**UNIVERSITE DE CARTHAGE**

**INSTITUT SUPERIEUR E GESTION DE BIZERTE**Une image contenant texte, Police, logo, Graphique

Description générée automatiquement

**MÉMOIRE DE FIN D’ETUDE**

En vue de l’obtention du diplôme de **Licence Business Computing**

PARCOURS : **Business Intelligence**

**Sujet**

Développement d'un système d'information décisionnel (BI) pour le contrôle l'aide à la décision et la prédiction des ventes avec intégration d'une data Mart

Soutenu le …juin 2024

Réalisé par : **CHEBBI Ichrak, BEN HASSEN Zohra**

Maître de Stage : **BEN RAIS Kamel, ELFOULADH**

Encadrant Académique : **AKKARI Wafa, ISGB**

**Année Universitaire 2023-2024**

**DEDICACES ICHRAK**

**DEDICACES Zohra**

**REMERCIEMENT**

Nous exprimons notre sincère gratitude à tous les enseignants de l'ISGB pour les connaissances qu'ils nous ont transmis tout au long de notre parcours académique.

Nous exprimons notre profonde gratitude à Mme Akkari Wafa pour son précieux soutien, ses conseils avisés et son aide tout au long du projet. Sa connaissance approfondie du terrain et sa capacité à nous guider ont été des facteurs déterminants dans la réussite de notre projet.

Nous exprimons notre profonde gratitude à tous ceux qui ont contribué à notre réussite dans des conditions favorables lors de leur stage au sein de l'entreprise sidérurgique tunisienne EFOULADH.

Nous tenons à remercier notre maître de stage, M. Ben Rais Kamel, pour ses conseils constants, son expertise partagée au quotidien et son soutien précieux tout au long de la réalisation du projet.

Enfin, nous tenons également à remercier les membres du jury d’avoir accepté d’évaluer notre travail. Nous espérons qu’ils trouveront la clarté et la motivation qu’ils désirent.

**Introduction Générale**

A l’heure actuelle, le monde connaît un progrès technologique considérable dans  
tous les secteurs, grâce aux technologies de l’information, c’est une science qui étudie les techniques de traitement automatique de l’information. Il joue un rôle important dans l’évolution de la société et d’autres établissements.

C’est dans ce cadre que s’inscrit notre projet de fin d’études qui vise à développer un système d'information décisionnel (BI) pour le contrôle l'aide à la décision et la prédiction des ventes avec intégration d'une data Mart au sein de la Société Tunisienne de Sidérurgie ELFOULADH.

L’objectif principal est de Concevoir un système d’information décisionnel pour le service de vente , il comprendra des fonctionnalités de contrôles des ventes ,d’aide à la prise de décision basée sur des données analytiques et l’intégration des algorithmes de prédiction pour anticiper les tendances de ventes futures .En outre, le projet impliquera la mise en place d’un Data Mart pour optimiser le stockage et l’accès aux données visant à fournir aux utilisateurs un outils facilitant la gestion , la compréhension et l’anticipation des performances de vente au sein de l’entreprise .

Ce rapport est structuré en 5 chapitres :

Dans le 1er chapitre, nous allons décrire le contexte de notre projet.

Le chapitre 2 sera consacré à la définition des besoins, les analyser et présenter la méthodologie de gestion de projet adaptée.

Dans le 3ème chapitre, nous allons créer l’ETL.

Le reporting sera mis en œuvre au chapitre 4 où nous présenterons nos tableaux de bord avec interprétation.

Enfin le chapitre 5 va clôturer notre travail. Nous allons proposer une application web.

**CHAPITRE 1 : CONTEXTE DU PROJET ET ANALYSE DES BESOINS**

**Introduction :**  
Dans ce premier chapitre, nous commencerons par présenter l'entreprise. Ensuite, nous analyserons et critiquerons la situation actuelle. Enfin, nous passerons à l'identification de la problématique que nous étudierons, suivie de la description de notre proposition.

1. **Présentation de l’organisme d’accueil :**

La Société Tunisienne de Sidérurgie EL FOULADH, établie en 1962 et entrée en production en 1965, est une société anonyme placée sous l'égide du ministère de l'Industrie et de la Technologie. Son objectif principal est la fabrication d'acier destiné au marché national des ronds à béton.

Pendant plus de quarante ans, elle a été la seule usine sidérurgique dans le pays, produisant de l'acier liquide et le transformant en ronds à béton. Avec une capacité de production actuelle de 200 000 tonnes par an, elle occupe une position unique sur le marché.

Cette société emploie environ 1000 personnes dans une usine implantée à « Menzel Bourguiba », au nord de la Tunisie.



Figure 1: Logo ELFOULADH

1. **Carte d’identité d’ELFOULADH :**

Nom : société Tunisienne de Sidérurgie EL FOULADH

Année de création : 1962

Adresse : Route Tunis km 3, Menzel Bourguiba 7050, Tunisie

SiteWeb : www.elfouladh.com.tn

Téléphone : [+216 72 473 222](tel:2167247322221672473100) /

E-mail : [commercial@elfouladh.com.tn](mailto:commercial@elfouladh.com.tn)

1. **Activité de « ELFOULADH » :**

"EL FOULADH" exploite une usine complète et diversifie ses activités. Son principal focus est la production d'acier en billettes, utilisant un four électrique d'une capacité de 100 000 tonnes, qui sera prochainement renforcé par un second four de même capacité. Cette matière première est ensuite utilisée dans la fabrication de divers produits, notamment des ronds à béton en barres et en couronnes, ainsi que du fer marchand comme le fer carré, rond, lisse et plat.

En plus de la production d'acier, l'entreprise fabrique également une gamme de tréfilés, comprenant le fil recuit et PICK-UP, le fil clair, le fil galvanisé et le fil PCW. Ces produits sont largement utilisés dans différents secteurs industriels tels que la construction, l'énergie, l'automobile et l'électronique. Par ailleurs, EL FOULADH est spécialisée dans la conception et la fabrication de structures métalliques en acier, telles que des pylônes de transport d'énergie électrique, des pylônes d'éclairage, des supports d'antenne et des charpentes métalliques.

En complément de ses activités de fabrication, l'entreprise offre des services de galvanisation sur mesure pour répondre aux besoins spécifiques de sa clientèle.

Cette société est reconnue pour la qualité de ses produits et services et est engagée dans l'amélioration constante de ses processus de production pour répondre aux exigences du marché et offrir une satisfaction optimale à sa clientèle[[1]](#footnote-1).

1. **L’organigramme de « ELFOULADH » :**

L'organigramme est une représentation graphique de la structure organisationnelle de l’entreprise. Il présente généralement les différents départements, les postes et les relations hiérarchiques entre eux. L'organigramme permet de visualiser la répartition des responsabilités et les flux de communication au sein de l'organisation.

Une image contenant texte, diagramme, Plan, schématique

Description générée automatiquement

Figure 2 : Organigramme de l'organisme d'accueil

1. **Cadre du projet**

Dans le cadre de la gestion d’une entreprise, l'optimisation du temps de travail est cruciale. Il est essentiel pour un gestionnaire de pouvoir économiser du temps sur les tâches administratives afin de se concentrer sur des activités à forte valeur ajoutée. C'est pourquoi nous nous concentrons désormais sur la création et la mise à disposition d'un site de vente en ligne, accompagné de tableaux de bord, pour simplifier la gestion d’entreprise.

1. **Étude de l’existant**

Nous avons essayé d’étudier l'existant. Cette étude nous permettra d'identifier les points forts et les points faibles des systèmes actuels, ainsi que les besoins spécifiques des clients, afin de les prendre en considération lors de la conception et de la réalisation de notre solution. Dans cette section, nous entreprendrons donc une analyse des différentes applications web spécialisées dans le domaine de la vente électronique. Nous mettrons en évidence leurs avantages et leurs faiblesses, ce qui nous permettra de déterminer les besoins à adresser dans notre projet.

1. **Problématique :**

Après une analyse approfondie, il est évident que la société "ELFOULADH" est en retard sur le plan technologique, n'ayant pas de site de vente en ligne ni de système décisionnel en place. Cette situation la met en danger de manquer des opportunités cruciales dans un environnement commercial de plus en plus numérique. L'absence de système décisionnel signifie également un manque d'utilisation efficace des données pour guider les prises de décision stratégiques. En résumé, l'entreprise doit rapidement combler ces lacunes pour rester compétitive et prospérer dans un marché en constante évolution.

1. **Solution envisagée :**

Pour répondre aux besoins spécifiques de "ELFOULADH", le projet vise à développer une solution générique qui répondra également aux exigences des utilisateurs sur le marché. Cette solution prendra la forme d'une application web divisée en deux parties distinctes : une plateforme de vente en ligne et un tableau de bord administratif. L'objectif est de fournir une solution complète qui permettra un traitement efficace, fiable et simplifié de la gestion de l'entreprise, tout en offrant un support client accessible et réactif.

1. **Identification des acteurs :**

Un acteur représente une entité externe qui engage des interactions avec ce système dans le but d'accomplir des tâches spécifiques ou d'obtenir des informations via une interface dédiée. Dans notre système, il y’a cinq acteurs :

- Direction Générale : Ses principales tâches consistent à consulter les tableaux de bord et la prise de décisions.

- Client : peut consulter le catalogue de produits, effectuer des recherches aisées, gérer son compte, ajouter des articles au panier, effectuer des paiements sécurisés, obtenir de l'assistance via un chatbot, et évaluer les livreurs.

- Administrateur : C’est le responsable des commandes, ses tâches consistent à gérer les utilisateurs, gérer leurs rôles.

- Sous-direction vente : C’est le responsable de la plateforme, ses tâches consistent à gérer les produits.

- Livreur : Marquer les commandes comme livrées et mettre à jour leurs statuts.

1. **Spécification des besoins :**
2. **Les besoins fonctionnels :**

Les besoins fonctionnels décrivent les actions que le système à développer doit entreprendre en réponse aux demandes des utilisateurs.  
Notre système doit posséder les différentes fonctionnalités, décrites ci-dessous :

* Besoin Fonctionnels d’un directeur général :

- Accès à des rapports et des tableaux de bord consolidés présentant le

performances globales de l'entreprise.

* Besoin Fonctionnels d’un client :

- Consultation du catalogue de produits avec des informations détaillées.

- Rechercher facilement des produits en fonction de critères spécifiques.

-Créer et de gérer leur compte sur le site

-Ajouter des produits à leur panier d'achat, de visualiser le contenu du

panier, de modifier les quantités

-Paiement sécurisées, telles que les cartes de crédit, les virements

bancaires

-Interaction avec un chatbot pour obtenir de l’aide et des réponses aux

questions fréquentes

-Evaluer et commenter les livreurs

* Besoin Fonctionnels d’un Admin :

- Crée, modifier, supprimer des comptes de sous-direction vente.

-Crée, modifier, supprimer des comptes livreurs.

-Crée, modifier, supprimer des comptes directeurs générales.

-Communication avec livreur.

* Besoin Fonctionnels d’un Sous-direction vente :

- Ajouter, modifier, supprimer des produits en fer : description, prix,

quantité en stock et promotion.

-Suivi des paniers.

-Gestion des modes de paiement.

* Besoin Fonctionnels d’un Livreur :

-Marquer les commandes comme livrées et mettre à jour leurs statuts.

1. **Les besoins non fonctionnels :**

Afin de répondre aux besoins fonctionnels mentionnés ci-dessus, les contraintes ci-dessous doivent être vérifiées :

- **La disponibilité :** Accessibilité du système quand nécessaire.

**- La sécurité :** Protection contre les accès non autorisés.

**- La facilité de maintenance :** les différents modules de l’application doivent être bien lisibles, compréhensibles et faiblement couplés pour faciliter leur maintenance.

**- La fiabilité** : l’application devra être opérationnelle d’une façon continue et puissante afin d’avoir des résultats performants

**Conclusion**

Ce chapitre introductif nous a permis de mettre le projet dans son contexte général.  
Nous avons pris connaissance du cadre d’élaboration du projet, à travers la présentation  
de la société tunisienne de la sidérurgie "EL FOULADH", une problématique et des solutions proposées.

**CHAPITRE 2 : CONCEPTS FONDAMENTAUX ET METHODOLOGIE DE TRAVAIL**

**Introduction :**  
Ce chapitre se concentrera sur les principes fondamentaux de la Business Intelligence et de l'apprentissage automatique, ainsi que sur les méthodologies associées. Nous commencerons par explorer les divers modèles de conception des entrepôts de données, puis nous aborderons les approches méthodologiques de travail.

1. **Concepts fondamentaux de BI :**
2. **La chaine décisionnelle :**

La chaîne décisionnelle représente un processus essentiel dans le traitement des données pour prendre des décisions éclairées et stratégiques, contribuant ainsi à la réussite et à la performance globale de l'entreprise.

Cette séquence comprend une variété d'éléments et d'outils, répartis en quatre catégories distinctes, chacune étant dédiée à une phase spécifique du processus :

**1.L’étape de collecte de données :** Toutes les bases de données, qu'elles soient internes ou externes à l'entreprise, doivent être rassemblées. Ces diverses sources de données seront ensuite consolidées à l'aide de l’ETL, puis nettoyées et mises en forme.

**2.L’étape de stockage**: Les données sont stockées dans un entrepôt de données.

**3. L’étape de distribution ou restitution des données :** Les données doivent être facilement accessibles et adaptées aux profils des utilisateurs.

**4. l’étape d’analyse et d’exploitation des données :** Après le traitement et le stockage des données, elles sont prêtes à être analysées par les utilisateurs finaux ou les analystes. Cette analyse est facilitée par l'utilisation d'outils divers, comme les cubes OLAP pour les analyses multidimensionnelles, le Data Mining pour découvrir des relations cachées ou prédire des valeurs futures, ainsi que des tableaux de bord offrant une visualisation claire des principaux indicateurs.

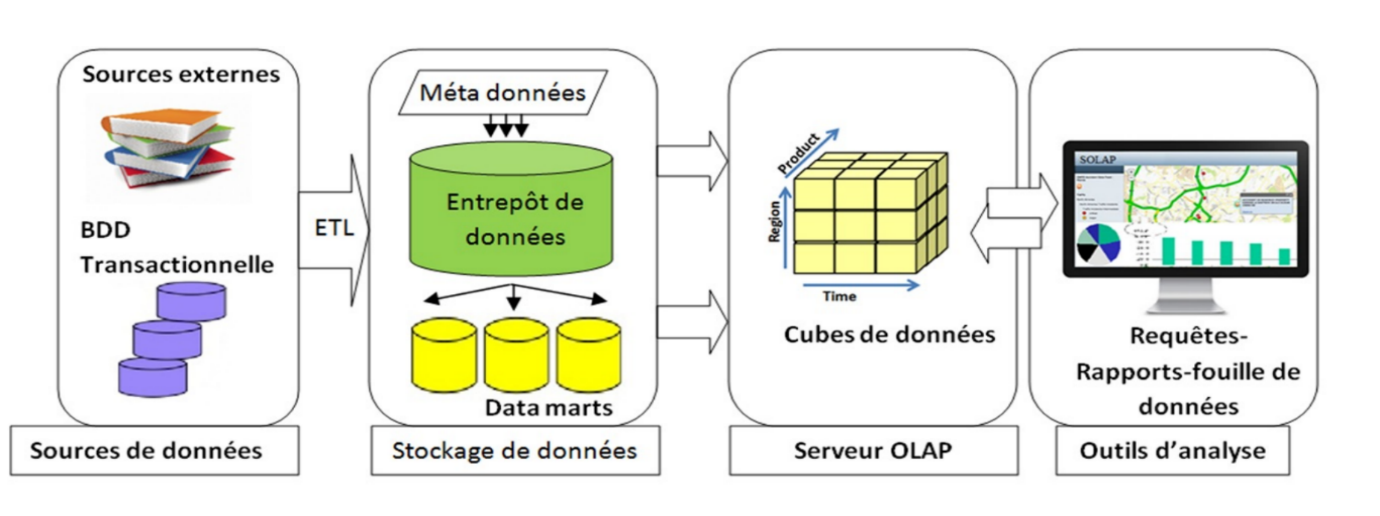
****

Figure 3: Processus de la chaine décisionnelle

1. **Les modèles de conception :**

Décrivent la structure, les relations et les interactions entre les éléments constitutifs, facilitant ainsi la compréhension et la mise en œuvre du projet.

Il existe 3 modèles :

|  |  |  |  |
| --- | --- | --- | --- |
| **Modèle de conception** | **Description** | **Avantages** | **Inconvénients** |
| En étoile (star schema) | -Une table de fait centrale et des dimensions  -Les dimensions n’ont pas de liaison entre elles | -Facilité de navigation  -Nombre de jointures limité | -Redondance dans les dimensions  -Toutes les dimensions ne concernent pas les mesures |
| En flocon de neige | - divise les tables de données en parties plus petites et plus détaillées, facilitant ainsi leur gestion et leur utilisation.  - organiser les informations de manière claire et précise et en améliorant la qualité de la base de données. | - réduire la redondance des données en divisant les tables en structures plus petites, ce qui économise de l'espace de stockage.  - simplifie la gestion des données en les organisant de manière hiérarchique, ce qui facilite les mises à jour et les requêtes | -La complexité peut augmenter avec le nombre de tables  - Les jointures multiples ralentissent les performances du système. |
| En constellation de faits (*fact constellation schema*) | **-** Les tables centrales permettent une analyse détaillée et flexible.  **-** Une structure modulaire facilite l'interrogation et la navigation des données. | **-** Flexibilité pour l'analyse détaillée des données multidimensionnelles.  **-** Facilité de navigation et d'interrogation grâce à sa modularité. | **-** Complexité accrue due à la multiplication des tables centrales.  **-** Possibilité de performances réduites lors de requêtes complexes. |

Tableau 1: Comparaison entre les différents modèles de conception d’entrepôt de données

1. **Présentation de la méthode de gestion de projet :**
2. **Comparaison entre les méthodologies classiques et agiles**

Deux principales méthodes se démarquent dans la gestion de projet : l'approche traditionnelle et l'approche Agile. Chacune présente ses propres avantages et inconvénients, et le choix entre les deux dépendra de divers facteurs tels que la taille, la complexité et la durée du projet, ainsi que des ressources disponibles.

Ce tableau comparatif examinera les distinctions entre l'approche traditionnelle et l'approche Agile afin d'aider à déterminer la méthode la mieux adaptée à votre projet.

|  |  |  |
| --- | --- | --- |
| **Thème** | **Approche traditionnelle** | **Approche Agile** |
| **Cycle de vie** | En cascade ou en V, phases séquentielles | Itératif et incrémental |
| **Planification** | Caractérisée par des plans détaillés | Adaptative avec plusieurs niveaux de planification |
| **Documentation** | Produite en quantité importante | Réduite au strict nécessaire |
| **Equipe** | Equipe avec des ressources spécialisées | Equipe responsabilisée et privilégiant l'initiative |
| **Qualité** | Contrôle qualité à la fin du cycle | Contrôle qualité précoce et permanent |
| **Communication** | Communication limitée | Communication fréquente et transparente |
| **Gestion de projet** | Le chef de projet est le leader centralisé | Le chef de projet est un facilitateur |
| **Flexibilité** | Les changements sont difficiles à intégrer | Les changements sont encouragés |
| **Priorisation** | La priorisation est basée sur le plan initial | La priorisation est basée sur la valeur pour le client |
| **Livraison** | Livraison à la fin du projet | Livraison régulière d'incréments |
| **Test** | Les tests sont effectués à la fin du projet | Les tests sont effectués en continu |
| **Rétrospective** | Effectuée à la fin du projet | Effectuée régulièrement pour améliorer le processus |

Tableau 2 : Comparaison des approches[[2]](#footnote-2)

1. **Méthode adoptée**

L’agilité est une priorité pour notre projet, compte tenu de notre besoin de collaboration et de communication quotidienne avec le reste de l’équipe afin de bien assimiler les  
besoins du client.

C’est pourquoi nous avons décidé de travailler avec la méthodologie SCRUM, qui représente une mise en œuvre de l’approche agile.

* **Les intervenants dans Scrum**

Dans un projet suivant la méthodologie SCRUM nous distinguons trois rôles :

**- Scrum master (Directeur du produit**) : Ce dernier agit en tant que facilitateur entre le responsable produit et  
l’équipe, il dirige l’équipe selon les règles et les processus de la méthodologie, gère  
la réduction des obstacles du projet et travaille avec Product Owner pour optimiser  
le retour sur investissement.

**- Product Owner (Propriétaire du produit) :** C’est le responsable du retour sur investissement du projet que  
concentre sur la partie commerciale. Il traduit la vision du projet à l’équipe, valide  
les bénéfices dans des histoires à intégrer dans le carnet de commandes du produit  
et les priorise régulièrement.

**- L’équipe de développement** : Un groupe de professionnels ayant les connaissances techniques nécessaires qui développent le projet qui sont chargés de transformer les besoins exprimés par le Product Owner en caractéristiques utilisables.

* **Les artéfacts du Scrum**

**-Backlog produit :** C'est une liste de tâches, évaluées en fonction de leur niveau de complexité, qui seront réalisées. Chaque tâche est définie en fonction des besoins du client.

**- Backlog de Sprint :** consiste en une sélection des éléments du Backlog Produit qui seront travaillés lors du prochain Sprint, offrant ainsi une vision anticipée des fonctionnalités à livrer ainsi que des tâches requises pour les réaliser.

**- Incrément Produit :** C’est un des artefacts SCRUM les plus importants de la culture Agile. Durant chaque Sprint, l’équipe de développement réalise un incrément de produit.[[3]](#footnote-3)

La figure ci-dessous illustre le processus de la méthodologie Scrum :

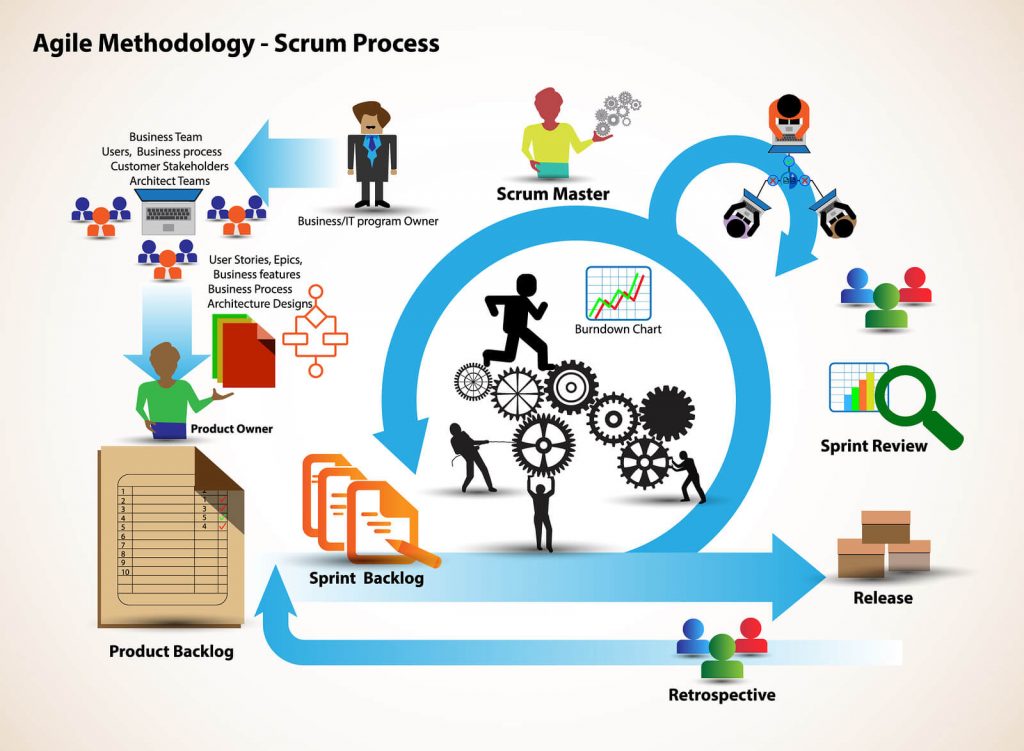


Tableau 3 : Le processus de la méthodologie Scrum[[4]](#footnote-4)

Dans le cadre de notre projet, voici notre équipe Scrum :

|  |  |  |
| --- | --- | --- |
| Une image contenant croquis, dessin humoristique, costume, silhouette  Description générée automatiquement | Une image contenant croquis, dessin humoristique, costume, silhouette  Description générée automatiquement |  |
| Product Owner | Scrum master | Team membres |
| **Berrais** Kamel | **Berrais** Kamel | **Chebbi** ichrak et **Ben Hassen** Zohra |

Tableau 4 : Equipe SCRUM

1. **Planification du travail :**

* **Backlog produit**

A l’aide du Product Owner, nous avons collecter tous les besoins de chaque acteur, que nous détaillerons dans le tableau ci-dessous :

|  |  |  |  |
| --- | --- | --- | --- |
| ID | Releases | User story | Priorité |
| 1 | Création de l’ETL | En tant qu'analyste, Il est nécessaire de concevoir des tableaux de bord, de planifier leur mise à jour et de les simplifier pour les utilisateurs. | Élevé |
| 2 | Développement un site web et intégration des tableaux de bord | En tant que Client, je dois consulter et rechercher des produits avant de m'authentifier.  En tant que Client, je dois demander de créer un compte et ajouter des produits aux panier.  En tant que Client, je dois interacter avec un chat bot et évaluer les livreurs.  En tant que Direction Générale, je dois accéder aux tableaux de bord.  En tant qu’administrateur, je souhaite Gérer les utilisateurs.  En tant que Sous-direction vente, je souhaite Gérer les produits.  En tant que livreur, je dois mettre à jour mon statut. | Élevé |

Tableau 5 : Backlog Général

**Conclusion**

Dans ce chapitre, nous avons exploré les principes de base de la Business Intelligence et du machine Learning, ainsi qu'une comparaison des différentes approches de gestion de projet. De plus, nous avons exposé les motivations qui ont conduit à notre choix de la méthodologie Scrum.

**RELEASE 1 : Création de l’ETL**

**Introduction :**

Ce chapitre sera dédié à la gestion du premier sprint de notre projet. Nous commencerons par définir le Backlog du sprint, puis nous procéderons à la mise en œuvre de notre solution. Pour ce premier sprint, nous nous attellerons d'abord à la conception, puis au développement de l'ETL (Extraction, Transformation, Chargement) nécessaire pour établir l'entrepôt de données.

1. **: Outils technologiques utilisés :**

Dans cette section, nous décrirons tous les outils de développement et les technologies utilisées durant le premier sprint :

|  |  |  |
| --- | --- | --- |
| **Logo** | **Nom** | **Description** |
| Une image contenant Arts créatifs, Papier couché, Papier de bricolage, Papier origami  Description générée automatiquement | Star UML | Star UML est un logiciel de modelage UML qui est entré récemment dans le monde de l'open source.[[5]](#footnote-5) |
| Une image contenant symbole, Bleu électrique, bleu, ligne  Description générée automatiquement | Visual Studio code | Visual Studio Code est un éditeur de code source qui peut être utilisé avec une variété de langages de programmation[[6]](#footnote-6) |
| Une image contenant clipart, symbole, Graphique, dessin humoristique  Description générée automatiquement | Python | C’est un [langage de programmation](https://fr.wikipedia.org/wiki/Langage_de_programmation) [interprété](https://fr.wikipedia.org/wiki/Langage_interpr%C3%A9t%C3%A9),  [multiparadigme](https://fr.wikipedia.org/wiki/Paradigme_(programmation)) et [multiplateformes](https://fr.wikipedia.org/wiki/Plateforme_(informatique)). [[7]](#footnote-7) |
|  | Microsoft Excel | C’est un [logiciel](https://fr.wikipedia.org/wiki/Logiciel) [tableur](https://fr.wikipedia.org/wiki/Tableur) de la [suite](https://fr.wikipedia.org/wiki/Suite_logicielle) [bureautique](https://fr.wikipedia.org/wiki/Logiciel_bureautique) [Microsoft Office](https://fr.wikipedia.org/wiki/Microsoft_Office) développé et distribué par l'éditeur [Microsoft](https://fr.wikipedia.org/wiki/Microsoft). [[8]](#footnote-8) |
| Une image contenant Graphique, Police, logo, capture d’écran  Description générée automatiquement | Docker desktop | Docker est un logiciel libre qui vous permet de simplifier la création de vos environnements d’exécution. [[9]](#footnote-9) |
|  | Apache Airflow | C’est une plateforme de planification de flux de travail open-source, très utilisée dans le domaine de l'ingénierie des données. [[10]](#footnote-10) |
|  | PostgreSQL | C’est une base de données open source réputée pour sa fiabilité, sa flexibilité et sa prise en charge des normes techniques ouvertes.[[11]](#footnote-11) |
| Une image contenant jaune, capture d’écran, Rectangle, conception  Description générée automatiquement | Power BI | C’est une solution d'[analyse de données](https://fr.wikipedia.org/wiki/Analyse_des_donn%C3%A9es) de [Microsoft](https://fr.wikipedia.org/wiki/Microsoft). Il permet de créer des [visualisations de données](https://fr.wikipedia.org/wiki/Visualisation_de_donn%C3%A9es) personnalisées et interactives avec une interface suffisamment simple pour que les utilisateurs finaux créent leurs propres rapports et tableaux de bord.[[12]](#footnote-12) |
|  | Jupyter Notebook | Jupyter Notebook est un outil de travail collaboratif à l'usage des programmeurs et des développeurs logiciels. Il offre à ces derniers la possibilité de partager du code, des graphiques ou des fichiers multimédias dans des blocs-notes interactifs.[[13]](#footnote-13) |

Tableau 6 : Technologies utilisées

1. **Release 1 :**

|  |  |  |  |
| --- | --- | --- | --- |
| **Release** | **Id** | **Sprint** | **Priorité** |
| Release1 | 1.1 | -Collecte des données à partir de la base de données de la société.  -Faire une analyse des données pour les comprendre. | Élevé |
| 1.2 | - Préciser les faits et les dimensions  -Conception de datamarts | Élevé |
| 1.3 | -Extraire et transformer les données.  -charger les tables de faits et de dimensions.  -planifier l’ETL. | Élevé |

Tableau 7 : Backlog de Release 1

1. **Cycle de vie de la solution BI :**

Notre solution est fondée sur trois sections principales :

* Identifier les différentes sources de données nécessaires pour alimenter le système.
* Le processus ETL consiste à extraire les données de ces sources, les transformer selon les besoins, puis les charger dans un datamart.
* Le processus de reporting et visualisation des tableaux de bord. Le sous-direction vente peut prendre des décisions se basant sur ces résultats.

Une image contenant capture d’écran, diagramme, texte

Description générée automatiquement

Figure 4 : Cycle de vie de la solution BI

1. **: Représentation conceptuelle des données :**

Dans cette section, nous expliquant les différentes tables de dimensions et de faits.

1. **Détermination des tables de dimension :**

Une dimension est une table, c’est un axe d'analyse avec lesquels on veut faire l'analyse.

|  |  |
| --- | --- |
| **Dimension** | **Attributs** |
| Dimension\_commande | Id\_commande  Id\_client  Section\_vente |
| Dimension\_facture | Id\_facture  Date  Mode\_payement  Id\_client |
| Dimension\_article | Id\_article  Designation  Unite\_mesure  Famille |

Tableau 8 : Les dimensions

1. **Détermination la table de fait :**

La table de fait contient les valeurs des mesures et les clés vers les tables de dimensions.

|  |  |
| --- | --- |
| **Fait** | **Champs** |
| Table\_de\_fait\_vente | Id\_facture  Id\_client  Id\_article  Id\_commande  Prix  Quantite |

Tableau 9: table de fait

1. **: Datamart de vente :**

Un datamart est fréquemment un sous-ensemble du datawarehouse de l'entreprise, obtenu par extraction et agrégation des données.

Une image contenant texte, capture d’écran, diagramme, nombre

Description générée automatiquement

Figure 5 : Datamart vente

1. **: ETL :**
2. **: Installation Airflow en Docker :**

La configuration d'Airflow sur Docker a été achevée en suivant une procédure précise. Nous avons utilisé un Dockerfile personnalisé pour configurer l'image Docker d'Airflow en fonction de nos exigences.

Une image contenant diagramme, conception, origami

Description générée automatiquement

Figure 6 : Installation Airflow en Docker

1. **La création du fichier Dockerfile**

Pour débuter, nous avons élaboré un Dockerfile afin de définir l'environnement d'exécution d'Airflow. Ce fichier renferme les directives instructions pour construire une image Docker sur mesure, comprenant les dépendances et configurations propres au projet.

Une image contenant texte, capture d’écran, logiciel, Logiciel multimédia

Description générée automatiquement

Figure 7 : DockerFile

Dans ce Dockerfile, nous avons déclaré l'image de base "apache/airflow :2.8.1" pour construire notre environnement.

Ensuite, nous avons copie la répertoire local "C:/projet\_bi\_elfouladh/bi/data\_elfouladh" dans la répertoire "/opt/airflow/data" de l'image Docker. Cette étape assure que les données nécessaires sont disponibles dans notre l'environnement.

Puis, nous avons procédé à l'installation de plusieurs dépendances nécessaires à notre projet :

* **markupsafe==2.0.1** : Installe la version spécifique de la bibliothèque Markupsafe quelle est indispensable pour le bon fonctionnement d’Airflow.
* **apache-airflow-providers-jdbc** : Installe le connecteur JDBC pour Apache Airflow, ce package a été ajouté pour permettre la connexion à des bases de données via JDBC.
* **Openpyxl**: Installe la bibliothèque Python openpyxl pour lire et écrire des fichiers Excel au format xlsx.
* **Pandas**: Installe la bibliothèque Python pandas pour importer, nettoyer, transformer et analyser des données.

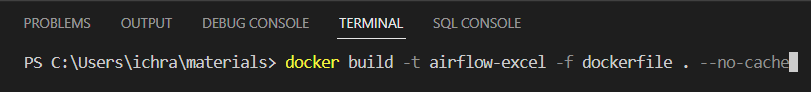
La création de l'image Docker sera effectuée en lançant la commande suivante :

Figure 8: construire une image Docker à partir d'un Dockerfile

1. **Démarrage des conteneurs Docker :**

En utilisant un fichier Docker Compose, nous avons défini les différents services nécessaires pour notre environnement Airflow, comme le serveur web et la base de données. En exécutant une commande simple, les conteneurs Docker ont été créés et lancés avec les configurations spécifiées.

Le démarrage du conteneur Airflow a été réalisé en lançant la commande suivante :

Une image contenant texte, capture d’écran, Police

Description générée automatiquement

Figure 9: Démarrage du conteneur Airflow

Dans cette capture d'écran, nous pouvons voir l'interface de Docker Desktop affichant l'image Airflow que nous avons construit :

Une image contenant capture d’écran, texte, logiciel, nombre

Description générée automatiquement

Figure 10 : Image d'Airflow en Docker

Dans cette capture d'écran, nous pouvons voir les conteneurs Airflow en cours d'exécution. Cette vue nous permet de gérer les différents composants de notre environnement Airflow :

Une image contenant texte, capture d’écran, logiciel, nombre

Description générée automatiquement

Figure 11: Conteneurs Airflow

1. **: Les avantages de l'utilisation d'Airflow sur Docker :**

**Portabilité** : Docker permet de créer des conteneurs déployés sur n'importe quel environnement, assurant ainsi que l’application Airflow fonctionne de la même manière partout.

**Isolation** : Chaque conteneur est isolé, c’est à dire que les dépendances et configurations d'Airflow n'entreront pas en conflit avec d'autres applications sur la même machine.

**Scalabilité** : Nous pouvons facilement ajouter ou supprimer des conteneurs.

**Facilité de déploiement** : Nous pouvons tester des changements de nouvelles versions d'Airflow dans un conteneur avant de les mettre en production.

**Gestion simplifiée** : Docker offre des outils (Docker Compose, Docker Dashboard …) pour la gestion des conteneurs, qui facilite la supervision et la maintenance d'Airflow.

1. **Développement de l’ETL :**

Après avoir préparé l'environnement de développement, nous passons à l'implémentation de notre module BI.

Premièrement on a créé un DAG (Graphe dirigé acyclique), est un fichier Python qui contient les fonctions et les tâches et qui permet de planifier et exécuter les tâches de notre système.

Cette capture d'écran contient un DAG, appelé ETL, a été conçu pour extraire, traiter et charger les données dans nos entrepôt. Ce DAG est planifié et s’exécuter pour traiter les données quotidiennement.

Une image contenant capture d’écran, texte, Police, ligne

Description générée automatiquement

Figure 12 : DAG dans l'interface Airflow

1. **: Connection aux bases de données :**

La première figure montre les fichiers Excel qui contient les données préparées par le maitre de stage.

Une image contenant texte, capture d’écran, logiciel, nombre

Description générée automatiquement

Figure 13 : les fichier Excel

La deuxième figure est pour garantir la connexion avec notre destination (entrepôt de donnée)

**Une image contenant texte, capture d’écran, logiciel, affichage

Description générée automatiquement**

Figure 14 : Connection à la base de données destination avec pgAdmin

1. **Extraction les données**

On extraite les données à partir des fichier Excel tels que les produit vendu, période de vente et commande passée.

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

Figure 15: Les données extraites

1. **transformation des données**

Après l’extraction, nous avons procédé au nettoyage des données avant le charger dans les tables de faits et de dimensions.

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

Figure 16 : Extraction et transformation des données

1. **Chargements des données**

La capture suivante montre la connexion à la base de données est établie avec succès :

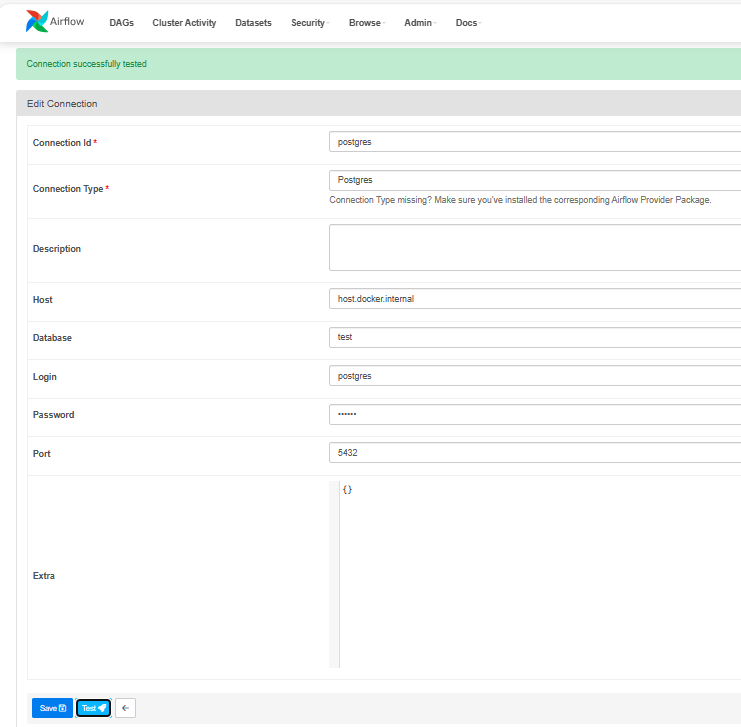


Figure 17 : connexion airflow à l'entrepôt de données

1. **Chargements des tables**

* Pour la table de dimension temps, nous avons récupéré les dates dans le fichier Excel ‘qua.xlsx’, à l’aide des fonctions de manipulation des dates nous avons obtenu le mois, le nom de mois, le jour et l’année, aussi nous modifions les noms des colonnes, nous avons sélectionner les colonnes à charger et nous éliminons les codes NULL.
* Pour la table de dimension article, nous avons sélectionner les colonnes à charger, nous remplaçons les valeurs existantes dans la colonne famille avec les valeurs correspondantes et nous avons modifié les noms des colonnes.
* Pour la table de dimension commande, nous avons sélectionner les colonnes à charger.
* Finalement après avoir chargé les tables de dimensions, nous avons chargé la table de fait vente à l’aide des jointures avec les tables de dimensions pour obtenir leurs clés étrangères et nous avons sélectionner les prix de vente et quantité vendu.

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

Figure 18: ETL

En fin, les tables sont chargées dans notre entrepôt de données :

Une image contenant texte, nombre, Police, capture d’écran

Description générée automatiquement

Figure 19 : Notre entrepôt de données

1. **Mise à jour les données de l’entrepôt de données :**

Il est essentiel de maintenir des données à jour dans l’entrepôt de données. Pour ce faire, différentes méthodes de rafraîchissement sont disponibles, telles que l'exécution manuelle de scripts ou l'utilisation des fonctionnalités offertes par Apache Airflow.

Apache Airflow propose un "Schedule" (planificateur de tâches) qui permet de planifier et configurer le processus de chargement des données. Pour le rafraîchissement des données de notre Data Warehouse, nous avons décidé d'utiliser Apache Airflow pour programmer une mise à jour quotidienne à 00h. Cette planification automatisée garantit que les données de l’entrepôt de données sont régulièrement mises à jour.

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

Figure 20: Planification de DAG

**Conclusion**

Au cours de ce chapitre, nous avons mis en œuvre notre premier release, qui comprenait des activités telles que la conception, l'intégration des données et l'implémentation des processus ETL.

**RELEASE 2 : Développement d’un site web et intégration des tableaux de bord**

**Introduction :**

Ce chapitre couvre deux sprints. Le premier sprint inclut la création des tableaux de bord et l'implémentation du modèle de prédiction. Le deuxième sprint se concentre sur la conception du fonctionnement du système et les interactions entre les utilisateurs, ainsi que la présentation des principales interfaces de notre application.

1. **Release 2 :**

|  |  |  |  |
| --- | --- | --- | --- |
| **Release** | **Id** | **Sprint** | **Priorité** |
| Release2 | 2.1 | - Elaboration des tableaux de bord et la mise en place d'un modèle de prédiction. | Élevé |
| 2.2 | - la gestion du panier d'achat, authentification et gestion des utilisateurs, ainsi que l’intégration des tableaux de bord | Élevé |

Tableau 10 : Backlog de Release 2

1. **Restitution et prédiction des données :**
2. **Restitution des données :**

La phase de restitution des données est essentielle pour les décideurs de chaque entreprise, car les rapports générés des indicateurs pertinents, des analyses détaillées, et des visualisations graphiques qui facilitent une prise de décision.

1. **Connexion de Power BI avec l’entrepôt de données :**

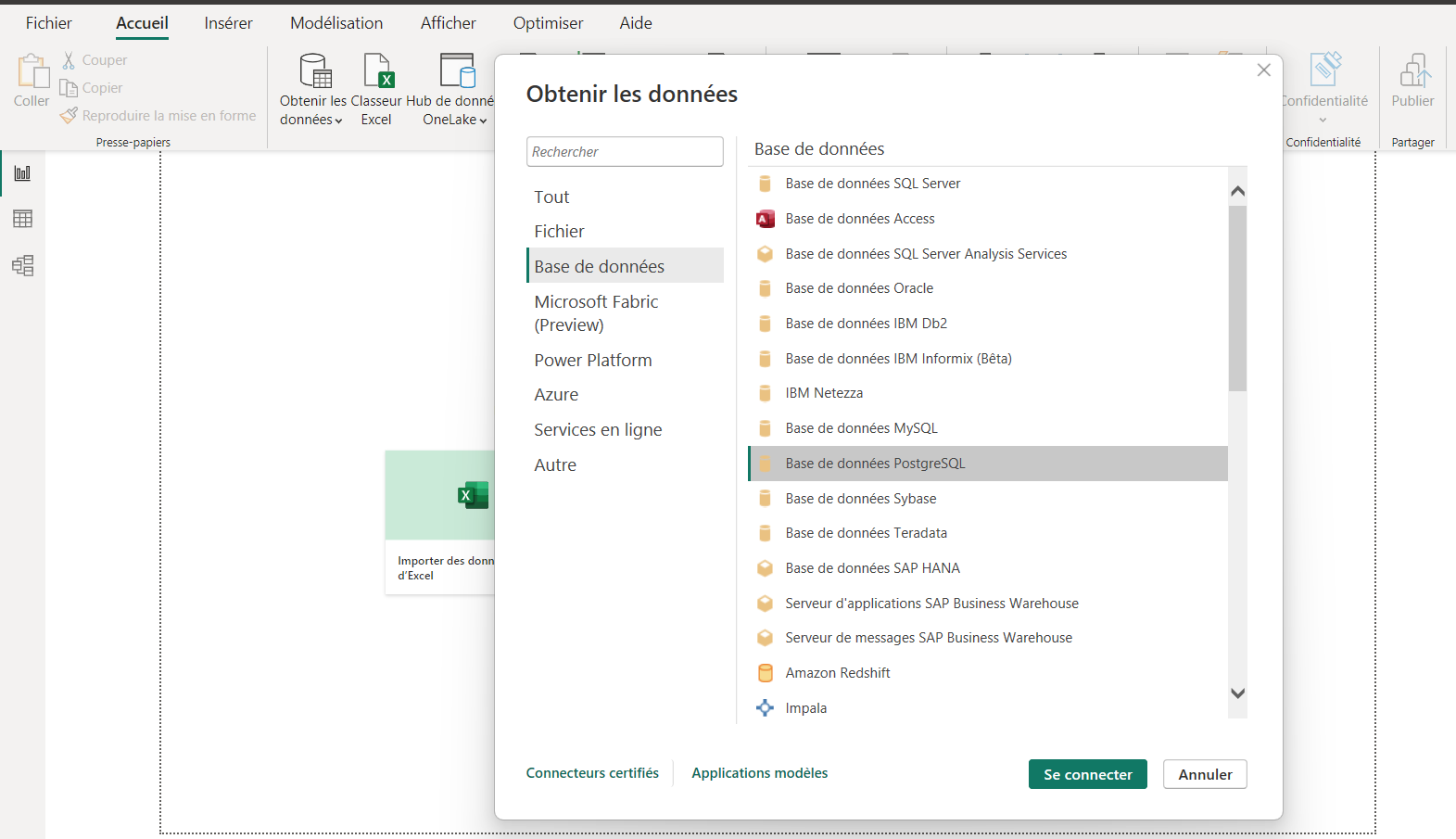
****

Figure 21: Connexion de Power BI à la base de données PostgreSQL

À cette étape, nous avons choisi la base de données PostgreSQL et nous sommes connectés à notre localhost en utilisant l'option «Direct Query» , Cette option permet de récupérer les données directement depuis la source, au moment de la requête

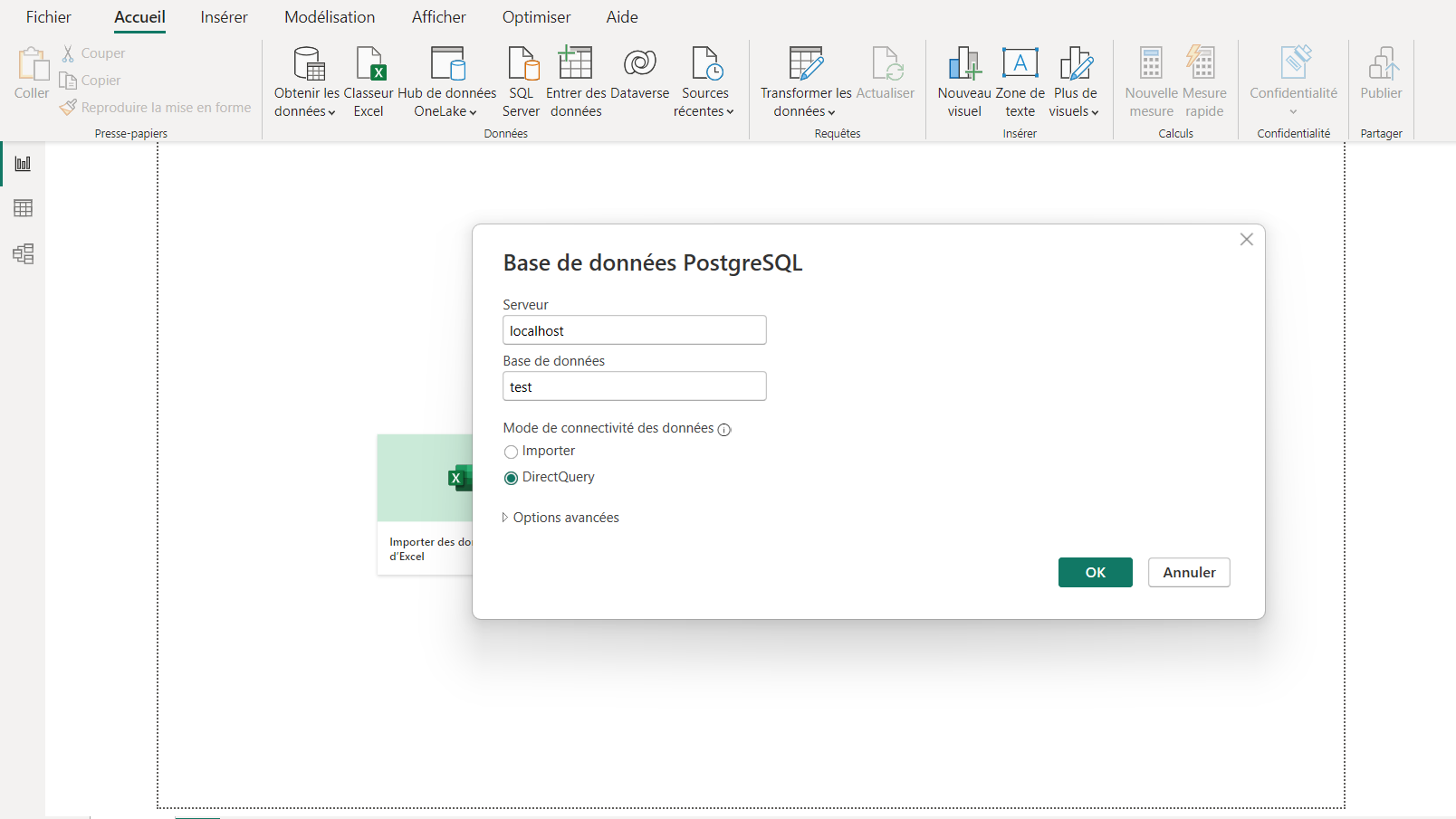


Figure 22 : Connexion à l'entrepôt de données avec l'option "Direct Query"

Après la connexion, nous avons sélectionné la table de fait et les tables de dimensions contenant les données pertinentes pour commencer à créer les tableaux de bord. La sélection des tables est illustrée dans la figure suivante :

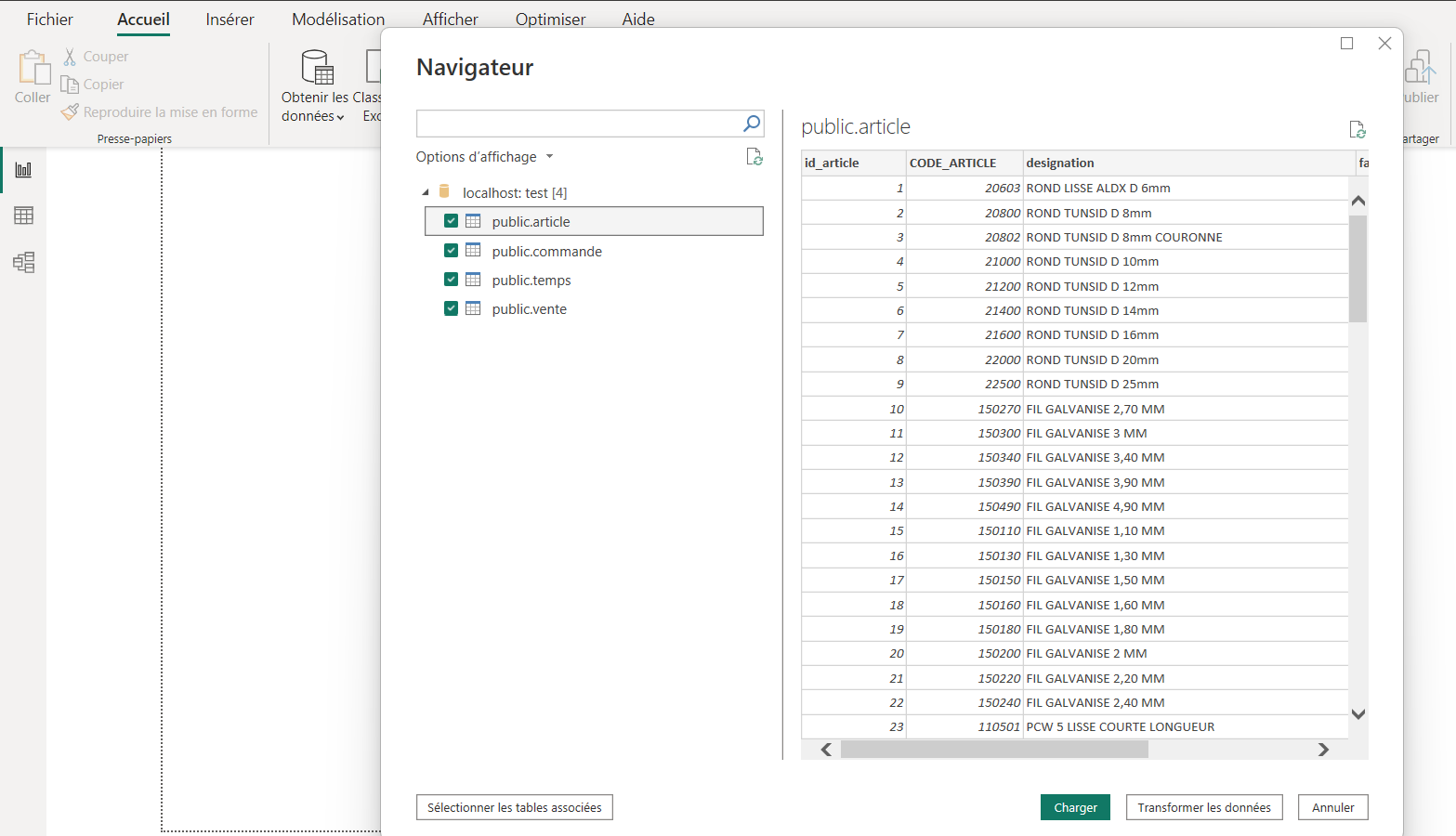


Figure 23 : Sélection des tables

1. **Elaboration des Dashboard :**

Les tableaux de bord des ventes offrent une vue globale de la performance financière et des ventes de l'entrepris.

* **Tableau de bord « Revenu des ventes »**

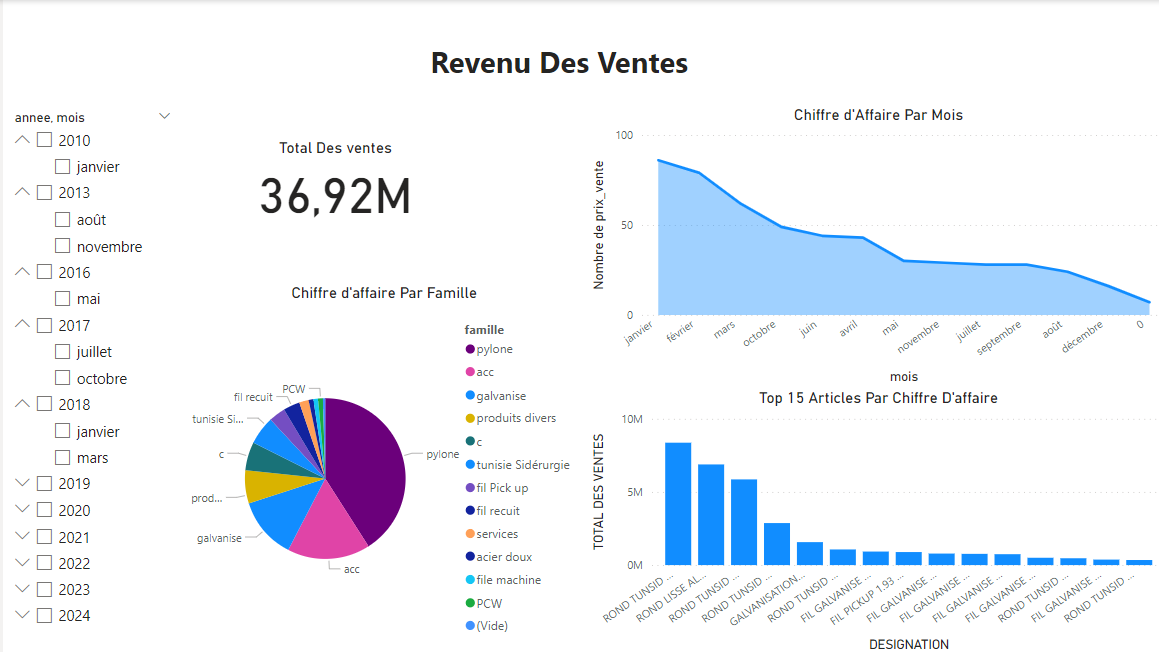
****

Figure 24 : Revenu des ventes

-Le chiffre d'affaires par famille aide les décideurs à concentrer leurs efforts sur les segments de produits les plus rentables pour maximiser les revenus de l'entreprise.

-Le chiffre d'affaires par mois permet aux décideurs de prendre des décisions éclairées et d'ajuster leur stratégie commerciale pour atteindre les objectifs de revenus mensuels fixés par l'entreprise.

Le top 15 articles par chiffre d'affaires fournissant aux décideurs une compréhension claire des produits contribuant le plus aux revenus globaux de l'entreprise.

* **Tableau de bord « Quantité vendu »**

Une image contenant texte, capture d’écran, diagramme, Tracé

Description générée automatiquement

Figure 25 : Quantité vendu

-Le total de la quantité vendue par famille de produits permet aux décideurs de comprendre quelles familles de produits sont les plus populaires et de prendre des décisions éclairées sur la gestion des stocks et de la production

-Le total de la quantité vendue par mois perme aux décideurs de mieux ajuster les ressources et de prévoir les besoins futurs en fonction des tendances mensuelles de la demande en termes de volume de ventes.

- Le top 15 des articles par total de quantité aide les décideurs à comprendre quels produits sont les plus demandés en termes de volume de ventes.

* **Tableau de bord « Commande Passée »**

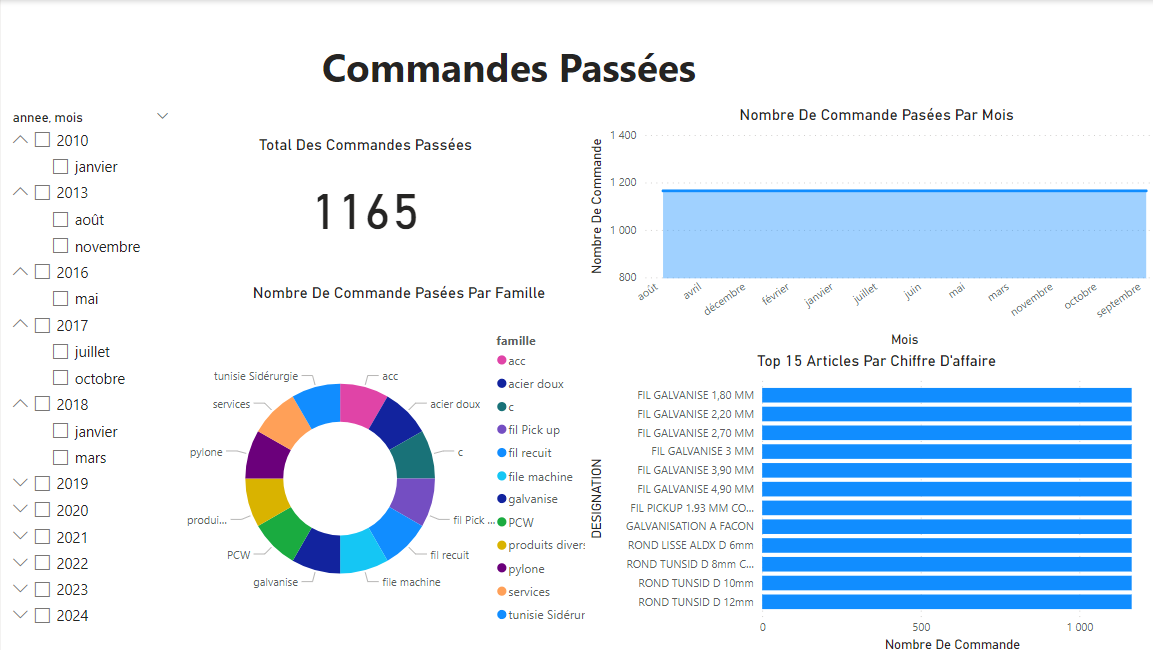
****

Figure 26 : Commande passée

-Le nombre de commandes passées par famille de produits fournisse une compréhension de la répartition de la demande par catégorie pour guide la gestion des stocks, la production et les stratégies de vente.

-Le nombre de commandes passées par mois permet aux décideurs de détecter les variations saisonnières et de prendre des décisions stratégiques pour répondre efficacement à la demande.

-Le top 15 articles par nombre de commandes aide les décideurs à ajuster leur stratégie de production et de gestion des stocks pour répondre à la demande et maximiser les ventes.

1. **Prédiction :**
2. **Concepts du machine Learning :**

Le Machine Learning est une sous-catégorie de l’intelligence artificielle , permet aux machines d'acquérir des compétences par l'expérience.

Elles peuvent répondre à des tâches complexes sans intervention humaine, en s'appuyant sur l'analyse de données pour comprendre et s'adapter à de nouvelles situations en fonction des modèles appris.

1. **Les types d’apprentissage**

Il y a deux types principaux :

-**Apprentissage supervisé** : Obtenir un modèle optimal avec les performances requises grâce à L’apprentissage basé sur des échantillons labellisés de catégories connues.

-**Apprentissage non supervisé** : pour les échantillons non labellisés, les algorithmes d'apprentissage modélisent directement les ensembles de données d'entrée.

1. **Les modèles de prédiction**

Les modèles d’analyse prédictive sont des algorithmes statistiques ou d’apprentissage automatique qui utilisent des données historiques pour prédire des événements ou des comportements futurs. Ces modèles analysent de grandes quantités de données pour identifier des tendances et des modèles, qu’ils utilisent ensuite pour prévoir des résultats futurs.[[14]](#footnote-14)

|  |  |  |  |
| --- | --- | --- | --- |
| **Modèle** | **Description** | **Avantages** | **Inconvénients** |
| Régression linéaire | - Utilise des variables explicatives pour prédire une variable cible.  - Modélise une relation linéaire entre les données d'entrée et la sortie | - Facile à interpréter et à expliquer aux non- experts.  - Calcul rapide et efficace, adapté à de grandes données | - Néglige les relations non linéaires entre variables.  - Sensible aux valeurs aberrantes et aux données manquantes. |
| Régression logistique | - Modèle la probabilité d'une variable binaire.  - Utilise une courbe en forme de S. | -Adapté à la prédiction de variables binaires.  -Génère des probabilités pour les prédictions | -Ne convient pas à la prédiction de variables continues.  -Sensible au sur-ajustement avec un grand nombre de variables. |
| Régression polynomiale | -Modèle une relation courbe entre les données.  -Utilise des polynômes pour ajuster les données. | -Peut capturer des relations non linéaires entre variables.  -Peut s'adapter à des données avec des tendances courbes. | -Peut être sensible au sur-ajustement avec des degrés élevés.  -Plus complexe à interpréter que la régression linéaire |
| Lissage exponentiel | -Utilise des moyennes pondérées pour lisser les données.  -Donne plus de poids aux données récentes dans les prédictions | -Capable de détecter les tendances et les modèles dans les données.  -Répond rapidement aux changements dans les séries chronologiques. | -Peut être sensible aux variations extrêmes des données.  -Nécessite une sélection judicieuse des paramètres pour éviter le sur-ajustement. |

Tableau 11 : Comparaison entre les différents modèles de prédiction

1. **Développement de la plateforme:**
2. **Architecture de la plateforme :**

On a utilisé l’architecture MVT pour le développement de notre plateforme. Elle s’inspire de l’architecture MVC, son objectif est de séparer les responsabilités d'une application web pour faciliter le développement.

L’abréviation MVT est :

-**M**odèle : interagit avec la base de données et représente les structures de données

-**V**ue : interagit avec les modèles pour récupérer les données et le renvoie au client sous forme de réponse http.

-**T**emplate : Gère la logique de présentation.

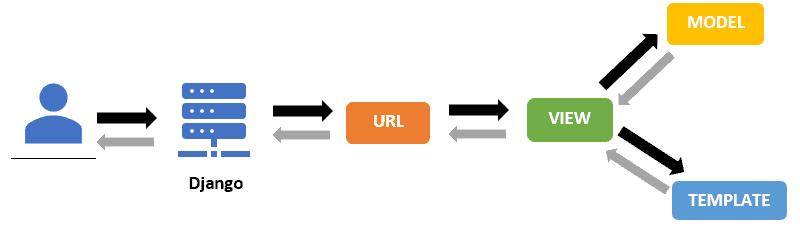


Figure 27 : Architecture MVT de la plateforme

1. **Conception de la plateforme :**
2. **Diagramme de cas d’utilisation par chaque acteur**

* Directeur général

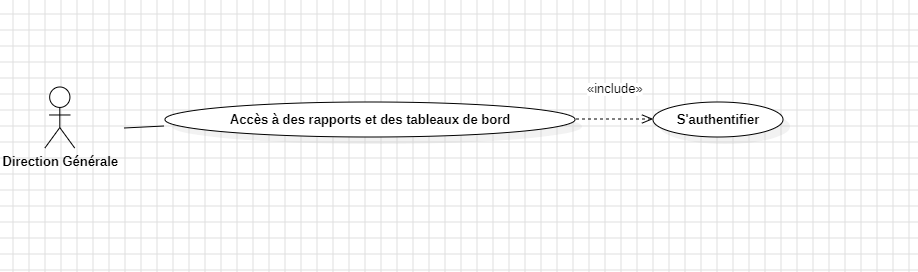


Figure 28 : Diagramme de cas d'utilisation du directeur générale

* Client

**Une image contenant diagramme, dessin, ligne, texte

Description générée automatiquement**

Figure 29 : Diagramme de cas d'utilisation du client

* Administrateur

**Une image contenant diagramme, dessin, croquis, blanc

Description générée automatiquement**

Figure 30 : Diagramme de cas d’utilisation de l’administrateur

* Sous-direction vente

Une image contenant diagramme, ligne, blanc

Description générée automatiquement

Figure 31 : Diagramme de cas d'utilisation du sous-direction vente

* Livreur

**Une image contenant diagramme, blanc, ligne, croquis

Description générée automatiquement**

Figure 32 : Diagramme de cas d'utilisation du livreur

1. **Diagramme de séquence :**

C’est un outil pour comprendre les interactions de notre plateforme , il permet de visualiser comment les différents composants interagissent entre eux.

1. : disponible sur : http://www.elfouladh.com.tn [Consulté le 16/04/2024] [↑](#footnote-ref-1)
2. : disponible sur : https://www.consultrade.info/gestion-de-projet/la-gestion-de-projet-methode-classique-vs-agiles/ [Consulté le 17/04/2024] [↑](#footnote-ref-2)
3. : https://blog-gestion-de-projet.com/agilite-et-scrum-fondamentaux/artefacts-scrum/ [Consulté le 17/04/2024] [↑](#footnote-ref-3)
4. : https://acute360.com/pdf/guide-rd.pdf [Consulté le 17/04/2024] [↑](#footnote-ref-4)
5. : <https://matthieu-brucher.developpez.com/tutoriels/conception/staruml/> [Consulté le 13/05/2024] [↑](#footnote-ref-5)
6. : <https://fr.wikipedia.org/wiki/Visual_Studio_Code> [Consulté le 13/05/2024] [↑](#footnote-ref-6)
7. : https://fr.wikipedia.org/wiki/Python\_(langage) [Consulté le 13/05/2024] [↑](#footnote-ref-7)
8. : <https://fr.wikipedia.org/wiki/Microsoft_Excel> [Consulté le 13/05/2024] [↑](#footnote-ref-8)
9. : <https://codedesign.fr/docker-desktop-windows-debutant/> [Consulté le 13/05/2024] [↑](#footnote-ref-9)
10. : <https://datascientest.com/apache-airflow> [Consulté le 13/05/2024] [↑](#footnote-ref-10)
11. : <https://www.ibm.com/fr-fr/topics/postgresql> [Consulté le 13/05/2024] [↑](#footnote-ref-11)
12. : <https://fr.wikipedia.org/wiki/Microsoft_Power_BI> [Consulté le 13/05/2024] [↑](#footnote-ref-12)
13. https://www.lebigdata.fr/jupyter-notebook-definition[Consulté le 18/05/2024] [↑](#footnote-ref-13)
14. : disponible sur : <https://www.voxco.com/fr/blog/types-de-modeles-danalyse-predictive-et-leurs-applications/> [Consulté le 17/04/2024] [↑](#footnote-ref-14)