  | 0  | 0    | 0 | 1 | 0  | 0 | 0         | 1 | 0  | 0 | 0 |     |
| 14      | 0     | 1  | 0 | 0  | 0  | 0    | 0 | 1 | 0  | 0 | 0         | 1 | 0  | 0 | 0 |     |
| 15      | 1     | 0  | 0 | 0  | 0  | 0    | 1 | 0 | 0  | 0 | 0         | 1 | 0  | 0 | 0 |     |
| 16      | 1     | 0  | 0 | 0  | 0  | 0    | 0 | 0 | 1  | 1 | 0         | 0 | 0  | 0 | 1 |     |
| 17      | 1     | 0  | 0 | 0  | 0  | 0    | 0 | 0 | 1  | 0 | 1         | 0 | 0  | 0 | 0 |     |
| 18      | 0     | 1  | 0 | 0  | 1  | 0    | 0 | 0 | 0  | 0 | 1         | 0 | 0  | 0 | 0 |     |
| 19      | 1     | 0  | 0 | 0  | 1  | 0    | 0 | 0 | 0  | 0 | 1         | 0 | 0  | 0 | 0 |     |
| 20      | 1     | 0  | 0 | 0  | 0  | 0    | 0 | 1 | 0  | 0 | 1         | 0 | 0  | 0 | 0 |     |
| 21      | 1     | 0  | 0 | 1  | 0  | 0    | 0 | 0 | 0  | 1 | 0         | 0 | 0  | 0 | 0 |     |
| 22      | 1     | 0  | 0 | 0  | 0  | 0    | 1 | 0 | 0  | 1 | 0         | 0 | 0  | 0 | 0 |     |
| 23      | 1     | 0  | 0 | 0  | 0  | 0    | 0 | 1 | 0  | 1 | 0         | 0 | 0  | 0 | 0 |     |
| 24      | 1     | 0  | 0 | 1  | 0  | 0    | 0 | 0 | 0  | 1 | 0         | 0 | 0  | 0 | 0 |     |
| 25      | 0     | 0  | 1 | 0  | 0  | 0    | 0 | 0 | 1  | 1 | 0         | 0 | 0  | 0 | 0 |     |
| 26      | 0     | 0  | 1 | 0  | 1  | 0    | 0 | 0 | 0  | 1 | 0         | 0 | 0  | 0 | 0 |     |
| 27      | 0     | 0  | 1 | 1  | 0  | 0    | 0 | 0 | 0  | 1 | 0         | 0 | 0  | 0 | 0 |     |

Client<sub>1</sub>=σ<sub>ville=Poitiers</sub>(Client)

| RID <sup>C</sup> | CID | Nom    | Ville    |
|------------------|-----|--------|----------|
| 6                | 616 | Gilles | Poitiers |
| 2                | 212 | Eric   | Poitiers |
| 1                | 111 | Pascal | Poitiers |

Temps<sub>6</sub>=σ<sub>mois=juin</sub>(Temps)

| RID <sup>T</sup> | TID | Mois | Année |
|------------------|-----|------|-------|
| 1                | 66  | Juin | 2003  |

Produit<sub>1</sub>=σ<sub>cat=Beauté</sub>(Produit)

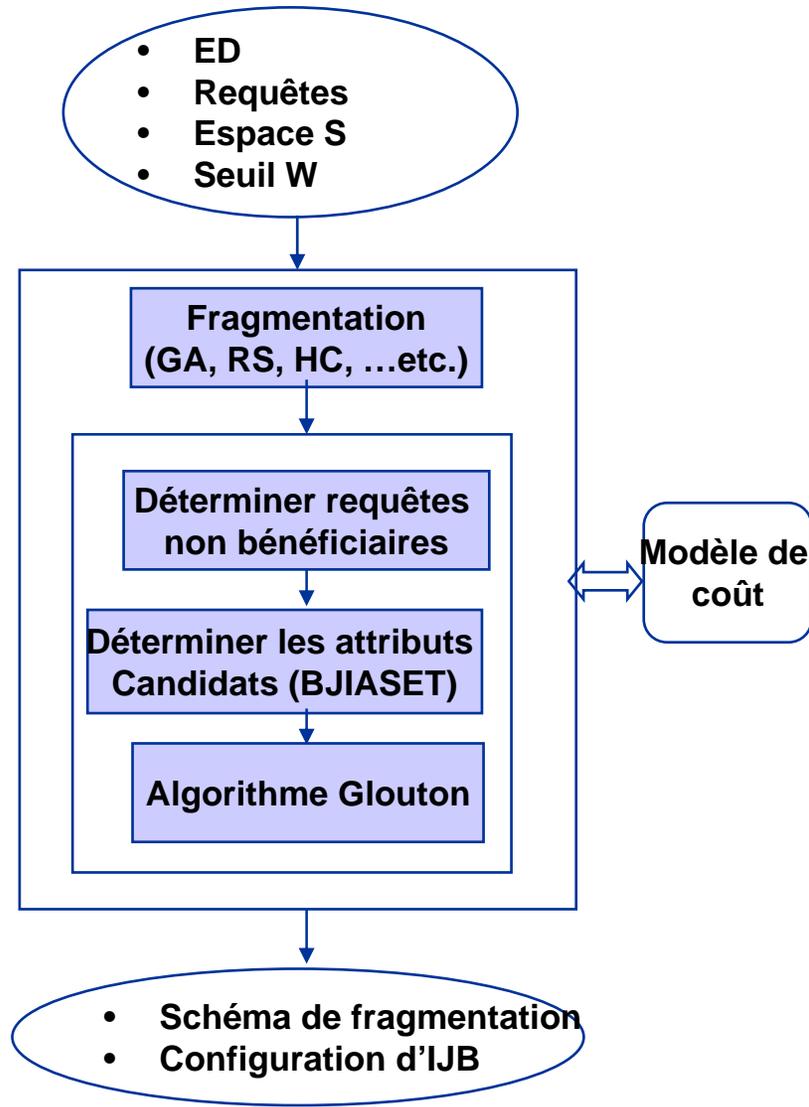
| RID <sup>P</sup> | PID | Nom       | Catégorie |
|------------------|-----|-----------|-----------|
| 6                | 106 | Sonoflore | Beauté    |
| 5                | 105 | Clarins   | Beauté    |

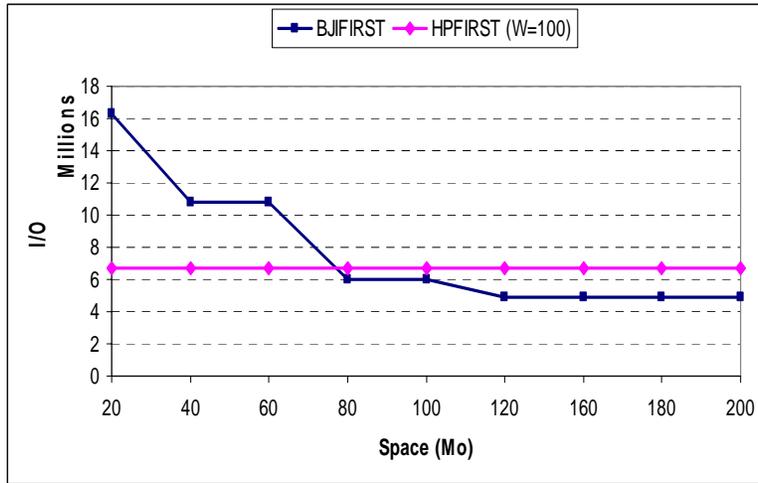
| Ventes_BJP       |     |     |     |        |
|------------------|-----|-----|-----|--------|
| RID <sup>S</sup> | CID | PID | TID | Amount |
| 2                | 616 | 106 | 66  | 28     |
| 16               | 111 | 105 | 66  | 44     |

1. Sélectionner un schéma de fragmentation
2. Déterminer les requêtes non bénéficiaires
  - Utiliser le taux  $\lambda$  fixé par ADMIN
  - Si  $\text{Coût}(Q, FS) / \text{Coût}(Q, \Phi) \leq \lambda$  alors Q est bénéficiaire

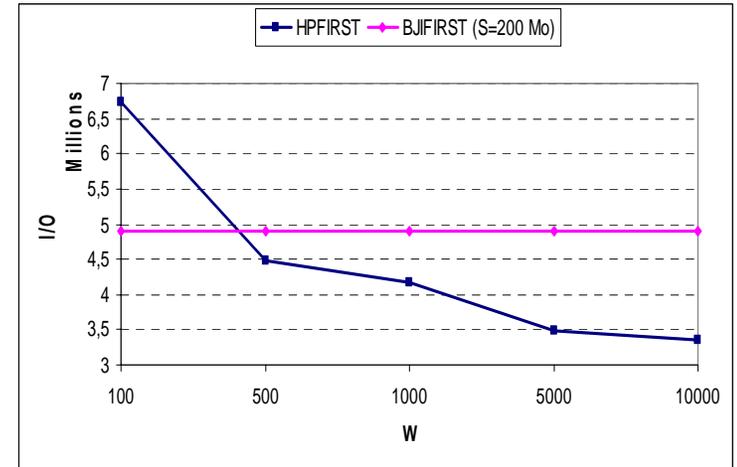
$\text{Coût}(Q, \phi)$  ,  $\text{Coût}(Q, FS)$  : coût avant et après fragmentation

3. Déterminer les attributs candidats (BJIASET)
  - Attributs candidats = Attributs de faibles cardinalités dans les requêtes non bénéficiaires et non utilisés pour fragmenter l'entrepôt
4. Sélectionner un ensemble de BJIs

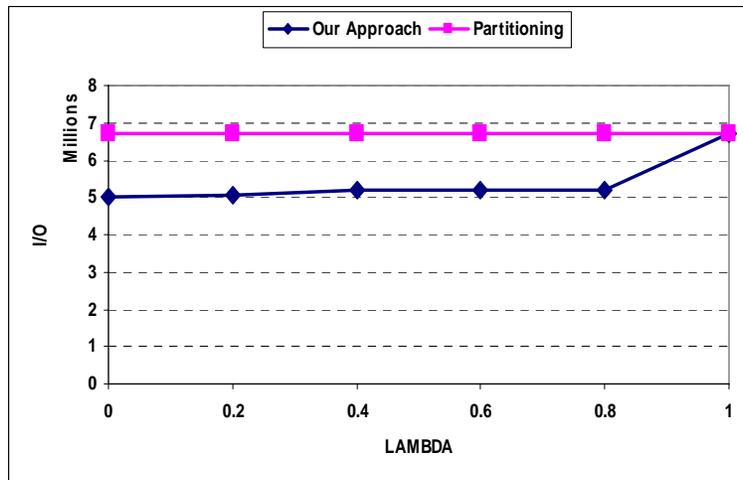




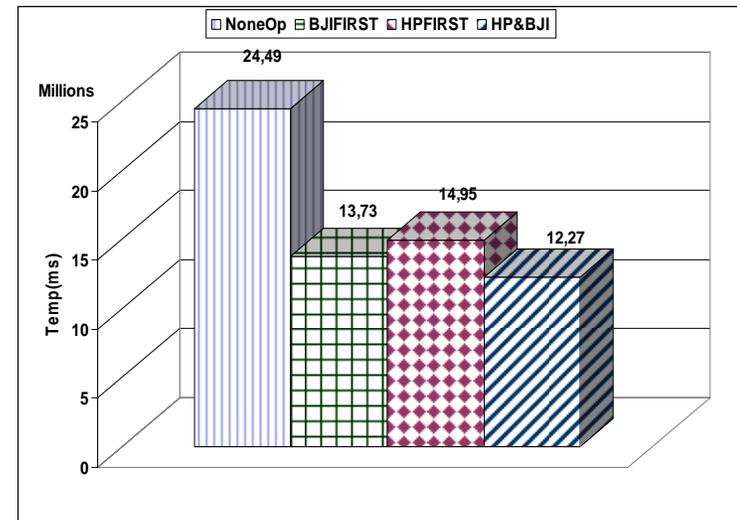
Sélection séparée (S)



Sélection séparée (W)



Paramètre de tuning  $\lambda$



Validation Oracle 10g

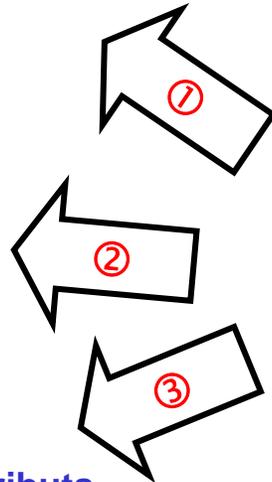
## Sélection des Techniques d'optimisation

- Vues
- Index
- Fragmentation Horizontale
- Clustering

## Mode de sélection

- Isolé
- Combinée

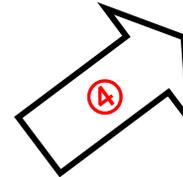
## Choix des tables/attributs



Administrateur

## Algorithmes de sélection

- Moins de paramétrage
  - Gloutons
- Plus de paramétrage
  - Algorithmes génétiques
  - Recuit simulé
  - ....



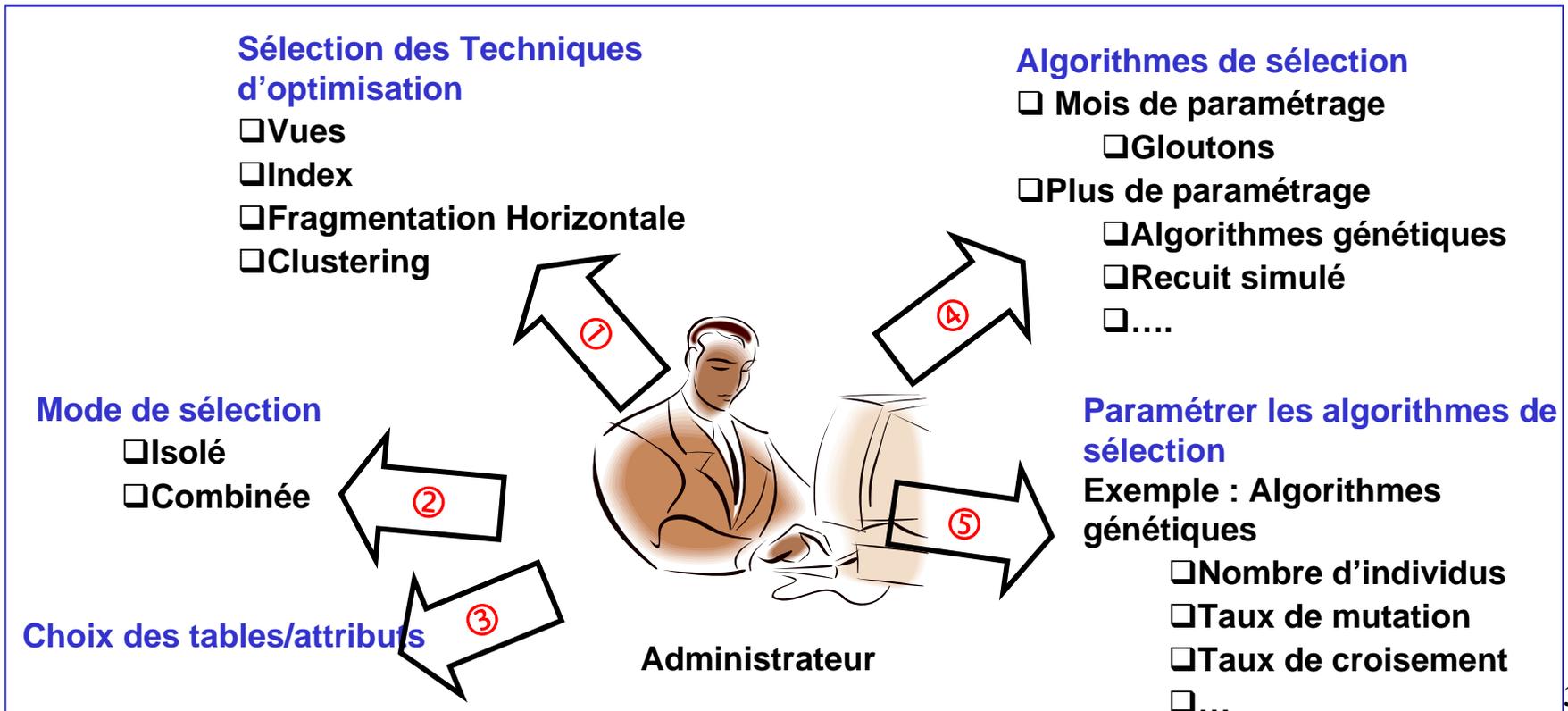
## Paramétrer les algorithmes de sélection

Exemple : Algorithmes génétiques

- Nombre d'individus
- Taux de mutation
- Taux de croisement
- ...

**Nécessité d'une assistance lors de la conception physique**

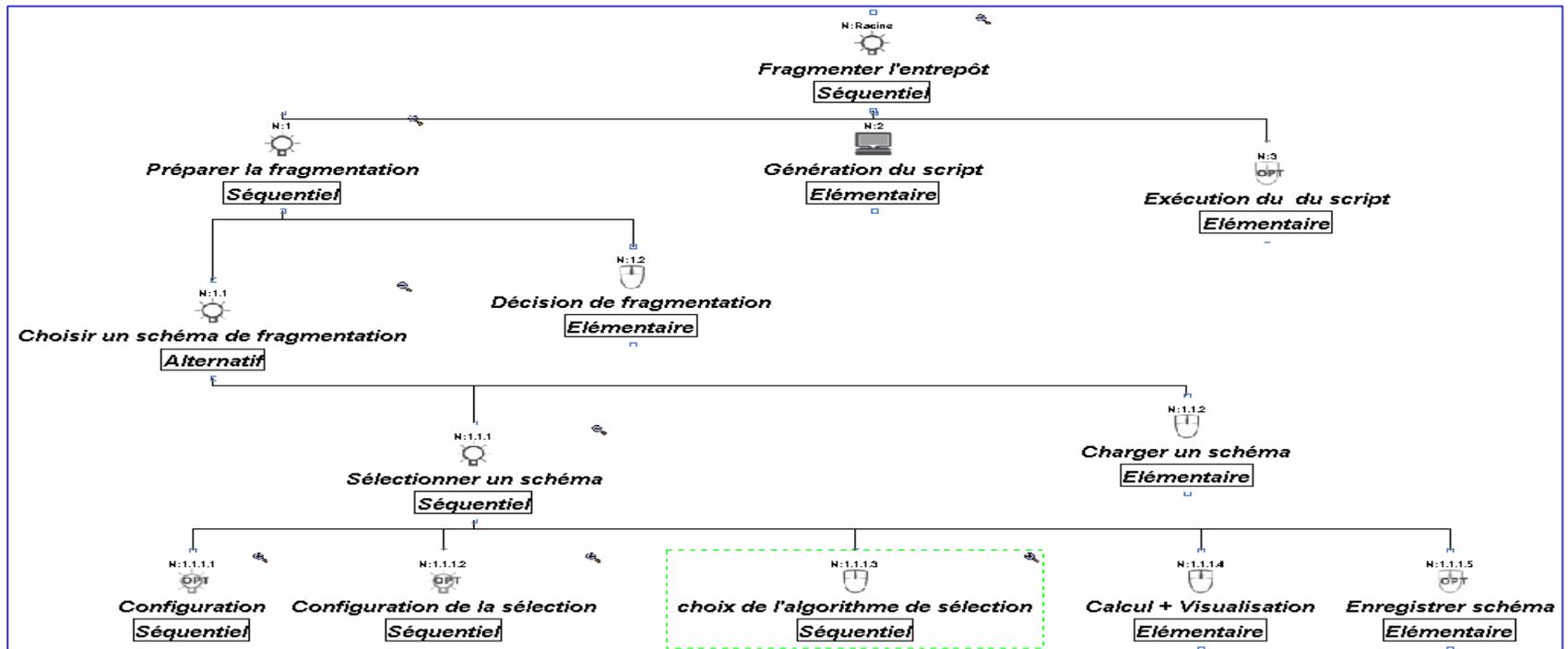
- Etude des Besoins
- Techniques d'optimisation
- Conception
- Validation

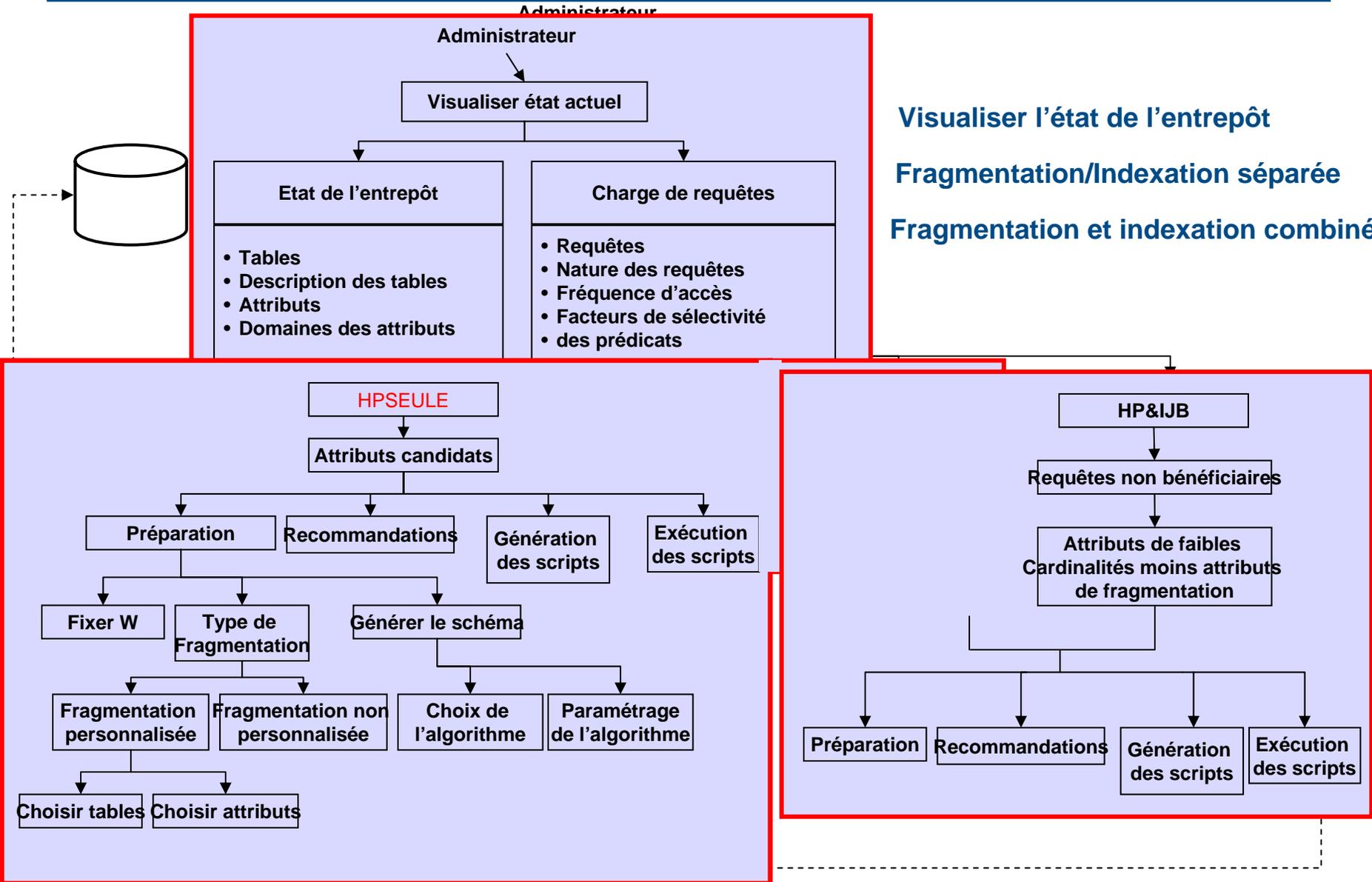


- Fragmentation horizontale primaire (FHP)
  - Appliquée sur les tables de dimension
- Fragmentation horizontale dérivée (FHD)
  - Appliquée sur la table des faits
- Index de jointure binaires (IJBs)
- Sélection isolée
  - FHSEULE : sélection d'un schéma de fragmentation.
  - BJISEULS : sélection d'une configuration d'index de jointure
- Sélection combinée (FH&IJBs)

- Fragmentation
  - Hill climbing
  - Recuit Simulé (RS) [BNCOD'06]
    - Température, décroissance de la température, équilibre, gel
  - Algorithme génétique [DAWAK'05]
    - Taux de mutation, taux de croisement, nombre d'individus, temps d'exécution
- Indexation
  - Glouton [DEXA'07]
  - Approche data mining [DAWAK'07]

- Modélisation par K-MADE [INRIA]
  - Exprimer les activités utilisateurs
  - Valider l'interface par un simulateur de scénarii
  - Post et pré conditions
- Exemple d'un modèle de tâches





Visualiser l'état de l'entrepôt  
 Fragmentation/Indexation séparée  
 Fragmentation et indexation combiné

- Charge de requêtes
  - Description de chaque requête
  - Fréquence d'utilisation.
- Tables de l'entrepôt
  - Attributs, nombre de tuples,... etc.
- Attributs de sélection
  - Sous domaines
  - Taille.

ETAT DE L'ENTREPOT

Workload Tables **Attributs** Configuration

| TABLES    | CARDINALITES | ATTRIBUTS      |               |               |
|-----------|--------------|----------------|---------------|---------------|
| Actvars   | 24786000     | CUSTOMER_LEVEL | PRODUCT_LEVEL | CHANNEL_LEVEL |
| prodlevel | 9000         | Class_level    | Group_level   | Family_level  |
| timelevel | 24           | Year_level     | Month_level   | Quarter_level |
| custlevel | 900          | Retailer_level | Gender_level  | City_level    |

Afficher les tables

- Fragmentation non personnalisée
  - Faire confiance à l'outil
- Fragmentation personnalisée
  - Participation de l'administrateur aux différents choix

**FRAGMENTATION**

Administration | Algorithmes | Résultats | **Recommandadions** | Générations des Scripts

|                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                         |                                                                                                                                                                                                                                     |                                                                                                     |                                                                         |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|
| <input checked="" type="checkbox"/> <b>Prodlevel</b><br><b>Attributs</b><br><input checked="" type="checkbox"/> Class_level<br><input type="checkbox"/> Group_level<br><input type="checkbox"/> Family_level<br><input checked="" type="checkbox"/> Line_level<br><input checked="" type="checkbox"/> Division_level | <input checked="" type="checkbox"/> <b>Timelevel</b><br><b>Attributs</b><br><input checked="" type="checkbox"/> Year_level<br><input type="checkbox"/> Month_level<br><input checked="" type="checkbox"/> Quarter_level | <input checked="" type="checkbox"/> <b>Custlvel</b><br><b>Attributs</b><br><input checked="" type="checkbox"/> Retailer_level<br><input checked="" type="checkbox"/> Gender_level<br><input checked="" type="checkbox"/> City_level | <input type="checkbox"/> <b>Chanlevel</b><br><b>Attributs</b><br><input type="checkbox"/> All_level | <b>Seuil W</b><br><input style="width: 50px;" type="text" value="100"/> |
| <b>Personnalisation</b><br><input type="radio"/> Non Personnalisée <input checked="" type="radio"/> Personnalisée                                                                                                                                                                                                    |                                                                                                                                                                                                                         |                                                                                                                                                                                                                                     |                                                                                                     |                                                                         |

## Sélection combinée

### 1. Sélectionner un schéma de fragmentation

Schéma de fragmentation final

| Coût par requête | Attributs de Fragmentation                                      | Tables Fragmentées |
|------------------|-----------------------------------------------------------------|--------------------|
| Class_level      | {{ P00HV1RICH5W},{ CI493YZ9KZUJ},{ FDXAQ1N5U026},{ ELSE}}       | ---> 4 Partitions  |
| Group_level      | { E4NJTW0ZR9FN, ELSE}                                           | ---> 1 Partitions  |
| Family_level     | { BEMFVK0N8125, UJHZ4TZMJT6V, SHDF8QT29KFF, AGG214DG271Q, ELSE} | ---> 1 Partitions  |
| Line_level       | { MJ1F1U1EG009, ELSE}                                           | ---> 1 Partitions  |
| Division_level   | { BCR2T4K2K9D3, XRLXY6H61SLC, RC5406URP1IE, G4HA5YITG3H7}       | ---> 1 Partitions  |

### 2. Sélectionner une configuration d'index.

INDEXATION

Select IJBs Administration Algorithmes Résultats **Recommandations** Générations des Scripts

Coût par requête Résultat BJI

|                     |                |                              |    |                                      |          |
|---------------------|----------------|------------------------------|----|--------------------------------------|----------|
| Attributs candidats |                | Nombre de requêtes de départ | 60 | Nombre de requêtes non bénéficiaires | 39       |
| Index créés         |                |                              |    | Espace disponible                    | 100 Mo   |
| Group_level         | Division_level |                              |    | Espace Occupé                        | 81 Mo    |
| Family_level        | Month_level    |                              |    | E/S Avant Fragmentation              | 62160727 |
| Line_level          | Quarter_level  |                              |    | E/S après fragmentation              | 36594992 |
| Division_level      | Gender_level   |                              |    | E/S après indexation                 | 8048871  |
| Month_level         | All_level      |                              |    |                                      |          |
| Quarter_level       |                |                              |    |                                      |          |
| Gender_level        |                |                              |    |                                      |          |
| All_level           |                |                              |    |                                      |          |

- ParAdmin fournit un ensemble de recommandations
  - Schéma de fragmentation et/ou configuration d'index
  - Temps d'exécution avant et après la sélection
  - Attributs de fragmentations
  - Tables fragmentées
- ADMIN peut valider et générer les scripts.
- Il pourra revenir en arrière pour reconsidérer ses choix.

| Schéma de fragmentation final | Coût par requête | Attributs de Fragmentation                                                                                                                                                                                                     | Tables Fragmentées                                                                                                                                                                                                                                                                                                                                                                                               |
|-------------------------------|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                               |                  | <b>Attributs de Fragmentation</b><br>Class_level : 4 Partitions sur 4 SD stables<br>Year_level : 2 Partitions sur 2 SD stables<br>Retailer_level : 4 Partitions sur 4 SD stables<br>City_level : 3 Partitions sur 4 SD stables | <b>Les autres Attributs</b><br>Group_level : 1 PARTITION SUR 2 SD stables<br>Family_level : 1 PARTITION SUR 5 SD stables<br>Line_level : 1 PARTITION SUR 2 SD stables<br>Division_level : 1 PARTITION SUR 4 SD stables<br>Month_level : 1 PARTITION SUR 12 SD stables<br>Quarter_level : 1 PARTITION SUR 4 SD stables<br>Gender_level : 1 PARTITION SUR 2 SD stables<br>All_level : 1 PARTITION SUR 5 SD stables |

| Schéma de fragmentation final | Coût par requête | Attributs de                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |         |        |       |       |      |    |   |        |     |        |    |   |        |      |        |    |    |        |     |        |    |    |        |        |       |    |    |         |         |       |    |   |        |        |      |    |   |        |        |      |    |    |         |         |       |    |   |        |        |      |     |    |         |         |       |
|-------------------------------|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|--------|-------|-------|------|----|---|--------|-----|--------|----|---|--------|------|--------|----|----|--------|-----|--------|----|----|--------|--------|-------|----|----|---------|---------|-------|----|---|--------|--------|------|----|---|--------|--------|------|----|----|---------|---------|-------|----|---|--------|--------|------|-----|----|---------|---------|-------|
|                               |                  | <table border="1"> <thead> <tr> <th>Req</th> <th>Freq</th> <th>Avant</th> <th>Après</th> <th>Gain</th> </tr> </thead> <tbody> <tr><td>Q1</td><td>3</td><td>251982</td><td>237</td><td>251745</td></tr> <tr><td>Q2</td><td>7</td><td>587958</td><td>2184</td><td>585774</td></tr> <tr><td>Q3</td><td>10</td><td>839940</td><td>790</td><td>839150</td></tr> <tr><td>Q4</td><td>10</td><td>936470</td><td>923160</td><td>13310</td></tr> <tr><td>Q5</td><td>12</td><td>1009392</td><td>1017576</td><td>-8184</td></tr> <tr><td>Q6</td><td>3</td><td>300699</td><td>295644</td><td>5055</td></tr> <tr><td>Q7</td><td>3</td><td>254388</td><td>254778</td><td>-390</td></tr> <tr><td>Q8</td><td>16</td><td>1383216</td><td>1372896</td><td>10320</td></tr> <tr><td>Q9</td><td>4</td><td>339336</td><td>339752</td><td>-416</td></tr> <tr><td>Q10</td><td>12</td><td>1018224</td><td>1019328</td><td>-1104</td></tr> </tbody> </table> | Req     | Freq   | Avant | Après | Gain | Q1 | 3 | 251982 | 237 | 251745 | Q2 | 7 | 587958 | 2184 | 585774 | Q3 | 10 | 839940 | 790 | 839150 | Q4 | 10 | 936470 | 923160 | 13310 | Q5 | 12 | 1009392 | 1017576 | -8184 | Q6 | 3 | 300699 | 295644 | 5055 | Q7 | 3 | 254388 | 254778 | -390 | Q8 | 16 | 1383216 | 1372896 | 10320 | Q9 | 4 | 339336 | 339752 | -416 | Q10 | 12 | 1018224 | 1019328 | -1104 |
| Req                           | Freq             | Avant                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Après   | Gain   |       |       |      |    |   |        |     |        |    |   |        |      |        |    |    |        |     |        |    |    |        |        |       |    |    |         |         |       |    |   |        |        |      |    |   |        |        |      |    |    |         |         |       |    |   |        |        |      |     |    |         |         |       |
| Q1                            | 3                | 251982                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 237     | 251745 |       |       |      |    |   |        |     |        |    |   |        |      |        |    |    |        |     |        |    |    |        |        |       |    |    |         |         |       |    |   |        |        |      |    |   |        |        |      |    |    |         |         |       |    |   |        |        |      |     |    |         |         |       |
| Q2                            | 7                | 587958                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 2184    | 585774 |       |       |      |    |   |        |     |        |    |   |        |      |        |    |    |        |     |        |    |    |        |        |       |    |    |         |         |       |    |   |        |        |      |    |   |        |        |      |    |    |         |         |       |    |   |        |        |      |     |    |         |         |       |
| Q3                            | 10               | 839940                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 790     | 839150 |       |       |      |    |   |        |     |        |    |   |        |      |        |    |    |        |     |        |    |    |        |        |       |    |    |         |         |       |    |   |        |        |      |    |   |        |        |      |    |    |         |         |       |    |   |        |        |      |     |    |         |         |       |
| Q4                            | 10               | 936470                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 923160  | 13310  |       |       |      |    |   |        |     |        |    |   |        |      |        |    |    |        |     |        |    |    |        |        |       |    |    |         |         |       |    |   |        |        |      |    |   |        |        |      |    |    |         |         |       |    |   |        |        |      |     |    |         |         |       |
| Q5                            | 12               | 1009392                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 1017576 | -8184  |       |       |      |    |   |        |     |        |    |   |        |      |        |    |    |        |     |        |    |    |        |        |       |    |    |         |         |       |    |   |        |        |      |    |   |        |        |      |    |    |         |         |       |    |   |        |        |      |     |    |         |         |       |
| Q6                            | 3                | 300699                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 295644  | 5055   |       |       |      |    |   |        |     |        |    |   |        |      |        |    |    |        |     |        |    |    |        |        |       |    |    |         |         |       |    |   |        |        |      |    |   |        |        |      |    |    |         |         |       |    |   |        |        |      |     |    |         |         |       |
| Q7                            | 3                | 254388                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 254778  | -390   |       |       |      |    |   |        |     |        |    |   |        |      |        |    |    |        |     |        |    |    |        |        |       |    |    |         |         |       |    |   |        |        |      |    |   |        |        |      |    |    |         |         |       |    |   |        |        |      |     |    |         |         |       |
| Q8                            | 16               | 1383216                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 1372896 | 10320  |       |       |      |    |   |        |     |        |    |   |        |      |        |    |    |        |     |        |    |    |        |        |       |    |    |         |         |       |    |   |        |        |      |    |   |        |        |      |    |    |         |         |       |    |   |        |        |      |     |    |         |         |       |
| Q9                            | 4                | 339336                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 339752  | -416   |       |       |      |    |   |        |     |        |    |   |        |      |        |    |    |        |     |        |    |    |        |        |       |    |    |         |         |       |    |   |        |        |      |    |   |        |        |      |    |    |         |         |       |    |   |        |        |      |     |    |         |         |       |
| Q10                           | 12               | 1018224                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 1019328 | -1104  |       |       |      |    |   |        |     |        |    |   |        |      |        |    |    |        |     |        |    |    |        |        |       |    |    |         |         |       |    |   |        |        |      |    |   |        |        |      |    |    |         |         |       |    |   |        |        |      |     |    |         |         |       |

1. Fragmentation verticale et mixte
2. Enrichissement de ParAdmin
3. Traitement Parallèle : Fragmentation + Allocation
4. Considération d'autres algorithmes de sélection
5. Real Time Data Warehouse
6. Web Service Driven Approach for DB Administration
7. Self Tuning
8. ...

1. Bellatreche, L. **Optimization and tuning in Data warehouses**. To Appear in Encyclopedia of Database Systems, Tamer Özsu, Springer, December 2008
2. Bellatreche, L. **Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction** (Advances in Data Warehousing and Mining Book Series), December 2008
3. Bellatreche, L., Boukhalfa K. and Caffiau, S. **ParAdmin: Un Outil d'Assistance à l'Administration et Tuning d'un Entrepôt de Données**, 4<sup>èmes</sup> Journées sur les Entrepôts de Données et Analyse en Ligne (EDA), June 2008, Toulouse
4. Boukhalfa, K., Bellatreche, L. and Richard, P. **Horizontal Partitioning in Data Warehouse: Hardness Study, Selection Algorithms and Validation on ORACLE10G**, to appear in DAWAK'08, LNCS, September 2008, Turin Italy
5. Benkrid S., Bellatreche, L., Driass, H. **A Combined Algorithm for Data Warehouse Fragmentation and Allocation in a Parallel Environment**, to appear in 4th International Workshop on Data Management in Global Data Repositories (GRep'08), September 2008, Turin Italy
6. Bellatreche L., Missaoui R., Necir H. and Drias H. **A Data Mining Approach for Selecting Bitmap Join Indices**. Journal of Computing Science and Engineering, 1(2), pp. 177-194, December 2007.
7. Boukhalfa K., Bellatreche L., Alimazighi, Z. **HP&BJI: A Combined Selection of Data Partitioning and Join Indexes for Improving OLAP Performance**, Information Systems Journal, December 2008
8. Boukhalfa, K. and Bellatreche, L. **Sélection de schéma de fragmentation horizontale dans les entrepôts de données : formalisation et algorithmes**. Ingénierie des Systèmes d'Information (ISI), 11(6/2006), pp. 55-82, Novembre, 2006.
9. Bellatreche L., Boukhalfa, K. and Mohania, M. **Pruning Search Space of Physical Database Design**. 18th International Conference on Database and Expert Systems Applications (DEXA'07), pp. 479-488, LNCS, September, 2007, Germany.
10. Bellatreche L., Missaoui R., Necir H. and Drias H. **Selection and Pruning Algorithms for Bitmap Index Selection Problem using Data Mining**, DaWaK'07, pp. 221-230, LNCS, September, 2007, Regensburg, Germany.
11. Bellatreche, L., Boukhalfa, K. Abdalla, H. **SAGA : A Combination of Genetic and Simulated Annealing Algorithms for Physical Data Warehouse Design**. 23rd British National Conference on Databases (BNCOD'06), pp. 212-219, LNCS, July, 2006, Belfast, Northern Ireland.
12. Bellatreche L. **Selection of Redundant and non Redundant Optimization Structures in VLDBs**. 3rd International Workshop on Data Management in Global Data Repositories (GRep'2007), pp. 819-824, IEEE Computer Society Press, September, 2007, Regensburg, Germany.