Université Constantine 2-Abdelhamid Mehri

Faculté des NTIC Département TLSI Licence : L3 GL

UE: Techniques Avancées des Bases de Données(TABD)

TD3

Optimisation des requêtes

Objectifs:

- Manipuler les opérateurs algébriques,
- Exprimer des requêtes dans le langage algébrique,
- Utiliser les arbres d'expression pour optimiser les requêtes,
- Calculer le coût d'une requête.

Exercice 1:

Soit le schéma de base de données suivant :

Employe(<u>numero</u>, nom, age, salaraire)

Superviser(responsable, employe)

Superviser[responsable] ⊆ Employes[numero]

 $Superviser[employe] \subseteq Employes[numero]$

Questions:

- 1. que signifie le fait que **employe** soit une clé de superviser
- 2. Est-ce que ce schéma permet d'avoir un employé n'ayant pas de responsable?
- 3. Si l'attribut responsable devient clé de la relation Superviser, comment sera notre domaine de valeurs?
- 4. Comment sera notre domaine de valeurs si le couple (responsable, employe) devient clé de la relation superviser?
- 5. Ecrire l'expression algébrique permettant de retrouver les noms et les salaires de tous les responsables ayant un employé avec un salaire > 100.

Exercice 2:

Soit le schéma de base de données contenant deux relations : R(A, B) et S(B, C), telle que toutes les valeurs sont des entiers ; ne pas faire d'hypothèse sur les clés.

Considérons les trois expressions algébriques suivantes :

- a) $\pi_{A,C}(R \bowtie \sigma_{B=1}(S))$
- b) $\pi_A(\sigma_{B=1}(R)) \times \pi_C(\sigma_{B=1}(S))$
- c) $\pi_{A,C}(\pi_A(R) \times \sigma_{B=1}(S))$

Deux expressions parmi les trois sont équivalentes (produisent le même résultat sur toutes les bases de données), alors que la troisième donne un résultat différent.

Questions:

- 1. Quelle expression algébrique donne un résultat différent ?
- 2. Donner une instance de la base de données permettant de donner un résultat différent.

Exercice 3

Soit le schéma de base de données suivant :

- Employe(NSS, Nom, Prenom, DN, Adresse, Sexe, Salaire, SuperNSS, NumDep)
- Departement(<u>NumD</u>, NomD, , NSSResp, DateEntreeResp)
- EmplacementDept(<u>NumD, Emplacement</u>)
- Projet(NProjet, NomProjet, Emplacement, NumD)
- TravailleSur(<u>TNSS</u>, <u>NumProjet</u>, Heures)

Soient les deux requêtes Q1, Q2:

Q1: SELECT Nom

FROM Employe, TravailleSur, Projet

WHERE NomProjet = 'Tramway' AND NProjet = NumProjet

AND TNSS = NSS AND DN > '14-01-1968'

Q2: SELECT E1.Nom, Projet.NomProjet

FROM Employe E1, Employe E2, Departement, TravailleSur, Projet

WHERE E1.SuperNSS = E2.NSS AND Projet = TravailSur.Numprojet

AND TravailleSur.TNSS = E1.NSS AND E2.NSS = Departement.NSSRes
AND E2.Sexe = 'H' AND E2.Salaire > 10000

AND TravailleSur.Heures > 20

Ouestions:

- 1. Quel est le résultat de chaque requête ?
- 2. Traduire chaque requête en une expression de l'algèbre relationnelle
- 3. Optimiser la requête en appliquant les règles de transformation

Exercice 4:

Soit une relation R(A, B) contenant r tuples, et une relation S(B, C) contenant s tuples; on suppose que r, s > 0. Ne pas faire d'hypothèse sur les clés.

Pour chacune des expressions algébriques ci-dessous, donner le nombre minimal et maximal de tuples du résultat en fonction de *r* et *s*.

- a) $\pi_{A,C}(R \bowtie S)$
- b) $\pi_B R (\pi_B R \pi_B S)$
- c) $(R \bowtie S) \bowtie R$
- d) $\sigma_{A>B}R \cup \sigma_{A< B}R$

Exercice 5:

On considère trois relations $r_1(A, B, C)$, $r_2(C, D, E)$, et $r_3(E, F)$) avec A, C, et E leurs clés primaires respectivement. Supposons que r1 comporte 1000 tuples, r2 comporte 1500 tuples et r3 comporte 750 tuples.

Question:

- 1. Donner une estimation de la taille de $r_1 \bowtie r_2 \bowtie r_3$
- 2. donner une stratégie efficace pour calculer la jointure.

Exercice 6:

Soit le schéma relationnel suivant :

Etudiants (Eid, nom, age, adresse)

Livre (Lid, titre, auteur)

Pret (Eid, Lid, date_Pret)

Soit la requête SQL suivante :

SELECT E.nom

FROM Etudiant E, Livre L, Pret P

WHERE E.Eid = P.Eid AND L.Lid = P.Lid AND L.auteur = 'Agatha Christie' AND E.age > 12 AND E.age < 20

- 1. Que fait cette requête?
- 2. Ecrire l'expression algébrique correspondant exactement à cette requête.
- 3. Donner un plan logique optimisé de cette requête.
- 4. On considère maintenant les données suivantes sur les tables:
 - La table Etudiant comporte 10000 enregistrements stockés sur 1000 pages
 - La table Livre comporte 50000 enregistrements stockés sur 5000 pages
 - La table Prêt comporte 300000 enregistrements stockés sur 15000 pages

Question : En supposant qu'il n'existe aucun indexe et que les tables ne sont pas triées, donner un plan physique (préciser l'algorithme à utiliser pour chaque opération et la cardinalité de chaque table de jointure).

- 5. Afin d'optimiser plus la requête, proposer deux indexes et un plan physique optimal exploitant ces indexes.
- 6. Ecrire la requête SQL qui renvoie le nombre de prêts effectués par chaque étudiant.

Pour exécuter la requête précédente, doit-on utiliser de la matérialisation des résultats ou du pipeline ? Justifier.

Exercice 7:

On considère le schéma de base de données suivant :

Produit (Pid, Pnom, Prix)

Entreprise (Eid, Enom, Adresse, Pid#)

Les caractéristiques de cette BD sont comme suit :

- Cardinalité(Produit) = 20000 tuples, Taille(enregistrement) = 20 octets
- Cardinalité(Entreprise) = 1000 tuples, Taille(enregistrement) = 25 octets
- Taille(page) = 5000 octets
- Le tampon(buffer) mémoire réservé à cette base de données comporte 102 frames (un frame peut contenir une page)
- 1. Calculer le nombre de pages des tables Produit et Entreprise
- 2. Soit la requête

SELECT Pnom, Enom

FROM Produit, Entreprise

WHERE Produit.Pid = Entreprise.Pid AND Adresse = 'Constantine'