

Rapport d'Analyse sur les solutions et les préconisations d'utilisation de la base de données Bottleneck

Date: 03/03/2025 François JOLLY.

1 Introduction

1.1 Objectif du rapport

L'objectif du rapport est d'explorer et de comparer plusieurs solutions **d'utilisation de la base de données** de Bottleneck, afin de recommander celle qui est la plus adaptée pour permettre aux équipes **d'accéder aux données** via un outil de **data visualisation**.

1.2 Contexte de l'analyse

Bottleneck est une entreprise spécialisée dans la vente de vin. Actuellement, elle rencontre des **difficultés** dans **l'utilisation** et le **suivi** des **données** à sa disposition, ce qui impacte sa **compétitivité** par rapport à ses concurrents. Cette situation affecte également son **organisation** dans le suivi des différentes données disponibles et limite sa capacité à **croître** en explorant de nouveaux segments de marché.

2 État des lieux et évaluation de la pertinence

2.1 Description de la situation actuelle

Les données provenaient de différentes sources externes, notamment l'extraction de l'ERP et du site web. Cependant, les bases de données devenaient vieillissantes, entraînant de nombreuses erreurs. Un fichier externe était même nécessaire pour lier ces différentes bases. Suite à cela, une refonte des bases de données a été effectuée. Un consultant dédié a corrigé, simplifié, fiabilisé et optimisé les données. Aujourd'hui, l'ensemble des données de Bottleneck est maintenant regroupé au sein d'une base de données unique et fiable.

2.2 Évaluation de la pertinence des données existantes

Les données de Bottleneck sont désormais nettoyées, fiables et regroupées au sein **d'une seule base de données cohérente**. Cette base est organisée en différentes **tables**, chacune **dédiée** à un **domaine** ou une thématique :

- Web: Contient les informations et caractéristiques des produits vendus en ligne.
- Finance: Indique quel produit a été vendu à telle date, avec son prix de vente et son prix d'achat.
- Sales: Fournit le nombre de produits vendus au cours du mois, ainsi que l'état des stocks et les commandes des fournisseurs.
- Promo: Répertorie le nombre de produits vendus pendant les périodes promotionnelles, ainsi que les prix appliqués.

2.3 Identification des principaux problèmes ou enjeux

L'un des enjeux majeurs est de déterminer quelle **solution optimale** permettrait à Bottleneck de tirer le meilleur parti de ses données pour chaque département.

« Quelles sont, parmi les différents outils et méthodes, la meilleure solution et la plus **adéquate** pour Bottleneck pour **utiliser les données** de la base de données et les **mettre à disposition**, les **visualiser** et les **analyser** pour chaque département de **l'entreprise** ? »

Chaque département doit avoir accès à des visualisations détaillées pour soutenir leurs décisions quotidiennes, telles que :

- visualiser et suivre des KPIS dans le temps (marge en valeur, prix HT, rotation des stocks, etc.)
- visualiser l'évolution des prix d'achat par rapport à l'inflation.
- Avoir des informations et des analyses des performances sur les différents marchés (vin, spiritueux, gamme sans alcool, etc.) et sur les produits (évolution du CA, quantités, les tops produits, les flops produits, etc.) pour les chef de produits.
- Suivre les tendances sur les nouveaux marchés : comme la gamme de produits sans alcool qui viennent d'être lancée.
- Suivre la performance promotionnelle : analyser le ROI (Return On Investment) sur les offres promotionnelles

Ainsi, cela permettra **d'optimiser** la compétitivité, l'organisation et la croissance de l'entreprise dans un marché en constante **évolution**.

3 Besoin d'outils

3.1 Identification des outils nécessaires pour collecter les données

Différents outils pourrait être utiliser pour collecter/extraire les données :

- des bases de données (SQL, NoSQL)
- des fichiers plats (CSV, Excel)
- des API pour récupérer des données externes

3.2 Identification des outils nécessaires pour traiter les données

Différents outils pourrait être utiliser pour **traiter les données** nettoyer, éliminer les incohérences, les doublons, et les rendre exploitables, normaliser les formats, etc...:

- des scripts en **Python** (avec pandas, NumPy..)
- des ETL intégrés tel que **Power Query** (intègre dans Power Bl , Excel..)
- des ETL standalone comme Knime
- des logiciels **spécifiques** de traitement de données

3.3 Identification des outils nécessaires pour analyser les données

Différents outils pourrait être utiliser pour analyser les données en créant des visualisations graphique ou des tableaux de bord :

- Des Tableurs comme Excel
- Des scripts R ou Python avec des librairies dédiées à la visualisation (Matplotlib, Seaborn, Plotly...) ou d'analyses statistiques / modèles prédictifs (scikit-learn...)
- Des logiciels de visualisation dédiés tel que Tableau, Power BI, Google Data Studio

4 Solutions d'extraction, de traitement et de visualisation

4.1 Solutions proposées pour l'extraction des données

Pour l'extraction des données, la solution choisie est une **extraction direct des données en CSV** de la base de données de Bottleneck :

 □ une extraction direct de chaque Table Web/Finance/Sales/Promo en utilisant par exemple dans SQLite Studio la fonction/Outils « Exporter » directement disponible sous un format CSV encodé en UTF8

4.2 Solutions proposées pour le traitement des données

Pour le traitement des données, la solution choisie consiste à utiliser directement l'ETL Power Query intégré à l'outil de Visualisation PowerBI afin de pouvoir traiter/préparer /transformer chaque fichier CSV préalablement exporté dans l'étape d'extraction et par la suite éliminer les incohérences, normaliser les formats et ainsi rendre les données exploitables pour la Visualisation.

4.3 Solutions proposées pour la visualisation des données

Pour la Visualisation des données, la solution choisie consiste à utiliser l'outil de Visualisation Power BI couplé a son ETL Power Query retenue précédemment pour le traitement de Donnés, afin de fournir des visuelles de qualité interactifs et dynamiques pour chaque utilisateurs.

5 Cohérence des solutions avec le besoin

5.1 Explication détaillée de chaque solution proposée

5.1.1 Au niveau de l'extraction des données :

La méthode d'utiliser un ETL seul afin d'extraire les fichiers CSV automatiquement et de produire un fichier global à plat contenant toutes les données n'a pas été retenue dans le contexte de BottleNeck car la séparation par fichier pour chaque table Web/Finance/Sales/Promo fait déjà du sens fonctionnellement pour dissocier chaque spécificité lié à l'activité.

Même si avec cette option on n'aurait pas besoin de gérer le schéma en étoile au niveau de l'outil PowerBI, ni de maintenir une connexion permanente à la base de donnée, 1 seul table unique aurait été plus lourd à manipuler dans l'outil de Visualisation, moins souple avec des redondances d'information, tout en complexifiant également la mise à jour des données.

La méthode d'utiliser une connexion directe à la base de données (Direct Query) n'a pas été retenue car le besoin avant tout de l'entreprise n'est pas un suivi temps réel de l'activité de vente et des stocks contrairement à une activité boursière : le besoin est plutôt une analyse des ventes sur une période données pour en tirer des conclusions sur des tendances et des directions commerciales à prendre (pas une analyse temps réelle de ce qui se vend)

Cette méthode est également avantageuse lorsque l'on manipule une grosse volumétrie de donnée ce qui n'est pas le cas à l'heure actuel pour BottleNeck. Aussi des contraintes telles que fonctionnalités DAX très limitée, l'impossibilité de créer des colonnes calculées , un accès a ETL Power Query très limité et des performances moindre (lenteur) dues au performance du réseau et de l'accessibilité du server de BDD (connexion permanente), restent des limitations importantes dans le contexte de l'entreprise.

La méthode d'utiliser une extraction direct des données en CSV de la base de données (Mode Import) de Bottleneck a été retenue car elle offre de nombreux avantages :

- La volumétrie des données de l'entreprise reste globalement faible pour ne pas recourir à une approche Direct Query
- La possibilité d'effectuer une export régulier de chaque fichier CSV nommé par domaine afin de pouvoir mettre à jour les visualisation à une fréquence souhaité (ou le temps réel n'a pas d'utilité pour les analyses Bottleneck). Cette solution est adaptée aux données plus statiques qui ne nécessitent pas de mises à jour fréquentes c'est bien le cas de BottleNeck.
- La possibilité de pouvoir profiter pleinement des capacités de traitement / préparation et transformation des données de l'ETL Power Query.
- Des données rapidement accessibles avec de bonne performances d'affichage des rapports
- Toutes les fonctionnalités DAX disponibles dans l'outil PowerBI afin de pourvoir créer des colonnes calculées par exemple
- ⇒ Cela reste donc une solution très souple et flexible même si une actualisation des données est nécessaire et parfois limité (dépend de l'abonnement Power Bi) afin de réimporter des nouvelles données plus récentes.

Ci-dessous un tableau **comparatif** établi par notre équipe pour ce projet comparant les **Mode Import** vs **Mode Direct Query** principalement utilisés avec l'outil Power BI :

Méthode	Description	Avantages	Inconvénients
Extraction des données en CSV de la base de données : Mode Import	Mise en cache des données / copie des donnée totalité du contenu (table) de la base de données sera importé et stocké directemen t dans Power BI	Souple et flexible donnée rapidement accessible / meilleures performances toutes les fonctionnalités DAX disponibles accès complet a ETL Power Query possibilité d'importer des données en provenance de plusieurs sources vitesse de calcul est plus élevée / affichage plus rapide des rapports adapté au volume de données faible ou moyen	 actualisation nécessaire et limité (dépend abonnement PBI) afin de réimporter les nouvelles données nécessite un rafraîchissement manuel ou programmé pour la maj des données adapté au données plus statiques qui ne nécessitent pas de mises à jour fréquentes Pas de données actualisé en temps réel Plus la taille du jeux de donné augmente plus la taille fichier power bi augmente pas adapté aux gros volumes de données
Connexion directe à la base de données: Mode Direct Query (pilote dédié ODBC sqlite nécessaire)	connexion direct aux données de la BDD pas de mise en cache ou de données stockés en locale la source de données est interrogée (requête) à chaque nouvelle requête et interaction avec le logiciel de Data Viz Power BI	 jeux de données en temps réel (reflet de la base actuelle / activité boursière) nécessitent pas de mises à jour manuelles ou programmés fréquentes / => adapté aux données qui changent fréquemment contrairement au mode import. Gros volumes de données (difficile à charger dans un fichier) très dynamiques Pas de stockage des données (fichier Dataviz .pbi allégé) 	une seule source de données à la fois performances moindres fonctionnalités DAX très limitée (fonctions temporelles pas dispo) pour optimiser et garantir l'instantanéité de l'information et garantir l'optimisation du temps des requêtes pas de possibilité de créer des colonnes calculées ou de faire des Calculation Group accès a ETL Power Query très limité car modifications traduites dans le langage de requête de la base de données avoir une haute dispo du server de BDD et du réseaux (ressources suffisantes) sinon lenteur d'affichage maintenir une connexion permanente à la base de donnée

5.1.2 Au niveau du traitement des données :

L'ETL Power Query intégré à l'outil de Visualisation PowerBI a été choisi afin de de pouvoir traiter/ préparer /transformer chaque donnée de chaque fichier CSV et surtout de pouvoir rejouer automatiquement toutes les étapes de nettoyage/normalisation à chaque nouveaux jeux de données exportés de la base de données.

Ainsi des traitements ont été effectués dans chacun des fichiers CSV/Tables:

Web:

- Suppression des colonnes redondantes ou sans information additionnelles pertinentes (ie redondantes ou avec valeurs uniques) => (Virtuals / Downloadable / Post_author / Post_date / comment status / ping status / Post_type / post_parent etc...)
- Normalisation des Données: Création d'une nouvelle colonne « Contenance Modifiée» qui contient la contenance manquante principalement des produits sans alcool mais pourtant disponible à la fin de la description « Post title » (traitement PowerQuery)
- Normalisation des Donnés : Remplacement des mauvaises dates et format dans « Post_modified» / « Post_modified_gmt » : 29/02/2022 15:25:02 => 28/02/2022

Finance:

- Normalisation des Données : Remplacement des mauvaises valeurs « on Sales web » 4 =>0
- Unification nom de colonne / tables pour « Key_Finance »)=> « key » afin d'autoriser les Jointures/liaisons entre table dans l'outil PowerBI

Sales:

 Unification nom de colonne / tables pour « Key_Sales »)=> « key » afin d'autoriser les Jointures/liaisons entre table dans l'outil PowerBI

Promo:

• Unification nom de colonne / tables pour « Key_Promo »)=> « key » afin d'autoriser les Jointures/liaisons entre table dans l'outil PowerBI

5.1.3 Au niveau de la Visualisation des données

L'outil de Visualisation Power BI couplé à son ETL Power Query a été choisi comme solution de Visualisation des données car en plus de fonctionner conjointement avec les traitement Power Query, il ne nécessite pas de licence spécifique contrairement à l'outil Tableau pour fournir un tableau de bord directement accessible pour l'équipe, c'est également l'outil de Data visualisation le mieux maitrisé par nos équipe Data-Viz.

Sa gestion par défaut du Mode Import est entièrement compatible avec l'export des Fichier CSV vu précédemment et permet d'être directement paramétré pour lire ces données exportées régulièrement et rafraichir les tableaux de bord en conséquence.

Au niveau de la visualisation sa **facilité** d'utilisation et ses capacités de navigation **simple** et de **filtrage intuitif** entre les différents **graphiques**, à la fois **interactifs** et **dynamiques** sont un grand atout pour répondre à la demande d'analyse des ventes des équipes BottleNeck.

5.2 Alignement des solutions avec les besoins identifiés

Les solutions retenues sont donc les plus adéquates pour Bottleneck et **les plus alignées en terme de besoin utilisateurs** d'analyse et de visualisation fonctionnelles pour les différentes équipes de chaque département de l'entreprise :

- Complexité moindre avec l' Extraction direct des données en CSV de la base de données permettant un Access régulier paramétrable au différentes données stockées et collectées, l'équipes BottleNeck ne nécessitant pas de connaître des information de vente de produit à la minute près.
- ⇒ Les tableaux générés avec l'outil Power BI permettront une analyse visuelle accessible, une navigation simple, fonctionnelle et user friendly pour visualiser et suivre les évolutions de données détaillées telles que le prix d'achat, la marge en valeur, le prix HT, la rotation des stocks...

Des pages dédiées permettront l'accès direct à des informations et des analyses de performances des ventes et des promotions sur les différents marchés pour les chef de produits avec suivi des tendances.

Les fonctionnalité intuitives de PowerBI permettront également à un **personnel moins technique** d'ajouter facilement des indicateurs de performance **KPIS** avec la fonctionnalité de création de **mesures** Power BI.

5.3 Avantages et limites de chaque approche

Méthodes	Avantages	Limites
Extraction direct des données en CSV de la base de données	 Simple, Souple et flexible donnée rapidement accessibles dans l'outil de DataViz Bonnes performances: vitesse de calcul plus élevée Traitements plus court pas de connexion permanente à la BDD Accès complet a DAX et ETL Power Query 	 Snapshot de la BDD a un instant T figé Pas d'accès aux données temps réel Nécessite un paramétrage d'extraction des fichiers et un rafraîchissement manuel ou programmé pour la maj des données pas adapté au Gros volumes de données
Traitement des données avec Power Query	 Interface intuitive Simple et Directement intégré à l'outil de DataViz Power Bi Répétabilité: Automatisation de chaque transformation /normalisation Transformations avancées des données avec DAX / M 	 Nécessite un paramétrage pour accéder aux fichiers de données plus récents afin de rejouer les étapes de nettoyage/transformation Les grandes quantités de données peuvent ralentir les processus dans Power Query Dépendance à la source de données : Si la structure de la source de données change, il peut être nécessaire de modifier les processus de nettoyage et de transformation
Visualisation des données avec Power Bl	 Avec la méthode d'extraction Direct Import, les donnée sont rapidement accessibles Bonnes performances et vitesse de calcul des les affichage de graphique Interface intuitive Tableaux de bord interactif, filtrable et dynamique en fonction de la sélection utilisateur Richesse des visualisations Schéma en étoile de modèle de données directement disponible et personnalisable en utilisant des relations entre différentes tables, facilitant ainsi des analyses plus complexes par le biais de jointures et d'agrégations 	 Affichage des données a un instant T figé Pas d'accès aux données temps réel Nécessite une Licence Power BI Server pour le partage de tableau de bords dans l'entreprise Requêtes DAX : le langage DAX peut être difficile à maîtriser pour des utilisateurs non techniques La gestion des permissions et de la sécurité des données peut devenir complexe lorsqu'on travaille avec de nombreux utilisateurs.

6 Conclusion

6.1 Synthèse des principales conclusions de l'analyse

Pour Bottleneck, la solution retenue s'articule autour d'une extraction simplifiée des données en CSV, facilitant l'accès régulier aux informations nécessaires sans exiger une précision excessive en temps réel. L'intégration de Power Query pour l'automatisation de la transformation des données optimise la normalisation, rendant les fichiers directement exploitables dans Power BI.

Les tableaux de bord créés avec **Power BI** favorisent une analyse visuelle claire et intuitive, permettant aux utilisateurs de naviguer facilement à travers des données critiques, telles que les **prix d'achat** et la **rotation des stocks**. Des pages dédiées garantiront aux **chefs de produits** un accès direct à des analyses de performances de **ventes** et de **promotions**, tout en suivant les **tendances** des différents marchés.

De plus, la convivialité de Power BI facilite l'implication du personnel moins technique dans l'ajout d'indicateurs de performance, rendant l'outil accessible à tous au sein des équipes. En somme, cette approche collaborative et automatisée vise à optimiser les performances commerciales de Bottleneck en répondant aux besoins des utilisateurs tout en assurant une visualisation fonctionnelle et efficace des données.

6.2 Recommandations pour la mise en œuvre des solutions proposées

Interface avec equipe IT BottleNeck:

Il faudra travailler conjointement avec l'équipe IT BottleNeck afin de s'assurer l'accessibilité à la base de donnée pour le paramétrage des extraction des tables en fichier CSV, et également l'accès réseaux aux serveur Power BI pour les différents départements tout cela en prenant bien en compte l'aspect sécurité des données et des accès/droits de chaque type utilisateurs.

Formation des utilisateurs :

Il faudra organiser des sessions de formation pour les utilisateurs des tableaux de bord et de Power Bl. Cela permettra de s'assurer que tout le monde comprend comment utiliser l'outil pour tirer le meilleur parti des données disponibles.

Normalisation/standardisation des données:

Il faudra mettre en place des normes pour l'importation et l'exportation des données. Cela garantit la cohérence et la qualité des données utilisées dans les tableaux de bord surtout si la base de données évolue au cours du temps.

Documentation et guides d'utilisation :

Il faudra créer des guides d'utilisation détaillés et de la documentation sur les processus, les rapports, et les bonnes pratiques. Cela facilitera l'adoption et l'utilisation des outils par les utilisateurs.

Feedback utilisateurs:

Il faudra lancer un projet pilote avec un groupe restreint d'utilisateurs pour tester les tableaux de bord. Des retours d'expérience seront les bienvenus pour améliorer les fonctionnalités avant une mise en œuvre à plus grande échelle. Aussi instaurer des sessions de retour d'expérience régulières pour comprendre les difficultés rencontrées par les utilisateurs et ajuster les formations ou la documentation en conséquence.

Mise à jour régulière des données :

Il faudra bien s'assurer que les données sont mises à jour régulièrement dans Power Bl. La mise en place du processus d'automatisation avec Power Query facilitera cette étape

Intégration avec d'autres outils :

A terme il faudra étudier l'intégration de Power BI et Power Query avec d'autres outils et systèmes utilisés dans l'entreprise pour maximiser leur utilisation et faciliter les interconnexion (automatisation de commande de stock chez le fournisseurs etc..)

Annexes

Liste des sources de données utilisées :

4 fichier CSV contenant chacune des tables et ses données extraites de la base Bottleneck.db:

- Web.csv
- Finance.csv
- Sales.csv
- Promo.csv

Description détaillée des outils recommandés

MS Power Query

Description : Power Query est un outil de préparation et de transformation de données intégré dans Microsoft Excel et Power Bl. Il permet aux utilisateurs de connecter, transformer et enrichir des données provenant de diverses sources avant de les analyser.

Fonctionnalités principales :

- Récupération de données : Il facilite l'importation et la combinaison de données provenant de plusieurs sources.
- Connexion à des sources de données variées : Power Query peut se connecter à des sources comme des fichiers Excel, des bases de données SQL, des services cloud, des fichiers textuels et bien d'autres.
 - Bases de données : SQL Server, Oracle, MySQL, Access, etc.
 - Fichiers: CSV, Excel, JSON, XML, etc.
 - Services web: APIs REST, Azure, Salesforce, etc.
 - D'autres sources : SharePoint, folders, listes Internet.

• Éditeur de requêtes :

L'éditeur de requêtes offre une interface graphique qui permet de réaliser plusieurs opérations, telles que :

- Filtrage des lignes : Ne garder que les données pertinentes.
- Transformation des colonnes : Modifier le type, renommer ou séparer des colonnes.
- Fusion et ajout de requêtes : Combiner plusieurs tables dans une seule.
- Groupement de données : Résumer ou agréger les données (calculer des moyennes, sommes, etc.).

• M (langage de formule):

Power Query utilise un langage de formule, M, qui permet de créer des étapes de transformation plus complexes. Qui plus est, les transformations

créées sont enregistrées sous forme de script, permettant de les reproduire facilement.

• Actualisation des données :

Une fois que la requête est configurée, les utilisateurs peuvent actualiser les données à tout moment, rendant le flux de travail très efficace.

- Transformations de données : Vous pouvez nettoyer et transformer les données sans avoir besoin de compétences en programmation. Des opérations telles que la suppression des doublons, le changement de type de données, la fusion de colonnes, etc., sont facilement réalisables.
- Automatisation : Une fois une requête configurée, elle peut être actualisée automatiquement pour importer les données mises à jour.
- Interface intuitive: L'interface graphique conviviale permet de réaliser des transformations simples tout en offrant la possibilité d'écrire des requêtes plus avancées en M (langage de formule de Power Query).

MS Power BI

Description : Power BI est une suite d'outils d'analyse de données qui permet de visualiser des données et de partager des insights au sein de votre organisation ou de les publier sur le web. Il se compose de plusieurs éléments, notamment Power BI Desktop, Power BI Service et Power BI Mobile.

Fonctionnalités principales :

• Visualisation des données : Power BI permet de créer des tableaux de bord interactifs et des rapports à partir des données importées. Avec des graphiques variés et des options dynamiques, les utilisateurs peuvent explorer les données en profondeur.

• Visualisations interactives:

Power BI offre une variété de visualisations (graphiques, cartes, tableaux, etc.) qui peuvent être personnalisées. Les utilisateurs peuvent interagir avec les rapports en sélectionnant des données spécifiques pour obtenir des informations plus détaillées.

• Modelage des données :

Power BI permet de construire des modèles de données robustes en utilisant des relations entre différentes tables, facilitant ainsi des analyses plus complexes par le biais de jointures et d'agrégations

- Partage et collaboration: Les rapports peuvent être publiés sur le cloud via Power BI Service, ce qui permet aux utilisateurs de collaborer, de partager des insights et d'accéder aux rapports de n'importe où.
 Intégration avec d'autres outils: Power BI s'intègre parfaitement à d'autres outils Microsoft, comme Excel, Teams et Azure, ainsi qu'avec de nombreuses applications tierces, rendant l'analyse des données encore
- DAX (Data Analysis Expressions) :

plus fluide.

- DAX est un langage de formule puissant utilisé dans Power BI pour créer des mesures, des colonnes calculées et des calculs avancés. Il permet d'exécuter des calculs complexes, des analyses temporelles, et d'effectuer des opérations sur des ensembles de données
- Mises à jour en temps réel : Power BI peut se connecter à des sources de données en temps réel, permettant aux utilisateurs d'avoir accès à des informations actualisées instantanément.

• Power BI Service:

Il s'agit d'un environnement cloud où les utilisateurs peuvent partager leurs rapports et collaborer. Il permet également d'interagir avec les rapports depuis différents appareils, favorisant la mobilité.

• Options de sécurité :

Power BI propose des options de sécurité avancées pour contrôler qui peut voir et interagir avec certaines données, en mettant l'accent sur la confidentialité des informations sensibles.

Exemples de visualisations de données proposées











