Entrepots de Données et Big Data : Optimisation de requête - partie 1

TD/TP en 2 parties (18/09 et 25/09) à faire en binômes ou trinômes. Rendu facultatif avant le 01/10 : un seul document pdf par groupe (attention : soigné, clair, et synthétique - max 5MB) sera à déposer dans l'espace Moodle dédié au cours.

1 Cout de plans d'exécution logiques

Soit le modèle relationnel composé des relations suivantes :
ETUDIANTS(IDE, NOM, AGE) – la relation contenant tous les étudiants
MODULES(IDM, RESPONSABLE, INTITULE) – la relation contenant tous les modules
IP(#IDE, #IDM) – la relation contenant la liste des inscripions pédagogiques (inscription d'un étudiants à un module)
FORMATION(IDF, NOMF) – la relation contenant toutes les formations
IA ((#IDE, #IDF) – la relation contenant la liste des inscripions administratives (inscription d'un étudiants à une formation)

Nous supposons qu'il y ait 200 étudiants, 70 modules, 4200 IP (inscriptions pédagogiques), 50 formations et 70 IA (des étudiants peuvent être inscrits à plusieurs formations).

Nous souhaitons éxécuter la requête suivante :

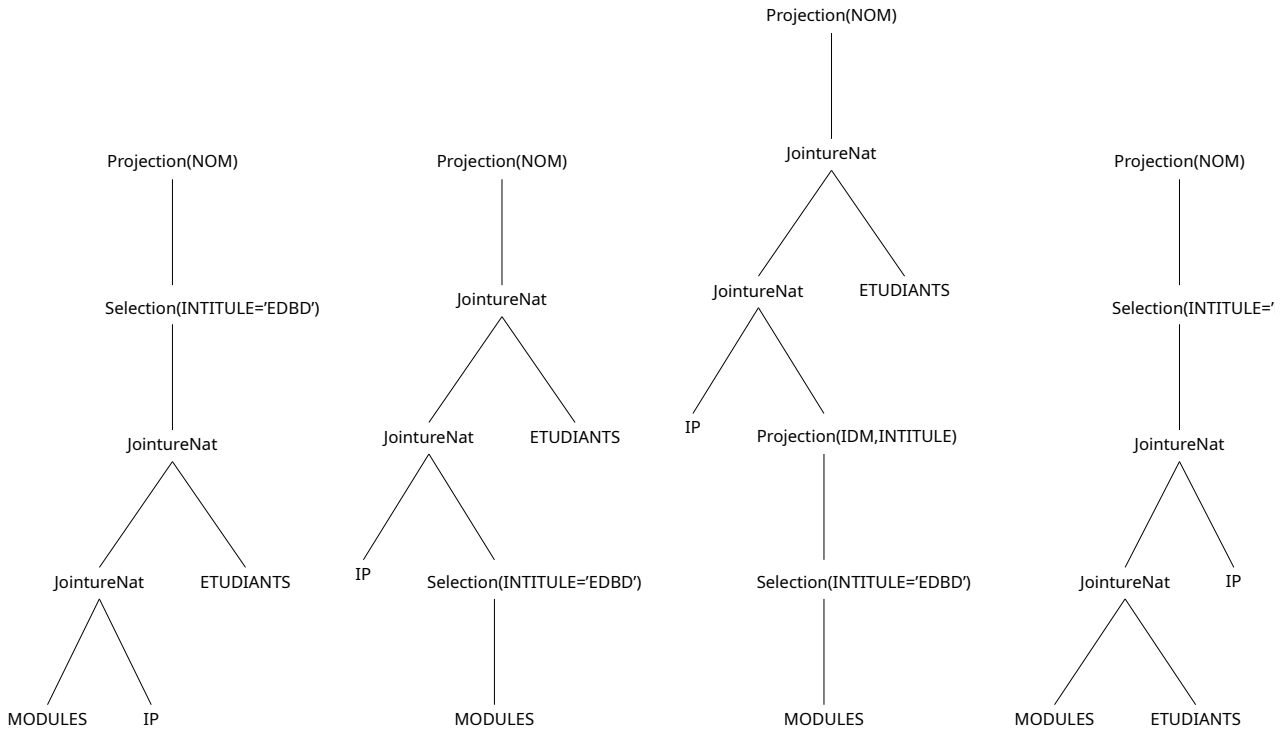
```
SELECT NOM
FROM ETUDIANTS E ,MODULES M ,IP I
WHERE E.IDE = I.IDE AND M.IDM=I.IDM
AND INTITULE = "EDBD" ;
```

Question 1 Que permet d'obtenir la requête ci-dessus ?

Pour cette requête, nous proposons 4 plans d'exécution logiques représentés par des arbres algébriques ci-dessous.

Question 2 Pour chaque plan d'exécution logique, calculer le nombre de lignes intermédiaires créées.

Question 3 Quel est le plan d'exécution logique optimal ? Pourquoi ?



2 Définition de plans d'exécution logiques

Question Pour chacune des requêtes ci-dessous :

- indiquer ce qu'elles permettent d'obtenir,
- donner différents plans d'exécution logique,
- indiquer le plan optimal.

Requete 1 :

```
SELECT RESPONSABLE
FROM ETUDIANTS E ,MODULES M ,IP I
WHERE E.IDE = I.IDE AND M.IDM=I.IDM
AND NOM = "DUPOND" AND INTULE LIKE "HMIN
```

Requete 2 :

```
SELECT NOM
FROM ETUDIANTS E ,FORMATION F ,IA A, IP I, MODULE M
WHERE E.IDE = A.IDE AND F.IDF=A.IDF AND I.IDE=E.IDE AND M.IDM=I.IDM,
AND NOMF = "MASTER AIGLE" AND INTULE = "EDBD";
```

Requete 3 :

```
SELECT INTILULE
FROM ETUDIANTS E ,MODULES M ,IP I
WHERE E.IDE = I.IDE AND M.IDM=I.IDM
AND AGE = (select min (AGE) from ETUDIANTS);
```

3 Réécriture de plans d'exécution logiques

Soit le schéma relationnel suivatn :
JOURNALISTE (IDJ, NOM, PRENOM) – La relation contenant tous les journalistes
JOURNAL (TITRE, REDACTION, #REDACTEUR_ID) – La relation contenant tous les journaux rédigés par des journalistes

On considère la requête suivante :

```
SELECT NOM
FROM JOURNAL, JOURNALISTE
WHERE TITRE='Le Monde' AND IDJ=IDJOURNALISTE AND PRENOM='Jean' ;
```

Voici deux expressions algébriques :

$$\pi_{nom}(\sigma_{titre='Le\ Monde' \wedge prenom='Jean'}(Journaliste \bowtie_{jid=redacteur\_id} Journal))$$

et

$$\pi_{nom}(\sigma_{prenom='Jean'}(Journaliste) \bowtie_{jid=redacteur\_id} \sigma_{titre='Le\ Monde'}(Journal))$$

Question 1 Les deux expressions retournent-elles le même résultat (sont-elles équivalentes) ? Justifiez votre réponse en indiquant les règles de réécriture que l'on peut appliquer.

Question 2 Une expression vous semble-t-elle meilleure que l'autre si on les considère comme des plans d'exécution ?

4 Tous les plans d'exécution logiques

Soit le modèle relationnel suivant :
ACTEUR (idA, nom, prenom, nationalite) – la relation contenant tous les acteurs
FILM (idF, titre, annee, nbspectateurs, #idRealisateurs, #idGenre) – la relation contenant tous les films
JOUER (#idActeur,#idFilm, salaire) – la relation contenant la listes des acteurs et des films dans lesquels ils jouent
REALISATEUR (idR, nom, prenom, nationalite) – la relation contenant tous les réalisateurs
GENRE (idG, description) – la relation contenant tous les genres des films (horeur, comédie, ...)

Soit la requête suivante :

```
SELECT acteur.nom,acteur.prenom
FROM acteur,jouer,film,genre, realisateur
WHERE (idA=idActeur) AND (idFilm=idF) AND (idGenre=idG) AND (idRealisateur=idR)
AND (nationalite='France') AND (description='comédie') AND (realisateur.nom = "Les frères Coen") ;
```

Question 1 Pour la requête ci-dessus, donner TOUS les plans d'exécution logique.

Question 2 Parmi les plan d'exécution , quel est le plan optimal ? Justifier votre choix.