

1) Quel sont les 2 types de jointures ?

Jointure avec index (très courante) : on effectue naturellement la jointure sur les clé primaires/étrangères (non bloquant).

jointure sans index : 3 algorithmes : jointure par boucle imbriquées, jointure par tri-fusion, jointure par hachage.

2) Que fait indexJoin ?

À partir de l'opérateur fullscan sur la première table il récupère le nuplet et la valeur du foreignKey qu'il va chercher dans l'index de clé primaire de l'autre table, puis avec cette valeur d'index il fait un directAcces pour récupérer le nuplet de l'autre table.

Cela permet un pipelining complet (non bloquant).

3) Comment fonctionne la jointure par boucle imbriquée ?

Ne fonctionne que sur de petites tables.

Pour chaque nuplet de la première table on parcourt la deuxième table pour trouver la liaison. $O(n^2)$.

Si une des deux tables tient en mémoire, l'algorithme chargera alors cette table en mémoire et ne devra parcourir la deuxième table qu'une fois pour faire la jointure.

Si aucune table ne tient en mémoire, on peut découper une des deux tables en fragments, par conséquent le nombre de lectures de la seconde table est divisé par le nombre de fragments. Ceci est plus rentable que de parcourir n fois les deux tables.

4) Comment est choisie l'algorithme de jointure ?

Le système choisit l'algorithme le plus efficace en fonction de statistiques : il regarde la taille des tables, l'espace en mémoire ... Si un index est présent, la jointure avec index est souvent préférée.

5) Qu'est-ce que la jointure par tri fusion ?

Jointure à 2 phases :

- trie les deux tables sur l'attribut de jointure
- Puis on fusionne

6) Qu'est-ce que la jointure par hachage ?

On décompose les 2 tables en fragments par une fonction de hachage puis on associe les fragments.