



RAPPORT SUR LES SOLUTIONS DE DATA VISUALISATION

Introduction

Objectif du rapport

L'objectif de ce rapport est d'analyser différentes solutions pour exploiter nos données à travers des outils de data visualisation. Nous explorerons trois approches principales : la connexion directe à la base de données, l'extraction des données en CSV, et l'utilisation d'un ETL (PowerQuery, Knime, etc.). Chaque solution sera évaluée en termes de pertinence, efficacité et adaptation aux besoins de notre entreprise.

Contexte de l'analyse

Notre entreprise accuse un retard significatif dans l'utilisation des données, ce qui impacte notre compétitivité, notre organisation et notre croissance. Récemment, nos bases de données ont été refondues pour devenir plus fiables et unifiées. Le défi actuel est de mettre ces données à disposition via un outil de data visualisation afin d'aider les équipes à mieux piloter leurs activités.

État des lieux et évaluation de la pertinence

2.1 Description de la situation actuelle

Actuellement, les données de l'entreprise sont stockées dans une base de données centralisée. Les équipes n'ont pas un accès direct à ces données et utilisent principalement des fichiers Excel ou CSV pour leurs analyses. Les processus sont manuels et souvent inefficaces, entraînant des erreurs et des retards.

2.2 Évaluation de la pertinence des données existantes

Les données sont désormais propres et bien structurées. Elles couvrent divers aspects du business, notamment les ventes, les stocks, et les prix. Cependant, l'accès et la visualisation de ces données sont limités par les outils et méthodes actuellement utilisés.

2.3 Identification des principaux problèmes ou enjeux

- Manque d'accès direct et en temps réel aux données.
- Processus d'extraction et de traitement des données manuels et chronophages.
- Difficulté à intégrer différentes sources de données pour une analyse cohérente.
- Besoin d'une solution fiable et évolutive pour répondre aux besoins de visualisation des données.

Besoin d'outils

3.1 Identification des outils nécessaires pour collecter les données

- Connexion directe à la base de données via des outils de BI.
- Extraction de données en CSV pour des analyses hors ligne.
- Utilisation d'ETL pour automatiser l'extraction et la transformation des données.

3.2 Identification des outils nécessaires pour traiter les données

- ETL (Extract, Transform, Load) pour automatiser et fiabiliser le traitement des données.
- Scripts SQL pour des manipulations directes dans la base de données.
- PowerQuery pour des transformations de données au sein d'Excel.

3.3 Identification des outils nécessaires pour analyser les données

- Outils de data visualisation comme Power BI, Tableau ou Looker.
- Tableaux de bord interactifs pour les KPI et autres métriques.
- Filtres et options d'exploration de données pour des analyses approfondies.

Solutions d'extraction, de traitement et de visualisation

4.1 Solutions proposées pour l'extraction des données

1. Connexion directe à la base de données

Avantages	Inconvénients
Accès en temps réel aux données	Dépendance à la performance de la base de données
Moins de duplication et risque d'erreurs réduit	Nécessite des connaissances techniques pour les requêtes SQL
Configuration initiale requise mais maintenance limitée	

2. Extraction des données en CSV

Avantages	Inconvénients
Simplicité et compatibilité universelle	Nécessite des processus manuels réguliers
Facilité de partage des données	Risque d'erreurs et de versioning des fichier

3. Utilisation d'un ETL

Avantages	Inconvénients
Automatisation de l'extraction et transformation des données	Configuration initiale complexe mais grande flexibilité
Peut intégrer diverses sources de données	Besoin de compétences techniques pour la mise en place
Flexibilité et évolutivité	

4.2 Solutions proposées pour le traitement des données

1. Connexion directe à la base de données

Avantages	Inconvénients
Utilisation de requêtes SQL pour des traitements dynamiques	Capacités de traitement limitées par la performance de la base de données
Données toujours à jour	Dépendance aux compétences en SQL

2. Extraction des données en CSV

Avantages	Inconvénients
Traitement manuel via Excel ou autres outils	Manque d'automatisation et risques d'erreurs humaines
Facilité d'utilisation pour les utilisateurs non techniques	Non adapté aux grands volumes de données

3. Utilisation d'un ETL

Avantages	Inconvénients
Transformation des données automatisée	Configuration initiale complexe
Traitement en batch ou en temps réel selon les besoins	Nécessite des compétences techniques pour la mise en place
Intégration de multiples sources de données	

4.3 Solutions proposées pour la visualisation des données

1. Connexion directe à la base de données via Power BI/Tableau

Avantages	Inconvénients
Tableaux de bord interactifs et en temps réel	Dépendance à la performance de la base de données
Flexibilité pour créer diverses visualisations	Complexité initiale de la configuration
Mise à jour automatique des données	

2. Utilisation de fichiers CSV importés dans Power BI/Tableau

Avantages	Inconvénients
Analyse hors ligne	Nécessite une mise à jour régulière des fichiers
Facilité d'utilisation pour les utilisateurs non techniques	Données potentiellement obsolètes

3. ETL intégrant Power BI/Tableau

Avantages	Inconvénients
Données toujours à jour et intégrées	Configuration initiale complexe
Flux de données automatisé et réduit les interventions manuelles	Besoin de compétences techniques pour la mise en place
Flexibilité et évolutivité	

Cohérence des solutions avec le besoin

5.1 Explication détaillée de chaque solution proposée

Connexion directe

- **Description :**
 - La connexion directe à la base de données permet un accès en temps réel aux données. Cela offre une grande flexibilité pour la création de tableaux de bord dynamiques et interactifs. La configuration initiale est nécessaire, mais la maintenance est ensuite limitée.

Extraction en CSV

- **Description :**
 - L'extraction des données en CSV est une méthode simple et rapide à mettre en place. Elle permet une compatibilité universelle avec divers outils d'analyse, mais elle est peu évolutive et peut entraîner des risques de données obsolètes et des erreurs humaines.

ETL

- **Description :**
 - L'utilisation d'un ETL est une solution complète et flexible, permettant une automatisation et une intégration optimale des données. Elle requiert une expertise technique pour la configuration initiale, mais offre ensuite une grande flexibilité et réduit les interventions manuelles.

5.2 Alignement des solutions avec les besoins identifiés

Connexion directe

- **Alignement :**
 - La connexion directe répond au besoin d'accès en temps réel et de visualisation dynamique, facilitant la prise de décisions rapide basée sur des données à jour.

Extraction en CSV

- **Alignement :**
 - L'extraction en CSV est utile pour des analyses ponctuelles ou pour les utilisateurs non techniques, mais elle n'est pas adaptée aux besoins quotidiens de l'entreprise en raison de sa nature manuelle et non automatisée.

ETL

- **Alignement :**
 - L'utilisation d'un ETL est idéale pour un processus automatisé et intégré, répondant à la fois aux besoins de traitement et de visualisation. Elle permet une gestion efficace des données et une flexibilité pour des analyses approfondies.

5.3 Avantages et limites de chaque approche

Connexion directe

Avantages	Limites
Données en temps réel	Dépendance à la performance de la base de données
Maintenance minimale après la configuration initiale	Configuration initiale requise

Extraction en CSV

Avantages	Limites
Simplicité et compatibilité universelle	Manque d'automatisation
Facilité de partage des données	Risque d'erreurs et de données obsolètes

ETL

Avantages	Limites
Automatisation et intégration de multiples sources	Configuration initiale complexe
Flexibilité et évolutivité	Besoin d'expertise technique

Conclusion

Synthèse des principales conclusions de l'analyse

- **Connexion directe à la base de données :**
 - Offre un accès en temps réel mais dépend de la performance de la base de données.
- **Extraction en CSV :**
 - Solution simple mais peu évolutive et sujette à des erreurs.
- **Utilisation d'un ETL :**
 - Solution la plus flexible et évolutive, permettant une automatisation complète du flux de données.

Recommandations pour la mise en œuvre des solutions proposées

Pour répondre aux besoins actuels de l'entreprise en matière de data visualisation, la meilleure solution serait d'adopter un ETL comme PowerQuery ou Knime pour automatiser l'extraction et le traitement des données, couplé à un outil de visualisation comme Power BI ou Tableau pour la création de tableaux de bord interactifs et en temps réel. Cette approche permettrait de maximiser l'efficacité, réduire les erreurs et offrir une flexibilité optimale pour des analyses approfondies.

Annexes

Liste des sources de données utilisées

1. Base de Données Centrale

- Type : SQL Database
- Contenu : Données de ventes, stocks, prix, produits,
- Fréquence de mise à jour : Quotidienne

2. Fichiers CSV/Excel

- Type : Fichiers plats
- Contenu : Données historiques, rapports de ventes mensuels, inventaire
- Fréquence de mise à jour : Ponctuelle, selon les besoins

Description détaillée des outils recommandés

1. Power BI



Description : Power BI est un outil de business intelligence (BI) développé par Microsoft. Il permet aux utilisateurs de créer des rapports et des tableaux de bord interactifs en connectant, modélisant et visualisant des données provenant de différentes sources.

Fonctionnalités Clés:

- **Connexion à diverses sources de données:** Power BI permet de se connecter à une grande variété de sources de données, incluant SQL Server, Excel, SharePoint, ainsi que des services cloud comme Azure, Google Analytics, et plus encore. Cette flexibilité garantit que les utilisateurs peuvent centraliser leurs données pour une analyse complète.
- **Création de visualisations interactives:** L'outil offre une large gamme de visualisations telles que des graphiques en barres, en lignes, en aires, des cartes géographiques, des tableaux, des radars et bien d'autres. Ces visualisations sont interactives et permettent aux utilisateurs de creuser dans les données pour obtenir des insights détaillés.
- **Partage et collaboration:** Power BI permet le partage des rapports et tableaux de bord via le service Power BI en ligne. Les utilisateurs peuvent collaborer en temps réel, commenter et discuter des rapports, ce qui facilite la prise de décision collective.

- **Mises à jour automatiques** : Les rapports et tableaux de bord peuvent être configurés pour se mettre à jour automatiquement selon un planning défini, assurant que les utilisateurs travaillent toujours avec les données les plus récentes.

Avantages	Limites
Intégration fluide avec les outils Microsoft tels qu'Excel, Azure, et Office 365	Les fonctionnalités avancées peuvent nécessiter une formation supplémentaire
Interface utilisateur intuitive, réduisant la courbe d'apprentissage	Peut être coûteux pour les grandes équipes si toutes les fonctionnalités premium sont nécessaire
Capacité de créer des visualisations complexes sans besoin de programmation	

2. Tableau



Description : Tableau est un outil de visualisation de données réputé pour sa facilité d'utilisation et ses puissantes capacités analytiques. Il permet aux utilisateurs de transformer des données brutes en insights exploitables via des visualisations interactives.

Fonctionnalités Clés:

- **Connexion à une large gamme de sources de données**: Tableau peut se connecter à des bases de données relationnelles, des fichiers, des services cloud, et bien plus encore. Cette polyvalence permet une intégration facile des données provenant de diverses sources.
- **Drag-and-drop pour créer des visualisations**: L'interface utilisateur de Tableau est conçue pour être intuitive. Les utilisateurs peuvent créer des visualisations simplement en glissant et déposant des champs dans des zones de graphique, rendant l'analyse rapide et accessible même pour les non-techniciens.
- **Tableau Server**: Pour le partage et la collaboration, Tableau Server permet de publier des tableaux de bord sur une plateforme sécurisée où d'autres utilisateurs peuvent interagir avec les rapports, filtrer les données et explorer les visualisations selon leurs besoins.
- **Gestion de grands ensembles de données**: Tableau est capable de traiter et d'analyser de grands volumes de données, offrant des performances robustes même avec des ensembles de données complexes et volumineux.

Avantages	Limites
Excellente flexibilité et personnalisation des visualisations	Coût élevé des licences pour des équipes nombreuses
Forte communauté d'utilisateurs et de nombreuses ressources d'apprentissage disponibles	Les fonctionnalités avancées peuvent nécessiter une expertise technique approfondie
Capacités analytiques avancées avec des fonctionnalités de calcul intégrées	

3. PowerQuery (Excel)



Description: PowerQuery est un outil d'extraction, de transformation et de chargement des données intégré à Microsoft Excel. Il permet aux utilisateurs de préparer des données pour analyse de manière efficace et automatisée.

Fonctionnalités Clés:

- **Connexion et transformation de données depuis plusieurs sources:** PowerQuery permet d'importer des données de différentes sources comme des bases de données SQL, des fichiers Excel, des API web, etc. Les utilisateurs peuvent ensuite transformer ces données pour les rendre adaptées à l'analyse.
- **Automatisation des tâches de nettoyage:** Les utilisateurs peuvent créer des scripts de transformation pour nettoyer et préparer les données. Ces transformations peuvent être sauvegardées et réutilisées, permettant une automatisation des tâches répétitives.
- **Intégration fluide avec Excel:** Une fois les données préparées, elles peuvent être analysées directement dans Excel, offrant aux utilisateurs une transition fluide entre la préparation des données et l'analyse.
- **Interface utilisateur intuitive:** PowerQuery propose une interface conviviale qui permet aux utilisateurs de définir les étapes de transformation via des actions simples de glisser-déposer et des menus contextuels.

Avantages	Limites
Idéal pour les utilisateurs familiers avec Excel, réduisant la courbe d'apprentissage	Capacité limitée pour traiter de très grands ensembles de données
Automatisation des processus de préparation des données, augmentant l'efficacité	Moins puissant que des outils dédiés comme Power BI ou Tableau pour la visualisation de données avancée
Gratuit pour les utilisateurs d'Excel, offrant un excellent rapport coût-bénéfice	

4. Knime



Description: Knime est une plateforme open-source pour l'analyse de données et le machine learning. Elle offre des capacités avancées d'extraction, transformation et chargement (ETL) des données.

Fonctionnalités Clés:

- **Interface de flux de travail visuel:** Knime utilise une interface de flux de travail visuel où les utilisateurs peuvent assembler des blocs de traitement de données pour modéliser leurs processus analytiques. Cette approche rend le développement et la maintenance des workflows plus intuitifs et visuels.
- **Intégration de nombreuses sources de données:** Knime peut se connecter à une vaste gamme de sources de données, y compris des bases de données SQL, des fichiers plats, des API web, et des outils de big data comme Hadoop.
- **Automatisation des processus ETL:** Les utilisateurs peuvent automatiser les tâches d'extraction, transformation et chargement des données, rendant les processus répétitifs plus efficaces et réduisant le risque d'erreurs humaines.
- **Extensible avec des plugins:** Knime dispose d'une riche bibliothèque de plugins et extensions qui permettent d'étendre ses capacités, y compris des outils de machine learning, des algorithmes statistiques avancés, et des intégrations avec d'autres plateformes analytiques.

Avantages	Limites
Grande flexibilité et extensibilité grâce à une architecture de plugins	La courbe d'apprentissage peut être raide pour les utilisateurs novices
Interface visuelle intuitive pour la création de workflows analytiques	Nécessite une certaine expertise technique pour tirer pleinement parti de ses fonctionnalités avancées
Capacité de gérer des ensembles de données volumineux et complexes	

En conclusion, chaque outil présente des avantages et des limites spécifiques. Power BI et Tableau se distinguent par leurs puissantes capacités de visualisation et d'interactivité, tandis que PowerQuery et Knime offrent des solutions robustes pour la préparation et la transformation des données. Le choix de l'outil dépendra des besoins spécifiques de l'entreprise, du budget, et du niveau d'expertise disponible en interne.

Exemples de visualisations de données proposées – solution développée avec Power BI

1. Tableau de bord des KPI clés

- **Visualisations :**
 - Courbes d'évolution du chiffre d'affaires par catégorie de produits.
 - Graphique en barres pour la marge brute et le taux de marge brute par catégorie de produits.
 - Graphique en barres pour la rotation des stocks par catégorie de produits.
 - Différents KPI majeurs globaux : chiffre d'affaires, marge brute, total de ruptures de stocks, rotation moyenne des stocks, augmentation des prix d'achat par rapport à leur valeur originelle

2. Page d'Analyse des Prix d'Achat

- **Visualisations :**
 - Graphique en courbes pour suivre l'évolution des prix d'achat par catégorie de produits.
 - Graphique en barres pour visualiser les hausses de prix par catégorie de produits
 - Table permettant l'analyse précise de suivi d'évolution du prix d'achat par produit.

3. Page de Suivi des Stocks

- **Visualisations :**
 - Graphiques en courbes pour les niveaux de stock par catégorie de produits.
 - Graphique en barres pour les ruptures de stock par catégorie sur une période donnée.
 - Table permettant l'analyse précise de suivi des ruptures de stock par produit.

4. Page de Performance Commerciale

- **Visualisations :**
 - Graphiques en barres pour le chiffre d'affaires, la marge brute et le taux de marge brute par catégorie de produits.
 - Graphiques en courbes pour le volume de ventes par catégorie de produits.
 - Table permettant l'analyse précise de performance commerciale par produit.

5. Page de Suivi des Promotions

- **Visualisations :**
 - Graphique en barres pour le ROI des promotions par catégorie de produits.
 - Graphique en barres pour le volume des ventes en promotion.
 - Table permettant l'analyse précise du ROI par produit.

Ces exemples de visualisations sont conçus pour fournir des insights clairs et exploitables, facilitant la prise de décisions pour les différentes équipes de l'entreprise.