

Exploration et Analyse de AdventureWorks

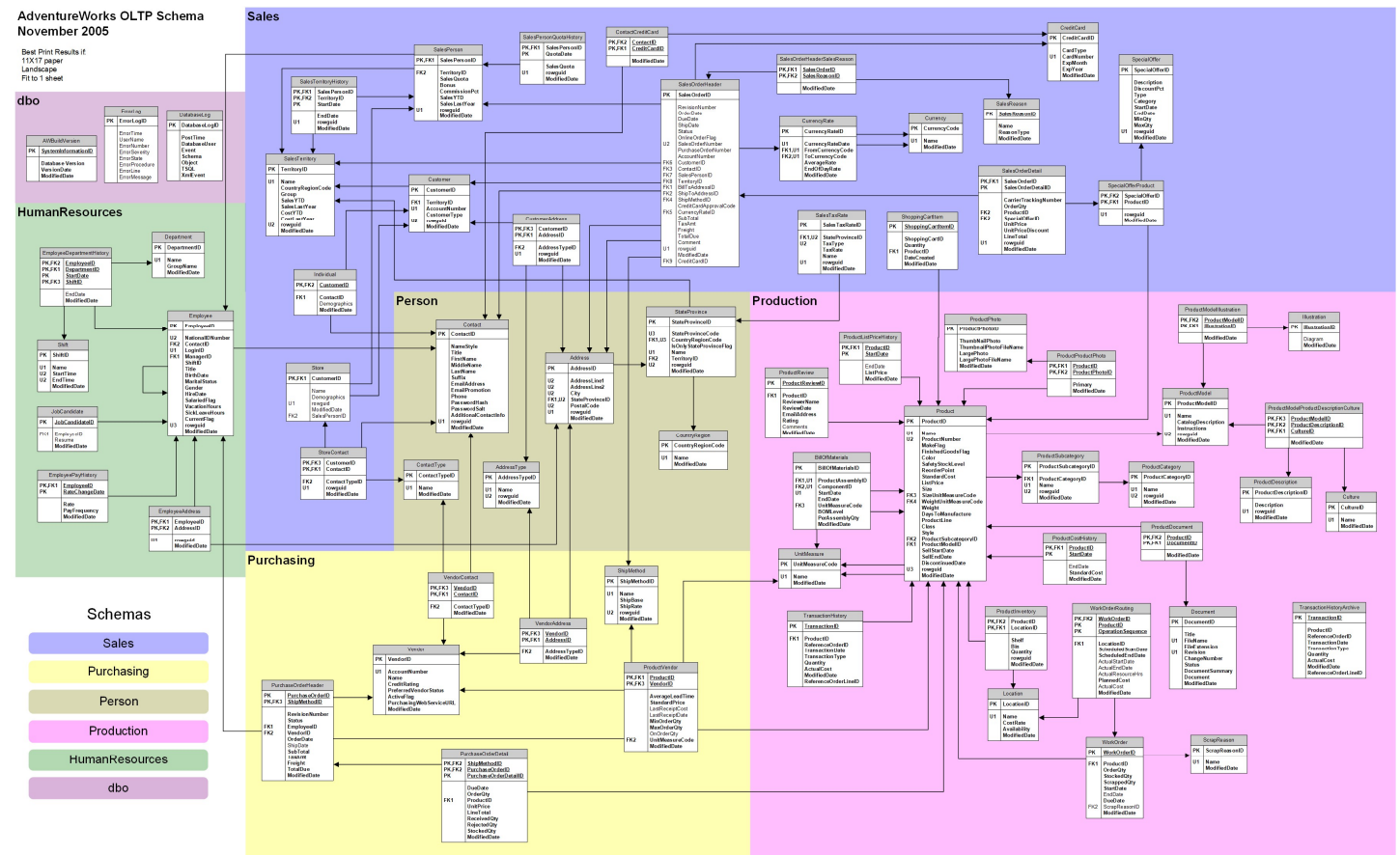
Objectif du Projet:

L'objectif de ce projet est de permettre aux équipes de découvrir et d'explorer les données d'une entreprise fictive, AdventureWorks, à travers plusieurs étapes pratiques allant de la manipulation des données à leur visualisation, en passant par la gestion de la qualité des données et la collaboration en équipe. Chaque équipe sera responsable d'un ensemble de tables et devra effectuer un nettoyage des données, puis les visualiser à l'aide de Power BI ou **Tableau**. La gestion des tâches sera assurée avec Taiga.io et la collaboration de code avec GitLab/GitHub.

Description de AdventureWorks:

AdventureWorks est un ensemble de bases de données d'exemples fournies par Microsoft qui simulent un scénario commercial réel. Ces bases de données sont largement utilisées pour l'apprentissage et la démonstration des fonctionnalités de SQL Server, ainsi que pour des projets liés aux données tels que l'ETL (Extract, Transform, Load), la modélisation des données et les tests de performance.

La base de données AdventureWorks couvre différents aspects d'une entreprise fictive multinationale de fabrication appelée AdventureWorks Cycles, qui produit et vend des vélos, des pièces et des accessoires. C'est une excellente ressource pour apprendre à gérer des systèmes de gestion de bases de données dans des scénarios réalistes.



Étapes du projet:

1. Téléchargement et Installation d'AdventureWorks:

- **Lien de téléchargement:** [AdventureWorks Databases](#)
- Les équipes devront télécharger la version **AdventureWorks2019** ou **AdventureWorksDW** en fonction des besoins du projet.
- Importez le fichier .bak dans SQL Server Management Studio (SSMS) pour installer la base de données.

2. Découverte des Tables:

- Les équipes exploreront les différentes tables dans la base de données. Voici une répartition possible:
 - **Équipe 1:** Tables liées aux **ventes** et aux **territoires** de vente (SalesOrderHeader, SalesOrderDetail, SalesTerritory).
 - **Équipe 2:** Tables liées aux **produits** et à la **production** (Product, ProductCategory, ProductionOrderHeader).

- **Équipe 3:** Tables liées aux **ressources humaines** (Employee, Department, Shift).
 - **Équipe 4:** Tables liées aux **achats** et aux **fournisseurs** (PurchaseOrderHeader, Vendor, PurchaseOrderDetail).
3. **Data Cleaning et Qualité des Données :**
- **Objectif :** Chaque équipe devra effectuer des tâches de nettoyage de données en identifiant les anomalies (données manquantes, doublons, formats incorrects).
 - **Outils :** Utiliser SQL Server Management Studio (SSMS) pour écrire des scripts SQL afin de nettoyer les tables ou bien utiliser un outil externe comme KNIME ou Pentaho.
4. **Visualisation des Données avec Power BI ou Tableau :**
- Chaque équipe créera des visualisations pertinentes dans **Power BI, Tableau** pour analyser les données propres à leur domaine.
 - **Exemples de visualisations :**
 - **Équipe 1 (Ventes):** Volume des ventes par territoire et par produit.
 - **Équipe 2 (Produits):** Répartition des produits par catégorie et performance des lignes de production.
 - **Équipe 3 (Ressources Humaines):** Distribution des employés par département et par shift.
 - **Équipe 4 (Achats):** Comparaison des commandes d'achat et des performances des fournisseurs.
 - Les tableaux de bord doivent fournir des insights clairs pour la prise de décision dans les domaines respectifs.
5. **Gestion des Tâches avec Taiga.io :**
- Créez un projet sur [Taiga.io](https://taiga.io) pour la gestion agile des tâches de chaque équipe.
 - Les **tickets** doivent être créés pour les différentes étapes du projet, comme :
 - Exploration et documentation des tables.
 - Nettoyage des données.
 - Création des visualisations **Power BI, Tableau**.
 - Revues de qualité et présentation des résultats.
 - Utilisez une méthode **Agile** pour planifier le travail de chaque équipe et suivre les progrès.
6. **Travail Collaboratif :**
- Créez un dépôt Git sur **GitLab** ou **GitHub** pour chaque équipe.
 - Partagez les scripts SQL de nettoyage de données et les rapports **Power BI, Tableau** dans le dépôt.
 - Effectuez des revues de code pour garantir la qualité du travail collaboratif.

Résultat attendu :

- Un rapport documenté pour chaque équipe sur l'état de leurs données avant et après le nettoyage.
- Un tableau de bord **Power BI, Tableau** contenant des visualisations pertinentes pour l'entreprise AdventureWorks.
- Des rapports de qualité des données avec des recommandations pour l'amélioration de la qualité.
- Utilisation efficace de **Taiga.io** pour la gestion des tâches et des tickets.
- Collaboration réussie et utilisation des branches et des merge requests sur **GitLab/GitHub**.

Livrables :

- Scripts **SQL** ou fichier **Pentaho** ou **KNIME** de nettoyage des données.
- Rapport de qualité des données avec indicateurs **KPI**.
- Tableau de bord **Power BI ou Tableau**.
- Rapport d'utilisation de **Taiga.io** avec répartition des tâches par sprint.
- Dépôt **GitLab/GitHub** avec l'historique du projet et les visualisations finales.

Remarques :

- Former des groupes de 2 ou 3 étudiants maximum;
- Le travail doit être envoyé vers la fin du semestre et contenant les éléments suivants :
 - Un rapport détaillé décrivant les différentes phases du projet
 - Les fichiers de chaque étape du processus décisionnel : les données sources, Le datamart, l'ETL, le(s) tableau(x) de bord, les rapports, ...
- La date d'évaluation orale sera communiquée ultérieurement.

NB : Vous serez noté sur la qualité de conception du DM et le niveau de détail de l'analyse.