

M1ISD, Examen: Entrepôts de données

25 octobre 2024

Nom :

Prénom :

Répondre sur cette feuille d'énoncé pour l'exercice 2. Barème donné à titre indicatif, susceptible de modifications. **Aucun document n'est autorisé.** Durée de l'examen : **2 heures**.

Exercice 1 (Généralités sur les entrepôts) (6 pts)

Répondez en quelques lignes :

1. Que signifie ETL (ETC en français) ? Expliciter le sigle, et expliquer de quoi il s'agit.
2. Expliquer pourquoi l'on utilise des entrepôts de données dédiés au lieu de travailler directement sur les bases de données opérationnelles. Plusieurs raisons sont attendues dans l'explication.
3. Qu'est-ce qu'un index de jointure bitmap et comment améliore-t-il les performances des requêtes dans un schéma en étoile ?

Exercice 2 (Questionnaire vrai/faux) (2 pts)

+0,5/-0,25 par bonne/mauvaise réponse (inutile de justifier). La note globale de l'exercice ne descendra pas sous 0 : s'il y a 2 blancs, 3 mauvaises réponses, et 4 bonnes cela fait $\max(0, 2 \times 0 - 3 \times 0.25 + 4 \times 0.5) = \max(0, 0 - 0.75 + 2) = 1.25$ points.

1. ✓ Vrai ✗ Faux Dans un schéma en flocon, toutes les tables de dimensions sont directement reliées à la table de fait.
2. ✓ Vrai ✗ Faux Dans un schéma en étoile, les tables de dimension sont généralement dénormalisées (ne sont pas en 3NF).
3. ✓ Vrai ✗ Faux Dans un entrepôt de données, il ne peut y avoir qu'une seule table de fait.
4. ✓ Vrai ✗ Faux Un arbre B peut traiter efficacement les requêtes à dimensions multiples, telles que les requêtes d'étendue sur deux colonnes ou plus.

Exercice 3 (SQL : extensions group by) (2 pts)

Réécrire les requêtes suivantes en utilisant un GROUPING SETS plutôt que les extensions ci-dessous :

1. `SELECT a, b, c, sum(d) FROM t GROUP BY ROLLUP(a,b,c);`

2. `SELECT c, b, max(d) FROM t GROUP BY d, c, CUBE(b,a);`

Exercice 4 (SQL : OVER) (2 pts)

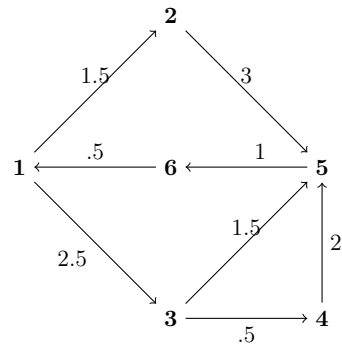
Pourquoi la requête à droite (quoique correcte) sur la table `t(a int, b text primary key, c int)` est-elle inutilement complexe ? Donner une requête un peu plus simple (mais toujours sans sous-requête) calculant le même résultat.

```
SELECT a,
       SUM(SUM(c)) OVER(Partition by a) s1,
       MAX(MAX(b)) OVER () s2
FROM t
GROUP BY a;
```

Exercise 5 (SQL : CTE) (2 pts)

Une instance possible de la table *t* :

p1	p2	d
1	2	1.5
2	5	3
1	3	2.5
3	4	0.5
3	5	1.5
4	5	2
5	6	1
6	1	0.5



- Donner le résultat de la requête Q1 : expliquer intuitivement ce qui est calculé, et donner le résultat sur l'instance ci-dessus.
- Expliquer intuitivement la différence entre les deux requêtes Q1 et Q2.

```

-- Requête Q1
WITH RECURSIVE temp1 (s1, s2, sd) AS (
  SELECT p1,p2,d FROM t
    WHERE p1=5 or p2=6
  UNION ALL
  SELECT p1, s2, d+sd
    FROM t JOIN temp1 ON t.p2=temp1.s1
    WHERE d+sd<7 AND p1 != 4
)
SELECT * FROM temp1;

```

```

-- Requête Q2
WITH RECURSIVE temp2 (s1, s2, sd) AS (
  SELECT p1,p2,d FROM t
    WHERE p1=5 or p2=6
  UNION ALL
  SELECT p1, s2, d+sd
    FROM t JOIN temp2 ON t.p2=temp2.s1
    WHERE d+sd<7
)
SELECT * FROM temp2 WHERE s1 != 4;

```

Exercise 6 (Schéma ROLAP) (6 pts)

Vous travaillez dans le département BI de FastEat, une entreprise internationale de livraison de repas. Votre objectif est de concevoir un schéma ROLAP pour analyser les données de commande, de livraison, et de performance des restaurants partenaires.

Contexte :

FastEat permet à ses utilisateurs de commander des repas auprès de restaurants partenaires et de se faire livrer à domicile. L'entreprise cherche à améliorer la qualité de ses services en analysant plusieurs aspects de ses opérations : performance des restaurants, efficacité des livreurs, et satisfaction des clients.

Problématique :

Votre mission est de modéliser un schéma ROLAP qui permettra à FastEat d'analyser les tendances suivantes :

- Quels sont les restaurants les plus performants par région et par type de cuisine ?
- Quelles sont les périodes (heures/jours) avec les pics de commandes pour les restaurants par catégorie ?
- Quels sont les délais moyens de livraison par zone géographique et par type de commande (sur place, à emporter, livraison standard ou express) ?
- Comment la satisfaction des clients évolue en fonction des temps de livraison et des restaurants ?

Tâches :

- Identifier et modéliser les faits et dimensions clés qui vous permettront d'analyser les performances des restaurants, des livreurs, et des clients. **2 pts**
- Proposer un schéma ROLAP. **2 pts**
- Justifier vos choix de schéma et de mesures, et organiser vos tables de faits et dimensions en conséquence. **2 pt**