



Date : 22/04/2025

## Introduction

- Objectif du rapport

Analyser les différentes méthodes d'exploitation des données dans un outil de data visualisation afin de recommander la solution la plus adaptée aux besoins de Bottleneck.

- Contexte de l'analyse

À la suite de la refonte des bases de données de Bottleneck, l'entreprise souhaite mettre ses données à disposition via un tableau de bord interactif. Trois approches techniques sont envisagées : connexion directe à la base, extraction en CSV ou usage d'un ETL.

## État des lieux et évaluation de la pertinence

### 2.1 Description de la situation actuelle

Les données sont désormais centralisées dans une base SQLite propre. L'objectif est de les rendre accessibles aux équipes via un outil de data visualisation.

### 2.2 Évaluation de la pertinence des données existantes

Les données couvrent un an d'activité (octobre 2022 à septembre 2023) et sont enrichies : ventes, promotions, prix d'achat, segments produits. Elles sont donc exploitables pour la création de KPI pertinents (marge, CA, etc.).

## 2.3 Identification des principaux problèmes ou enjeux

- Faciliter l'accès aux données pour des non-techniciens
- Mettre en place une solution maintenable et évolutive
- Assurer l'actualisation régulière des données dans le dashboard

## Besoin d'outils

### 3.1 Identification des outils nécessaires pour collecter les données

Outil	Avantages	Inconvénients	Adapté pour...
SQLite	Léger, portable, pas besoin de serveur, facile à manipuler	Non supporté nativement par Power BI, nécessite un connecteur externe	Stockage local simple, PME, test et prototypage
ODBC (Driver)	Permet la connexion directe à Power BI, simple à configurer, compatible	Pas de rafraîchissement automatique dans Power BI Service	Connexion directe avec Power BI Desktop
DB Browser for SQLite	Export manuel de tables ou requêtes, interface graphique	Manipulation manuelle, pas d'automatisation	Export ponctuel par utilisateur non technique
Python + sqlite3/pandas	Très flexible, automatisable, parfait pour exporter en CSV	Besoin de compétences techniques en Python	Automatisation avancée, script réutilisable

#### Autres solutions de collecte envisageables

Solution	Avantages	Inconvénients	Adapté pour...
API REST	Accès distant sécurisé, données filtrables, automatisation possible	Nécessite un développement backend, dépend d'une infrastructure API	Montée en charge future, architecture web/cloud
Stockage partagé (cloud)	Facile à mettre en œuvre avec OneDrive / SharePoint / Google Drive, compatible avec Power BI	Gestion manuelle ou semi-auto, dépend de la rigueur de dépôt	Scénarios semi-automatisés sans base distante
Base distante (ex. PostgreSQL)	Compatible nativement avec Power BI, performant et scalable	Nécessite configuration	

### 3.2 Benchmark des outils de traitement de données

Outil	Avantages	Inconvénients	Adapté pour...
PowerQuery	Intégré à Excel et Power BI, Interface simple et no-code, Parfait pour des traitements légers, Pas besoin d'installation	Limité à l'environnement Microsoft, Pas adapté aux traitements complexes multi-sources	Utilisateurs non techniques, Projets intégrés à Power BI
Knime	ETL visuel très puissant, Compatible avec de nombreux formats et bases, Automatisation et réutilisation de workflows	Besoin d'installation et configuration, Courbe d'apprentissage plus longue	Besoins complexes et automatisés, Traitement indépendant de Power BI
Python	Très puissant avec pandas, numpy, etc., Très flexible pour tout type de traitement, Parfait pour automatiser ou industrialiser	Nécessite des compétences en développement, Moins accessible aux utilisateurs métier	Analystes avancés, Automatisation poussée et sur mesure
Talend	Open source, puissant, très bon pour flux complexes et intégration API	Complexité, interface moins moderne que Knime	Projets ETL robustes à l'échelle entreprise
Apache Airflow	Orchestration de pipelines très puissante, open source, excellent pour planifier des traitements	Nécessite de coder (Python), configuration technique initiale importante	Architectures modernes, automatisation de flux complexes
Alteryx	Interface intuitive, flux visuels simples à maintenir	Coût élevé, version gratuite très limitée	Analystes métier avec budget
RStudio	Excellente bibliothèque statistique, bon pour modélisation prédictive	Moins BI-friendly que Python, nécessite des compétences statistiques	Projets orientés stats / machine learning
Pentaho	Suite ETL complète avec outils de reporting, bonne compatibilité avec bases SQL et fichiers plats	Interface un peu datée, courbe d'apprentissage, moins active en open source	Projets ETL structurés, entreprises cherchant une suite tout-en-un

### 3.3 Identification des outils nécessaires pour analyser les données

Outil	Avantages	Inconvénients	Recommandé pour
Power BI	Intégration MS, DAX, bon rapport qualité/prix, filtres puissants	Moins intuitif que Tableau	PME, utilisateurs Excel
Tableau	UX très intuitive, très bon pour storytelling	Coût élevé, moins bon sur data model	Analystes, data viz avancée
Looker Studio	Gratuit, facile à utiliser, intégration Google	Limité sur les grands volumes, peu de traitement complexe	Projets simples, petits budgets

Autres outils d'analyse envisageables (non retenus)

Outil	Avantages	Inconvénients	Recommandé pour
Qlik Sense	Moteur associatif très performant, très bon sur gros volumes, interface intuitive	Coût plus élevé, moins répandu dans les PME	Grandes entreprises, projets multi-sources
Metabase	Open source, interface simple, déploiement rapide	Moins puissant sur le plan analytique, visualisations limitées	Startups, reporting simple, budget limité
Excel (avec graphiques)	Facile à prendre en main, très répandu, déjà utilisé en interne	Peu adapté à l'exploration dynamique, pas scalable	Suivi ponctuel, utilisateurs Excel uniquement

## Solutions d'extraction, de traitement et de visualisation

### 4.1 Solutions proposées pour l'extraction des données

Pourquoi la connexion directe via **ODBC** est le meilleur choix pour **Bottleneck**

*Intégrée à Power BI*

- ✓ Permet une **connexion directe** à la base SQLite via un driver ODBC.
- ✓ Données **toujours à jour**, sans besoin d'export manuel.
- ✓ Compatible avec la stratégie de mise à jour continue du tableau de bord.

*Facile à configurer*

- ✓ Nécessite uniquement l'installation d'un driver ODBC (ex. : SQLite ODBC Driver).
- ✓ Une fois configurée dans Windows, la source est directement disponible dans Power BI.

- ✓ Aucune compétence technique avancée requise pour l'utiliser au quotidien.

#### *Limite les erreurs humaines*

- ✓ Évite les manipulations de fichiers CSV, souvent sources de doublons ou d'incohérences.
- ✓ Diminue les risques d'oublis ou de décalage temporel dans les données analysées.

#### *Adaptée à un usage local*

- ✓ Solution idéale pour Power BI Desktop, utilisé dans un environnement local sans infrastructure complexe.
- ✓ Compatible avec un usage autonome et ponctuel des collaborateurs.

## 4.2 Solutions proposées pour le traitement des données

Pourquoi **PowerQuery** est le meilleur choix pour **Bottleneck**

#### *Intégré à Power BI*

- ✓ Il s'utilise directement dans l'outil de visualisation choisi, sans outil supplémentaire à installer ou configurer.

#### *Facile à prendre en main*

- ✓ L'interface est graphique, intuitive, et bien connue des utilisateurs Excel. Elle permet aux chefs de produit ou managers de suivre ou modifier des flux simples.

#### *Suffisant pour les besoins actuels*

- ✓ Les transformations nécessaires (fusion de tables, suppression de doublons, normalisation) sont toutes réalisables avec PowerQuery.

#### *Gain de temps et faible coût*

- ✓ Pas de développement à maintenir ni de plateforme à héberger : parfait pour une PME sans équipe technique dédiée.

## 4.3 Solutions proposées pour la visualisation des données

Pourquoi **Power BI** est le meilleur choix pour **Bottleneck**

*Compatible avec toutes les méthodes de connexion*

- ✓ Power BI peut exploiter les données **via connexion directe, fichiers CSV ou ETL** (PowerQuery, Knime...), ce qui offre une **grande flexibilité** d'alimentation, quelle que soit l'architecture choisie.

*Visualisations riches et interactives*

- ✓ Il propose un **large éventail de visuels**, de filtres dynamiques et de fonctions DAX pour suivre efficacement les KPI (CA, marge, rotation des stocks...). Idéal pour des **tableaux de bord métiers opérationnels**.

*Intégré à l'écosystème Microsoft*

- ✓ Il s'intègre naturellement avec **Excel, OneDrive, SharePoint**, et permet aux équipes de **gagner du temps** dans le partage et la mise à jour des données.

*Facile à déployer et à maintenir*

- ✓ Une simple installation suffit pour démarrer. Il n'y a **pas besoin de serveur**, ce qui correspond parfaitement à une PME comme Bottleneck.

*Adapté aux profils non techniques*

- ✓ Les utilisateurs peuvent **naviguer facilement dans les rapports**, utiliser des filtres ou explorer les données sans connaissance technique.

## 4.4 Synthèse des outils choisis

Outil	Rôle dans le projet	Pourquoi retenu ?
Power BI	Visualisation et exploration des données	Intuitif, puissant, bien intégré à l'environnement Microsoft
PowerQuery	Traitement léger, transformation, nettoyage	Simple à utiliser, intégré à Power BI, rapide à mettre en œuvre
ODBC (Driver)	Connexion à la base SQLite	Permet de travailler directement sur les données sans export manuel

# Cohérence des solutions avec le besoin

## 5.1 Explication détaillée de chaque solution proposée

- **Connexion directe via ODBC** : permet d'interroger les données en temps réel directement depuis Power BI. Cette méthode réduit les erreurs manuelles et garantit des données actualisées.
- **Extraction en CSV** : méthode simple et rapide à mettre en œuvre, mais peu maintenable à long terme, et sujette aux erreurs humaines.
- **ETL avec PowerQuery / Knime** : offre un excellent compromis entre flexibilité, transformation automatisée, et contrôle sur la qualité des données.

## 5.2 Alignement des solutions avec les besoins identifiés

Besoin	Connexion directe	Extraction CSV	ETL (PowerQuery / Knime)
Simplicité pour utilisateurs métier	✓	✓	⚠ (selon l'outil)
Données toujours à jour	✓	✗	✓ (si automatisé)
Évolutivité / maintenabilité	⚠	✗	✓
Fiabilité et réduction des erreurs	✓	✗	✓

**Légende** : ✓ = adapté | ⚠ = adapté sous conditions | ✗ = peu adapté

## 5.3 Avantages et limites de chaque approche

### ➤ Connexion directe

- ◆ Mise à jour en temps réel
- ▼ Dépend de la structure et de la disponibilité de la base

### ➤ Extraction CSV

- ◆ Facile à manipuler
- ▼ Maintenance manuelle, non automatisée

### ➤ ETL (PowerQuery / Knime)

- ◆ Traitement puissant, automatisation possible
- ▼ Mise en place plus technique au départ

# Conclusion

## Synthèse des principales conclusions de l'analyse

Les trois approches d'intégration des données (connexion directe, extraction CSV, ETL) ont été évaluées selon leur simplicité, fiabilité, et maintenabilité. L'utilisation d'un ETL tel que PowerQuery ou Knime offre une solution robuste pour la transformation et l'automatisation des données. Power BI est l'outil de visualisation le plus adapté pour ce projet grâce à sa compatibilité avec toutes les méthodes d'alimentation, ses visualisations interactives et sa puissance analytique.

## Recommandations pour la mise en œuvre des solutions proposées

- Utiliser **Knime** pour automatiser les flux de données, ou **PowerQuery** si l'équipe est déjà familière avec l'écosystème Microsoft.
- Connecter **Power BI** à ces flux via fichiers plats ou base de données selon la fréquence de mise à jour souhaitée.
- Intégrer des règles de normalisation et de contrôle de qualité dans les flux pour garantir la cohérence.
- Prévoir une documentation d'usage du dashboard pour les utilisateurs finaux (chefs de produit, direction).



## Annexes – Bibliographie / Sources utilisées

- Microsoft – Documentation officielle Power BI : <https://learn.microsoft.com/fr-fr/power-bi/>
- Tableau Software – Guide des bonnes pratiques de data visualisation : <https://www.tableau.com/fr-fr>
- Looker Studio – Centre d'aide : <https://support.google.com/looker-studio/>
- SQLite ODBC Drivers :
  - <https://www.ch-werner.de/sqliteodbc/>
  - <https://www.devart.com/odbc/sqlite/>
- Knime Analytics Platform – Documentation officielle : <https://docs.knime.com/>
- Towards Data Science – "10 Best ETL Tools for Data Pipeline Management", <https://towardsdatascience.com/best-etl-tools-data-pipeline-2023>
- Apache Airflow : <https://airflow.apache.org/>
- Talend Open Studio : <https://www.talend.com/products/talend-open-studio/>
- Pentaho Data Integration : <https://community.hitachivantara.com/s/article/what-is-pentaho-data-integration>
- Alteryx : <https://www.alteryx.com/fr>
- RStudio : <https://posit.co/download/rstudio-desktop/>
- Qlik Sense : <https://www.qlik.com/fr-fr/products/qlik-sense>
- Metabase : <https://www.metabase.com/>
- Towards Data Science – Knime vs Alteryx vs Power BI : <https://towardsdatascience.com/>
- Medium – Comparatif Power BI vs Metabase : <https://medium.com/>

## Annexe – Glossaire des termes techniques

Terme	Définition
<b>ETL (Extract Transform Load)</b>	Processus d'extraction, de transformation et de chargement des données entre différentes sources et destinations, souvent automatisé.
<b>ODBC (Open Database Connectivity)</b>	Interface permettant à des logiciels comme Power BI ou Excel de se connecter à divers types de bases de données.
<b>SQLite</b>	Système de gestion de base de données léger et embarqué, fonctionnant dans un simple fichier.
<b>Driver ODBC</b>	Programme qui fait l'interface entre un logiciel (comme Power BI) et une base de données pour établir une connexion fiable.
<b>CSV (Comma-Separated Values)</b>	Format de fichier texte simple utilisé pour stocker des données tabulaires (souvent exporté depuis des bases de données).
<b>Python</b>	Langage de programmation polyvalent largement utilisé en data science et en automatisation de traitement des données.
<b>Power BI</b>	Outil de Microsoft permettant de visualiser et d'analyser les données à travers des tableaux de bord interactifs.
<b>PowerQuery</b>	Langage et interface intégrés à Excel et Power BI permettant de transformer les données avant analyse, sans coder.
<b>Knime</b>	Plateforme open source d'analyse de données et de machine learning basée sur des workflows visuels.
<b>Talend</b>	Plateforme ETL open source permettant de construire des pipelines de traitement de données à grande échelle.
<b>Apache Airflow</b>	Outil d'orchestration de flux de données open source permettant de planifier et exécuter des tâches automatisées.
<b>Alteryx</b>	Logiciel de préparation de data et d'ETL no-code, utilisé pour automatiser et visualiser les workflows analytiques.
<b>RStudio</b>	Environnement de développement pour le langage R, spécialisé en statistiques, data science et modélisation.
<b>Pentaho</b>	Suite logicielle d'intégration de données (ETL) avec reporting et analyses intégrées.
<b>Tableau</b>	Outil de data visualisation interactif utilisé pour le storytelling de données et l'analyse visuelle avancée.
<b>Looker Studio</b>	Outil gratuit de Google pour créer des rapports dynamiques à partir de sources de données variées.
<b>Qlik Sense</b>	Outil de BI basé sur un moteur associatif performant, adapté aux grands volumes de données.
<b>Metabase</b>	Plateforme de visualisation open source légère, simple d'utilisation, souvent utilisée dans les startups.