

Mini-projet : DATA WAREHOUSE

**Phase 2 :
Développement d'un CUBE OLAP
pour Analyse multidimensionnelle
des performances des établissements et formations supérieures**

Date de remise : la Mardi 17/12/2014

OBJECTIFS

1. Concevoir un Data Warehouse (DW) adapté aux données extraites de LinkedIn.
2. Créer un cube OLAP pour des analyses multidimensionnelles des performances académiques et professionnelles.
3. Développer et valider des requêtes analytiques pour répondre aux besoins des décideurs.
4. Synthétiser et intégrer les résultats pour générer des recommandations stratégiques.

ÉTAPES PRINCIPALES

Étape 1 : Conception du Data Warehouse

- Identifier les dimensions pertinentes (ex. : établissements, formations, promotions, industries).
- Définir les indicateurs de performance (KPIs), ex. : taux d'embauche, type d'industrie, salaire moyen.
- Normaliser les données extraites pour garantir leur cohérence.

Étape 2 : Modélisation du cube OLAP

- Déterminer les axes d'analyse : géographique, temporel, etc.
- Intégrer les relations hiérarchiques (promotion -> formation -> établissement).

Étape 3 : Validation des requêtes analytiques

- Définir des scénarios d'analyse (ex. : "Quel est le taux d'employabilité par formation ?").
- Implémenter des requêtes analytiques SQL pour répondre à ces scénarios.

Objectifs des requêtes analytiques

1. Répondre à des questions stratégiques sur les performances des établissements et des formations.
2. Permettre une exploration flexible des données grâce aux analyses multidimensionnelles.
3. Identifier les tendances et anomalies dans les indicateurs clés (KPI).

Étapes pour concevoir, valider et développer les requêtes analytiques

1. Définir les scénarios analytiques

Ces scénarios doivent refléter les besoins des décideurs. Voici des exemples concrets :

- **Performance des formations** : Quel est le taux d'emploi par formation pour une année donnée ?
- **Analyse géographique** : Quels établissements ont le plus d'employés dans une région spécifique ?
- **Tendances temporelles** : Comment les taux d'embauche ont-ils évolué sur les 5 dernières années ?
- **Analyse sectorielle** : Dans quels secteurs les diplômés d'une formation travaillent-ils majoritairement ?

2. Définir les KPI pour les requêtes

1. Questions sur le contexte et les objectifs

1. **Quel est l'objectif spécifique que le KPI doit mesurer ?**
 - Ex. : Évaluer le taux d'employabilité des diplômés d'une formation.
2. **Comment ce KPI s'aligne-t-il avec les objectifs stratégiques ?**
 - Ex. : Améliorer la performance des établissements supérieurs.
3. **Qui sont les utilisateurs finaux de ce KPI et pourquoi en ont-ils besoin ?**
 - Ex. : Décideurs, responsables académiques, recruteurs.

2. Questions sur les données disponibles

4. **Quelles sont les sources de données disponibles ?**
 - Ex. : LinkedIn, bases de données académiques, enquêtes.
5. **Les données disponibles sont-elles fiables et à jour ?**
 - Vérifier la qualité, l'exhaustivité et l'actualisation des données.
6. **Comment les données seront-elles collectées et transformées ?**

- Ex. : Utilisation d'outils ETL pour intégrer les données dans un Data Warehouse.

3. Questions sur les dimensions et le calcul

- 7. Quelles dimensions doivent être prises en compte pour ce KPI ?**
 - Ex. : Année de diplomation, formation, établissement, région.
- 8. Quelle formule ou méthode sera utilisée pour calculer ce KPI ?**
 - Ex. :
 - Taux d'employabilité (%) = (Nombre d'employés / Nombre total de diplômés) × 100.
 - Salaire moyen = Somme des salaires / Nombre de diplômés.
- 9. Ce KPI est-il simple à comprendre et interpréter ?**
 - Assurer que la métrique soit claire pour les utilisateurs finaux.

4. Questions sur la pertinence et la faisabilité

- 10. Ce KPI est-il vraiment utile pour la prise de décision ?**
 - Éviter les métriques qui ne génèrent pas de valeur ajoutée ou de conclusions exploitables.
- 11. Est-il possible de mesurer ce KPI de manière réaliste ?**
 - Vérifier que les données et outils nécessaires sont disponibles.
- 12. À quelle fréquence ce KPI doit-il être mis à jour ?**
 - Ex. : Mensuelle, annuelle, en temps réel.

5. Questions sur l'interprétation et l'impact

- 13. Quels seuils ou objectifs doivent être fixés pour ce KPI ?**
 - Ex. : Objectif = atteindre un taux d'employabilité supérieur à 85 %.
- 14. Que signifiera une variation de ce KPI ?**
 - Ex. : Une baisse du taux d'employabilité indique un besoin d'ajuster les formations.
- 15. Quelles actions peuvent être prises en fonction des résultats du KPI ?**
 - Ex. : Adapter le contenu des formations ou renforcer les partenariats avec les entreprises.

6. Questions sur le suivi et l'évolution

- 16. Comment ce KPI sera-t-il présenté et suivi ?**
 - Ex. : Via des tableaux de bord interactifs (Power BI, Tableau).
- 17. Ce KPI doit-il être combiné à d'autres pour une analyse plus complète ?**
 - Ex. : Croiser le taux d'employabilité avec le salaire moyen pour mieux évaluer la qualité de la formation.
- 18. Ce KPI devra-t-il évoluer avec le temps ?**

- Prévoir des ajustements selon les changements de stratégie ou de données.

Checklist pour valider un KPI

- **Spécifique** : Le KPI mesure-t-il un aspect précis et ciblé ?
- **Mesurable** : Dispose-t-on des données nécessaires pour le calculer ?
- **Atteignable** : Les objectifs associés sont-ils réalistes ?
- **Pertinent** : Le KPI est-il aligné avec les objectifs stratégiques ?
- **Temporellement défini** : La période de mesure est-elle claire et adaptée ?

Exemples

- **Taux d'embauche (%)** = $(\text{Nombre d'employés} / \text{Nombre total de diplômés}) * 100$
- **Durée moyenne pour le premier emploi** = Temps entre la date de fin des études et le premier emploi.
- **Répartition par secteur** = Pourcentage de diplômés dans un secteur donné.
- **Salaire moyen par formation** = Salaire moyen des diplômés d'une formation donnée.

Étape 4 : Synthèse et intégration des résultats

- Créer des tableaux de bord pour visualiser les résultats.
- Proposer des recommandations basées sur les analyses effectuées.

LIVRABLES :

Livrable 1 :

Contenu : WBS qui explicite le travail du groupe (y compris la collaboration avec un agent AI : usages , prompts pour chaque étape, analyse des sorties,)

Livrable 2 : Catalogue des requêtes développées

Contenu : Liste des requêtes documentées répondant aux besoins spécifiques avec leur syntaxe, objectif, et résultats attendus.

Livrable 2 : Synthèse et visualisation des résultats

Contenu : **Tableaux de bord interactifs**

- Dashboards construits avec Tableau.
- Visualisations dynamiques (graphiques, cartes, courbes temporelles).

Rapport analytique

- Synthèse des résultats des analyses (par région, secteur, formation, etc.).
- Identification des tendances, opportunités, et anomalies.
- Recommandations stratégiques pour les établissements.

EVALUATION :

20% de la note , répartie : 50% (apport individuel) 50% (Travail collaboratif)