



Laboratoire de la Communication dans les Systèmes Informatiques LCSi
Ecole nationale Supérieure d'Informatique ESI
Alger, Algérie

Sélection Statique et Incrémentale des Index de Jointure Binaires Multiples

Rima BOUCHAKRI

LCSI - Ecole nationale Supérieure
d'Informatique
ESI, Alger – ALGERIE
r_bouchakri@esi.dz

Ladje! BELLATRECHE

LISI/ENSMA - Université de
Poitiers – FRANCE
bellatreche@ensma.fr

EDA'11
Clermont Ferrand 8-9 Juin 2011

Contexte : Optimisation de requêtes en étoile

Schéma en étoile

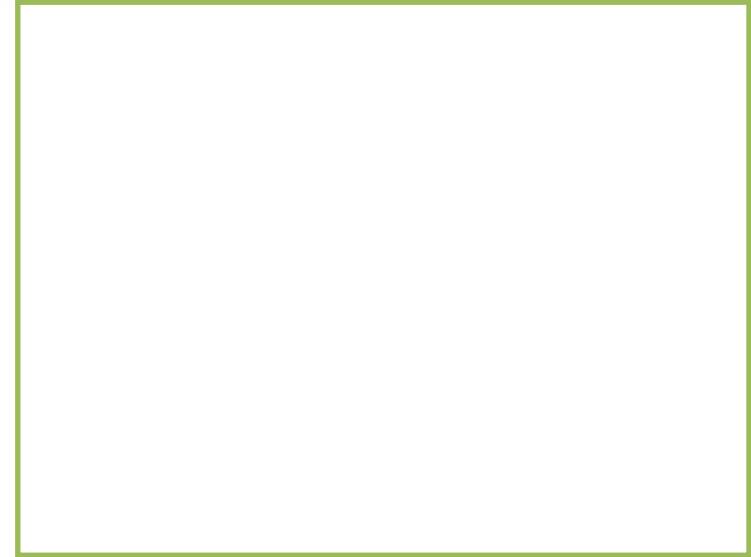
- ⇒ Une table de faits (**volumineuse**)
- ⇒ Plusieurs tables de dimension

Requêtes de jointure en étoile

- ⇒ **Sélection** sur les tables de dimension
- ⇒ **Multiple jointures** (tables de dimension et des faits)
- ⇒ Aucune jointure directe entre les tables de dimension

Exigences des décideurs

- ⇒ Temps de réponse raisonnable



```
SELECT  P.Marque, sum(montant)
FROM    Ventes V, Produit P, Temps T, Client C
WHERE   V.PID = P.PID                (jointure)
AND     V.TID = T.TID                (jointure)
AND     V.CID = C.CID                (jointure)
AND     T.année = 2006                (sélection)
AND     P.Catégorie = 'Beauté'       (sélection)
AND     C.Ville = 'Poitiers'         (sélection)
GROUP BY P.Marque
```

➔ **Importance extrême des structures d'optimisation**

Contexte : Index de Jointures Binaires IJB

- ❑ Stockage du résultat d'exécution d'une jointure sous forme de bitmaps
- ❑ Défini sur un ou plusieurs attributs dimensions

Vente

IJB *simples*

Client

RID ^c	CID	Nom	Ville	Genre
6	616	Gilles	Poitiers	F
5	515	Yves	Paris	F
4	414	Patrick	Nantes	M
3	313	Didier	Nantes	F
2	212	Eric	Poitiers	M
1	111	Pascal	Poitiers	M

RID ^s	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

RID	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

CREATE BITMAP INDEX

Ventes_PG_bjix

ON Ventes(Client.Ville, Client.Genre)

FROM Ventes, Client

WHERE Ventes.CID = Client.CID

Deux opérations :

1. Lire dans les bitmaps correspondant
2. Compter le nombre de 1.

➔ Pas de lecture dans la table Ventes

Etude de Complexité

IJB Simples

- Charge de requêtes Q
 - Un ensemble de n attributs indexables
- ⇒ Le nombre de configurations possibles d'index pour optimiser une charge de requêtes Q est :

$$2^n - 1$$

5 attributs :
 $2^5 - 1 = 31$ config

IJB Multiples

- Charge de requêtes Q
 - Un ensemble de n attributs indexables
- ⇒ Le nombre de possibilités pour évaluer un seul IJB est : $2^n - n - 1$
- ⇒ Le nombre de configurations d'index pour optimiser Q est :

5 attributs :
 $2^{26} - 1$!!

$$2^{(2^n - n - 1)} - 1$$

Sélection Statique

Entrées :

- Un ensemble de tables de dimension $D = \{D_1, \dots, D_d\}$ et une table des faits F
- Charge de requêtes les plus fréquentes $Q = \{Q_1, \dots, Q_m\}$
- S : taille de espace de stockage des index

Sorties :

- Configuration finale d'index $\{IJB_1, \dots, IJB_n\}$

Objectif :

- Réduire le temps d'exécution de Q
- Respecter la contrainte d'espace de stockage

Problème NP-Complet

→ **Besoin des heuristiques**

Analyse de la formalisation initiale:
 Q préalablement connues



Evolution de la charge :

- Ajout de nouvelles requêtes
- Changement des fréquences de requêtes



**Obsolescence des Structures
d'optimisation**

→ **Sélection Incrémentale**

- 1. Sélection Statique des IJB**
- 2. Sélection Incrémentale des IJB**
- 3. Etude expérimentale**
- 4**

Approches de sélection statique d'IJB existantes

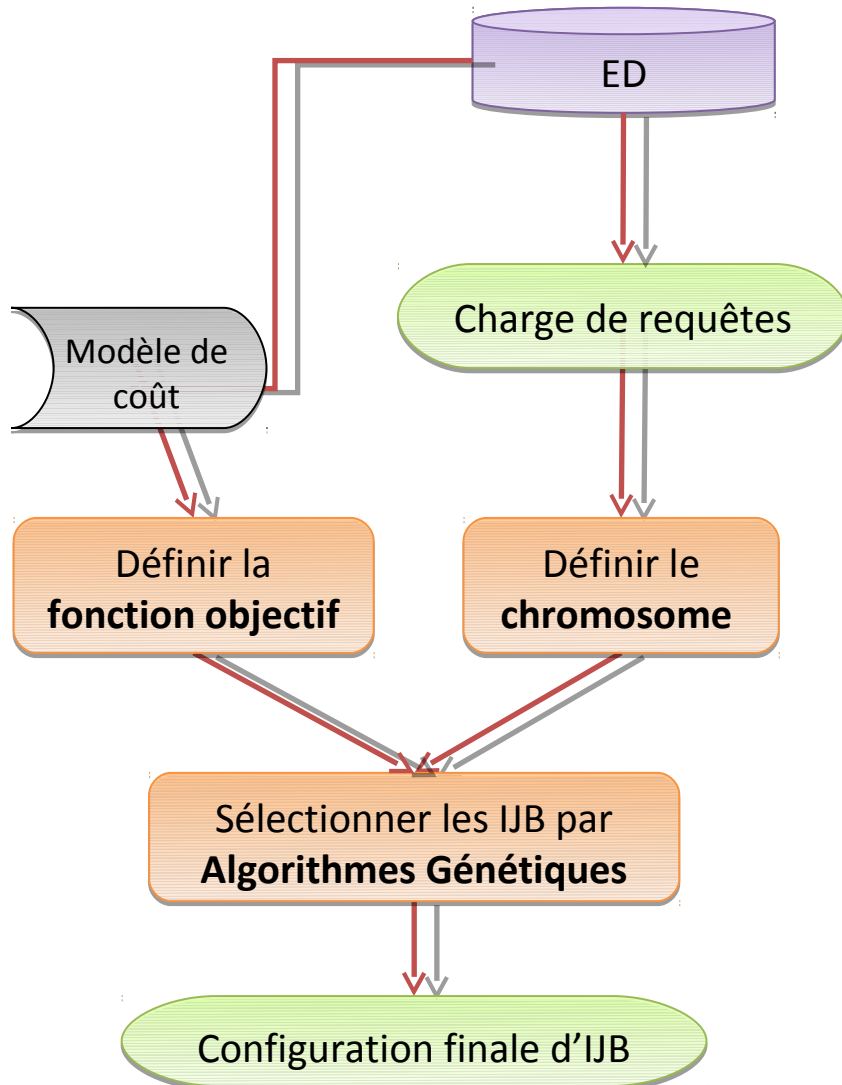
1. Sélection Statique des IJB
2. Sélection Incrémentale des IJB
3. Etude Expérimentale
4. Conclusion & Perspectives

	Algorithmes de sélection	Type de sélection
IJB Simples	<ul style="list-style-type: none"> - Gloutons (Bellatreche et al. 2008,2010) - Algorithmes génétiques (EDA'10) 	Statique
IJB Multiples	Fouille de données : <ul style="list-style-type: none"> - Aouiche et al. 2005 - Azefack et al. 2007 	Statique Incrémentale
	<ul style="list-style-type: none"> - Hill Climbing (Boukhalifa et al. 2010) 	Statique

- Peu de travaux sur la sélection des IJB multiples
- Peu de travaux sur la sélection incrémentale

Notre Approche de sélection d'IJB Multiples par AG (I)

1. Sélection Statique des IJB
2. Sélection Incrémentale des IJB
3. Etude Expérimentale
4. Conclusion & Perspectives

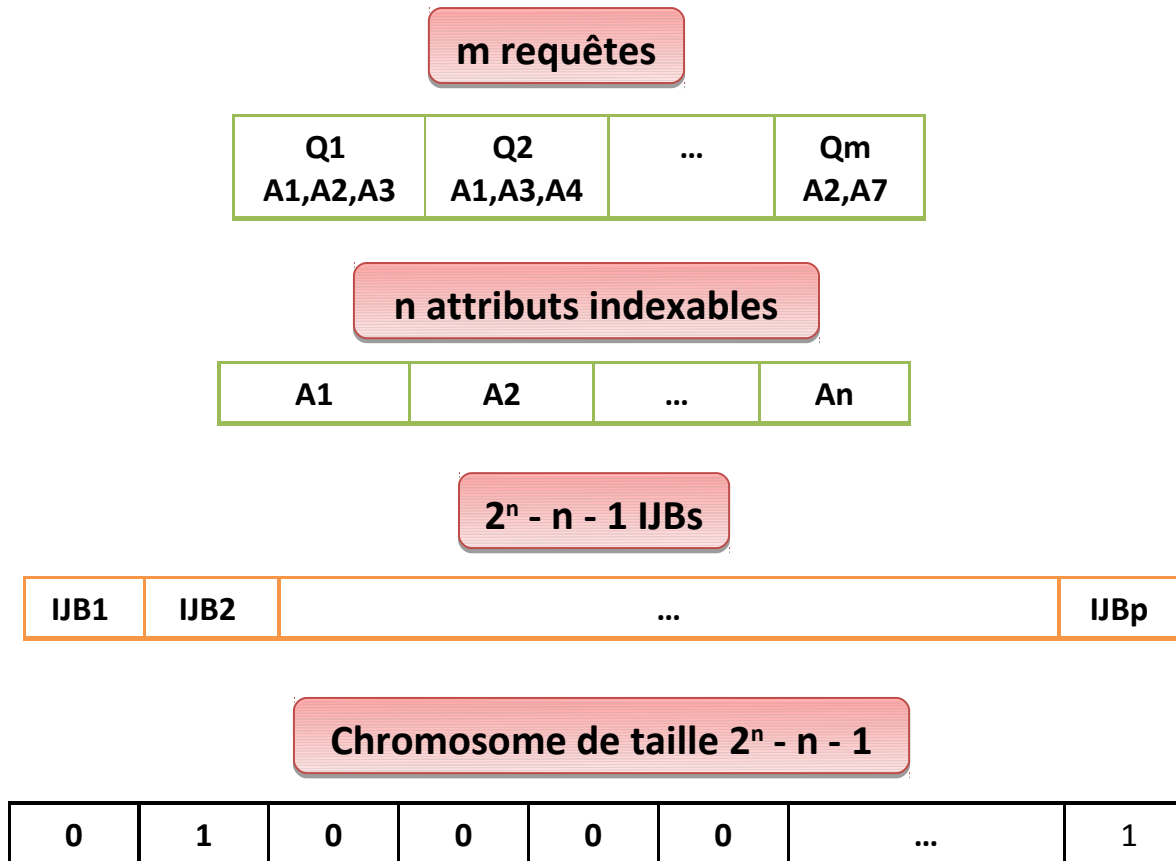


→ **Comment définir le chromosome ?**

Notre Approche de sélection d'IJB Multiples par AG (II)

1. Sélection Statiques des IJB
2. Sélection Incrémentale des IJB
3. Etude Expérimentale
4. Conclusion & Perspectives

Définir le chromosome : méthode naïve

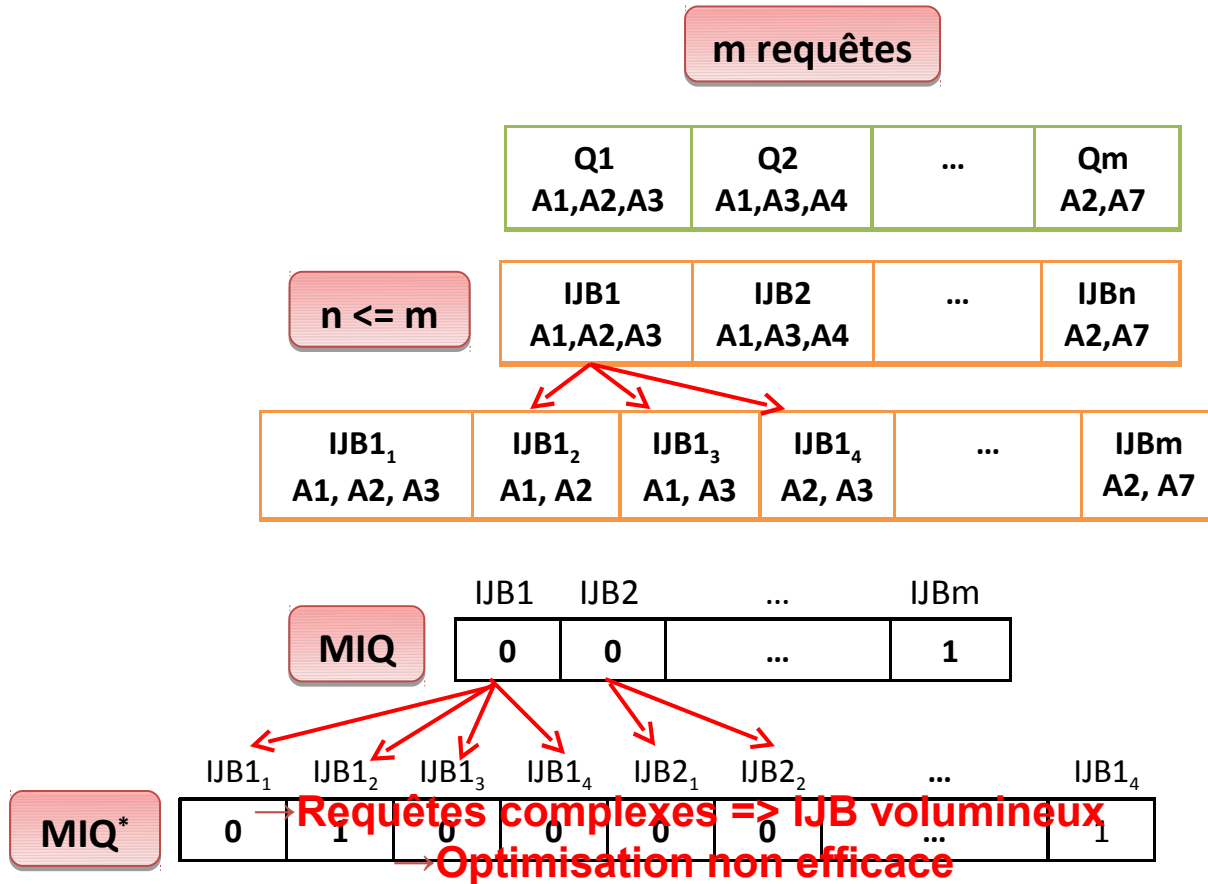


→ **Espace de recherche très complexe**

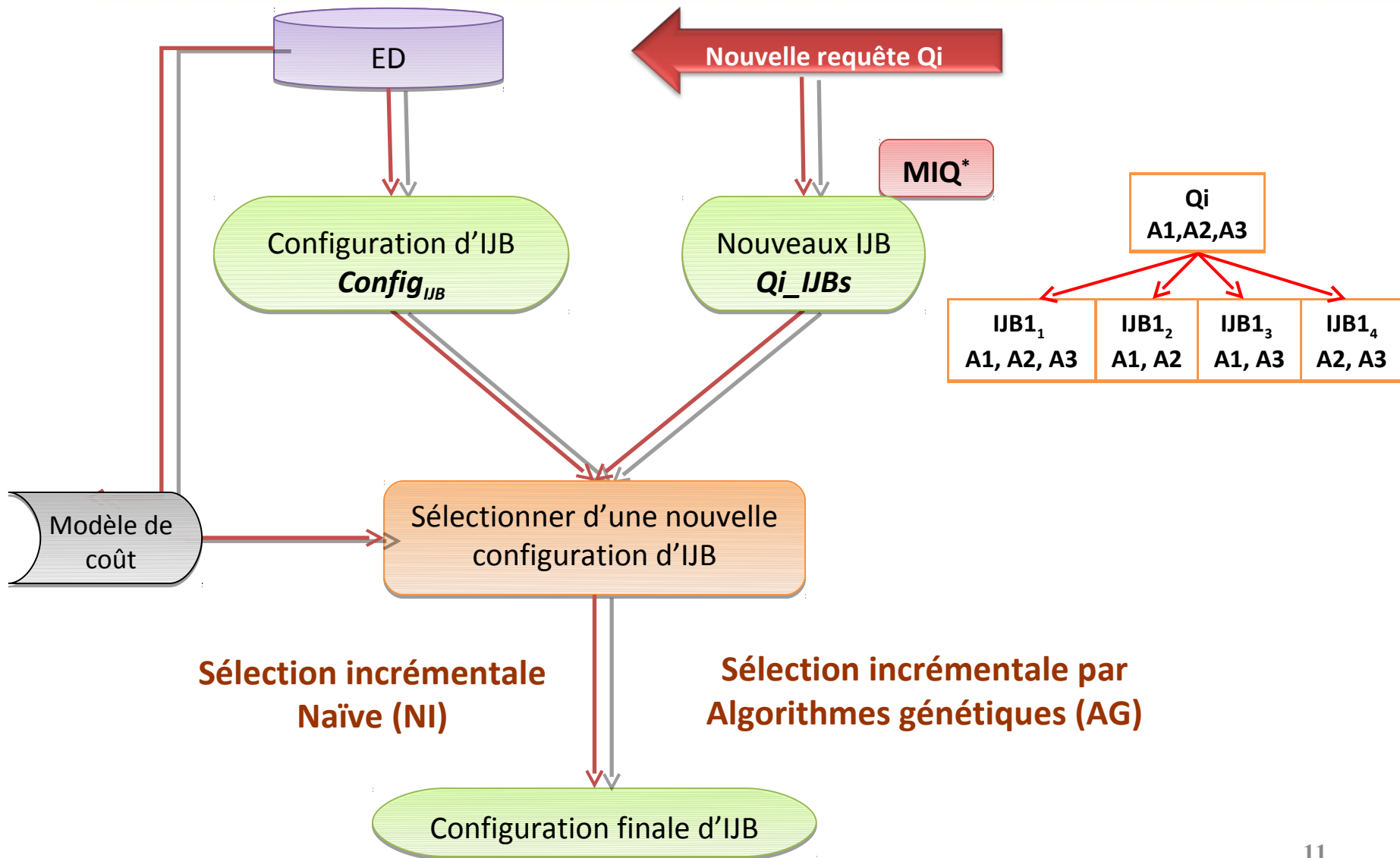
Notre Approche de sélection d'IJB Multiples par AG (III)

1. Sélection Statiques des IJB
2. Sélection Incrémentale des IJB
3. Etude Expérimentale
4. Conclusion & Perspectives

Définir le chromosome : méthodes basés requêtes

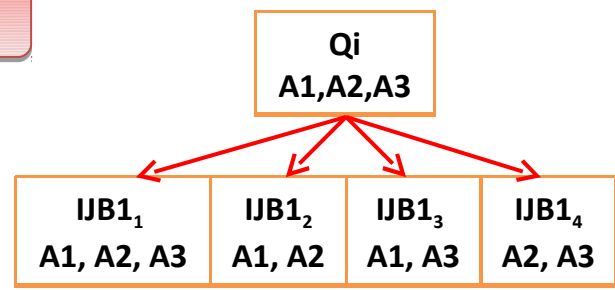
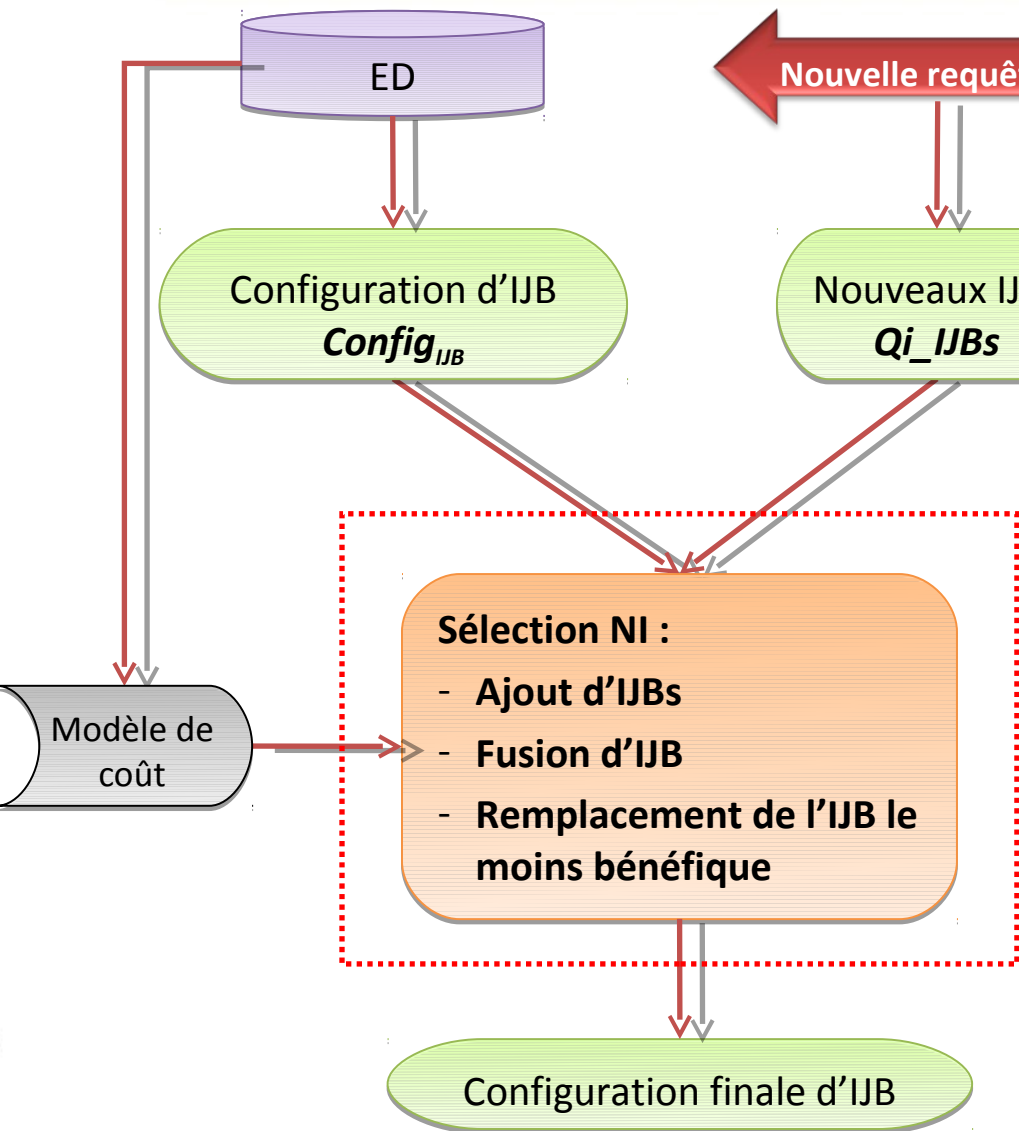


1. Sélection Statiques des IJB
2. **Sélection Incrémentale des IJB**
3. Etude Expérimentale
4. Conclusion & Perspectives



Sélection incrémentale Naïve (NI)

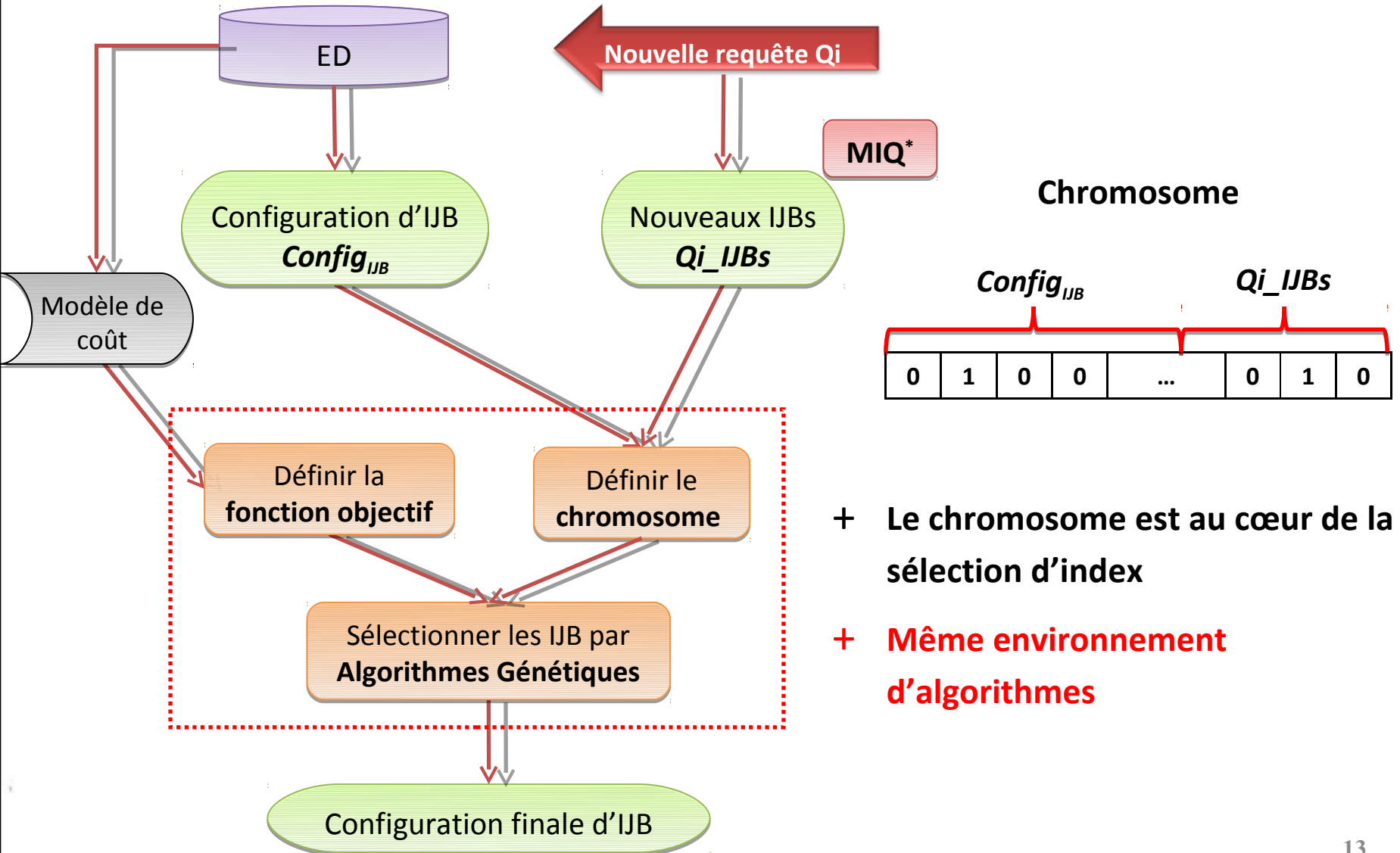
1. Sélection Statiques des IJB
2. **Sélection Incrémentale des IJB**
3. Etude Expérimentale
4. Conclusion & Perspectives



- a) Si l'espace de stockage d'index S n'est pas saturé, ajouter à Config_IJB un index de Qi_IJB
- b) Tant que S est saturé, fusionner des index de Config_IJB en un index de Qi_IJBj
 $IJB(A1,A2) + IJB(A2,A3) = IJB(A1,A2,A3)$.
- c) Si aucune fusion n'est possible, remplacer dans Config_IJB l'index le moins bénéfique

Sélection incrémentale par Algorithmes Génétiques (AG)

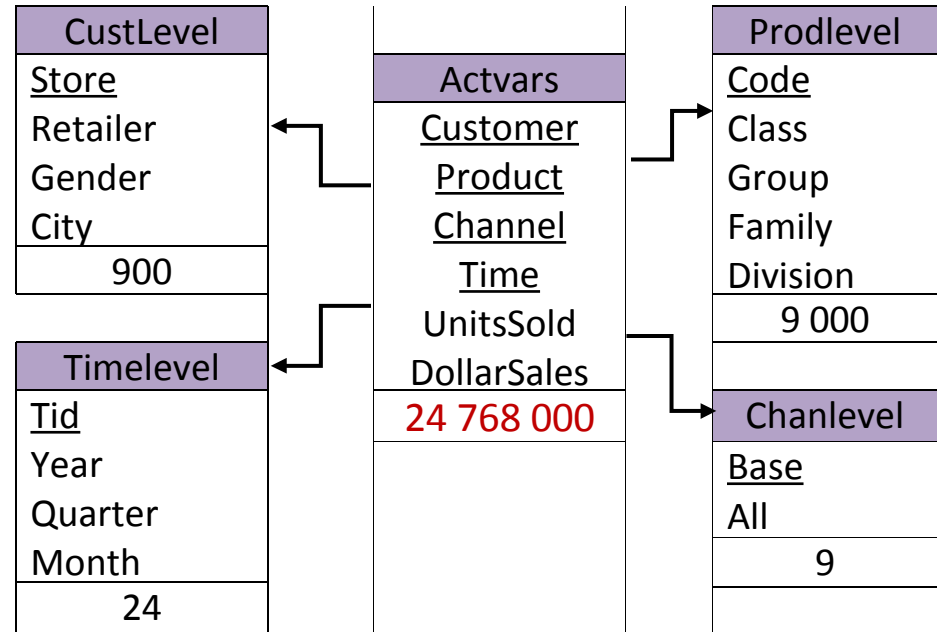
1. Sélection Statiques des IJB
2. **Sélection Incrémentale des IJB**
3. Etude Expérimentale
4. Conclusion & Perspectives



1. Sélection Statiques des IJB
2. Sélection Incrémentale des IJB
3. **Etude Expérimentale**
4. Conclusion & Perspectives

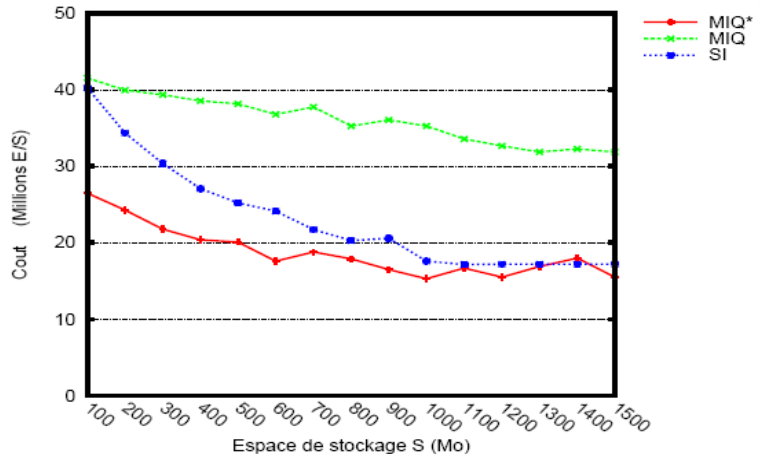
Environnement expérimental

- APB1 Benchmark (Council, 1998) sous le SGBD **Oracle 11g**
- Une charge de **70 requêtes**.
- **18 attributs** de sélection
- JAVA, API JGAP (JAVA Genetic Algorithms Package)
- Machine Core2Duo, 2 Go de RAM

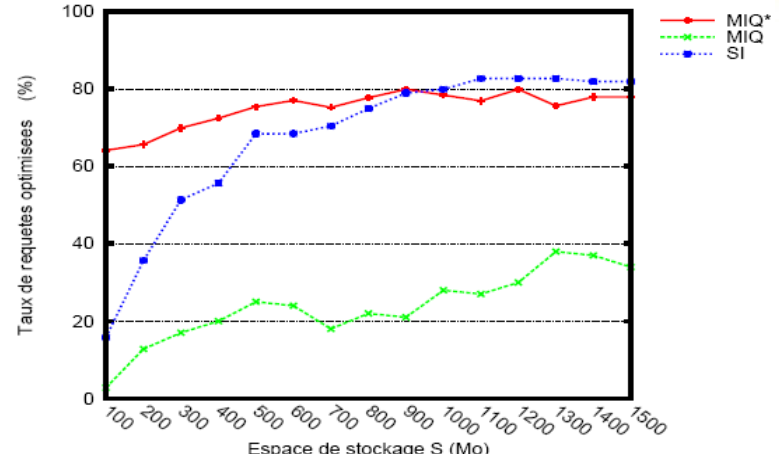


Evaluation théorique de la sélection statique des IJB

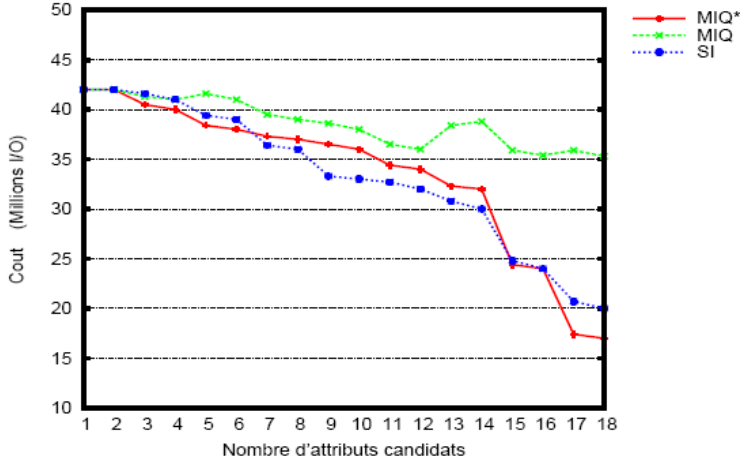
1. Sélection Statiques des IJB
2. Sélection Incrémentale des IJB
3. Etude Expérimentale
4. Conclusion & Perspectives



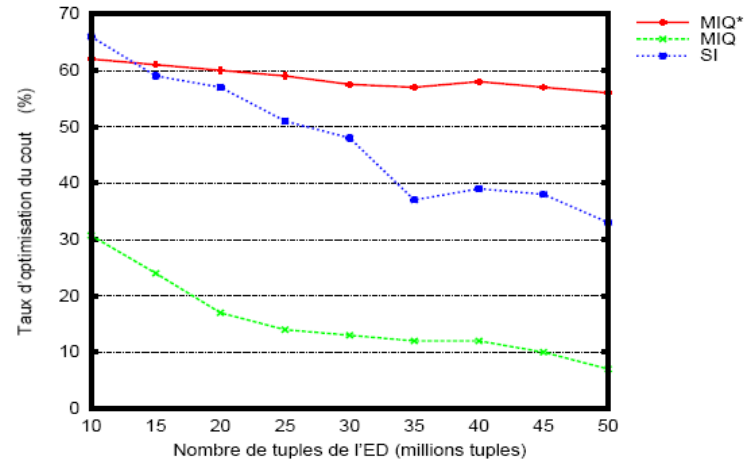
Coût d'exécution Vs. espace de stockage S



Taux d'optimisation des requêtes Vs. espace de stockage S



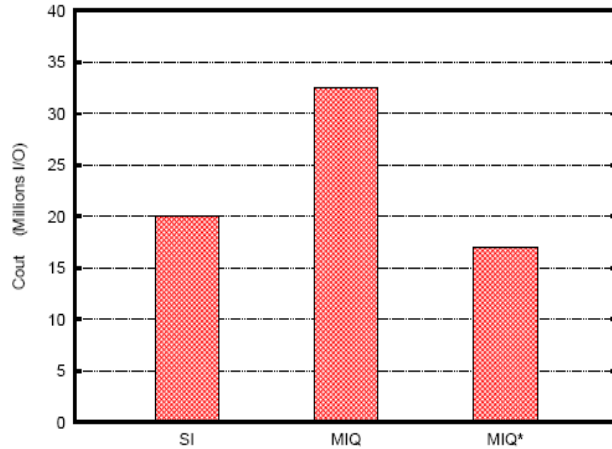
Coût d'exécution Vs. Nombre d'attributs candidats à la sélection



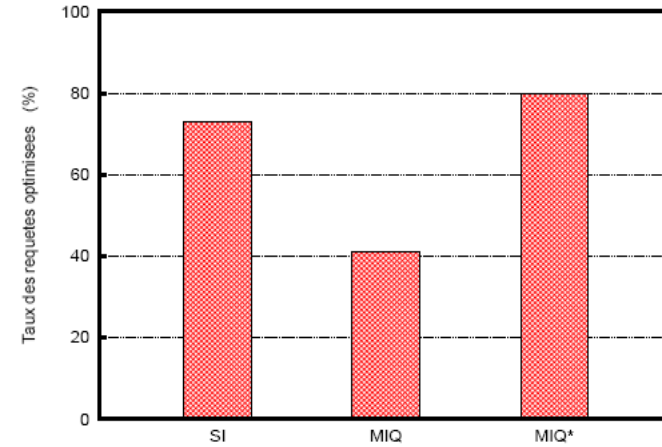
Coût d'exécution Vs. Taille de l'ED

Evaluation pratique de la sélection statique des IJB

1. Sélection Statiques des IJB
2. Sélection Incrémentale des IJB
3. **Etude Expérimentale**
4. Conclusion & Perspectives



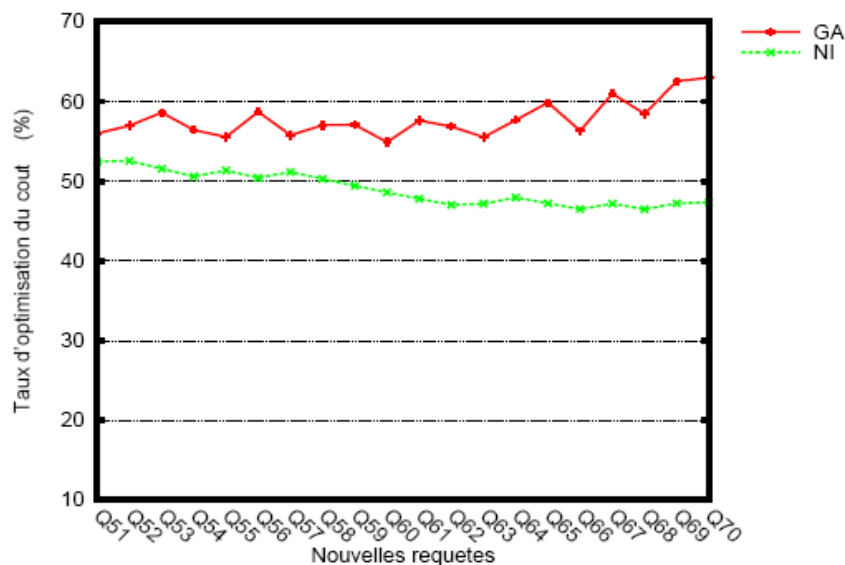
Le coût réel des requêtes sous Oracle11g (SI, MIQ et MIQ*)



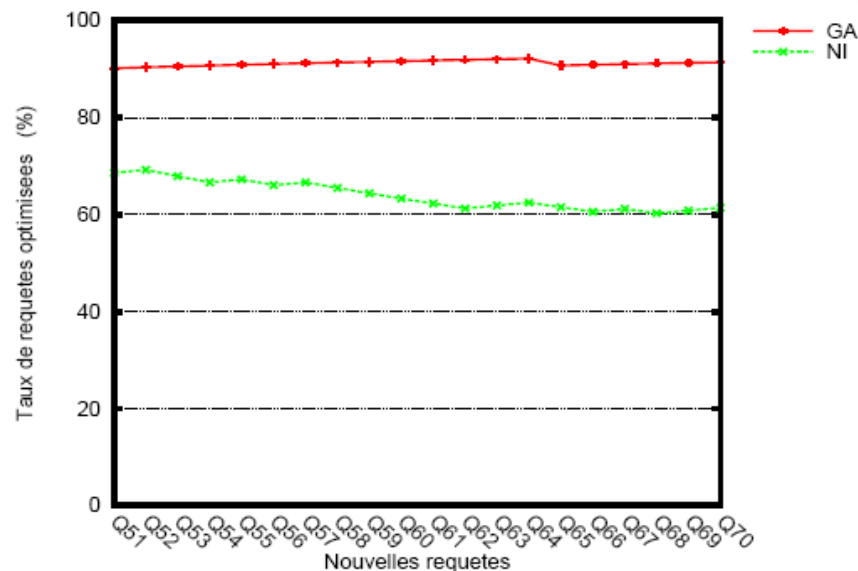
Le taux d'optimisation des requêtes sous Oracle11g (SI, MIQ et MIQ*)

Evaluation théorique de la sélection incrémentale des IJB

1. Sélection Statiques des IJB
2. Sélection Incrémentale des IJB
3. **Etude Expérimentale**
4. Conclusion & Perspectives



Taux de réduction du coût d'exécution des requêtes : NI vs. GA



Taux d'optimisation de la charge de requêtes : NI vs. GA

- ❑ Proposition d'une nouvelle approche de sélection statique des index de jointure binaire multiples par algorithmes génétiques
- ❑ Définition d'un codage de chromosome flexible (statique/incrémental)
- ❑ Proposition d'une démarche de sélection incrémentale
- ❑ Tests théoriques et pratiques sous Oracle 11g
- ❑ Définir une architecture incrémentale
- ❑ Effectuer des tests pratiques pour la sélection incrémentale

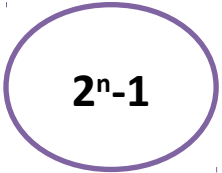
*Merci pour votre
attention
??*

IJB Simples vs. IJB Multiples

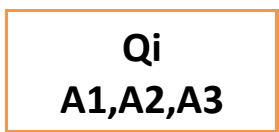
Espace de recherche

Espace de stockage

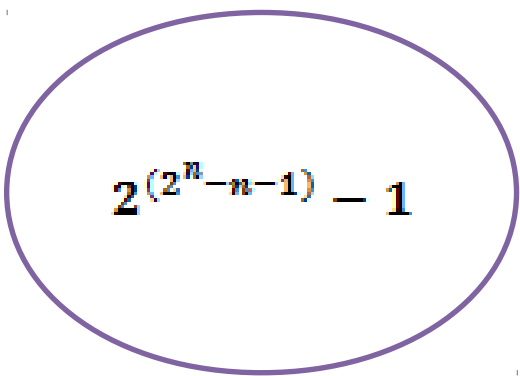
IJB simples



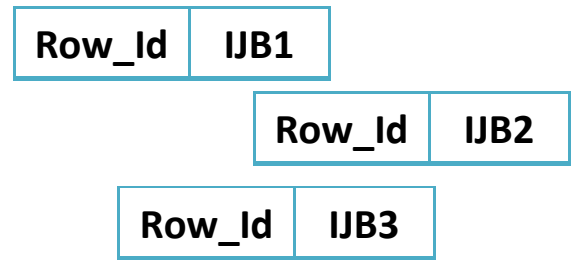
Chaque index contient un champs d'identifiant Row_Id (16 octets ORACLE 11g)



IJB multiples



IJB simples



IJB multiples



Choix appartient à l'administrateur :

- Espace vs. Complexité : IJB Simple vs. IJB Multiples