

Université Paris Nanterre  
Master 1 – MIAGE  
BI  
EXAMEN (2 Heures) - Avril 2025  
Aucun Document Autorisé

A. Exercice1 : Questions de cours RI (4 points)

- 1- Dans un contexte de systèmes de recherche d'informations, donnez un exemple de cas pratique où le rappel est important à maximiser et un autre exemple de cas pratique où la précision est importante à maximiser, expliquez pourquoi.
- 2- Expliquez la différence de **traitement** d'Elasticsearch d'un champ keyword et un champ texte, donnez des exemples
- 3- Dans une requête de recherche dans Elasticsearch, expliquez la différence entre Query et Filter
- 4- Expliquez la différence entre dimensions conformes et dimensions multi rôles.

Exercice2 : OLAP (5,5 points)

On s'intéresse à l'étude du nombre de succès des étudiants par matière, par école (Centrale ECP, Normale Sup ENS, Polytechnique) et par année (96, 97 et 98).  
Voici le graphique généré par un cube sur un jeu de données :



Donnez le résultat graphique ou en chiffre selon le cas de chacune de ces requêtes. Les données manquantes seront représentées par le symbole '?'.  
1. Un pivot sur les écoles  
2. Un slice sur 1998, un slice sur ENS, un slice relatif à l'année Droit.  
3. Un drill-down sur l'année.  
4. Un roll-up sur toute la période, un roll-up (\*) sur toutes les dimensions.  
5. Un drill-up sur les matières (matières scientifiques (MS) / matières littéraires (ML)).

### Exercice3 : Elasticsearch (6 points) :

Un institut chargé d'analyser l'opinion publique collecte des articles parus dans la presse en ligne, ainsi que les commentaires déposés par les internautes sur ces articles. Ces informations sont stockées dans une base relationnelle, puis mises à disposition des analystes sous la forme de documents JSON dont voici deux exemples.

```
{"index":{"_index": "articles", "_type": "article", "_id": 1}}
{"fields" :
{
  "source": "lemonde.fr",
  "date": "07/02/2017",
  "titre": "Le président Trump décide d'interdire l'entrée de tout
étranger aux USA",
  "contenu": "Suite à un décret paru ....",
  "internaute": "chocho@monsite.com",
  "note": 5,
  "commentaire": "Les décisions du nouveau président sont inquiétantes
..."
}
}
```

```
{"index":{"_index": "articles", "_type": "article", "_id": 2}}
{"fields" :
{
  "source": "nimportequoi.fr",
  "date": "07/02/2017",
  "titre": "La consommation de pétrole aide à lutter contre le
réchauffement",
  "contenu": "Contrairement à ce qu'affirment les media officiels ....",
  "internaute": "alain@dugenou.com",
  "note": 5,
  "commentaire": "Enfin un site qui n'a pas peur de dire la vérité ..."
}
}
```

Supposons que nous sommes sous Elasticsearch, donnez l'expression de la requête pour :

1. Une requête de recherche avec bool/should
2. Une requête de recherche avec Bool/must
3. Une requête de recherche avec must/must not
4. Une requête avec match\_phrase
5. Une requête avec une agrégation
6. Proposez une requête supplémentaire



**Exercice 4 : SQLAnalytique (4,5 points) :**

Soit la table Ventes suivantes :

	SALESID	DATEID	SELLERID	BUYERID	QTY	QTY_SHIPPED
↓	30001	8/2/2003	3	B	10	10
	10001	12/24/2003	1	C	10	10
⌘	10005	12/24/2003	1	A	30	
⌘	40001	1/9/2004	4	A	40	
	10006	1/18/2004	1	C	10	
↺	20001	2/12/2004	2	B	20	20
↻	40005	2/12/2004	4	A	10	10
	20002	2/16/2004	2	C	20	20
⌘	30003	4/18/2004	3	B	15	
↻	30004	4/18/2004	3	B	20	
	30007	9/7/2004	3	C	30	

**Questions :**

Donnez la requête en SQL analytique et **son résultat qui sera affiché** pour les requêtes suivantes :

1- Pour chaque vente, calculez la moyenne mobile des quantités vendues à 3 dates (moyenne des ventes du jour en cours ainsi que les 2 précédentes dates), pour le vendeur 3 (sellerid 3)

2- Pour chaque vente, calculez le maximum et le minimum de quantités vendues sur les 3 dernières dates de commande, pour l'acheteur 'B'.

3- Pour chaque vente, calculez la quantité vendue de la date précédente, ainsi que la quantité vendue de la date suivante pour l'acheteur 'A'.