

## Exercice n°1 : SQL et plan d'exécution

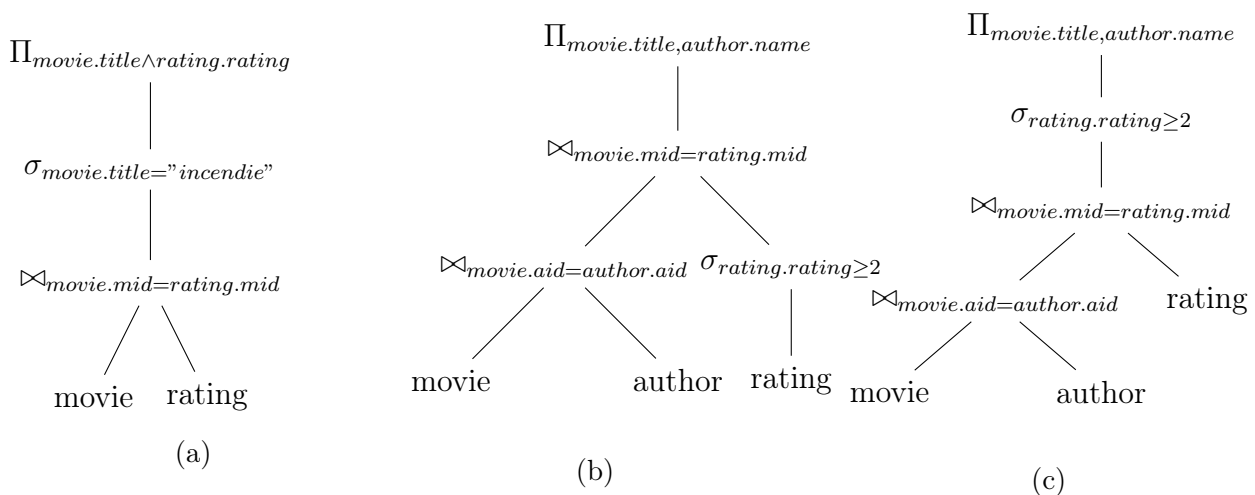


FIGURE 1 – Trois arbres de requête algébrique

**Question n°1 :** Proposez pour les trois arbres de requêtes algébriques (figure 1) une requête SQL possible.

**Question n°2 :** Pour les trois arbres de la figure 1, dites si oui ou non des matérialisations sont nécessaires. On rappelle que toutes les boucles externes de jointure correspondent au sous arbre gauche (si jointure par itération).

**Question n°3 :** Dessinez un arbre de requête algébrique pour :

**SELECT** movie.title, rating.rating **FROM** movie **INNER JOIN** rating **ON** movie.mid=rating.mid **WHERE** movie.year = 2000

## Exercice n°2 : Estimation du coût d'un plan

On considère dans cet exercice une base de données contenant trois relations :

- library (lid, name, location)
- book (bid, title, author, publisher, publication\_year )
- stock (bid, lid, quantity) (bid et lid sont des clefs étrangères)

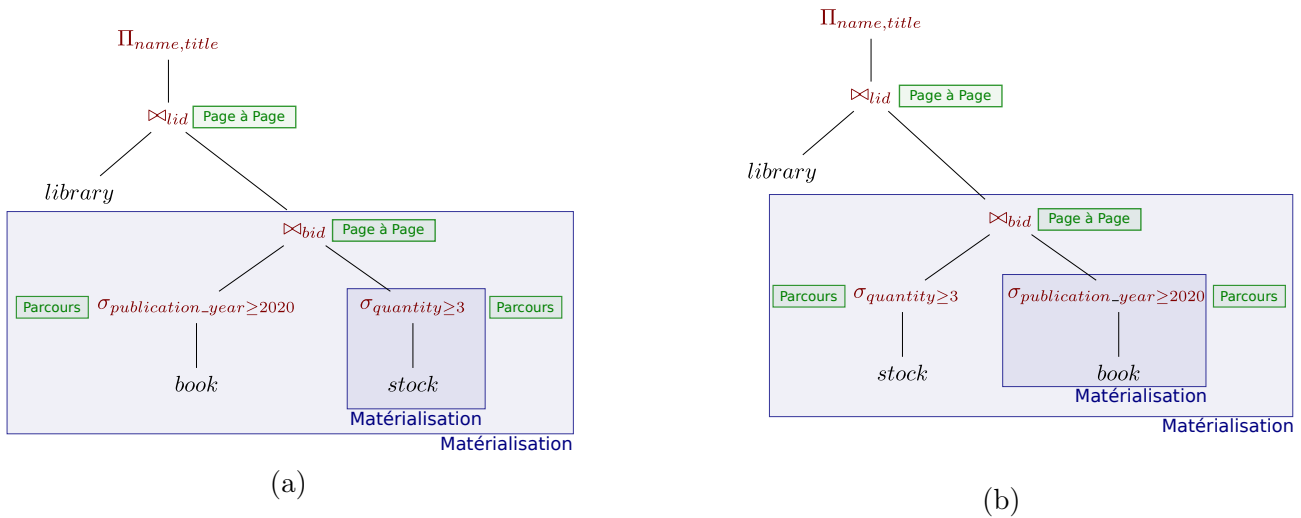
La base de données contient 250.000 livres, 55 bibliothèques et la relation **stock** contient 1.000.000 d'entrées. Une page peut contenir 250 livres, 500 bibliothèques et 1000 entrées de la relation stock. Il existe un index B+ de type 2 sur les clefs primaires de la table. On dispose aussi d'un index B+ (de type 2) pour les champs *publication\_year* et *quantity*, chaque page de l'index peut contenir (pour les deux index) 2000 entrées et la profondeur des arbres vaut respectivement 2 et 3.

**Question n°1 :** On cherche les bibliothèques et le nom des livres tel que la bibliothèque contient au moins 3 exemplaires où le livre a été publié après 2020. La requête qui peut s'exprimer en SQL comme

**SELECT** name, title **FROM** library **INNER JOIN** stock **ON** library.lid=stock.lid **INNER JOIN** book **ON** book.bid=stock.bid **WHERE** publication\_year  $\geq$  2020 **AND** quantity  $\geq$  3

Expliquez quelles sont les données dont vous auriez besoins pour calculer le nombre d'E/S du plan d'exécution.

**Question n°2 :** En considérant que l'évaluation de  $c_1 = \sigma_{publication\_year \geq 2020}$  produit 5000 enregistrements,  $c_2 = \sigma_{quantity \geq 3}$  500 enregistrements et  $c_1 \wedge c_2$  100 enregistrements. De plus en pipelining la jointure page à page attend de remplir une cellule du buffer/cache (l'équivalent d'une page) pour chaque itération. Calculez et comparez le temps d'exécution des plans suivants :



**Question n°3 :** Si maintenant on matérialise tous les résultats des conditions, quel est le coût en terme d'E/S des deux plans de la question précédente ?

**Question n°4 :** Même question que la précédente, mais cette fois-ci le cache disponible pour la requêtes contient 100 cellules et l'on considère donc des jointures par paquets de pages.

**Question n°5 :** Dans les deux cas précédents, deux matérialisations sont nécessaires, proposez un plan d'exécution ne nécessitant qu'une seule matérialisation en gardant les méthodes d'évaluation des opérateurs de la question 3. Calculez le nombre d'E/S si 3 cellules sont disponibles dans le cache.

**Question n°6 :** On se propose pour le plan d'exécution précédent d'utiliser la sélection via les index B+ disponibles. Donner le coût minimum possible en utilisant ou non ces index (les fichiers de données ne sont pas triés selon ces index).