

## TP Bases de données avancées : indexation

### **Création d'une relation, indexation, mesure du temps et explication de plans d'exécution (source : Laurent Amsaleg)**

1. Copiez les fichiers du tp sur votre compte.
2. Vérifiez que vous avez bien accès à Oracle en suivant les instructions de connexion (documentation disponible dans les fichiers du tp).

Le fichier scramble.data contient les données qui vont plus tard être insérées dans une relation nommée XXXX. Le schema de cette relation est :

```
nss INTEGER,  
nom VARCHAR(100) NOT NULL,  
dist_10 INTEGER,  
dist_100 INTEGER,  
dist_1000 INTEGER,  
dist_10000 INTEGER,  
random FLOAT
```

3. Sans charger la base dans Oracle, combien de tuples contient cette relation ? (commande Unix `wc -l <nom du fichier>`) ?
4. Il est plus facile de se rendre compte de ce que contient cette relation en faisant, dans une fenêtre plein écran la commande Unix :  

```
$ sort -n scramble.data | more
```
5. L'attribut `dist_10` par exemple prend des valeurs répétées tous les 10 n-uplets. Cela permet de jouer avec la sélectivité d'une requête (i.e. la fraction de la base total qu'elle retourne). De tête, combien de n-uplets seraient rendus si vous exécutiez  

```
$ select * from XXX where dist_10 =1  
$ select * from XXX where dist_1000 =1    ?
```

  
Vous avez donc ici un moyen de jouer synthétiquement sur la sélectivité.

Le but du TP est d'essayer de créer des index sur la relation XXXX et de voir l'effet qu'a leur utilisation sur les performances du système. Le but est aussi d'essayer de voir ce qui est exécuté par la machine. A partir de maintenant, on va charger les données dans Oracle.

### **Création d'une table en mode non indexé (HEAP)**

6. Regardez le scripts "heap\_create\_table.sql". Il crée une table **non indexée** dans Oracle. Exécuter ce script (sqlplus est l'interface de commande avec Oracle) :  

```
$ sqlplus / @heap_create_table.sql
```
7. Chargez les données dans la table XXXX :  

```
$ sqlldr / control=load.sql data=scramble.data
```

## Visualisation du plan d'exécution choisi

8. Lancez l'interpréteur interactif sqlplus

*\$ sqlplus /*

9. Activez le traçage des exécutions :

*SQL> SET AUTOTRACE ON*

10. Lancez la requête de sélection ponctuelle suivante :

*SQL> select \* from XXXX where nss=21954;*

Qu'obtenez vous ? Essayez d'interpréter chaque ligne du résultat (une recherche avec les mots clés ORACLE EXPLAIN sur votre engin de recherche favori sera utile). Notez les performances obtenues.

### **Vous êtes arrivé au checkpoint, bravo ! Prévenir votre chargé de TP !**

11. En utilisant *SET AUTOTRACE TRACEONLY*, vous pouvez-faire l'économie de l'affichage du résultat de la requête.

12. Refaites de même avec la requête de sélection par intervalle suivante :

*SQL> select \* from XXXX where nss>16873 and nss<16973*

13. Faire *drop table XXXX;*

## Création d'une table avec différents index

14. Regardez maintenant le script *hash\_create\_table.sql* et refaites la séquence de tests (questions 7 à 13).

15. Cherchez dans la documentation en ligne ce que crée *hash\_create\_table.sql*.

16. Que penser des temps de réponse comparés a ceux du stockage HEAP ?

### **Vous êtes arrivé au checkpoint ! Prévenir votre chargé de TP !**

17. Supprimez la table et l'Index en faisant

*SQL> drop table XXXX;*

*SQL> drop cluster h\_primaire XXXX;*

18. Refaire les mêmes tests (questions 7 à 13) avec le script *bTree\_create\_table.sql*.

19. Quel est le type d'index utilisé ?

20. Proposez une conclusion générale sur l'indexation.

### **Encore un checkpoint ! Prévenir votre chargé de TP !**

21. Supprimer l'index avec *drop index XXXX\_ix\_nss* et effacez *scramble.data* de votre répertoire.

## Passage à l'échelle (bonus très apprécié !)

22. Trouvez une taille de base de données permettant de provoquer des lectures en mémoire externe (physical reads). Une possibilité est de faire un script de création de données, où d'utiliser une requête *INSERT INTO ... SELECT* qui double la taille de XXXX à chaque fois (pas complètement évident à trouver).

23. Tracez la courbe du nombre de lectures physiques en fonction de la taille de la base. Comparez les différents index sur les requêtes de sélection ponctuelle et d'intervalle.

### **Vous en avez trop fait. Prévenir votre chargé de TP !**