



Faculté : Sciences exactes
Département : Informatique

كلية : العلوم الدقيقة
قسم : الإعلام الآلي

مسابقة الدخول لدكتوراه الطور الثالث، ل م د 2021/2020

Concours d'accès au doctorat 3^e cycle, LMD 2020/2021

Spécialité : Systèmes d'Information Avancées الاختصاص :

Variante : 01 الخيار رقم :

Epreuve : Distribution de données اختبار :

Durée : ساعتان المدة : Coefficient : 02 المعامل :
Date : 06/03/2021 التاريخ : Heure : 15:00 التوقيت :

Exercice 01 : (4 points)

On considère la table suivante Etudiant :

| NumInscription | Note | Université |
|----------------|------|-------------|
| E01 | 15 | El Oued |
| E03 | 7 | Constantine |
| E04 | 8 | Biskra |
| E02 | 16 | Biskra |
| E05 | 9 | El Oued |
| E06 | 18 | El Oued |
| E07 | 19 | Constantine |

Questions :

- Proposer une bonne **fragmentation horizontale** de la table **Etudiant** selon la condition : chaque fragment (table) contient uniquement les informations des étudiants d'une seule université.
- Donner les tables résultantes de cette fragmentation.
- Montrer que la fragmentation est correcte.

Exercice 02 : (10 points)

Soit le schéma global

Artiste (A, nom, prénom, nation) un artiste est identifié par son numéro A
l'artiste numéro A joue le rôle *nom_rôle* dans le film
Rôle (nom_rôle, F, A) numéro F.

Les données sont fragmentées de la manière suivante :

$\text{Artiste}_i = \sigma_{pi}(\text{Artiste})$ où pi est de la forme $nation = N$ avec $N \in \{ 'D', 'F', 'US' \}$

$\text{Rôle}_i = (\text{Rôle} \bowtie_A \text{Artiste}_i)$

Les données sont réparties sur 4 sites dans 3 pays. Chaque pays contient les données en lien avec les artistes citoyens du pays :

Berlin (**B**) contient Artiste1 et Rôle1.

Paris (**P**) contient Artiste2 et Rôle2.

New York (**NY**) et Los Angeles (**LA**) contiennent les mêmes données : Artiste3 et Rôle3.

Il y a 20 rôles par acteur, 10 rôles par film. Il y a 100 artistes allemands, 100 artistes français et 1000 artistes américains. On considère que tous les artistes sont acteurs. Un acteur n'a jamais 2 rôles dans le même film.

Il n'y a pas d'homonyme : $\text{card}(\Pi_{\text{nom}}(\text{Artiste})) = \text{card}(\text{Artiste})$

Le coût d'un traitement local sur un site est négligeable par rapport au coût d'un transfert intersite. Le coût de transfert dépend de la distance entre les sites. La fonction $\text{tr}(T, X, Y)$ donne le coût pour transférer le résultat de l'expression algébrique T depuis le site X vers le site Y . On a :

$\text{tr}(T, \text{NY}, P) = \text{card}(T)$

$\text{tr}(T, \text{LA}, \text{NY}) = 2 * \text{card}(T)$

$\text{tr}(T, \text{LA}, P) = 3 * \text{card}(T)$

$\text{tr}(T, X, Y) = \text{tr}(T, Y, X)$

Q1)

a) Quelles sont les cardinalités de Artiste, Rôle, $\Pi_F(\text{Rôle})$ $\sigma_{\text{nation} = 'US'}(\text{Artiste})$?

b) Que vaut le facteur de sélectivité $\text{SF}(\sigma_{\text{nation} = 'US'}(\text{Artiste}))$?

Q2)

a) Requête **R1** posée à Los Angeles (LA) : Donner le nom des acteurs français qui ont joué avec l'acteur américain nommé Pitt (*i.e.*, dans le même film que lui).

select a2.nom

from Artiste a1, Artiste a2, Role r1, Role r2

where a1.nom = 'Pitt' **and** a1.nationalité = 'US' **and** a2.nationalité = 'F'

and a1.a = r1.a **and** r1.f = r2.f **and** r2.a = a2.a

b) La requête **R2** est posée à Paris (P) : Donner le numéro des films dans lesquels jouent au moins un acteur français et au moins un acteur américain. Quel est le plan optimal pour R2 ?

Exercice 03 : (6points)

Une application au site A veut calculer la jointure de (Etudiant \bowtie Transcription) tel que :

Etudiant(EtudId, Spécialité) au site B

Transcript(EtudId, CrsCode) au site C.

Le résultat doit être retourné au site B

Supposons que les différentes longueurs des attributs sont:

– EtudID : 8 octets;

– Spécialité : 3 octets;

– CrsCode : 6 octets.

• Etudiants a 15,000 tuples.

• 6,000 étudiants sont enregistrés à un cours au moins. Au moyenne chaque étudiant est enregistré dans 5 cours.

Quel est le plan d'exécution optimisé de la requête de jointure (Etudiant \bowtie Transcription)?