TD1

### **Exercice 1**

1. **La Business Intelligence** : b) désigne les technologies, infrastructures et applications permettant l'analyse de données pour une meilleure prise de décision.
2. **Synonymes d'un entrepôt de données** : b) Base de données décisionnelles, d) Data Warehouse.
3. **Ensemble de données pour un métier spécifique** : a) Data Mart.
4. **Synonymes d'un Magasin de données** : a) Data Mart, b) Data cube, c) Comptoir de données.
5. **Le cube OLAP** : b) une structure de données qui permet une analyse rapide des données selon les multiples dimensions.
6. **Synonymes de l'ODS** : b) Magasin de données opérationnelles.
7. **L'intégration de données** : c) un processus de combinaison, de consolidation et de fusion de données de plusieurs sources.
8. **KPIs** : c) Key Performance Indicators.
9. **Données du Data Mart fortement dénormalisées** : a) Vrai.
10. **Aide à prendre des décisions stratégiques** : a) Data Warehouse.
11. **But du Data Mart** : c) Métier.
12. **Cube avec plus de trois dimensions** : a) Hyper cube.
13. **Approche de Ralph Kimball** : a) Vraie.
14. **Formats de données des systèmes opérationnels** : a) Très normalisé.
15. **Cohérence et fiabilité des données** : b) L'intégrité des données.
16. **BI embarqué** : c) l'intégration de rapports, tableaux de bord et visualisations dans une application.
17. **Data Warehouse mis à jour en temps réel** : b) Faux.
18. **Propriété incorrecte du Data Warehouse** : b) Volatile.
19. **DSS** : b) Decision Support System.
20. **Synonymes du Système OLTP** : c) Système opérationnel.
21. **Emplacement entre OLTP et Data Warehouse** : a) Staging Area.
22. **Système OLTP** : b) un type d'application informatique qui sert à effectuer des modifications d'informations en temps réel.
23. **But d'un système OLAP** : d) Permettre une analyse multidimensionnelle.
24. **Collection de données pour le support d'une activité opérationnelle** : b) ODS.
25. **Approche de compromis pour organisations décentralisées** : c) architecture data warehouse fédérée.
26. **Opération OLAP pour résumer** : d) Roll up.
27. **Opération OLAP pour accéder aux détails** : e) Drill down.
28. **Systèmes OLTP comparés à OLAP** : b) orientés transaction.
29. **Étapes de la chaîne décisionnelle** : Collecte, Intégration, Analyse, Restitution.
30. **Caractéristiques du Data Mart** : Centré sur un sujet, Agrégé, Facile d’accès, Dédié à un métier.
31. **Méthodes de rapatriement de données vers Staging Area** : ETL, ELT.
32. **Méthodes de chargement sur le Data Warehouse** : Extraction, Transformation, Chargement.

### **Exercice 2**

1. **Décisions de haut niveau** : c) Décisions stratégiques.
2. **Décisions fréquentes et prévisibles** : b) Décisions opérationnelles.
3. **Business Intelligence** : c) Outils pour exploiter des données pour la prise de décisions.
4. **Intégration de données** : b) L'informatique décisionnelle.
5. **Données du Data Warehouse normalisées** : b) Faux.
6. **Nombre de dimensions d'un cube de données** : c) Beaucoup ; limité seulement par des ressources.
7. **Cube avec plus de trois dimensions** : a) Hyper cube.
8. **Rotation de l'hypercube** : b) Pivot.
9. **Sous-cube en choisissant des valeurs spécifiques** : e) Dice.
10. **Outil ETL** : c) Logiciel pour extraire, transformer et charger les données.
11. **Synonymes du Système OLTP** : c) Système opérationnel.
12. **Emplacement entre OLTP et Data Warehouse** : c) Staging Area.
13. **Niveau de détails** : d) Granularité.
14. **Data Mining** : b) Ensemble d'outils pour établir des corrélations.
15. **Analyse ad hoc** : b) Processus du BI pour une question précise.

### **1. Qu'est-ce qu'un Data Lake ?**

* Un Data Lake est un système de stockage qui permet de conserver une grande quantité de données brutes et non structurées ou semi-structurées, de façon très flexible. Ces données peuvent être stockées en l'état, sans nécessiter de transformation ou de structuration préalables, afin de permettre diverses analyses plus tard.

### **2. Avantages du Data Lake :**

* **Scalabilité** : Peut stocker d'énormes volumes de données de différentes structures.
* **Flexibilité** : Permet de conserver les données dans leur format brut.
* **Faible coût** : Le stockage de données dans des Data Lakes peut être plus économique que des systèmes de stockage structurés.
* **Polyvalence** : Les données peuvent être utilisées pour divers cas d'utilisation, y compris l’analyse avancée, l’intelligence artificielle et le machine learning.

### **3. Inconvénients du Data Lake :**

* **Complexité de gestion** : Peut devenir un "marécage de données" (data swamp) si les données ne sont pas bien gérées.
* **Manque de gouvernance** : Sans une gouvernance adéquate, les données peuvent être difficiles à trouver et à utiliser.
* **Sécurité** : Des défis de sécurisation des données peuvent se poser, surtout lorsque de nombreuses sources et types de données sont stockés.

### **4. Enjeux du Data Lake :**

* **Gouvernance des données** : Besoin d'une gestion efficace pour garantir l'intégrité et la qualité des données.
* **Accessibilité** : Permettre l'accès approprié aux données pour les utilisateurs tout en protégeant les informations sensibles.
* **Interrogation efficace** : Développer des moyens efficaces d'interroger des données massives.
* **Conformité** : Assurer la conformité avec les réglementations sur les données (comme le RGPD).

### **5. Différences entre Data Lake et Data Warehouse :**

* **Structure** : Le Data Warehouse stocke des données structurées, tandis que le Data Lake stocke des données brutes de tous types.
* **Utilisation** : Les Data Warehouses sont optimisés pour les analyses business et les rapports, alors que les Data Lakes sont utilisés pour les analyses avancées et le machine learning.
* **Stockage** : Le Data Warehouse est plus coûteux en termes de stockage structuré, tandis que le Data Lake est plus économique pour de grandes quantités de données brutes.
* **Temps de préparation** : Le Data Warehouse nécessite une préparation des données en amont, tandis que le Data Lake conserve les données telles quelles.

### **6. Qu'est-ce qu'un Data Lakehouse ?**

* Un Data Lakehouse est une architecture qui combine les avantages des Data Lakes et des Data Warehouses, en permettant de stocker des données brutes tout en offrant des fonctionnalités de gestion et d'analyse structurées. Il offre ainsi le meilleur des deux mondes.

### **7. Avantages du Data Lakehouse :**

* **Unified Architecture** : Combine des charges de travail de traitement de données structurées et non structurées.
* **Performance** : Optimisé pour des requêtes plus rapides par rapport aux Data Lakes traditionnels.
* **Gestion simplifiée** : Simplifie la gouvernance et la gestion des données.

### **8. Qu'est-ce qu'un Data Mesh ?**

* Un Data Mesh est une approche décentralisée de la gestion des données, basée sur des principes d'organisation par domaine, où chaque équipe gère ses propres données comme un produit, en suivant des standards communs. Cela vise à rendre les données plus accessibles et utilisables à grande échelle.

### **9. Avantages du Data Mesh :**

* **Scalabilité organisationnelle** : Permet une gestion des données plus évolutive, en rendant chaque domaine responsable de ses propres données.
* **Autonomie des équipes** : Chaque équipe ou domaine peut gérer et développer ses capacités de gestion de données.
* **Standardisation** : Assure une interopérabilité grâce aux standards communs.

### **10. Comparaison entre Data Mesh et Data Lake :**

* **Architecture** : Le Data Mesh est plus orienté vers une gestion décentralisée, alors que le Data Lake est un dépôt centralisé.
* **Responsabilité** : Data Mesh responsabilise chaque domaine pour ses propres données, tandis que Data Lake centralise les données sans responsabilisation par domaine.

### **11. Brainstorming sur les métiers de la data :**

1. **Analyste BI (BI Analyst)**
   * **Rôle** : Utiliser des outils BI pour analyser et présenter des données d'entreprise pour faciliter la prise de décision.
   * **Compétences** : SQL, reporting, outils BI (Tableau, Power BI), compréhension des KPI.
2. **Architecte Data (Data Architect)**
   * **Rôle** : Concevoir et gérer l'architecture de données pour assurer la fiabilité, la scalabilité et la sécurité.
   * **Compétences** : Modélisation des données, gestion de base de données, technologies cloud, big data.
3. **Analyste DW (DW Analyst)**
   * **Rôle** : Analyser et gérer les données d'entrepôts pour des analyses de performance et des rapports.
   * **Compétences** : ETL, SQL, conception de Data Warehouse, optimisation de performance.
4. **Analyste Data (Data Analyst)**
   * **Rôle** : Explorer et analyser les données pour extraire des insights significatifs.
   * **Compétences** : Python/R, SQL, statistiques, data visualization.
5. **Consultant BI (BI Consultant)**
   * **Rôle** : Conseiller les entreprises sur les solutions de BI et mettre en place des outils d'analyse.
   * **Compétences** : Stratégie BI, outils de BI, gestion de projet, communication.
6. **Manager DW (DW Manager)**
   * **Rôle** : Superviser la gestion des entrepôts de données, assurer la disponibilité et la qualité des données.
   * **Compétences** : Management, gouvernance des données, ETL, sécurité des données.
7. **Data Scientist**
   * **Rôle** : Appliquer des techniques de machine learning pour prédire des tendances et extraire des insights.
   * **Compétences** : Modélisation statistique, machine learning, programmation (Python, R), data wrangling.