

# Entrepôts de Données et Big Data : Optimisation de requête - partie 2

TD/TP en 2 parties (22/09 et 29/09). Rendu facultatif avant le 08/10 (attention : soigné, clair, et synthétique - max 5MB) à déposer dans l'espace Moodle dédié au cours.

## 1 Les plans d'exécution sous ORACLE

Téléchargez sur le Moodle du module les fichiers `script_table.sql` et `script_remplissage.sql`. Connectez-vous sur ORACLE et agrandir les tailles de lignes sous sqlplus afin de visualiser les plans d'exécution dans leur globalité (`SET LINESIZE 200`).

### 1.1 Sélection

Exécutez les 2 scripts sql téléchargés (commande : `@nom_fichier.sql`).

**Question 1** Examinez les scripts pour comprendre ce qu'il font.

Remarque : vous pouvez activer l'autotrace (`set autotrace on`) afin de visualiser certains éléments de statistiques.

**Question 2** Explicitez le plan d'exécution choisi par l'optimiseur et les statistiques obtenues lors de l'exécution d'une requête permettant d'afficher le nom des villes dont le numéro insee est 1293 : testez avec `insee=1293` et `insee='1293'`, puis comparez.

**Question 3** Ajoutez une clé primaire sur la table ville (utiliser l'attribut insee).

**Question 4** Explicitez à nouveau le plan d'exécution choisi par l'optimiseur et les statistiques obtenues lors de l'exécution de la requête précédente (afficher le nom des villes dont le numéro insee est 1293 en testant `insee=1293` et `insee='1293'`). Quelles sont les différences observées par rapport à la question 2 ?

### 1.2 Jointure

**Question 5** Explicitez maintenant le plan d'exécution choisi par l'optimiseur et les statistiques obtenues lors de l'exécution d'une requête permettant d'afficher le nom du département pour la ville dont le numéro insee est `insee='1293'`.

**Question 6** Faites de même avec la requête permettant d'afficher le nom des départements de toutes les villes. Quelles sont les différences observées par rapport à la question 5 ?

### 1.3 Modification du comportement de l'optimiseur

**Question 7** Essayez maintenant la requête de la question 6 mais en forçant l'utilisation de boucles imbriquées (*nested loops* par la directive `/*+ use_nl(table1 table2)*/`) et explicitez le plan d'exécution choisi par l'optimiseur et les statistiques obtenues

## 1.4 Utilisation d'index

**Question 8** Créer un index secondaire sur l'attribut *dep* de la table *ville* : *create index idx\_dep\_ville on ville (dep)*. Ré-exécutez les requêtes des questions 5 et 6 et explicitez le plan d'exécution choisi par l'optimiseur et les statistiques obtenues.

**Question 9** Exécutez la requête suivante et explicitez le plan d'exécution choisi par l'optimiseur et les statistiques obtenues : afficher le nom des villes, de leurs départements et de leurs régions.

**Question 10** Créer un index secondaire sur l'attribut *reg* de la table *departement*. Ré-exécutez la requête précédente et explicitez le plan d'exécution choisi par l'optimiseur et les statistiques obtenues.

**Question 11** Exécutez maintenant la requête suivante : afficher le nom des villes, de leurs départements et de la région pour la région dont le numéro (*id*) est 91. Explicitez le plan d'exécution choisi par l'optimiseur et les statistiques obtenues.

**Question 12** Exécutez maintenant la requête suivante : afficher le nom des villes dont le numéro de département (*dep*) commence par '7'. Explicitez le plan d'exécution choisi par l'optimiseur et les statistiques obtenues. Qu'en est-il de l'utilisation de l'index secondaire ?

## 1.5 Les statistiques des tables

**Question 13** Regardez les données disponibles dans la table *USER\_TAB\_COL\_STATISTICS* pour les tables précédentes. Est-ce que les statistiques correspondent bien aux données présentes dans vos tables ?

**Question 14** Demandez maintenant à Oracle de recalculer les statistiques sur les tables précédentes en utilisant la commande suivante :

*exec dbms\_stats.gather\_table\_stats('login','nom\_table');* sur une table

ou *exec dbms\_stats.gather\_schema\_stats('login');* sur votre schéma.

Regardez à nouveau les données disponibles dans la table *USER\_TAB\_COL\_STATISTICS* pour les tables précédentes.

Ces informations correspondent aux données utilisées par Oracle pour évaluer le coût des différents plans d'exécution afin de choisir le moins.

Vous pouvez ré-exécuter certaines requêtes précédentes (question 5, 6, 10, 11, 13 par exemple) afin de voir si l'optimiseur fait d'autres choix d'opérateurs maintenant qu'il connaît les statistiques sur vos données.