



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



ParAdmin: Un Outil d'Assistance à l'Administration et Tuning d'un Entrepôt de Données

Ladjel Bellatreche, Kamel Boukhalfa et Sybille CAFFIAU
Université de Poitiers - LISI/ENSMA
{bellatreche, boukhalk, sybille.caffiau}@ensma.fr

Contexte

Caractéristiques des entrepôts de données

❑ 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

Besoins Principaux d'Administration

Sélection des Techniques d'optimisation

- Vues
- Index
- Fragmentation Horizontale
- Clustering

Mode de sélection

- Isolé
- Combinée

Choix des tables/attributs



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

Exemple

Sélection des Techniques d'optimisation

- Index de jointure

Algorithmes de sélection

- Glouton basée sur la fréquence d'accès

Mode de sélection

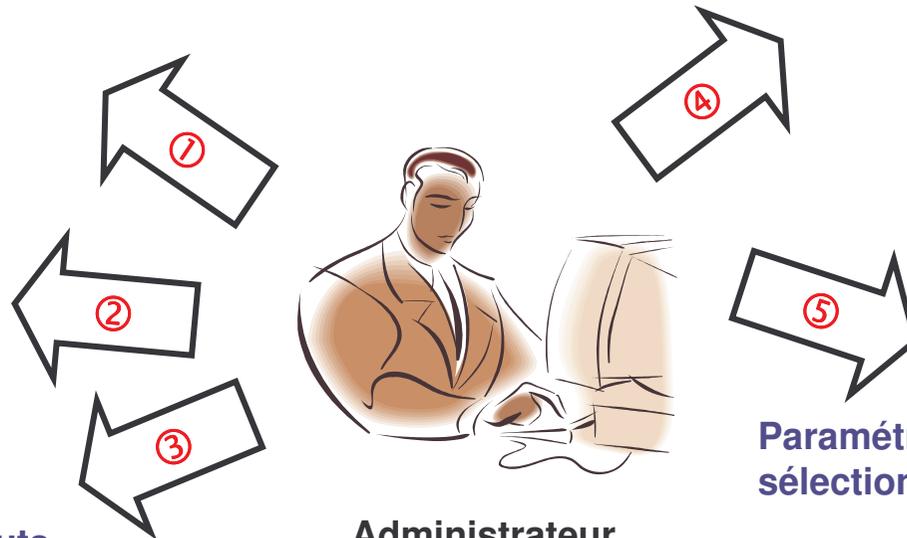
- Isolé

Choix des tables/attributs

- Tables volumineuses
- Attributs les plus utilisés



Administrateur



PLAN

- Existant
- ParAdmin
 - Conception
 - Mise en oeuvre
- Démonstration
- Conclusion et perspectives

Existant: Outils d'Administration

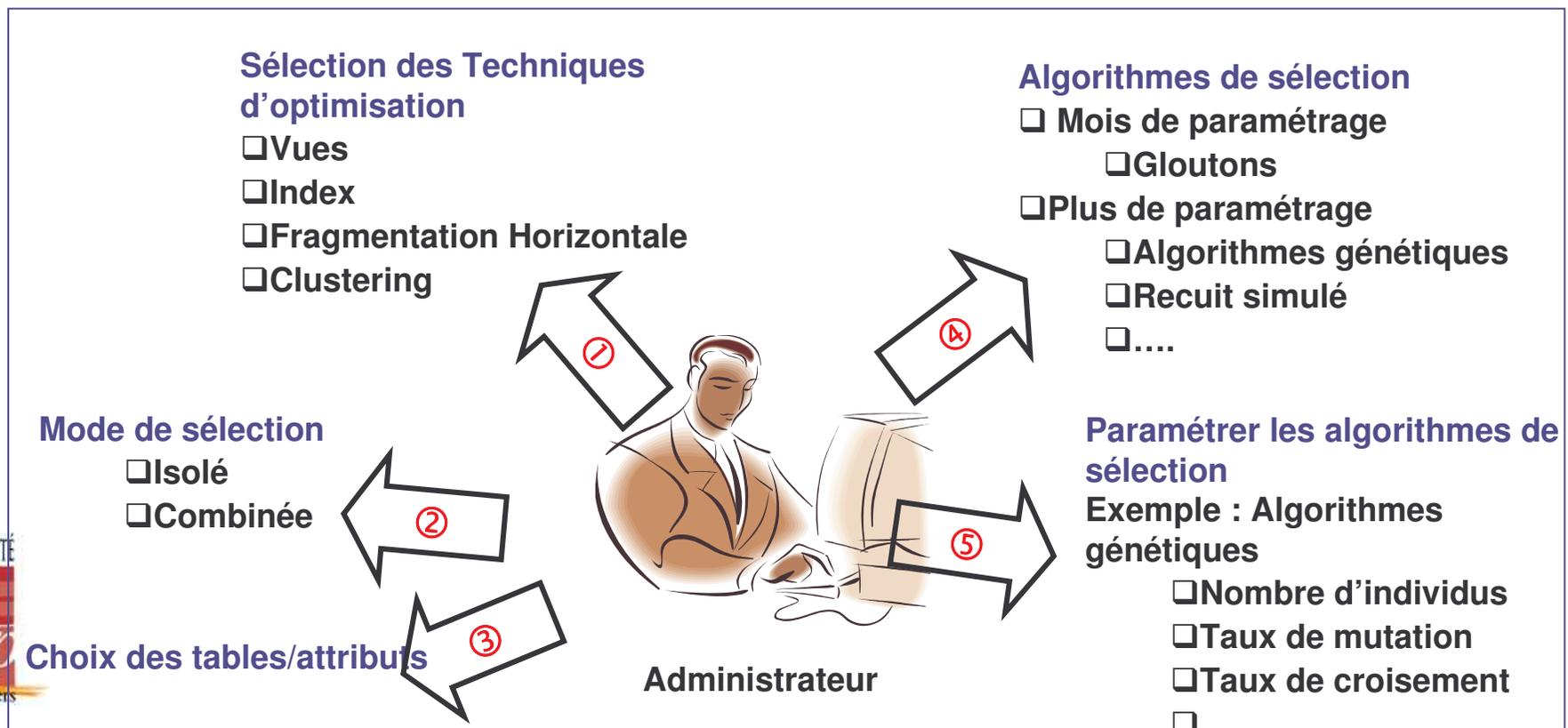
Outil	Techniques Supportées
SQL Database Tuning Advisor (DTA)	Fragmentation par intervalle Vues matérialisées Index mono attribut
Oracle SQL Access Advisor	Fragmentation: List, Range, Hash, Composite, Référence, etc. Vues matérialisées Index
DB2 Index Advisor tool	Fragmentation (modes simples) Vues matérialisées, index et groupement

Inconvénients

- Outils propriétaires
- Aucun paramétrage des algorithmes
- Absence d'un consensus sur les techniques d'optimisation supportées
- Non prise en considération le feedback

Etapes principales de mise en œuvre de ParAdmin

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



Techniques d'optimisation

- 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)
 - Pré-calculent des jointures entre la table des faits et les tables de dimension
 - Définis sur la table des faits en utilisant des attributs de dimension
 - Similarités entre FHD et IJBs
 - Pré-calculent des jointures
 - Définis sur les attributs de sélection
 - Optimisent les requêtes de jointure en étoile
- Etude des Besoins
 - Techniques d'optimisation
 - Conception
 - Validation

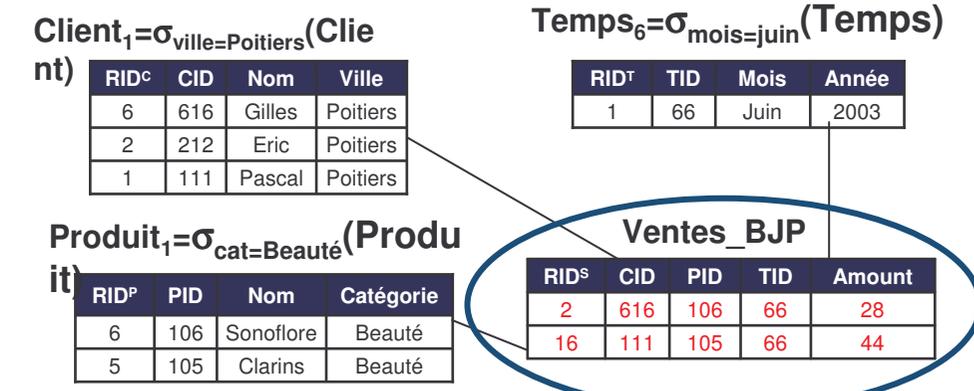
Similarité : exemple

Index de jointure binaire

Client				Ventes												AND
RID ^C	CID	Nom	Ville	RID ^P	PID	Nom	Catégorie	RID ^T	TID	Mois	Année	Montant	AND			
6	616	Gilles	Poitiers	16	106	Sonoflore	Beauté	2	66	Juin	2003	28	1			
5	515	Eric	Poitiers	5	105	Clarins	Beauté	5	66	Juin	2003	44	1			
4	414	Pascal	Poitiers	17	107	Clarins	Beauté	17	66	Juin	2003	20	1			
3	313	Eric	Poitiers	18	108	Clarins	Beauté	18	66	Juin	2003	20	1			
2	212	Eric	Poitiers	19	109	Clarins	Beauté	19	66	Juin	2003	60	1			
1	111	Pascal	Poitiers	20	110	Clarins	Beauté	20	66	Juin	2003	10	1			

```

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'
    
```



Sous schéma en étoile

- IJB** : Que fait l'optimiseur?
- Lire le bitmap
 - Effectuer un AND logique
 - Compter le nombre de tuples

- FH** : Que fait l'optimiseur?
- Lire le fragment Ventes_BJP
 - Compter le nombre de tuples
 - Éviter trois jointures

Modes de sélection supportées

- Sélection isolée
 - FHSEULE : sélection d'un schéma de fragmentation.
 - BJISEULS : sélection d'une configuration d'index de jointure
- Sélection combinée (FH&IJBs)
 1. Sélectionner un schéma de fragmentation (SF)
 2. Identifier les requêtes non bénéficiaires du processus de fragmentation
 3. Identifier les attributs indexables
 4. Sélectionner une configuration d'index.

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

Algorithmes supportés

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]

- Etude des Besoins

- Techniques d'optimisation**

- Conception

- Validation

Modèle de tâches pour la modélisation ParAdmin

❑ Modélisation par K-MADE [INRIA]

- ❑ Exprimer les activités utilisateurs
- ❑ Valider l'interface par un simulateur de scénarios
- ❑ Post et pré conditions

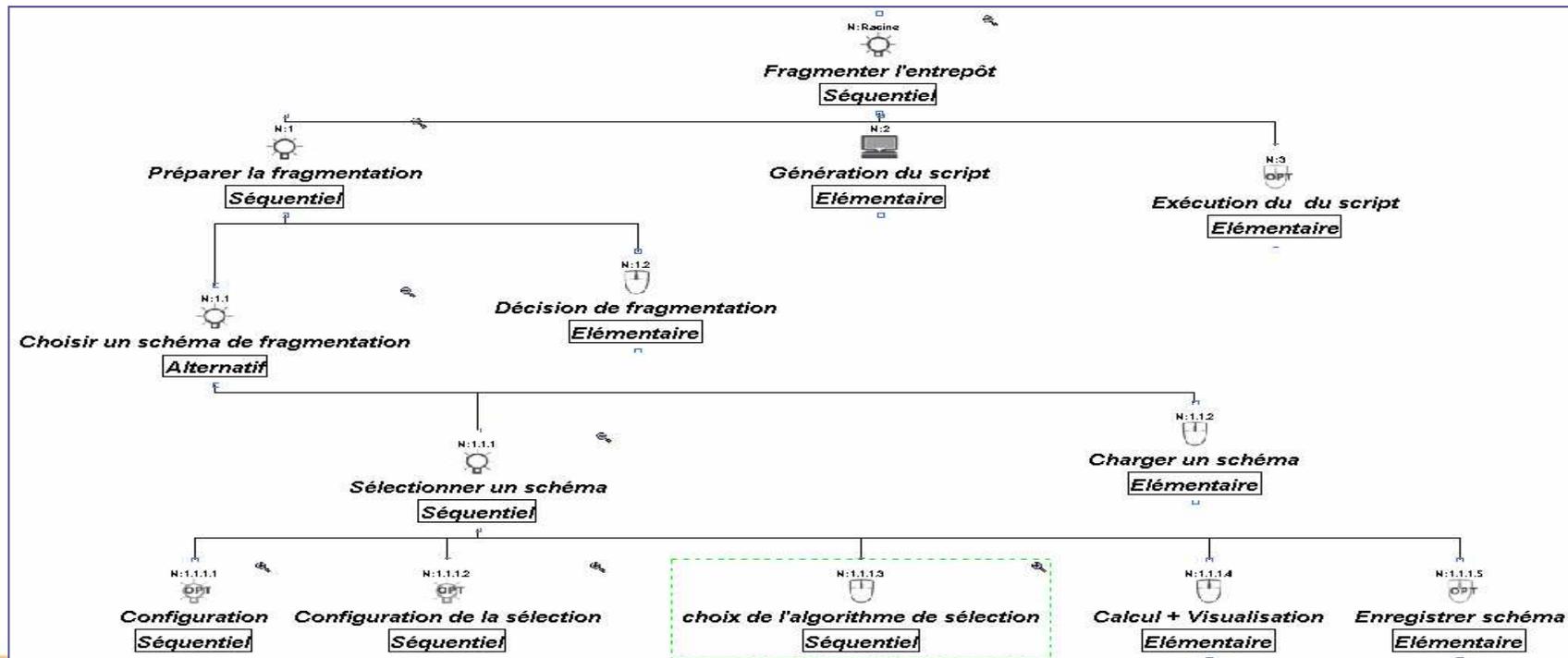
❑ Etude des Besoins

❑ Techniques d'optimisation

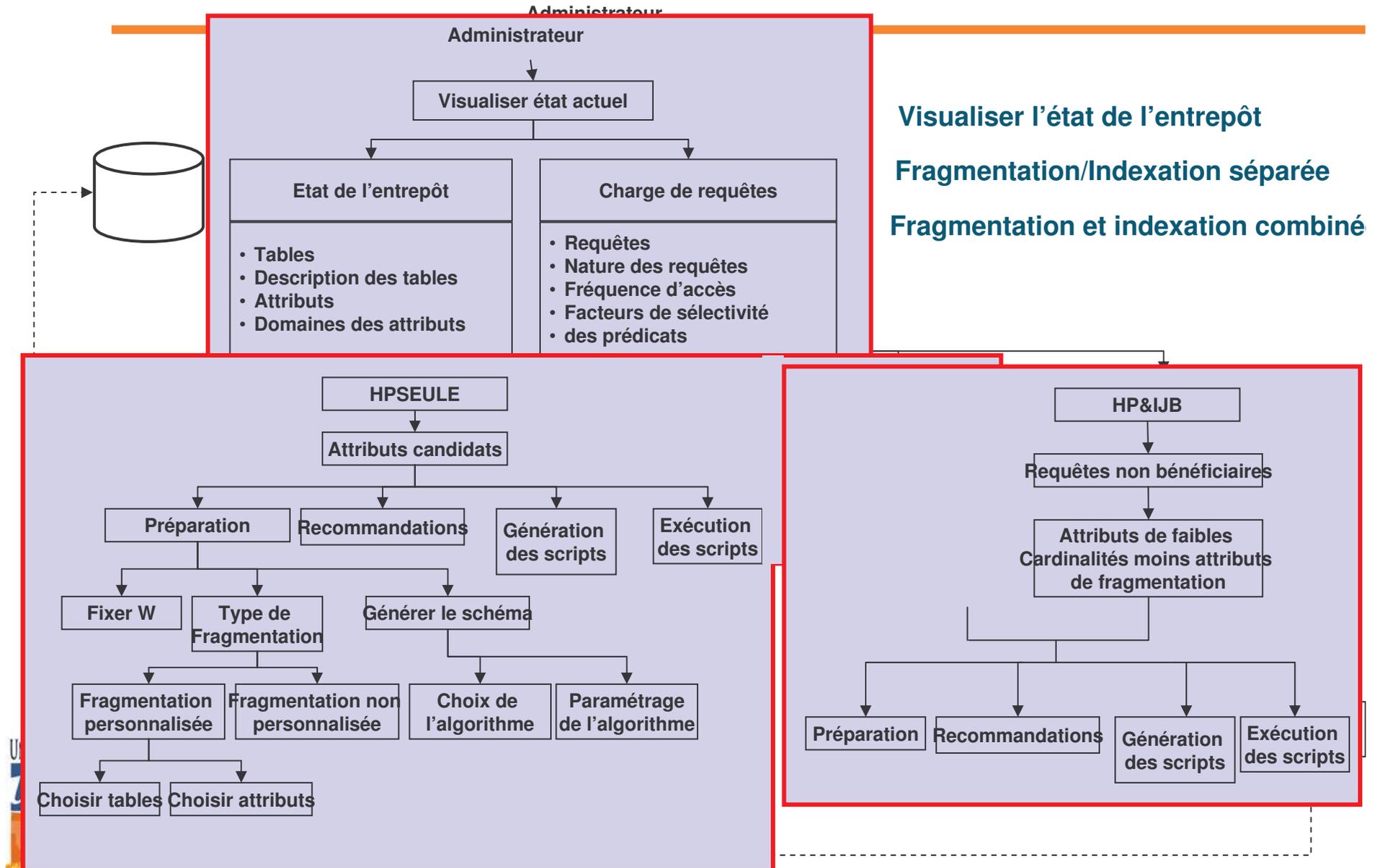
❑ Conception

❑ Validation

❑ Exemple d'un modèle de tâches



Architecture



Visualiser l'état de l'entrepôt

Fragmentation/Indexation séparée

Fragmentation et indexation combiné

Visualisation de l'état de l'entrepôt

- 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.

- Etude des Besoins
- Techniques d'optimisation
- Conception
- Validation : Visual C

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 de l'entrepôt

Fragmentation non personnalisée

Faire confiance à l'outil

Fragmentation personnalisée

Participation de l'administrateur aux différents choix

Etude des Besoins

Techniques d'optimisation

Conception

Validation : Visual C

FRAGMENTATION

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

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

Tuning de la conception physique

Sélection combinée

1. Sélectionner un schéma de fragmentation

Etude des Besoins

Techniques d'optimisation

Conception

Validation : Visual C

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		Génération des Scripts	
Select IJBs	Administration	Algorithmes	Résultats
Coût par requête		Résultat BJI	
Attributs candidats		Index créés	
Group_level Family_level Line_level Division_level Month_level Quarter_level Gender_level All_level		Division_level Month_level Quarter_level Gender_level All_level	
Nombre de requêtes de départ		60	Nombre de requêtes non bénéficiaires
			39
		Espace disponible	100 Mo
		Espace Occupé	81 Mo
		E/S Avant Fragmentation	62160727
		E/S après fragmentation	36594992
		E/S après indexation	8048871

Recommandations

- L'outil 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
- L'administrateur peut valider et générer les scripts.
- Il pourra revenir en arrière pour reconsidérer ses choix.

- Etude des Besoins
- Techniques d'optimisation
- Conception
- Validation : Visual C

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

Schéma de fragmentation final	Coût par requête	Attributs de				
		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

Conclusion et perspectives

- Développement d'un outil générique d'administration
- Prise en compte d'une administration personnalisée et non personnalisée.
- Utilisation des modèles de tâches (IHM) pour concevoir ParAdmin.
- Considération des feedback

- Enrichissement de l'outil (prise en considération d'autres techniques d'optimisation)