

# Solutions d’utilisation des données dans un outil de visualisation

Date: 02/2025

Table des matières

[Solutions d’utilisation des données dans un outil de visualisation 1](#_Toc190780055)

[Introduction 2](#_Toc190780056)

[Objectif du rapport 2](#_Toc190780057)

[Contexte de l'analyse 3](#_Toc190780058)

[État des lieux et évaluation de la pertinence 3](#_Toc190780059)

[2.1 Description de la situation actuelle 3](#_Toc190780060)

[2.2 Évaluation de la pertinence des données existantes 3](#_Toc190780061)

[2.3 Identification des principaux problèmes ou enjeux 3](#_Toc190780062)

[Besoin d'outils 4](#_Toc190780063)

[3.1 Identification des outils nécessaires pour collecter les données 4](#_Toc190780064)

[3.2 Identification des outils nécessaires pour traiter les données 4](#_Toc190780065)

[3.3 Identification des outils nécessaires pour analyser les données 4](#_Toc190780066)

[Solutions d'extraction, de traitement et de visualisation 5](#_Toc190780067)

[4.1 Solutions proposées pour l'extraction des données 5](#_Toc190780068)

[4.2 Solutions proposées pour le traitement des données 6](#_Toc190780069)

[4.3 Solutions proposées pour la visualisation des données 7](#_Toc190780070)

[Cohérence des solutions avec le besoin 8](#_Toc190780071)

[5.1 Explication détaillée de chaque solution proposée 8](#_Toc190780072)

[5.2 Alignement des solutions avec les besoins identifiés 10](#_Toc190780073)

[5.3 Avantages et limites de chaque approche 13](#_Toc190780074)

[Conclusion 16](#_Toc190780075)

[Synthèse des principales conclusions de l'analyse 16](#_Toc190780076)

[Recommandations pour la mise en œuvre des solutions proposées 16](#_Toc190780077)

## Introduction

### Objectif du rapport

Ce rapport a pour objectif d'analyser les différentes solutions permettant d’exploiter les données issues de la base de données dans un outil de visualisation. L’objectif est d’identifier la solution la plus adaptée en tenant compte des contraintes techniques et des besoins fonctionnels.

### Contexte de l'analyse

L’entreprise cherche à améliorer sa prise de décision grâce à une meilleure exploitation de ses données. L’utilisation efficace des données dans des outils de visualisation est cruciale pour la prise de décision stratégique. Les données actuelles incluent des informations riches telles que les performances produits (ventes, prix, prix d’achat…), les avis clients, les promotions. Cependant leur exploitation est limitée en l’absence de mise en place de solution pérenne pour leur visualisation.

## État des lieux et évaluation de la pertinence

### 2.1 Description de la situation actuelle

Actuellement les données sont stockées dans une base relationnelle. L’accès aux données se fait par des requêtes SQL ponctuelles ou des extractions manuelles en CSV ce qui limite la rapidité d’exploitation ce celle-ci.

### 2.2 Évaluation de la pertinence des données existantes

Les données présentes sont riches et variées, incluant des indicateurs de performance, des retours clients et des métriques financières. Elles sont globalement fiables et exploitables. Toutefois, des problèmes de qualité des données (données manquantes, incohérences) ont été identifiés. De plus toutes ne sont pas pertinentes à une utilisation à des fins de reporting.

### 2.3 Identification des principaux problèmes ou enjeux

Différents problèmes et enjeux ont été constatés :

* Une difficulté d’accès aux données : aucun accès n’est automatisé.
* Besoin de transformation des données pour répondre aux besoins des outils de visualisation.
* Nécessité de garantir la conformité RGPD.
* Difficulté à traiter de grands volumes de données en temps réel.

## Besoin d'outils

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

Différents outils sont disponibles pour la collecte de données tel que :

* Power BI Gateway pour la connexion directe à la base de données.
* ETL (Knime, Talend) pour automatiser l’extraction des données.
* Les extractions en CSV
* API d’extraction
* Replication (Data Warehouse ou Data Lake)

### 3.2 Identification des outils nécessaires pour traiter les données

Différents outils sont aussi disponibles pour le traitement de données :

* Power Query pour les transformations simples.
* ETL (Knime, Talend, Alteryx) pour les traitements complexes.
* Scripts Python/SQL pour les besoins spécifiques.
* Data Warehouse / Data Lake
* Outils d’automatisation (Airflow, Azure Data Factory)

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

Différentes solutions de visualisations s’offrent à nous :

* Power BI pour la visualisation interactive.
* Tableau pour les analyses exploratoires.
* Google Data Studio pour les rapports simples.
* Excel (Power Pivot)
* Google Data Studio
* Outils spécialisés

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

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

Voici un **tableau comparatif** pour synthétiser les différentes solutions d'extraction des données :

| **Solution** | **Principe** | **Avantages** | **Limites** | **Cas d’usage recommandé** |
| --- | --- | --- | --- | --- |
| **Connexion directe (Power BI Gateway)** | Connexion temps réel à la base de données source | - Mise à jour en continu  - Pas d’étape intermédiaire | - Dépend des performances de la base  - Charge élevée | Suivi en temps réel et tableaux de bord opérationnels |
| **Extraction en CSV** | Export manuel ou automatisé sous forme de fichier CSV | - Facile à mettre en place  - Compatible avec tous les outils | - Processus manuel  - Données rapidement obsolètes | Analyses ponctuelles ou archivage |
| **ETL (PowerQuery, Knime, Talend, etc.)** | Automatisation de l’extraction et transformation des données | - Automatisation complète  - Traitement des gros volumes | - Complexité et coût d’implémentation  - Besoin d’expertise | Flux réguliers et besoins de transformation complexe |
| **API d’extraction** | Utilisation d’API pour accéder aux données | - Sécurisé et automatisé  - Adapté aux environnements cloud | - Limites des API (requêtes, structure figée)  - Dépendance | Données SaaS ou issues de plateformes externes |
| **Replication (Data Warehouse ou Data Lake)** | Centralisation des données dans une base dédiée | - Gestion des gros volumes  - Historisation des données | - Coût et complexité élevés  - Infrastructure requise | Projets longs nécessitant une gestion avancée |

### 4.2 Solutions proposées pour le traitement des données

De même voici un tableau comparatif des différents outils de traitement des données :

| **Solution** | **Principe** | **Avantages** | **Limites** | **Cas d’usage recommandé** |
| --- | --- | --- | --- | --- |
| **Power Query (Power BI)** | Nettoyage et transformation dans Power BI | - Intégré et facile à utiliser  - Automatisation possible | - Moins adapté aux gros volumes  - Transformations limitées | Traitement rapide avant visualisation |
| **ETL (Knime, Talend, Alteryx)** | Collecte, transformation et chargement automatisés | - Supporte les gros volumes  - Multi-sources et automatisé | - Complexité et coût pour certaines solutions | Traitement complexe et flux régulier |
| **Scripts personnalisés (Python, SQL, R)** | Automatisation avec des scripts sur mesure | - Flexibilité maximale  - Puissant pour besoins spécifiques | - Maintenance complexe  - Nécessite des compétences en code | Cas spécifiques non pris en charge par les outils |
| **Data Warehouse / Data Lake** | Traitement dans une base centralisée (Snowflake, Synapse) | - Performant pour très grands volumes  - Historisation | - Infrastructure nécessaire  - Complexité initiale | Gouvernance et projets à grande échelle |
| **Outils d’automatisation (Airflow, Azure Data Factory)** | Orchestration et automatisation des pipelines de données | - Automatisation avancée  - Évolutivité et monitoring | - Besoin d’expertise DevOps  - Coût parfois élevé | Pipelines de données automatisés |

### 4.3 Solutions proposées pour la visualisation des données

Ci-dessous le tableau comparatif des outils de datavisualisation :

| **Solution** | **Principe** | **Avantages** | **Limites** | **Cas d’usage recommandé** |
| --- | --- | --- | --- | --- |
| **Power BI** | Rapports interactifs et dynamiques | - Intuitif  - Intégration Microsoft 365 | - Performance limitée sur gros volumes | Visualisation opérationnelle et analytique |
| **Tableau** | Exploration visuelle des données | - Visualisations avancées  - Large communauté | - Coût élevé  - Courbe d’apprentissage | Analyses exploratoires et storytelling |
| **Excel (Power Pivot)** | Tableaux croisés dynamiques | - Déjà connu  - Simple à utiliser | - Non adapté aux gros volumes | Analyses rapides |
| **Google Data Studio** | Visualisation cloud gratuite | - Gratuit - Partage facile | - Moins puissant - Peu de personnalisation | Rapports marketing et données Google |
| **Outils spécialisés** | Visualisation sur mesure | - Cas d’usage complexe  - Reporting temps réel | - Coût élevé  - Courbe d’apprentissage | Projets spécifiques (SAP, Google Cloud) |

## Cohérence des solutions avec le besoin

### 5.1 Explication détaillée de chaque solution proposée

Voici un panorama des solutions proposées pour l’extraction des données, en fonction du contexte et des besoins spécifiques :

#### 1. Connexion directe à la base de données (via Power BI Gateway)

**Principe :**

Connexion en temps réel à la base de données source via Power BI Gateway.

**Avantages :**

* Mise à jour des données en continu.
* Éviter les étapes intermédiaires d’extraction.
* Centralisation et cohérence des données.

**Limites :**

* Dépend des performances de la base de données.
* Charge importante sur la base en cas de nombreuses requêtes simultanées.

**→ Recommandé pour :** Suivi en temps réel et tableaux de bord opérationnels.

#### 2. Extraction en CSV (Export manuel ou automatisé)

**Principe :**

Export des données sous forme de fichier CSV à partir de la base de données, puis import dans Power BI ou d’autres outils.

**Avantages :**

* Facile à mettre en place.
* Permet de travailler hors ligne sur un instantané des données.
* Pas de dépendance directe à la base source.

**Limites :**

* Processus manuel répétitif ou nécessitant une automatisation par script.
* Risque d’erreurs ou d’obsolescence rapide des données.
* Non adapté pour les gros volumes de données (> 1 million de lignes).

**→ Recommandé pour :** Analyses ponctuelles ou besoin d’archivage.

#### 3. Utilisation d’un ETL (Extract, Transform, Load)

**Principe :**

L’ETL collecte les données depuis la base, les transforme selon les besoins (nettoyage, agrégation, etc.), puis les charge dans une base dédiée ou directement dans Power BI.  
**Outils possibles :** PowerQuery (Power BI), Talend, Knime, Alteryx.

**Avantages :**

* Automatisation complète du processus d’extraction et transformation des données.
* Idéal pour des volumes importants.
* Permet d’harmoniser des données issues de plusieurs sources.

**Limites :**

* Complexité de mise en œuvre initiale.
* Coût élevé pour certaines solutions (Talend, Alteryx).
* Nécessite une expertise technique pour la configuration.

**→ Recommandé pour :** Projets nécessitant des flux de données réguliers et complexes.

#### 4. API d’extraction des données

**Principe :**

Extraction via une API (Application Programming Interface) pour accéder aux données en temps réel ou quasi temps réel.

**Avantages :**

* Accès contrôlé et sécurisé aux données.
* Adapté aux environnements cloud et aux bases SaaS (par exemple : HubSpot, Salesforce).
* Permet d’automatiser l’extraction.

**Limites :**

* Dépendance à la disponibilité et aux limites de l’API (limite de requêtes, structure figée des données).
* Configuration parfois complexe.

**→ Recommandé pour :** Données provenant d’applications SaaS ou plateformes externes.

#### 5. Replication dans une base dédiée (Data Warehouse ou Data Lake)

**Principe :**

Les données sont extraites et centralisées dans une base intermédiaire (Data Warehouse ou Data Lake) avant d’être utilisées dans Power BI.

**Avantages :**

* Gestion efficace de très grands volumes de données.
* Préparation des données en amont pour optimiser les performances des tableaux de bord.
* Historisation et gestion des versions de données.

**Limites :**

* Coût et complexité d’implémentation.
* Nécessite une infrastructure supplémentaire (Snowflake, Azure Synapse, BigQuery, etc.).

**→ Recommandé pour :** Projets nécessitant une gestion avancée des données sur le long terme

Nous privilégierons la solution : Power BI Gateway qui nous permettra des :

* Mise à jour des données en continu.
* Éviter les étapes intermédiaires d’extraction.
* Centralisation et cohérence des données.

Ce qui nous permettra un suivi en temps réel et tableaux de bord opérationnels.

### 5.2 Alignement des solutions avec les besoins identifiés

Voici les principales solutions proposées pour le traitement des données, adaptées selon le volume, la complexité des données et les objectifs du projet.

#### 1. Power Query (Intégré à Power BI)

**Principe :**

Power Query permet de transformer les données directement dans Power BI : nettoyage, fusion, suppression des doublons, ajout de colonnes calculées, etc.

**Avantages :**

* Intégré à Power BI, facile à utiliser.
* Idéal pour des transformations simples et moyennes.
* Automatisation des étapes de traitement via les requêtes M.

**Limites :**

* Moins adapté pour des volumes très importants (> 10 millions de lignes).
* Complexité croissante pour des transformations avancées.

**→ Recommandé pour :** Nettoyage et traitement rapide de données avant visualisation.

#### 2. Utilisation d’un ETL (Knime, Talend, Alteryx)

**Principe :**

Les ETL permettent de collecter, transformer et charger les données dans une base ou directement dans Power BI après traitement.  
**Outils :** Knime (open source), Talend, Alteryx (payant).

**Avantages :**

* Automatisation avancée du traitement.
* Supporte les gros volumes et flux de données complexes.
* Intégration facile de multiples sources (bases de données, fichiers plats, API).

**Limites :**

* Besoin d’expertise technique.
* Coût pour certaines solutions (Alteryx, Talend).

**→ Recommandé pour :** Traitement massif et transformation complexe de données multi-sources.

#### 3. Scripts personnalisés (Python, SQL, R)

**Principe :**

Utilisation de scripts pour automatiser le traitement des données : nettoyage, agrégation, enrichissement, etc.  
**Langages courants :** Python (Pandas, NumPy), SQL (requêtes complexes), R (analyses statistiques).

**Avantages :**

* Flexibilité maximale.
* Puissant pour les besoins spécifiques.
* Large communauté de support.

**Limites :**

* Maintenance plus complexe.
* Besoin de compétences en programmation.

**→ Recommandé pour :** Cas spécifiques non pris en charge par les outils ETL ou Power Query.

#### 4. Traitement dans un Data Warehouse ou Data Lake

**Principe :**

Le traitement est effectué directement dans la base de données centralisée (ex. : Snowflake, Azure Synapse), avant que les données soient envoyées dans Power BI.

**Avantages :**

* Performant pour de très grands volumes.
* Historisation et versioning des données.
* Simplifie l’architecture du traitement.

**Limites :**

* Dépendance à l’infrastructure existante.
* Complexité d’implémentation initiale.

**→ Recommandé pour :** Projets à grande échelle nécessitant une gouvernance robuste des données.

#### 5. Automatisation via des outils de workflow (Airflow, Azure Data Factory)

**Principe :**

Orchestration et automatisation des pipelines de données, assurant l’enchaînement des tâches de traitement (extraction, transformation, chargement).

**Avantages :**

* Automatisation des processus complexes.
* Notifications et suivi des erreurs.
* Évolutivité élevée.

**Limites :**

* Besoin d’expertise DevOps.
* Coût pour certaines solutions (Azure Data Factory).

**→ Recommandé pour :** Traitement de données dans un cadre automatisé et scalable.

Les besoins identifiés en termes d’actualisation des données, de traitement et de visualisation sont en phase avec les solutions proposées, notamment l’approche hybride entre connexion directe et ETL. L’utilisation de Power Query semble à privilégier puisque :

* Intégré à Power BI, facile à utiliser.
* Il permet de transformer les données directement dans Power BI : nettoyage, fusion, suppression des doublons, ajout de colonnes calculées, etc.
* Idéal pour des transformations simples et moyennes.
* Automatisation des étapes de traitement via les requêtes M.

Ce qui nous permettra un nettoyage et traitement rapide de données avant visualisation.

### 5.3 Avantages et limites de chaque approche

Voici les solutions proposées pour la visualisation des données, en fonction de l’objectif, des utilisateurs finaux et des besoins en termes d’interactivité et de performance.

#### 1. Power BI (Desktop & Service)

**Principe :**

Solution complète pour créer, partager et consulter des rapports interactifs et tableaux de bord dynamiques.

**Avantages :**

* Interface intuitive.
* Nombreux connecteurs de données.
* Visualisations personnalisables et interactivité avancée.
* Intégration facile avec Microsoft 365 (Excel, Teams, SharePoint).

**Limites :**

* Performance limitée sur les très gros volumes sans optimisation.
* Coût de la version Premium pour les grandes entreprises.

**→ Recommandé pour :** Visualisation opérationnelle et analytique pour des utilisateurs métiers.

#### 2. Tableau

**Principe :**

Solution leader dans la visualisation avancée, idéale pour l’exploration visuelle des données.

**Avantages :**

* Visualisations interactives et esthétiques.
* Performant pour l’exploration visuelle et les analyses ad hoc.
* Large communauté d’utilisateurs.

**Limites :**

* Coût élevé pour les licences.
* Courbe d’apprentissage plus longue que Power BI.

**→ Recommandé pour :** Analyses exploratoires et besoins en storytelling visuel.

#### 3. Excel (avec Power Pivot)

**Principe :**

Visualisation via des tableaux croisés dynamiques et graphiques dans Excel, avec possibilité d’intégration des modèles de données Power Pivot.

**Avantages :**

* Déjà connu de nombreux utilisateurs.
* Simple pour les petits jeux de données.
* Intégration avec Power BI.

**Limites :**

* Non adapté pour des volumes importants.
* Moins interactif que Power BI ou Tableau.

**→ Recommandé pour :** Analyses rapides sur des données simples ou pour complément dans Power BI.

#### 4. Google Data Studio

**Principe :**

Outil de visualisation gratuit pour créer des rapports interactifs, principalement orienté cloud (Google Analytics, BigQuery, etc.).

**Avantages :**

* Gratuit.
* Facilement partageable via un simple lien.
* Idéal pour les données provenant de l’écosystème Google.

**Limites :**

* Moins puissant que Power BI pour les gros volumes ou les transformations complexes.
* Moins de personnalisation.

**→ Recommandé pour :** Rapports marketing et analyses rapides sur des données Google.

#### 5. Outils spécialisés (Qlik Sense, SAP Analytics Cloud, Looker)

**Principe :**

Solutions spécifiques pour des cas d’usage précis : reporting avancé, BI en temps réel, ou visualisation intégrée dans des applications.

**Avantages :**

* Adaptés aux contextes complexes.
* Reporting en temps réel possible.
* Intégration dans les écosystèmes spécifiques (SAP, Google Cloud, etc.).

**Limites :**

* Coût élevé pour certaines solutions.
* Courbe d’apprentissage selon l’outil.

**→ Recommandé pour :** Cas d’usage spécifiques dans des environnements métiers complexes.

## Conclusion

### Synthèse des principales conclusions de l'analyse

L’analyse montre que chaque solution a ses points forts selon les besoins spécifiques de l’entreprise. Dans notre cas et aux vues de nos problématiques nous privilégierons les solutions suivantes :

* **Power BI Gateway** pour les données en temps réel, offrant une mise à jour automatique des rapports et un accès direct aux données de production. Ce qui permettra aux chefs de produits de pouvoir effectuer un suivi journalier.
* **Power Query** pour les transformations simples et rapides, directement dans Power BI.
* **Power BI** comme outil principal de visualisation, déjà pour ses différents connecteurs de données mais aussi en raison de son interface intuitive, de ses visualisations personnalisables et interactivité avancée ainsi que de son intégration avec l’écosystème Microsoft

Ainsi Power BI nous permet de gérer l’ensemble des problématiques rencontré sans pour autant nécessiter la mise en place d’une multitude de solutions et des différentes limitations inhérentes

### Recommandations pour la mise en œuvre des solutions proposées

#### Adopter une approche progressive :

* + Mettre en place rapidement la **connexion directe via Power BI Gateway** pour les besoins immédiats en reporting opérationnel. Pour cela il va falloir utiliser les drivers ODBC pour connecter la Base de Données SQLITE avec Power BI
  + Déployer une solution **ETL** pour automatiser les processus de traitement les plus complexes et récurrents.

#### Renforcer la gouvernance des données :

* + Mettre en place des règles claires de nettoyage et d’enrichissement des données. Fusionner les tables en relation 1 to 1 et mettre en place différentes mesures nécessaires au reporting. Ainsi que la mise en place de certaines transformations pour pallier les erreurs de saisies et autres problèmes liés aux données
  + Désigner des responsables de la qualité des données et documenter les processus.

#### Former les équipes :

* + Former les utilisateurs clés à Power BI pour favoriser l’adoption et maximiser l’exploitation des données.
  + Former les équipes techniques à l’utilisation des outils ETL pour garantir leur autonomie.

#### Surveiller les performances :

* + Mettre en place un monitoring des performances des rapports Power BI.
  + Adapter les rapports Power BI selon les besoins/préférences des utilisateurs
  + Prévoir une réévaluation régulière afin qu’ils persistent dans le temps via l’apport/ mise à jour de différentes visualisations