

## **Examen Final**

Durée : 1h30

### **Questions de cours :**

1. Quelle est la différence entre les approches ROLAP et MOLAP pour la gestion de données dimensionnelles ? **2 pts**

ROLAP qui vient de Relational OLAP emploie les bases de données relationnelles pour réaliser les opérations d'analyse multidimensionnelles (OLAP). Le ROLAP utilise typiquement les vues matérialisées pour accélérer les requêtes nécessitant l'agrégation d'un grand nombre de lignes de la table de faits.

En revanche, MOLAP veut dire Multidimensional OLAP. Ce type de BD repose plutôt sur une structure de cube multidimensionnel (tableau à D dimensions) pour stocker et manipuler les données efficacement.

2. Dans le cas d'un schéma en étoile comportant plusieurs tables de fait, comment sont-elles reliées ? Quel est l'intérêt qu'elles soient reliées ? **2 pts**

Elles sont reliées par les tables de dimensions communes, on appelle ce modèle : le modèle en constellation. Ce dernier se compose donc de plusieurs tables des faits avec leurs tables de dimensions respectives. Les tables de dimensions communes aux différentes tables des faits ne font pas l'objet de redondances : c'est l'un des principaux avantages cette modélisation. Cela permet de réduire l'espace de stockage nécessaire. Idéalement, il faut que les tables de dimensions partagées soient identiques et contiennent les mêmes valeurs, les mêmes attributs.

3. Définir les termes suivants : pré-agrégation et datamart ?

- ✓ La pré-agrégation des faits correspond à pré-calculer des agrégations des données (par exemple : le total de ventes d'une année de manière à accélérer les requêtes analytiques nécessitant de faire ces agrégations).
- ✓ Les Data Marts sont des compartiments du Data Warehouse conçus pour répondre à un besoin métier spécifique. Un Data Mart regroupe et organise toutes les données nécessaires au traitement d'une problématique précise. Par exemple : toutes les données relatives aux ventes, ou bien toutes celles relatives à la logistique, au marketing, à la production, etc.

4. Quels sont les avantages de la modélisation dimensionnelle par rapport au modèle entités-relations dans le contexte analytique ? **1.5 pts**

La modélisation dimensionnelle facilite la compréhension des données par les décideurs. Au lieu d'avoir des informations et des relations divisées en plusieurs tables, les données

sont représentées comme une table de faits représentant le sujet analysé et les dimensions représentant les axes de l'analyse de ce sujet.

5. Combien y a-t-il d'agrégats distincts dans un cube portant sur n dimensions? **1 pt**

$$2^n$$

### Exercice 1: (Modélisation décisionnelle d'un DW)

**Date**(CléDate, Année, Mois, JourDeMois, JourDeSemaine, TrancheHoraire, DrapeauVacances)

**Assuré**(CléAssuré, MoisNaissance, AnnéeNaissance, MoisDécès, AnnéeDécès, Région, Département, District, Ville, Quartier, RevenuAssuré, RevenuFoyer, CatégorieSocioProfessionnel, SousCatégorieSocioProfessionnel, DomaineActivité, CaissePrimaire, CaisseComplémentaire, DrapeauAssuréPrincipal)

**Pratitien**(CléPraticien, Spécialité, SousSpécialité, Région, Département, District, Ville, Quartier, MoisNaissance, AnnéeNaissance, DrapeauConventionné)

**Acte**(CléDate, CléAssuré, CléPraticien, CléPathologie, MontantActes, MontantPriseEnChargeCaissePrimaire, MontantPriseEnChargeCaisseComplémentaire, NombreMedicamentsPrescrits, MontantPharmacologieGenerique, MontantPharmacologieNonGenerique, MontantDesActesComplémentaires, DrapeauActesComplémentairesBiologie, DrapeauActesComplémentairesChirurgie, DrapeauActesComplémentairesKinésithérapie, DrapeauActesComplémentairesRadiologie, NombreDeJoursDArrêtDeTravail, CoutJoursDArret).

**Pathologie**(CléPathologie, DesignationNormalisé, Spécialité, SousSpécialité, TauxDIncapacité, DuréeTraitement, Chronicité, DrapeauMaladieProfessionnelle)

- 1) Dessinez le schéma conceptuel correspondant à cet entrepôt. Donnez quels sont les faits, les dimensions et les mesures. Justifiez votre réponse. **2.5 pts**

Acte est la table de fait(car au centre des dimensions)

Assuré, Date, Pathologie et Partitien sont les dimensions. Chacune de ces tables a une seule clé qui est une clé primaire.

Il faut utiliser un modèle en étoile :

- 2) Donnez la requête SQL qui donne le top 10 des sous spécialités des pathologies qui ont entraîné le plus de dépenses (montant des actes + montant des pharmacologies) ? **1.5 pt**

```
SELECT P.Spécialité, P.SousSpécialité, SUM(MontantActes+ MontantPharmacologieGenerique
+ MontantPharmacologieNonGenerique) AS MontantTotalActes
FROM Pathologie P JOIN Acte A USING (CléPathologie)
GROUP BY P.Spécialité, P.SousSpécialité
SORT BY MontantTotalActes DESC
TOP(10)
```

- 3) Faites une représentation du cube OLAP sans tenir compte des hiérarchies. **1 pt**  
Ce schéma a 4 dimensions, donc pour faire une représentation en cube, il faut fixer une dimension et réaliser un cube de données pour chacune de ses valeurs. Par exemple pour chaque valeur d'assuré, il faut créer un cube de données pour les dimensions Date, Pathologie et Partitien.
- 4) Donnez le nouveau schéma du DW si on décide de diminuer la taille de la table Assuré par une mini-dimension Localisation (qui contient les détails d'adresse de l'assuré). **1 pt**  
Il faut ajouter dans la table Assuré la clé étrangère (n°loc) de la nouvelle mini-table de dimension Localisation et mettre les attributs (Région, District, ville, Quartier) dans la nouvelle table.

### Exercice 2: (OLAP)

- a) Combien de n-uplets ont les résultats des requêtes suivantes (justifiez vos réponses) : **2pts**

**Q1.** Select A, B, C, sum(M) From R Group by A ROLLUP(B, C)

Pour cette requête, on a les agrégats :

- A, B,C (2\*2\*2 tuples)
  - A,B 2\*2 tuples
  - All 1\*2 tuples
- Donc le résultat contient 14 n-uplets

**Q2.** Select A, B, C, sum(M) From R Group by GROUPING SETS((A,B),())

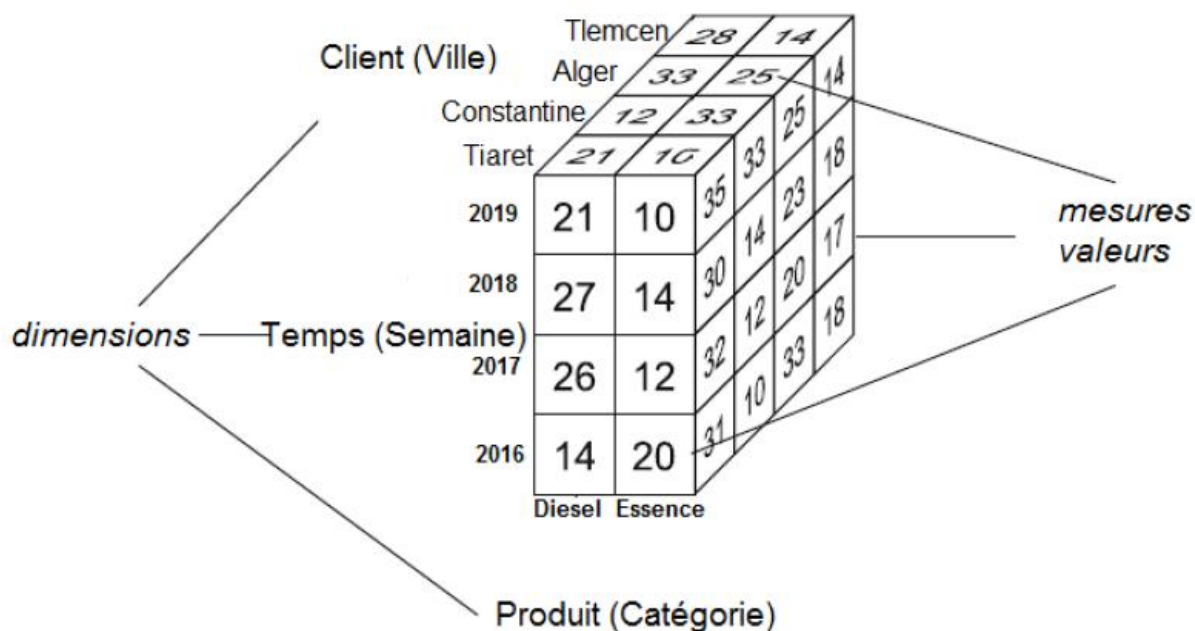
Pour cette requête, on a les agrégats :

- A,B (2\*2 tuples)
  - All 1 tuple
- Donc le résultat contient 5 n-uplets

- b) Si l'on ne dispose pas de l'opérateur ROLLUP, combien faudrait-il de requêtes group by pour **Q1** ? **1pt**

**Il faut 3 requêtes ( pour A,B,C, pour A,B et pour A).**

### Exercice 3: (Cubes de données)



1. L'opération à effectuer pour calculer la dimension client au niveau pays et le résultat de la requête dans ce cas : **1pt**

**Il faut utiliser ROLLUP, le résultat dans ce cas sera :**

2. L'opération à effectuer pour calculer la dimension Temps au niveau mois. **1pt**  
**Dans ce cas, il faut utiliser l'opération Drill down**
3. Expliquez comment afficher le total des ventes par semaine, ville et de type d'énergie de voiture. **1pt**

Il faut utiliser l'emboîtement des valeurs de l'une des dimensions dans une autre, parce qu'on ne peut pas visualiser des tableaux avec plus de 2 dimensions. Par exemple :

		Constantine	Alger	Oran
Sem1	Diesel			
	Essence			
Sem2	Diesel			
	Essence			