

Exercice 1 : Sélection

Soit le schéma Film(IdFilm: , Titre, Année, Genre, Résumé, IdMES, CodePays).

Question 1 : On suppose qu'un index de type 2 trié, B+ groupant et dense de clé de recherche Année a été créé sur Film. Expliquez comment le SGBD peut évaluer la requête

```
1 SELECT count(*) FROM Film GROUP BY Année ;
```

Question 2 : La même question lorsque l'index est non-groupant ?

Question 3 : Quel index doit être créé pour pouvoir évaluer la requête ci-dessous par une consultation de l'index seulement.

```
1 SELECT Année count(*) AS NbFilm FROM Film WHERE Année > 1984 AND année < 1999
   GROUP BY Année ;
```

Question 4 : Discuter de l'évaluation de la requête ci-dessous en supposant qu'il existe deux index sur Film, tous les deux ont des entrées de type 2, l'une des clef de recherche est Année et l'autre Genre.

```
1 SELECT Titre FROM Film WHERE Année=2017 AND Genre='Comédie '
```

Exercice 2 : Choix de l'index

Dans cet exercice on considère la relation suivante :

Restaurants(rid : **integer**, nom : **string**, note_moyenne : **float**)

Les données de la relation sont stockées dans un fichier sous forme d'un tas. On suppose qu'une page peut contenir 8000 octets de données, les entiers et les flottants occupent un espace 4 octets chacun, une chaîne de caractère est limitée à 192 octets et un pointeur (ou rid) occupe 6 octets. La relation contient 400 000 d'enregistrements avec des valeurs comprises entre 0 à 399 999. De plus on supposera que les notes sont uniformément réparties de 1 à 20.

Question 1 : Combien d'enregistrement peut contenir une page ?

Question 2 : Estimer la sélectivité de :

1. $\sigma_{restaurant.rid < 50.000}(Restaurants)$
2. $\sigma_{restaurant.rid = 50.000}(Restaurants)$
3. $\sigma_{restaurant.rid > 50.000 \wedge restaurant.rid < 50.020}(Restaurants)$
4. $\sigma_{restaurant.rid \neq 50.000}(Restaurants)$
5. $\sigma_{restaurant.moyenne = 20}(Restaurants)$
6. $\sigma_{restaurant.moyenne > 15}(Restaurants)$

Question 3 : Estimez le coût maximum (E/S sur les pages) des sélections précédentes pour une sélection utilisant un parcours du fichier

Question 4 : En considérant les index de type 2 suivants :

- Un index groupant B+ de hauteur 3 sur rid, les feuilles contiennent en moyenne 500 entrées
- Un index non-groupant B+ de hauteur 3 sur rid, les feuilles contiennent en moyenne 500 entrées
- Un index haché sur rid
- Un index non-groupant B+ de hauteur 3 sur la note_moyenne, les feuilles contiennent en moyenne 500 entrées

Estimez le coût maximum (E/S) des sélections précédentes pour chaque index

Question 5 : Quel serait l'impact sur le coût d'exécution des requêtes précédentes si les moyennes étaient réparties de la manière suivante :

- 5% des moyennes sont entre 0 et 5
- 90 % des moyennes entre 5 et 15
- 5 % des moyennes entre 15 et 20

On suppose aussi que sur chacun des trois intervalles les moyennes sont uniformément réparties.