



**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Jendouba**

Modélisation Modélisation multidimensionnelle et Entrepôt de Données

TD2 — OLAP : opérations et Analyse Multidimensionnelle

Objectifs

- Manipuler les opérations OLAP : roll-up, drill-down, slice, dice, pivot.
- Vérifier la cohérence des agrégations sur un cube simple.
- Pratiquer les fonctions fenêtrées et GROUP BY ROLLUP.

Prérequis rapides

- SQL : **GROUP BY, ROLLUP**, fonctions fenêtre (**LAG, RANK, PERCENT_RANK**).
- Compréhension : hiérarchies Date et Produit, mesure additive **montant**.

Cube de départ (conceptuel)

Dimensions : Date (jour → mois → année), Produit (prod → catégorie), Magasin.

Mesures : **montant, quantite**.

Dataset fourni (échantillon) : **fact_ventes, dim_date, dim_produit, dim_magasin** (via DDL/CSV dans le dossier du TD).

Travail demandé

- Requêtes SQL (à exécuter sur le dataset fourni) :
 - CA mensuel par produit et magasin (ROLLUP jour→mois), classé par CA décroissant.
 - Top 5 produits par région et par mois (**DENSE_RANK** ou **ROW_NUMBER**).
 - Variation MoM et YoY du CA par catégorie (**LAG** sur mois/année).
 - Pivot Produits en lignes, Magasins en colonnes, CA en valeur (PIVOT ou agrégation + CASE).
- Opérations conceptuelles :
 - Slice sur Magasin = M01 ; Dice sur Catégorie ∈ {C1, C2} et Mois ∈ {01,02}.
 - Expliquer roll-up vs drill-down sur la dimension Date.
- Contrôles de cohérence :
 - Vérifier que CA total (niveau année) = somme des CA mensuels.
 - Vérifier que CA magasin = somme des CA produits (même niveau).

Attendus (correction synthétique)

- Requêtes incluant ROLLUP/CTE + fonctions fenêtre (**LAG, DENSE_RANK**).
- Slice/dice correctement définis ; pivot produit x magasin.
- Contrôles d'égalité des totaux par niveau.

Déroulé (1h30)

- 10 min : rappel cube, dimensions/hierarchies, mesures additives.
- 30 min : requêtes d'agrégation (mois/produit/magasin) + top-N.
- 25 min : variations temporelles (LAG) + pivot.
- 15 min : contrôles de cohérence (totaux, égalités par niveau) + discussion.
- 10 min : restitution et justification (choix des partitions, niveaux, checks).

Questions de qualité (scénarios)

1. **Qualité du cube** : le grain de **fact_ventes** est-il constant (1 ligne = 1 vente produit-magasin-date) ? Quelles erreurs si le grain varie ?
2. **Qualité des agrégations** : **SUM(montant)** est-elle toujours valide ? Citer un exemple de mesure non additive.
3. **Qualité des requêtes** : comment vérifier que le **GROUP BY** ne double-compte pas (jointures) ? Proposer un test.
4. **Cohérence** : montrer un contrôle : total annuel = somme des totaux mensuels (même périmètre).
5. **Performance** : proposer 2 index utiles (et pourquoi) pour accélérer les requêtes (dimension temps, clés FK, colonnes groupées).

Diagramme des opérations

```
flowchart LR
A[Cube Ventes] --> B[Roll-up]
A --> C[Drill-down]
A --> D[Slice]
A --> E[Dice]
A --> F[Pivot]
```

Pour aller plus loin

- Calculer top 5 produits par mois et par magasin (SQL window ou TOP-N par partition).
- Discuter l'impact des agrégations pré-calculées (MOLAP) vs calculs à la volée (ROLAP).