

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université de la Manouba

Institut Supérieur des Arts Multimédias



MEMOIRE DE PROJET DE FIN D'ETUDES

Code 701

MISE EN PLACE D'UN DASHBOARD D'AIDE A LA DECISION AD-CAISSE

Réalisé par

Legleg Hend BDAD2

Trabelsi Moahmed BDAD2

Encadré par

Pr. Ben Hamouda Ahmed (ISAMM/ESC)

Mr. Dridri Oussama (TWINS DIGITAL LABS)

Année Universitaire 2021-2022

Dédicace

A mes chers parents, pour leurs efforts et sacrifices. Que dieu leur préserve bonne santé et longue vie.

A mon frère Mehdi, ma sœur Meriem, mon beau-frère Hamdi et ma belle-sœur Raghda qui m'ont beaucoup encouragé.

A mes amis, pour leur soutien et pour tous les bons moments qu'on a vécus ensemble.

Trabelsi Mohamed

Dédicace

Je dédie ce travail

À mon père Mahmoud

Je vous remercie pour votre confiance, vos sacrifices et votre amour et surtout tes encouragements tous au long de mes études, nulle dédicace ne peut exprimer ma grande reconnaissance, gratitude et mon respect.

À ma mère Saida

Pour votre soutien, votre confiance, bien vaillance, votre amour et surtout votre tendresse dès ma plus jeune enfance. , aucun mot ne peut vous exprimer ma reconnaissance pour vous.

À mon deuxième père Amor

Je vous remercie pour votre encouragement, votre soutien que vous m'avez toujours donné dès mon enfance. Mon amour pour vous dépasse la mer dont vous m'avez en parler dans la carte postale, que je tiens toujours à mon cœur.

À ma mère éternelle Rachida

Je ne trouve pas de mots pour vous remercier pour l'amour inconditionnel que vous m'avez offert, vos encouragements et votre soutien extraordinaire, merci d'être à mes côtés et merci d'être ma deuxième mère.

À ma sœur Imen et mon frère Mohamed

À ma tante Essia

À mes âmes sœurs Eya et Dina

Et à tous mes amis

Remerciements

Premièrement, nous voulons remercier Dieu de nous avoir donné la force d'achever ce travail.

Nous tenons à exprimer nos grandes reconnaissances envers les personnes qui nous ont apporté leurs soutiens.

Un immense merci à notre encadrant académique Pr. Ben Hammouda Ahmed pour les conseils qu'il nous a prodigués, pour sa pédagogie, sa qualité d'encadrement et son aide tout au long de ce projet.

Nos remerciements les plus chaleureux vont à notre encadrant à Twins Digital Labs Mr. Dridi Oussama, pour ses conseils directifs et les formations qui nous ont été d'une grande aide pour l'aboutissement de ce projet.

Nos vifs remerciements vont également à tous nos professeurs à l'Institut Supérieur des Arts Multimédia.

Enfin, nous avons l'honneur d'exprimer notre remerciement aux membres de jury pour avoir accepté d'évaluer ce travail.

Merci à tous.

Table des matières

Introduction générale	1
Chapitre 1 : Présentation du cadre du projet	3
1 Introduction.....	3
2 Présentation de la société TWINS DIGITAL LABS	3
3 Les services de la société	4
4 Etude de l'existant	4
5 Critique de l'existant	5
6 Solutions proposées	6
7 Les besoins fonctionnels	6
8 Les besoins non fonctionnels	7
9 Conclusion	7
Chapitre 2 : La méthodologie de travail et les outils adoptés	8
1 Introduction.....	8
2 Les méthodes classiques de gestion de projet	8
3 Les méthodes Agiles.....	10
3.1 La méthodologie Scrum.....	12
3.2 Processus Scrum.....	12
3.3 Les artefacts de Scrum.....	13
3.4 Les acteurs d'un Scrum	14
3.5 Les événements de la méthodologie Scrum.....	15
4 La méthode adoptée.....	17
4.1 La méthodologie Scrum appliquée sur notre projet	18
4.1.1 Pilotage du projet avec la méthodologie Scrum.....	18
4.1.2 Découpage des Sprints.....	18
4.1.3 Diagramme de Gantt.....	19
5 Les phases d'un projet décisionnel	20
6 Les choix techniques	21
6.1 Les outils de gestion de la base de données	21
6.1.1 MySQL	21
6.1.2 PostgreSQL.....	22
6.1.3 Oracle database	22
6.1.4 Comparaison entre les outils de gestion de la base de données	22
6.1.5 Choix adopté	23

6.2	Les différents outils de Extract Transform Load (ETL).....	24
6.2.1	Talend open source.....	24
6.2.2	Pentaho Data Integration	24
6.2.3	Comparaison entre les outils d'ETL.....	25
6.2.4	Choix adopté	26
6.2.5	Rédaction de Scripts	26
6.3	Les outils de visualisation de données.....	26
6.3.1	Microsoft power BI	26
6.3.2	Knowage	27
6.3.3	Comparaison entre les outils de visualisation de données.....	29
6.3.4	Choix adopté	30
6.3.5	Interprétation.....	30
7	Architecture globale	31
8	Conclusion	31
	Chapitre 3 : Conception et mise en place d'un entrepôt de données	32
1	Introduction.....	32
2	Le premier sprint : Conception d'un entrepôt de données.....	32
2.1	Etude de sources des données	33
2.2	Conception de la base de données.....	34
2.3	La modélisation multidimensionnelle	34
2.4	Conception de l'entrepôt de données	36
2.4.1	Choix des mesures.....	37
2.4.2	Choix des dimensions.....	37
2.4.3	Table des faits	38
2.4.4	Conception de l'entrepôt de données	39
3	Conclusion	40
	Chapitre 4 : La mise en place d'un processus ETL	41
1	Introduction.....	41
2	Le deuxième sprint : Mise en place d'un processus ETL.....	41
2.1	Backlog du sprint	41
2.2	Processus ETL.....	41
2.3	Réalisation	43
2.3.1	Extraction des données	44
2.3.2	Transformation des données et chargement.....	45
3	Conclusion	56

Chapitre 5 : Réalisation du tableau de bord	57
1 Introduction.....	57
2 Backlog du sprint.....	57
3 Définition des indicateurs	57
3.1 Définition d'un indicateur clé de performance	57
3.2 Les indicateurs choisis.....	58
4 Configuration de l'outil de création des tableaux de bord.....	59
4.1 Connexion.....	59
4.2 Gestion des utilisateurs et des rôles.....	60
5 Création des indicateurs par l'outil du reporting.....	62
5.1 Création des Data Sets.....	62
5.2 Création des chartes.....	63
5.2.1 Tableau de bord du module Général	63
5.2.2 Tableau de bord du module Produit	65
5.2.3 Tableau de bord du module Partenaire	67
5.2.4 Tableau de bord du module Secteur	68
6 Conclusion	69
Conclusion et perspectives	70

Liste des tableaux

Tableau 1: Diagramme de Gantt.....	19
Tableau 2: Comparaison entre les outils de création de data Warehouse.....	23
Tableau 3: Comparaison entre les outils d'ETL	25
Tableau 4: Comparaison entre les outils de visualisation de données.....	29
Tableau 5: les mesures de la table de faits	37
Tableau 6: Les tables de dimension.....	38
Tableau 7: Table des faits	39

Liste des figures

Figure 1: Logo Twins Digital Labs	4
Figure 2: Démarche méthode « Cycle en V »	10
Figure 3: Cycle de vie d'un Scrum	13
Figure 4: Les Sprints.....	15
Figure 5: Récapitulatif du processus Scrum.....	17
Figure 6: Equipe Scrum	18
Figure 7: Chaîne de projet décisionnel	21
Figure 8: MySQL.....	24
Figure 9: Talend Open source	26
Figure 10: Interface du Microsoft Power BI	27
Figure 11: Interface de Knowage	28
Figure 12: Logo Knowage	30
Figure 13: L'architecture totale de notre Projet	31
Figure 14: Base de données source.....	34
Figure 15: MYSQL WORKBENSH	37
Figure 16: Conception de l'entrepôt de données	40
Figure 17: Processus ETL	43
Figure 18: Les tables de l'entrepôt de données	44
Figure 19: Connexion de bases de données	44
Figure 20: fonction SOUNDEX appliquée sur la table produit (1).....	46
Figure 21: fonction SOUNDEX appliquée sur la table produit (2).....	46
Figure 22: Fonction UPDATE	47
Figure 23: Résultat des transformations	47
Figure 24: Transformation des adresses	48
Figure 25: Job produit.....	49
Figure 26: Job pointVente.....	50
Figure 27: Job partenaire.....	51
Figure 28: Job commandeDetails	51
Figure 29: Job categorieArticle	52
Figure 30: job Secteur.....	53
Figure 31: job time	54

Figure 32: job commande	54
Figure 33: job fait	55
Figure 34: tMap	56
Figure 35: Interface de connexion vers le datawarehouse	60
Figure 36: Gestion des rôles.....	60
Figure 37: Gestion des utilisateurs	61
Figure 38: Attribution d'un rôle à l'utilisateur.....	61
Figure 39: Création des datasets.....	62
Figure 40: les différents widgets	63
Figure 41: Le Module Générale	63
Figure 42: Classement des partenaires.....	64
Figure 43: Rapport des produits les plus vendus.....	64
Figure 44: Informations d'ADCAISSE	65
Figure 45: Module produit	66
Figure 46: comparaison des prix par une zone géographique.....	66
Figure 47: Interface de création de sélection	66
Figure 48: Consommé du produit.....	67
Figure 49: Rapport du module partenaire	67
Figure 50: Classement des produits pour un partenaire.....	68
Figure 51: rapport du module secteur	68
Figure 52: Classement des produits les plus chers pour un secteur	69

Liste des abréviations

BI	Business Intelligence
SQL	Structured Query Language
KPI	Key Performance Indicator
ETL	Extract-Transform-Load
TDL	Twins Digital Labs
SGBD	Système de Gestion de Base de Données Relationnel
SQL	Structured Query Language
PDI	Pentaho Data Integration
TOS	Talend Open Studio
SDK	Software Development Kit

Introduction générale

La plupart des projets implémentés par les entreprises sont destinés afin d'atteindre un certain nombre d'objectifs et de mieux s'adapter avec l'environnement externe. La réalisation du projet quel que soit la forme finale obtenue, représente pour ces entreprises l'atteinte totale ou partielle des objectifs déjà définis au départ.

Néanmoins, cette stratégie ne garantit pas la satisfaction des besoins des clients et n'accorde pas l'attention souhaitée à la mesure et au suivi des indicateurs et des effets du projet qui sont aussi importants tant pour les créateurs du projet que pour les bénéficiaires.

Afin de rectifier ces failles, ce projet est dédié précisément à créer un modèle qui permet d'assurer le suivi et l'évaluation des indicateurs du projet par les structures d'exécutions.

C'est dans ce cadre que nous nous sommes affectés au bureau TWINS DIGITAL LABS afin de mettre au point les fonctionnalités nécessaires au projet « AD-CAISSE », et produire des tableaux de bords qui tournent sur une analyse multidimensionnelle pour un usage simple et agréable de l'application.

Par ailleurs la notion du Business Intelligence est apparue. Elle propose des solutions qui se tournent vers l'analyse multidimensionnelle pour aider les membres de l'entreprise et les dirigeants dans leur pilotage.

La Business Intelligence facilite l'accès, l'interrogation et l'analyse de l'information issus des différents domaines fonctionnels et de différentes sources de données car elle contient les concepts d'entrepôt de données.

L'utilisation de ces systèmes d'information, qui englobent la notion d'entrepôt, constitue une source informationnelle fiable pour la prise des décisions au sein d'une entreprise

car elle permet de combiner les données, les synthétiser et les transformer en des informations clés pour aider et orienter les décideurs dans leur prise de décision.

Le rapport est divisé en cinq chapitres ainsi qu'une introduction générale et une conclusion :

Le premier chapitre appelé « Présentation du cadre du projet », c'est le chapitre qui décrit notre projet et notre entreprise d'accueil.

Le deuxième chapitre est nommé « La méthodologie de travail et les outils adoptés », ce chapitre intègre la méthode de travail utilisée ainsi que les outils Business Intelligence les plus appropriés pour notre projet.

Le troisième chapitre traite le premier sprint, il est appelé « Conception et mise en place d'un entrepôt de données » et il évoque principalement la conception d'un entrepôt de données.

Le quatrième chapitre traite le deuxième sprint « La mise en place d'un processus ETL » ce chapitre évoque l'implémentation de l'ETL.

Le cinquième chapitre appelé « Réalisation du tableau de bord » traite le troisième sprint, nous commençons par les indicateurs de clé performance (KPI), pour mener à un tableau de bord.

Chapitre 1 : Présentation du cadre du projet

1 Introduction

L'étude préalable consiste principalement à étudier l'existant, c'est-à-dire les solutions informatiques et les besoins que l'entreprise a mis en place, notamment en termes de fonctionnalités.

Au cours de ce premier chapitre, nous allons commencer par présenter la société TWINS DIGITAL LABS dans laquelle nous avons passé notre période de stage et puis nous allons parler de son domaine d'activité. Ensuite, nous allons présenter le projet et enfin nous tenons à critiquer l'existant afin de clarifier la problématique et donc de fournir des solutions et d'identifier les principaux besoins de l'entreprise.

2 Présentation de la société TWINS DIGITAL LABS

Créée en 2018, TWINS DIGITAL LABS est une société de services et d'ingénierie en informatique reconnue en tant qu'acteur national et international dans la mise en place de solutions digitales. Elle propose des services de formation et de coaching personnalisé dans tous les domaines de la technologie de l'information et de la communication, elle accompagne aussi les startups dans leurs projets web, mobile et business intelligence.

TWINS DIGITAL LABS est une startup du monde digital. Elle accompagne ses clients dans leurs transformations digitales et numériques à travers le développement et la mise en place de solutions digitales à forte valeurs ajoutées.



Figure 1: Logo Twins Digital Labs

Source : [W1]

3 Les services de la société

La société TWINS DIGITAL LABS garantie plusieurs services :

- Cadrage et stratégies.
- Développement agile.
- Test Automatiques, Tests d'intégration.
- Maintenance et support (Contrat de maintenance).

4 Etude de l'existant

L'entreprise possède actuellement une application AD-CAISSE qui permet de gérer l'ensemble des activités commerciales (transaction, règlements, passation de commande, gestion de stock, gestion des employées, etc...) toute en assurant un mécanisme de centralisation par rapport à la variance des activités pour un seul partenaire. A l'instant actuel, la solution est mise en production et exploitée par un réseau du marché local constitué de plusieurs partenaires réparties principalement sur les secteurs d'activités suivants : Restauration / vestimentaire / centre d'esthétique.

Cette application est dédiée pour : Les caissiers / clients / livreurs / chef cuisiniers

L'entreprise a un manque très important en matière décisionnelle par l'absence d'un système d'aide à la décision et le manque des moyens de reporting et de tableau de bord pour rassembler les bonnes informations dans les meilleurs délais.

La société TWINS DIGITAL LABS réalise le calcul de certains indicateurs manuellement, ce qui pose le problème d'un accès facile à des informations fiables et cohérentes.

Etant donnée l'augmentation des données, l'analyse par AD-CAISSE est devenue inévitable et une phase primordiale ayant comme objectif le cadrage de l'orientation du chemin évolutif du produit et la pénétration du domaine du consulting public pour la phase de pré-lancement des projets. Dans ce projet, notre équipe va prendre en charge la partie du Business Intelligence et elle sera la responsable de l'analyse décisionnelle.

5 Critique de l'existant

La société TWINS DIGITAL LABS a un manque important en matières décisionnelles par l'absence d'un système d'aide à la décision et l'indisponibilité des moyens de reporting et de tableaux de bords pour collecter les bonnes informations dans les meilleurs délais.

Les limites décelées sont :

- La société ne dispose d'aucun outil décisionnel.
- Une base de données qui n'est pas claire : manque de données, des différents formats de données etc.
- La diversité des données.
- La difficulté de trouver un outil de visualisation Open Source qui satisfait nos besoins.
- Manque d'une ETL (Extraction, Transformation, Chargement).

6 Solutions proposées

Pour améliorer le module d'analyse dans la société, nous avons proposé une solution Business Intelligence qui permet d'utiliser des données dans la base de données et créer des clés de performances à fin de construire des tableaux de bords :

- Mettre en place une solution décisionnelle au sein de la société.
- Mise en place d'un processus ETL.
- Transformez les données brutes des bases de données opérationnelles en visualisations d'informations alignées sur les objectifs de l'entreprise.
- Création des tableaux de bords afin de faciliter l'analyse.

Ainsi, on aura recours à :

- Un logiciel de gestion de la base de données : MYSQL WORKBENCH
- Un logiciel permettant la gestion des données hétérogènes ou homogènes au sein d'un système d'information : TALEND.
- Un logiciel de visualisation de données : KNOWAGE

7 Les besoins fonctionnels

Il s'agit des fonctionnalités du système. C'est l'ensemble des actions que le système fournit pour satisfaire les besoins de l'utilisateur :

- Assurer les fonctionnalités de reporting et de contrôle.
- Visualisation des données en temps réel.
- Génération des rapports.
- Comparateur des données.
- Génération d'une étude.
- L'accès rapide et simple aux informations.
- Création d'un ensemble des indicateurs clé performance qui répondent aux besoins de l'entreprise.

- Les données doivent être changeables selon les activités de l'application AD-Caisse.

8 Les besoins non fonctionnels

Il s'agit des besoins qui caractérisent le système. Ce sont des besoins en matière de comportement et de performance que le produit doit avoir.

- **L'ergonomie** : Améliorer l'ergonomie (les couleurs, le choix des dessins, l'organisation...) pour une meilleure perception des informations présentées
- **La disponibilité** : Les tableaux et les courbes doivent être disponibles à tout instant.
- **La simplicité** : La solution doit être facile à utiliser et à prendre en main, car l'utilisateur final n'est pas nécessairement un expert en informatique
- **La rapidité** : Le temps de réponse doit être acceptable.

9 Conclusion

Dans ce chapitre, nous avons présenté en premier lieu la société TWINS DIGITAL LABS dans laquelle nous avons passé notre période de stage. Ensuite nous avons établi une étude de l'existant, une critique de l'existant et nous avons proposé une solution pour étudier les besoins.

Chapitre 2 : La méthodologie de travail et les outils adoptés

1 Introduction

Au cours de ce deuxième chapitre, nous allons commencer par choisir la méthodologie de travail ainsi que les outils techniques que nous avons adoptés.

Dans la première section, nous allons faire une étude comparative entre les méthodes classiques de gestion de projet et les méthodes agiles.

Dans une deuxième partie, nous allons comparer les outils de travail que nous nous utiliserons.

2 Les méthodes classiques de gestion de projet

Les méthodes classiques sont les méthodes les plus utilisées dans la gestion des projets.

Une approche classique, la méthode en cascade, est une suite d'étapes consécutives, on les appelle aussi "cascades", puisqu'une étape de cette approche ne peut démarrer qu'après que l'ancienne étape soit terminée. La technologie en cascade essaie d'être accessible dans tous les cas pour éviter les retards qui font perdre du temps [W2].

Cette méthode permet une planification maîtresse du projet avant la phase de développement.

Par conséquent, le projet devrait être exécuté en plusieurs phases :

- **La phase de conception :** La phase de conception est une phase dans laquelle on conçoit le produit à mettre en œuvre et les outils ainsi que la détermination des objectifs. Cette phase comprend les études d'avant-projet et la définition des spécifications.
- **La phase de réalisation :** La phase de réalisation ou on peut aussi l'appeler la phase de développement d'un projet, au cours de laquelle on crée le produit.
- **La phase de recette :** C'est la phase de vérification du produit en se basant sur les spécifications définies dans la phase de conception. Ce n'est qu'à la fin de cette phase que les collaborateurs du projet décident de mettre en production ou de fournir des produits (logiciels) aux utilisateurs.

Les risques des méthodes classiques :

- Le manque de communication et d'échange.
- Le manque de l'autorégulation.
- La péremption du produit : tels projets peuvent être de longue durée et le résultat final est souvent difficile à adapter à l'évolution des besoins.
- Le manque de flexibilité.

Prenons l'exemple la méthode de la boucle en cascade (boucle en V) qui est la plus utilisée dans les projets de prise de décision, le risque n'est détecté dans la phase de test qu'une fois la phase de développement terminée.

La figure ci-dessous illustre la démarche du cycle en V :

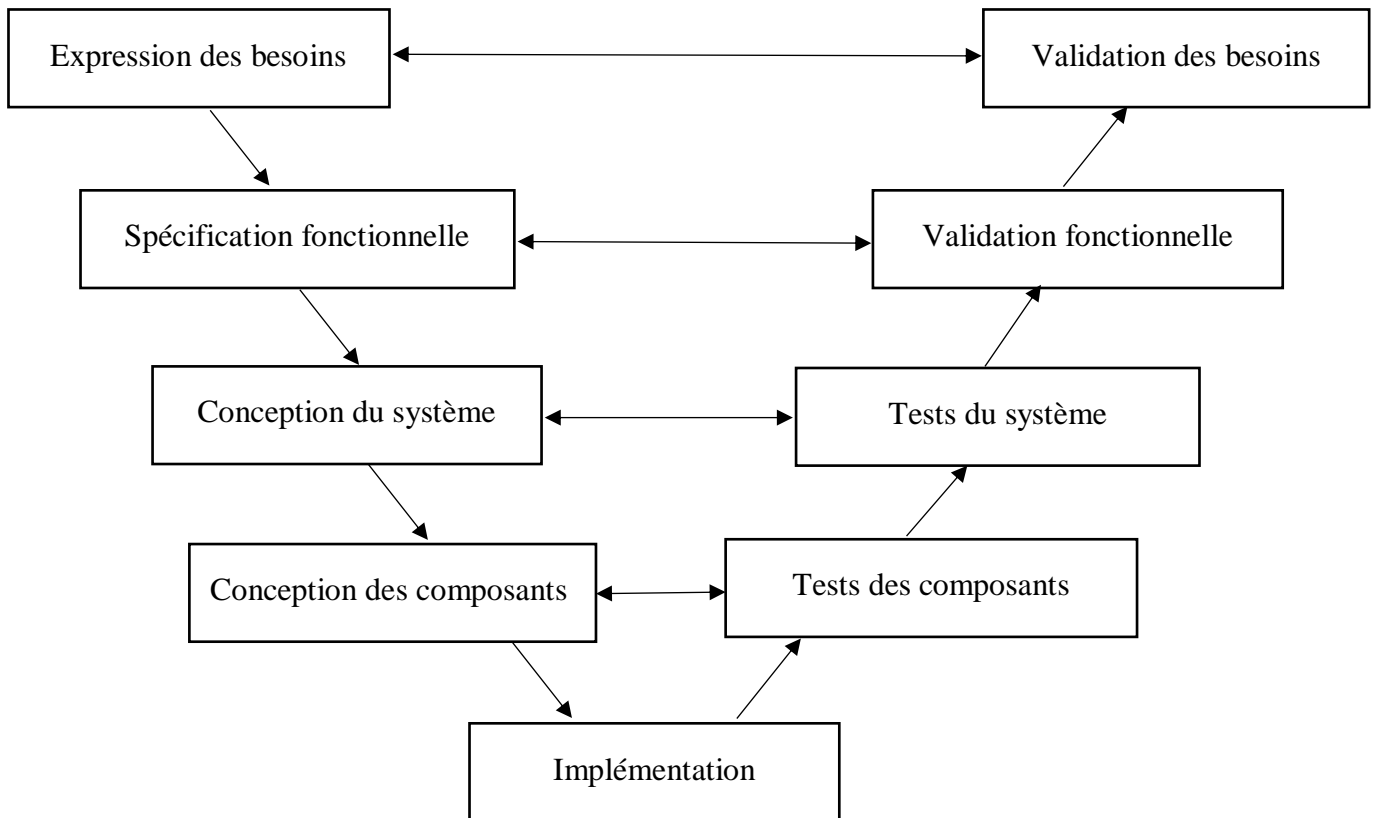


Figure 2: Démarche méthode « Cycle en V »

Source : [W3]

3 Les méthodes Agiles

La méthode agile est un outil qui regroupe plusieurs méthodologies, concepts et pratiques dédiés à la gestion des projets informatiques.

La méthode agile est une approche itérative qui tient son nom du Manifeste agile, elle privilège le dialogue entre toutes les parties prenantes du projet (clients, développeurs, directeur du projet,...) et elle consiste à fixer un objectif à court terme [W4].

Cette méthode agile implique :

- Des réponses rapides et flexibles.
- Des plans d'action flexibles.
- Le travail en équipe.
- Simplicité des processus.
- Une date limite est imposée.
- Engagement du client à la fin de chaque sprint.

Dans le cadre d'un projet de prise de décision, la méthode agile peut être utilisée pour garantir une communication efficace entre les parties prenantes du projet afin de corriger et modifier en fonction des besoins du client.

Les Principes des méthodes agiles

Les méthodes agiles s'articulent autour de douze principes, parmi lesquels on peut citer [W5]:

1. La satisfaction du client.
2. Accepter les demandes de changement.
3. Délivrer en continu des versions opérationnelles des applications.
4. Garantir une collaboration continue entre l'équipe et le client.
5. Construire les projets autour des individus motivés.
6. Favoriser la communication en face à face.
7. Estimer l'avancement du projet en fonction des fonctionnalités.
8. Le rythme de l'avancement du projet doit être constant.
9. Contrôler l'excellence technique.
10. Favoriser la simplicité.
11. La responsabilisation des équipes.
12. Ajuster régulièrement les comportements et les processus pour améliorer l'efficacité

Parmi les méthodes agiles, on peut citer l'exemple de la méthodologie Scrum.

3.1 La méthodologie Scrum

Scrum est une méthode de gestion itérative d'un projet informatique. Cette méthodologie a tendance à livrer rapidement des prototypes, fonctionnant par définition afin que les clients, les sous-traitants et les membres de l'équipe puissent l'évaluer [W6].

On peut distinguer trois phases dans la méthodologie Scrum:

- **La phase initiale:** au cours de laquelle les fonctionnalités du système sont listées et une architecture logicielle générale est définie.
- **La phase de suivie:** c'est une série de "sprints", chaque sprint correspondant à un incrément du système.
- **La phase de terminaison du projet:** cette phase permet de développer les artefacts et de tirer des leçons apprises durant le développement.

3.2 Processus Scrum

Le processus Scrum est le plus utilisé dans les méthodologies agiles, il fonctionne en cycles appelés sprints. Chaque sprint est d'une durée de 1 à 4 semaines afin de garder un rythme constant. Le concept est très facile, l'équipe qui avance est toujours prête à s'autodiriger au fur et à mesure de l'avancement du projet.

La méthodologie Scrum permet une meilleure adaptabilité, rapidité et flexibilité requises par l'environnement de travail qui continu d'être à la fois imprévisible et volatil. Les méthodes agiles remettent les personnes sous le feu des projecteurs et favorisent largement les interactions.

À la fin de chaque sprint, une version testable est fournie pour évaluation [W7].

On a présenté les processus Scrum comme l'illustre cette figure :

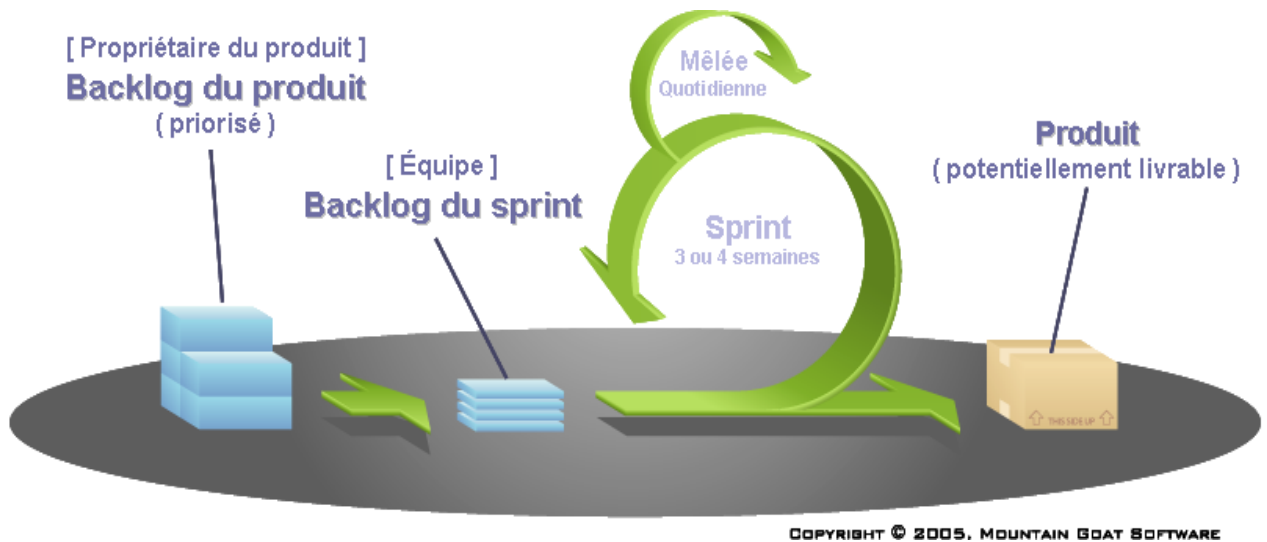


Figure 3: Cycle de vie d'un Scrum

Source : [W8]

3.3 Les artefacts de Scrum

- **Backlog des produits**

C'est une liste de tâches à faire que l'équipe Scrum doit aborder. Il peut s'agir de définitions de fonctionnalités pour le logiciel, d'exigences logicielles, d'histoires d'utilisateurs ou de descriptions de tâches supplémentaires nécessaires, telle que la définition de l'architecture ou la documentation utilisateur.

- **Sprint**

Le Scrum est une approche itérative, ces itérations sont appelées « sprints ». Chaque sprint dure entre 1 et 4 semaines selon la charge de travail de l'équipe, ainsi que le nombre de fonctionnalités voulues à la fin de cette durée.

- **L'Incrément produit**

L'incrément en Scrum correspond à l'ensemble des items du backlog de produit qui ont été accomplis pendant le sprint en cours.

3.4 Les acteurs d'un Scrum

La méthodologie Scrum réunit trois acteurs principaux : Le Product Owner, Le Scrum Master et l'équipe de développement. Dans cette approche, le Product Owner et le Scrum Master remplacent le chef de projet.

Dans cette section, nous allons examiner de plus près le rôle de chaque acteur [W9]:

- **Product Owner:**

Le Product Owner, ou le directeur de produit, est le stratège du projet. C'est lui qui a pour mission de définir les fonctionnalités du produit final. C'est lui aussi qui choisit les dates et les contenus des différentes versions. Il définit les priorités dans les fonctionnalités et valide ou non les travaux à chaque fin de Sprint. Le Product Owner a la responsabilité du retour sur investissement ainsi que l'acceptation ou le rejet des résultats.

C'est la personne responsable de la qualité fonctionnelle du produit, représente le client et la plus impliquée dans la gestion du « backlog de produit ».

➔ Dans ce projet le maitre de stage est le Product Owner.

- **Le Scrum Master :**

C'est la personne en charge du management du projet. Il doit faire appliquer les pratiques et les valeurs du Scrum. Il s'assure que l'équipe est totalement opérationnelle et productive. En cas de problème, c'est lui qui doit éliminer les obstacles pour que l'équipe puisse continuer le projet. Son but est de faciliter l'organisation de l'équipe et d'augmenter la productivité de l'équipe.

Il a pour rôle d'aider l'équipe à travailler et à s'améliorer constamment.

➔ Dans ce projet le maitre de stage est le ScrumMaster.

- **L'équipe de développement :**

L'équipe de développement est responsable de la livraison des projets qui ont été priorisés pour ce sprint à la fin de chaque sprint. L'équipe est habituellement composée de 2 à 10 personnes, et il est très important qu'elle s'auto-organise.

Rappelons que le Scrum Master n'est pas un client, mais un prestataire de services, il doit s'occuper de l'auto-organisation et l'orientation de l'équipe et le directeur de produit n'a pas à se débattre dans l'organisation de l'équipe.

➔ Dans ce projet, nous sommes l'équipe de développement

3.5 Les événements de la méthodologie Scrum

- **Le sprint**

Le Scrum est une méthode de gestion itérative. Ce dernier dure habituellement de 2 à 4 semaines pendant laquelle une version du produit est terminée et soit utilisable. Le nouveau sprint commence à la fin du précédent. Chaque sprint a un objectif et une liste de fonctionnalités à atteindre. Des modifications peuvent être appliquées pendant un sprint, mais cela va avoir une incidence sur les itérations suivantes.

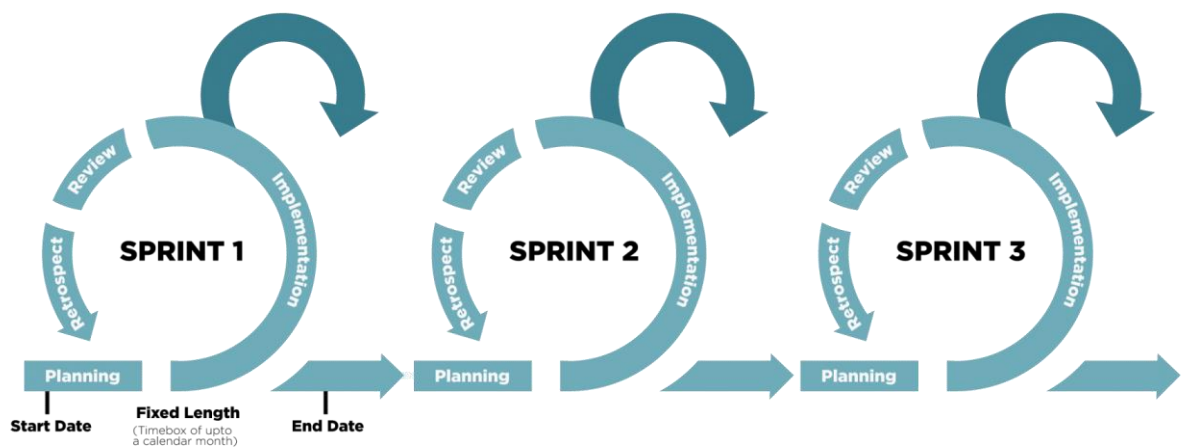


Figure 4: Les Sprints

Source : [W10]

- **Planification d'un sprint**

La planification d'un sprint est la phase la plus difficile de la méthodologie Scrum, c'est celle qui lance le Sprint.

Pendant la réunion, l'équipe définit ce qui peut être livré dans le Sprint et décide des travaux à réaliser lors de cette période. Nous déterminons ensuite le processus par lequel on va atteindre nos objectifs à travers les tâches à accomplir. À la fin de la réunion, l'équipe promet d'arriver aux objectifs dans les meilleurs délais.

Dans la méthodologie Scrum, toute l'équipe choisit les éléments qu'elle va définir dans le backlog de produit. Une fois la liste des tâches de sprint est créée, les tâches peuvent être identifiées et estimées. Enfin, nous pouvons lancer le Sprint.

- **Mêlée quotidienne**

Une réunion quotidienne de 15 minutes est indispensable. Elle se déroule au même moment et au même endroit tous les jours. Le Scrum master veille à ce que les réunions se tiennent à heures fixes.

Son objectif est de faire le point sur les progrès quotidiens de Sprint. Il permet à l'équipe de synchroniser ses activités, de planifier les prochaines 24 heures et de répartir les tâches à traiter. Chaque membre de l'équipe de développement doit répondre aux trois questions suivantes:

1. Qu'ont-ils réalisé la veille?
2. Que vont-ils faire aujourd'hui?
3. Quels sont les obstacles qui les retiennent ?

- **Revue du sprint**

À la fin du sprint, toute l'équipe se réunit pour un examen du sprint, qui dure jusqu'à 4 heures. Le but de la revue de sprint est de vérifier le logiciel généré pendant le sprint.

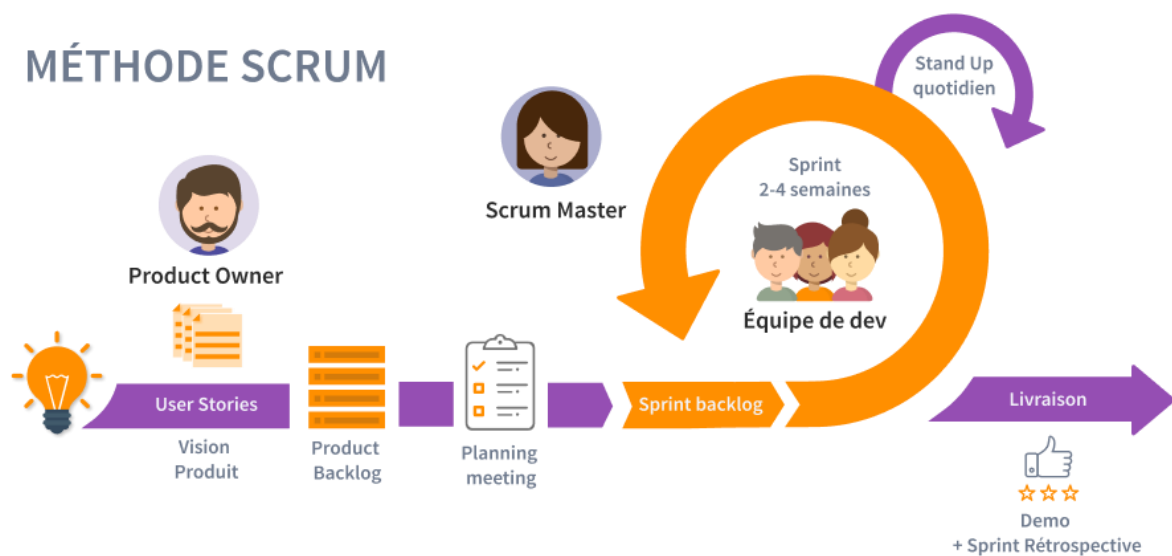
- **Rétrospective du sprint**

Le but de la réunion finale est de mener une évaluation post-mortem sur le sprint pour voir s'il y avait des bugs ou si certains points nécessitaient des améliorations.

La réunion est présidée par le responsable de Scrum (Maitre de stage), qui a guidé l'équipe pour apporter des changements et des améliorations.

Cette figure présente un récapitulatif d'un processus Scrum :

Figure 5: Récapitulatif du processus Scrum



Source : [W11]

4 La méthode adoptée

Après une recherche comparative entre les méthodes classiques de gestion de projet et les méthodes agiles, nous nous sommes mis d'accord à utiliser la méthode agile pour réaliser notre projet. Et plus précisément, nous allons se concentrer sur la méthode Scrum qui sera utilisée dans ce projet.

4.1 La méthodologie Scrum appliquée sur notre projet

4.1.1 Pilotage du projet avec la méthodologie Scrum

Dans ce projet, l'équipe se compose de trois membres :

- Le Product Owner est Mr Oussama DRIDI, Co-Fondateur de Twins Digital Labs.
- Le Scrum Master est Mr Oussama DRIDI.
- L'équipe de développement est composé de Hend LEGLEG et Mohamed TRABELSI

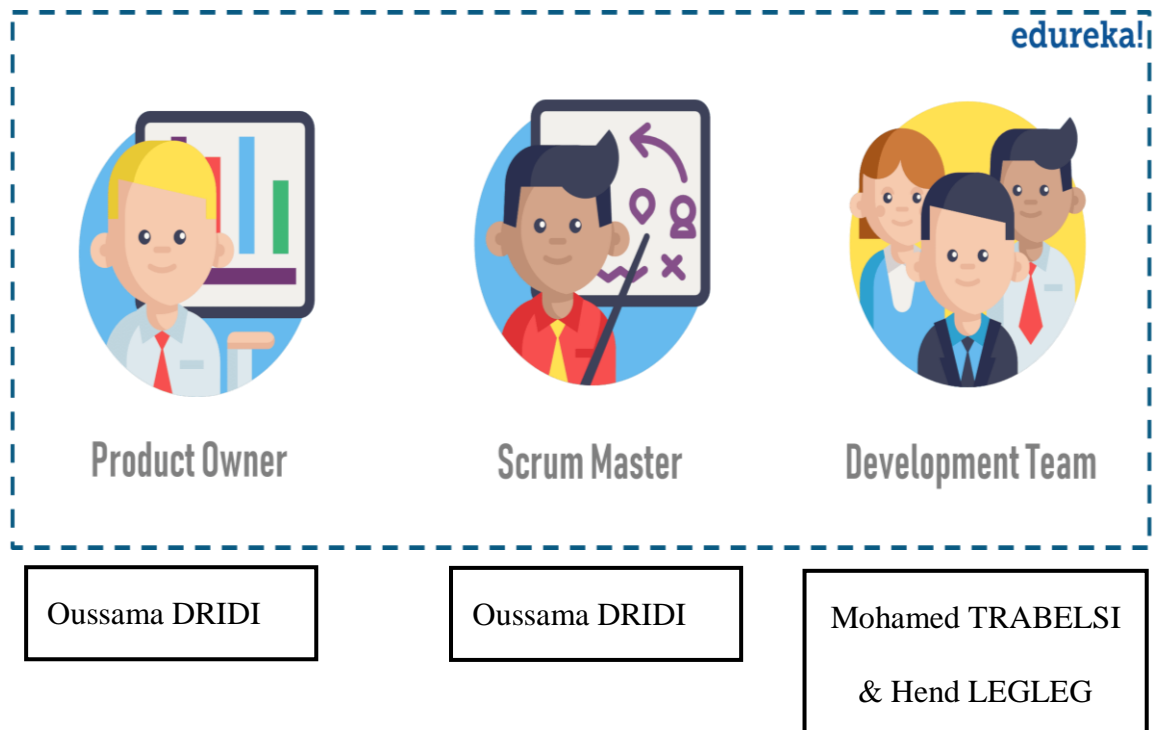


Figure 6: Equipe Scrum

Source : [W12]

4.1.2 Découpage des Sprints

Nous avons adopté le découpage suivant :

- Le premier Sprint : Conception et mise en place d'un entrepôt de données.
- Le deuxième Sprint : La mise en place d'un processus ETL.
- Le troisième Sprint : Tableau de bord.

4.1.3 Diagramme de Gantt

Afin d'exécuter notre projet nous avons utilisé le diagramme de Gantt pour avoir une vision globales des tâches à effectuer selon le calendrier suivant :

- Étape 1 : La montée en compétences.
- Étape 2 : Etude de l'existant et proposition des solutions.
- Étape 3 : Modéliser et développer le premier Sprint.
- Étape 4 : Modéliser et développer le deuxième Sprint.
- Étape 5 : Modéliser et développer le troisième Sprint.
- Étape 6 : Rédiger le rapport.



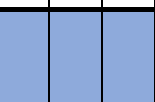

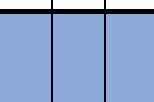

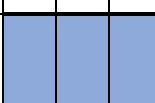
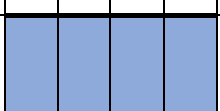
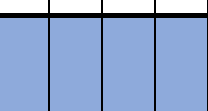
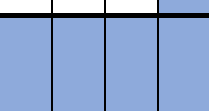
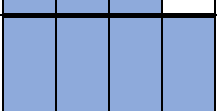
	<i>Février</i>	<i>Mars</i>	<i>Avril</i>	<i>Mai</i>
<i>Étape 1</i>				
<i>Étape 2</i>				
<i>Étape 3</i>				
<i>Étape 4</i>				
<i>Étape 5</i>				
<i>Étape 6</i>				

Tableau 1: Diagramme de Gantt

5 Les phases d'un projet décisionnel

Avant d'étudier les meilleures pratiques pour la réussite de notre projet, il est indispensable de rappeler qu'un projet décisionnel doit respecter les étapes qui permettent de structurer la chaîne décisionnelle :

Dans cette section, nous allons parler des phases d'un projet décisionnel :

- **Une phase de d'alimentation** : Les données opérationnelles (les données internes de la société) doivent être récupérées de manière pertinente à travers le processus ETL (Extract – Transport - Load), c'est-à-dire extraites de différentes sources disponibles, filtrées et adaptées à la prise de décision.
- **Une phase de stockage et de modélisation dimensionnelle** : Les données sont structurées, centralisées et stockées dans l'entrepôt de données. Celui-ci, doit être orienté sujet, intégré, non volatile et historisé. Cette aborde les notions des datamarts et de cubes qui doivent répondre aux besoins de la société.
- **Une phase de restitution / distribution** : Dans cette phase, divers outils de retour interviennent. On distingue des outils de reporting, des portails donnant l'accès aux tableaux de bord et des outils de navigation en cubes .
- **Une phase d'analyse et d'exploitation** : Après la génération des tableaux de bord, l'utilisateur final doit être en mesure d'analyser les informations et de tirer des conclusions. Cette phase peut inclure le datamining pour explorer les corrélations possibles et les analyses multidimensionnelles.

La figure ci-dessous représente le cycle de vie d'un projet décisionnel :

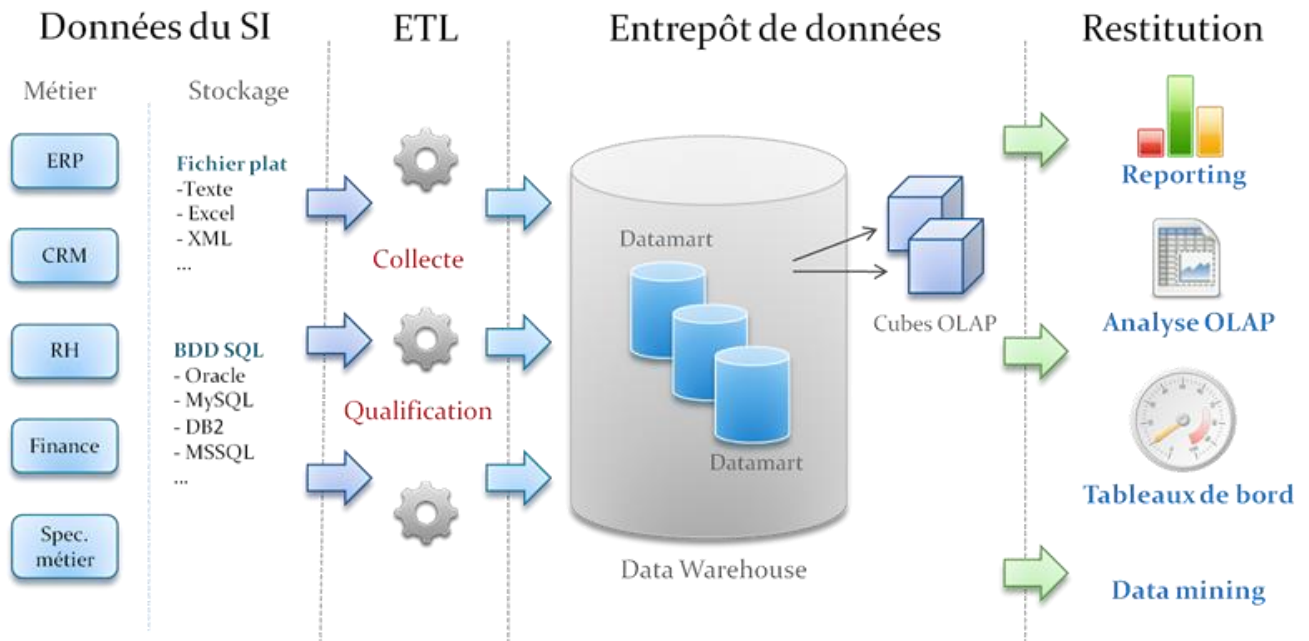


Figure 7: Chaîne de projet décisionnel

Source : [W13]

6 Les choix techniques

Dans cette section, on va parler des choix techniques utilisés au cours de notre projet

6.1 Les outils de gestion de la base de données

6.1.1 MySQL

Le MySQL (Structured Query Language), un système fourni par MYSQL AB, est un Système de Gestion de Base de Données Relationnelle (SGBDR) open source d'Oracle basé sur le langage de requête structuré (SQL). Il est connu pour ses fonctions puissantes, il possède une variété de fonctionnalités riches et avancées qui peuvent gérer de grandes quantités de données. Il donne la priorité aux performances et il permet d'interroger et de modifier les données d'une manière plus intuitive et graphique. [W14].

6.1.2 PostgreSQL

PostgreSQL est un Système de Gestion de Base de Données Relationnel (SGBDR) open source. Il peut prendre en charge en toute sécurité les charges de travail de données les plus compliquées. Il gère les transactions, les déclencheurs, les sous-sélections, les vues et l'intégrité référentielle de clés étrangères, ainsi que le verrouillage évolué. Il s'exécute sur de nombreuses plateformes dont Linux, la majorité des offres UNIX, Tru64, Solaris, Mac OS-X, Tru64 et Windows [W15].

6.1.3 Oracle database

Oracle database est essentiellement un Système de Gestion de Base de Données Relationnel (SGBDR). Il est connu comme une base de données d'entreprise, édité par la société Oracle-Corporation.

Oracle permet de gérer un gros volume de données et il peut gérer aussi les données géographiques [W16].

6.1.4 Comparaison entre les outils de gestion de la base de données

Cette table montre la comparaison entre les différents outils de création de datawarehouse.

	MySQL	PostgreSQL	Oracle
La description	SGBDR open source largement utilisé	SGBDR open source largement utilisé	SGBDR closed source largement utilisé
Modèle de base de données	SGBD relationnel	SGBD relationnel	SGBD relationnel
Développeur	Oracle Database	Groupe de développement mondial PostgreSQL	Oracle Database
Licence	Open source	Open source	Commercial
Systèmes d'exploitation	Windows -Solaris- Os - Linux	FreeBSD - Linux - NetBSD - OpenBSD - OS X - Solaris Unix	AIX - HP-UX - Linux - OS X -
Langage d'implémentation	C et C++	C	C et C++

Tableau 2: Comparaison entre les outils de création de data Warehouse

6.1.5 Choix adopté

Après une recherche comparative entre les outils de gestion de base de données, nous nous sommes mis d'accord à utiliser le MYSQL pour la réalisation de notre projet.



Figure 8: MySQL

Source : [W17]

6.2 Les différents outils de Extract Transform Load (ETL)

6.2.1 Talend open source

Talend est un logiciel open source, c'est un éditeur de solutions informatiques spécialisé dans la gestion des données. C'est un outil ETL (extraction, transformation, chargement) qui permet d'extraire des données hétérogènes, de les modifier et de les charger vers une base de données cible.

Cet outil est aussi fréquemment utilisé pour mettre en place des entrepôts de données dans les projets décisionnels.

Le Talend Open Source(TOS) peut gérer toutes les problématiques liées au traitement des données de la chaîne de décision et assurer le contrôle de la qualité des données [W18].

6.2.2 Pentaho Data Integration

Pentaho Data Integration(PDI) est une plate-forme d'aide à la décision de business intelligence open source développée par PENTAHO CORPORATION.

Son objectif principal est de permettre aux utilisateurs de transformer, concevoir et manipuler des données après qu'elles ont été récupérées à partir de sources hétérogènes dans différents formats. [W19]

6.2.3 Comparaison entre les outils d'ETL

La table suivante montre la comparaison entre les différents outils d'ETL

	Talend	Pentaho Data Integration
Licence	Open source	Open source
Création d'ETL	Oui	Oui
Outil d'analyse d'impact	Non	Oui
Exécution des requêtes	Oui	Oui
interface graphique	L'interface graphique de Talend est un peu difficile à saisir	L'interface graphique de Pentaho est assez modernisée et facile à comprendre
Déploiement	Nécessite un moteur Java indépendant pour s'exécuter sur une machine distincte	Le fichier Java et Perl peuvent s'exécuter indépendamment sur n'importe quelle machine
Vitesse	Talend est plus lent que Pentaho	Pentaho est presque deux fois plus rapide que Talend
Outil de création de requête	Oui	Non
Lecture/écriture de types complexes de données	Données cartographiques	Non

Tableau 3: Comparaison entre les outils d'ETL

6.2.4 Choix adopté

Après une recherche comparative entre les outils ETL, nous nous sommes mis d'accord à utiliser l'outil Talend pour faire le processus ETL.

La figure ci-dessous illustre le LOGO de Talend Open source:



Figure 9: Talend Open source

Source : [W20]

6.2.5 Rédaction de Scripts

Nous avons utilisé une solution scriptée pour des besoins simples en rédigeant des requêtes SQL pour simplifier, modifier, transformer les valeurs de champs dans les tables.

6.3 Les outils de visualisation de données

6.3.1 Microsoft power BI

Power BI défini par Microsoft est une solution Business Intelligence qui permet d'effectuer des analyses de données, il permet de créer des visualisations de données afin de permettre aux entreprises de créer leurs propres rapports et tableaux de bord (Dashboards) en toute sécurité.

La figure suivante illustre l'interface de Microsoft Power BI :

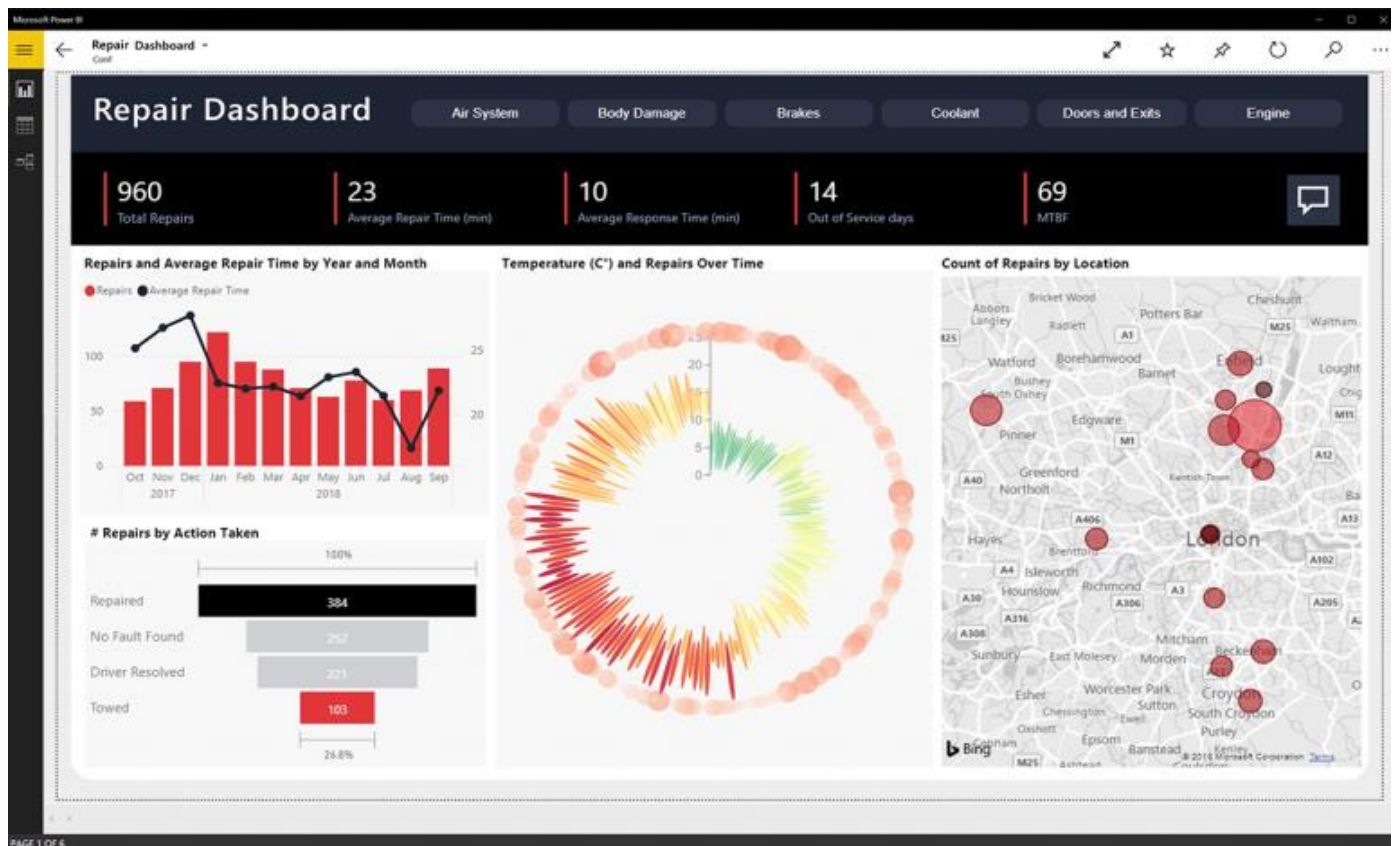


Figure 10: Interface du Microsoft Power BI

Source : [W21]

6.3.2 Knowage

Knowage est la nouvelle marque de SpagoBI résultant de nombreuses transformations, il s'agit d'une plate-forme de Business Intelligence (BI) qui permet des analyses open source couvrant tous les aspects de l'analyse de données. Il s'agit d'un outil de visualisation de données qui nous permet de créer des graphiques et des tableaux de bord en combinant données traditionnelles et Big Data [W22].

La figure suivante illustre l'interface de Knowage:

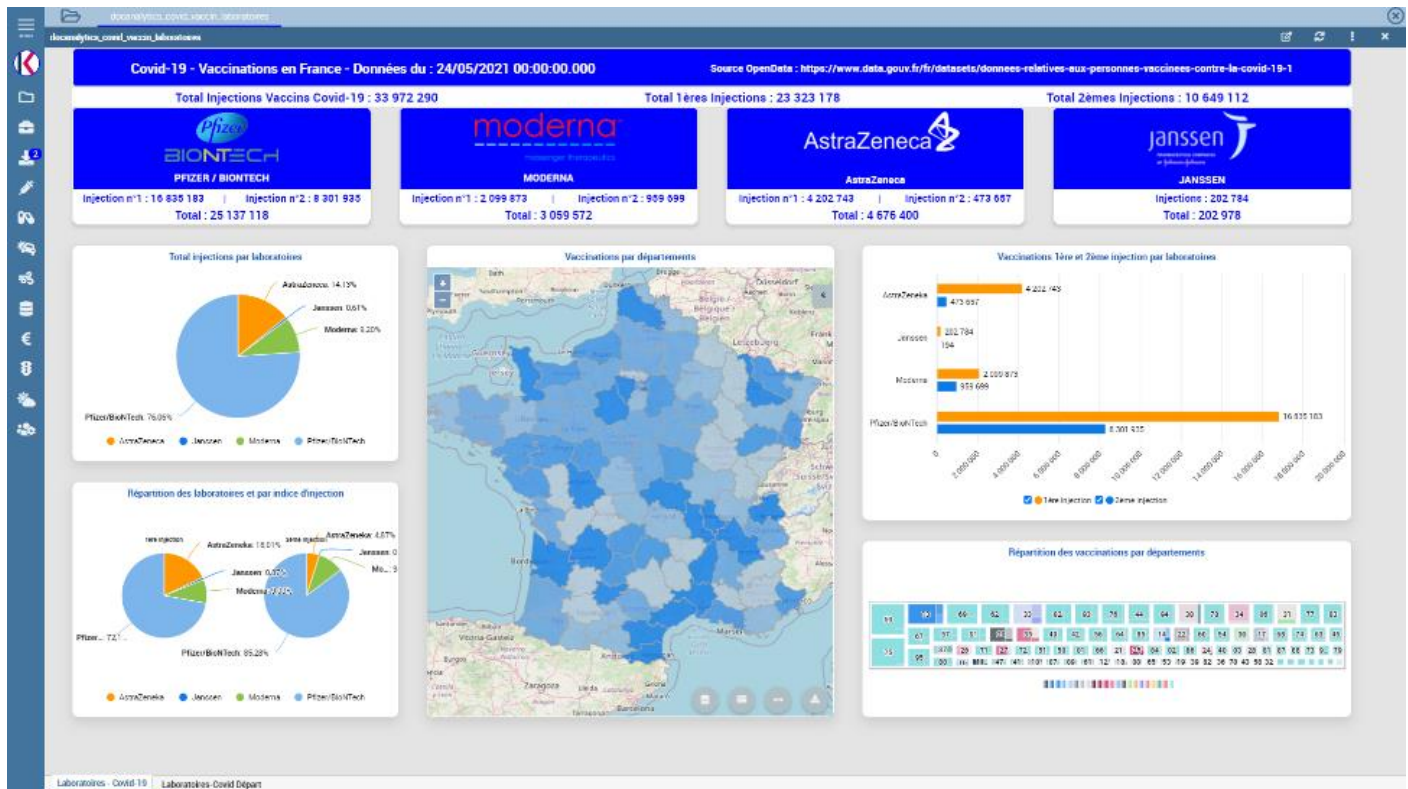


Figure 11: Interface de Knowage

Source : [W23]

Les Fonctionnalités de Knowage

- **Big Data Analytics (module BD)** : permet d'analyser des métadonnées dans des graphes de données et les bases de données NoSQL ainsi que la combinaison des données traditionnelles et Big Data en un modèle unique.
- **Smart Intelligence (module SI)** : permet de faire le reporting ad hoc et il est orienté vers les capacités de libre-service
- **Location Intelligence (module LI)** : relier les données d'entreprise à des informations spatio-géographique.

- **Performance Management (module PM)** : permet de gérer les indicateurs clés de performance et la formalisation de la stratégie des organisations afin de la déployer et de la mettre en œuvre plus efficacement
- **Predictive Analysis (module PA)**: simulations et prédictions pour des analyses avancées

6.3.3 Comparaison entre les outils de visualisation de données

	Knowage	Power BI
Licence	Open source	Open source
Tableau de bord	Oui	Oui
Analyse visuelle	Oui	Oui
Indicateurs clés de performance (KPI)	Oui	Non
Analyse prédictive	Oui	Non
Rapports Ad hoc	Oui	Oui
Publication / Partage	Oui	Non
Déploiement	Cloud, Saas, Web, Mac, Windows, On-Premise - Windows On-Premise – Linux	Cloud, Saas, Web, Windows, Android

Tableau 4: Comparaison entre les outils de visualisation de données

6.3.4 Choix adopté

Après une recherche comparative entre les outils de visualisation Microsoft Power BI et Knowage, nous nous sommes mis d'accord à utiliser l'outil Knowage pour la visualisation des tableaux de bords dans notre projet.

La figure ci-dessous illustre le LOGO de Knowage :



Figure 12: Logo Knowage

Source : [W24]

6.3.5 Interprétation

Knowage permet aux utilisateurs finaux d'avoir des vues modernes d'analyse de données et propose plusieurs fonctionnalités en libre-service.

Les utilisateurs peuvent créer leur propre analyse et explorer leur propre espace de données, vous pouvez également combiner des données provenant de différentes sources en un seul modèle.

7 Architecture globale

La figure ci-dessous présente l'architecture globale de notre projet qui englobe les outils techniques utilisés.



Figure 13: L'architecture totale de notre Projet

8 Conclusion

Dans ce chapitre et à après une étude comparative, nous avons présenté les différents outils informatiques qui seront utilisés dans la mise en œuvre de notre projet

Chapitre 3 : Conception et mise en place d'un entrepôt de données

1 Introduction

Après avoir évoqué les activités de la société TWINS DIGITAL LABS et des outils informatiques que nous allons utiliser dans la réalisation du projet, ce chapitre se concentre sur le premier sprint qui a pour but d'étudier les données sources et la conception du data Warehouse.

Dans la première partie, nous nous concentrerons sur la base de données et les tables de l'application AD-CAISSE qui sont mises en production et exploitées par un réseau local constitué de plusieurs partenaires réparties sur des différents secteurs d'activités et chacun d'eux vend des produits esthétiques ou bien alimentaires. Dans une deuxième partie, nous allons présenter l'entrepôt de données et son modèle en étoile.

2 Le premier sprint : Conception d'un entrepôt de données

Backlog du Sprint

Ce Sprint a pour objectif de livrer un modèle d'entrepôt de données qui enregistre les données du système opérationnel.

User Story

L'utilisateur a besoin d'une modélisation d'un entrepôt de données qui regroupe l'ensemble des données de l'entreprise afin de bien comprendre les relations entre les données.

Les tâches à faire

- Etude de données sources.
- Conception de la base de données.
- Choisir les dimensions et la table des faits pour le futur Datawarehouse.
- Modélisation de l'entrepôt de données.

2.1 Etude de sources des données

Cette section s'adresse à comprendre les données sources de l'entreprise qui ont été livrées sous formes de tables dans MYSQL WORKBENCH version 5.7 :

- **Table produit** : Cette table contient toutes les informations sur le produit : son nom, son type (service ou stockable), et le partenaire auquel il appartient.
- **Table partenaireBprice** : Cette table contient tous les détails des partenaires qui utilisent l'application AD-Caisse : leurs secteurs d'activité, leurs adresses etc...
- **Table commandeDetails** : Cette table comprend tous les produits dans chaque commande ainsi que ceux qui sont annulées avec le prix de chaque produit et sa quantité.
- **Table commande** : Cette table comprend toutes les commandes qui ont été prises dans chaque point de vente en donnant leur date de création.
- **Table pointVente** : Cette table contient tous les points de ventes des partenaires (un partenaire peut avoir un ou plusieurs points de ventes)
- **Table categorieArticle** : Cette table comprend toutes les informations concernant une catégorie (une catégorie contient plusieurs produits)
- **Table Sector** : Chaque partenaire travaille sur un secteur bien défini, cette table fait le lien entre les noms des secteurs et leurs id dans les autres tables.

2.2 Conception de la base de données

Cette section se focalise spécifiquement sur la modélisation des données au niveau relationnel ou logique.

La figure ci-dessous présente la modélisation de la base de données :

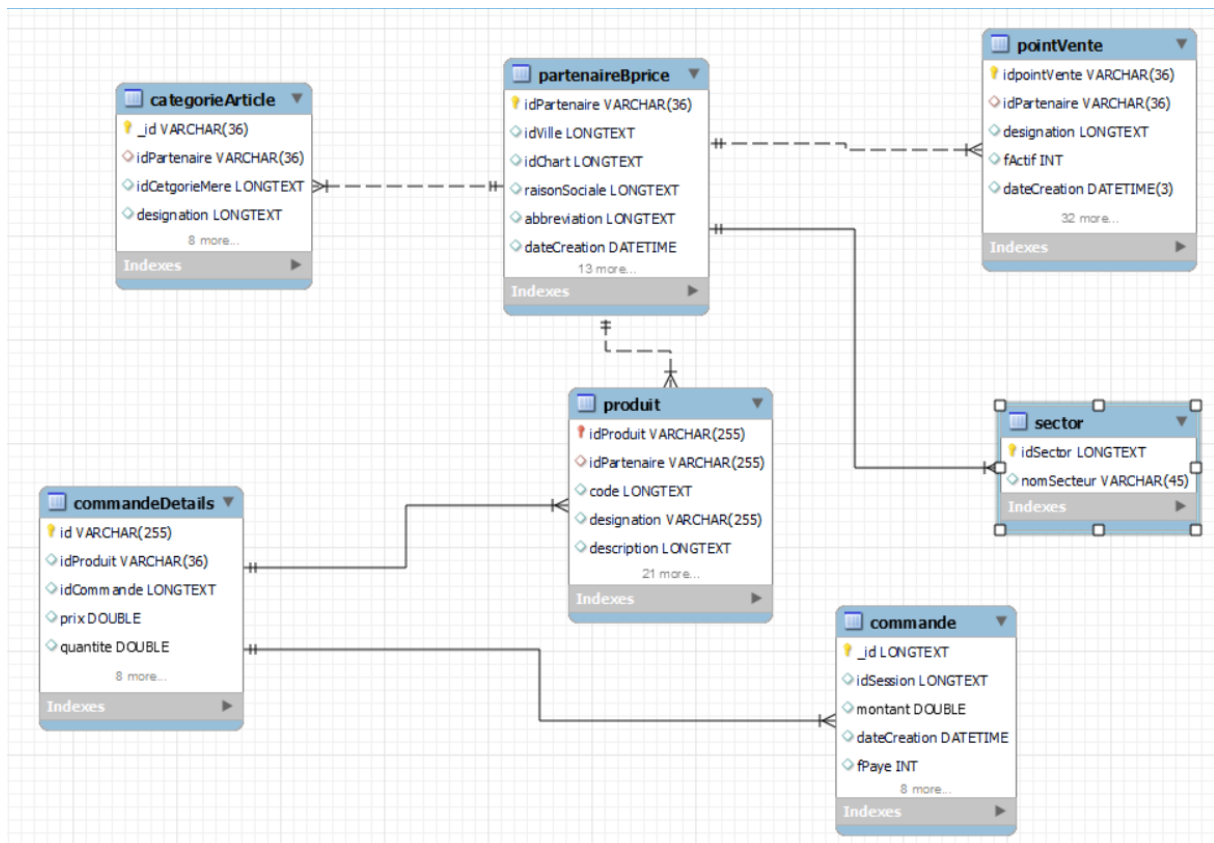


Figure 14: Base de données source

2.3 La modélisation multidimensionnelle

La modélisation multidimensionnelle est une méthode de conception logique qui vise à présenter les données sous une forme standardisée.

Un modèle décisionnel contient les mêmes informations qu'un modèle Entité/Relation, mais présente les données dans un format symétrique pour faire l'analyse de données.

Une modélisation multidimensionnelle permet de :

- Faciliter l'accès aux informations.
- Avoir un espace de stockage élevé
- Traiter les informations par lots.

L'entrepôt de données

Définition

William H. Inmon² a défini un entrepôt de données comme suit :

« Le Data Warehouse est une collection de données orientées sujet, intégrées, évolutives dans le temps, non volatiles et organisées pour le support d'un processus d'aide à la décision. »

[B1]

Un entrepôt de données est une base de données relationnelle qui permet de charger les données issues de différentes sources afin de les standardiser sous un format standard. [W25]

Un datawarehouse est :

- Orienté selon un sujet : Les données du Datawarehouse sont organisées.
- Historisé : Suivre les modifications apportées aux données dans la base de données et des dossiers, afin qu'un rapport de synthèse puisse être rédigé.
- Non volatile : Les données de l'entrepôt de données ne peuvent être ni modifiées ni supprimées, ils sont statiques et disponible en lecture seulement.
- Intégré : Les différents formats des données devraient être standardisés afin qu'elles aient la même apparence.

Les avantages de l'entrepôt de données

- Accès simple aux données.
- Améliorer les actions commerciales des entreprises.
- Obtenir une vision claire des données.

Le datawarehouse consiste en deux nouveaux concepts tels que les faits et les dimensions :

La table des faits :

- C'est la table principale du modèle décisionnel, elle modélise le sujet de l'analyse (Le fait).
- La table des faits ne contient que des valeurs numériques.
- Elle contient les clés étrangères des tables de dimensions.

Les tables des dimensions :

Le sujet analysé (le fait), est analysé suivant différentes axes caractérisant ses mesures de l'activité.

Ce sont les dimensions :

- Chaque dimension modélise les axes d'analyses.
- Chaque dimension comporte un ou plusieurs attributs.
- Les dimensions sont tout ce qui sera utilisé pour faire l'analyse.
- Chaque attribut de la dimension a ces propres caractéristiques.

La table des faits permet aux utilisateurs à analyser les dimensions de l'entreprise.

2.4 Conception de l'entrepôt de données

Dans cette section, nous allons présenter les étapes pour aboutir à un datawarehouse : le choix des dimensions, la table des faits et ses mesures.

L'outil de modélisation utilisé : Nous avons utilisé MySQL WORKBENCH pour la modélisation du Datawarehouse



Figure 15: MYSQL WORKBENCH

2.4.1 Choix des mesures

La construction de la table des faits nécessite une préparation et recherche de données. Puis, il faut préciser les composants dont on veut les mesurer.

Dans le tableau suivant, nous indiquons les mesures dans la table de faits :

Mesure	Libellé
prix	Prix des produits
quantité	Quantité des produits vendus

Tableau 5: les mesures de la table de faits

2.4.2 Choix des dimensions

Nous présentons dans cette partie les différentes dimensions qui vont constituer notre modèle. La table de dimension contient des attributs ainsi qu'une clé primaire, ce sont les axes d'analyse.

Dans le tableau suivant nous indiquons les dimensions :

Dimension	Libellé
dimcatégorie	Dimension Catégorie Article
dimDetailCommande	Dimension DétailCommande
dimcommande	Dimension commande
dimpointvente	Dimension pointVente
dimpartenaire	Dimension partenaire
dimsecteur	Dimension secteur
dimtime	Dimension temps
dimproduit	Dimension produit

Tableau 6: Les tables de dimension

2.4.3 Table des faits

Une fois que nous avons la liste des indicateurs pour la modélisation du modèle dimensionnel, nous avons alors besoin d'une table des faits.

Une table des faits est une table qui contient les données observables (les faits) que l'on possède sur un sujet et que l'on veut étudier, selon divers axes d'analyse (les dimensions)

	Attribut	Nécessité
Les clés étrangères	idDimCatégorie	Identifiant de la dimension catégorie
	idDimcommandeDétails	Identifiant de la dimension commandeDétails
	idDimPartenaire	Identifiant de la dimension partenaireBprice
	idDimPoint Vente	Identifiant de la dimension pointVente
	idDimProduit	Identifiant de la dimension produit
	idDimSecteur	Identifiant de la dimension secteur
	idDimTime	Identifiant de la dimension temps
	idDimcommande	Identifiant de la dimension commande
Les mesures	prix	Prix des produits
	quantité	Nombre des produits

Tableau 7: Table des faits

2.4.4 Conception de l'entrepôt de données

La figure ci-dessous illustre la conception du datawarehouse, notre datawarehouse se compose de huit dimensions et d'une table des faits.

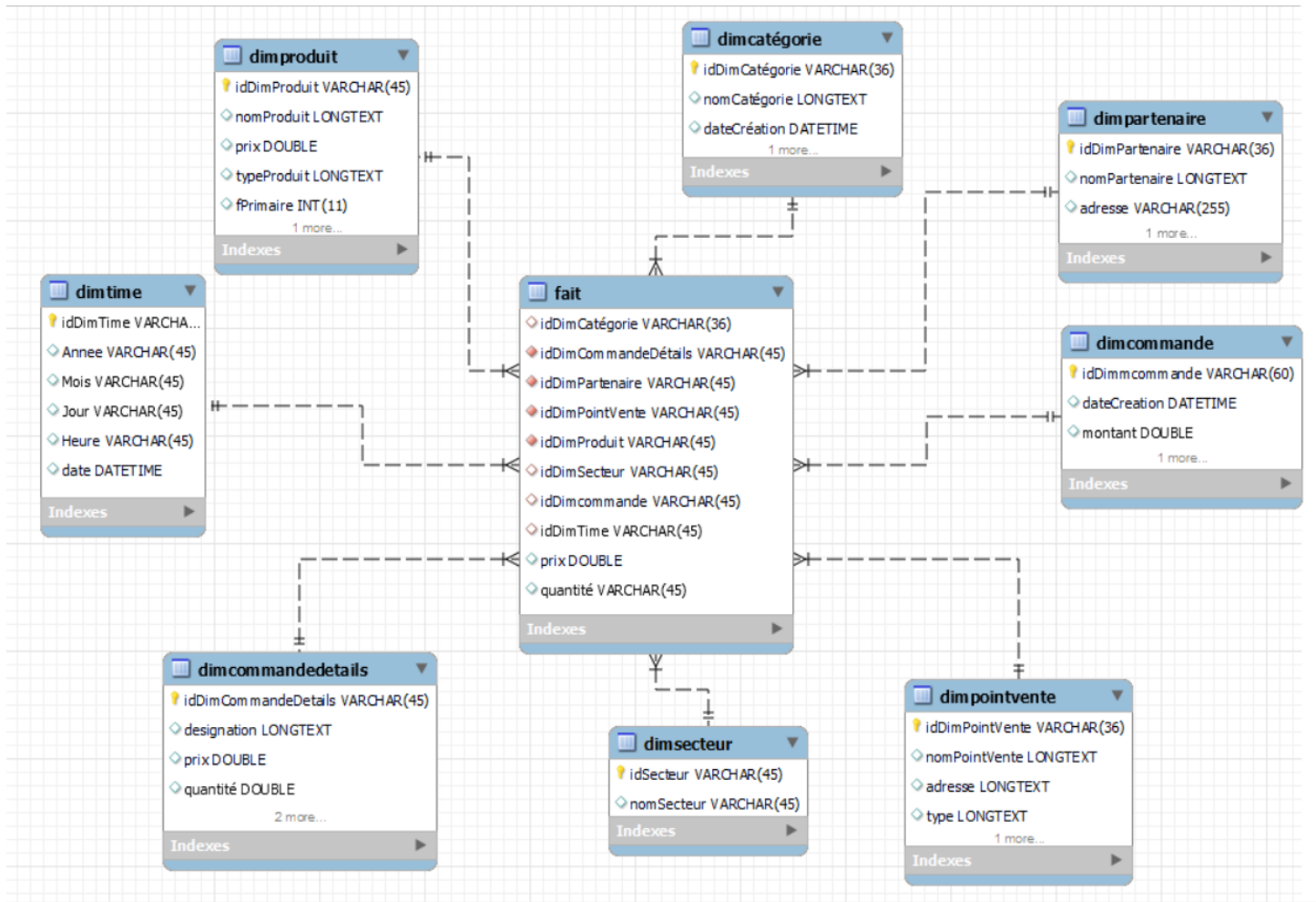


Figure 16: Conception de l'entrepôt de données

3 Conclusion

Au cours de ce chapitre, nous avons présenté en premier la base de données fournie par l'entreprise en précisant les tables et le contenu de chaque table, puis nous avons défini la modélisation multidimensionnelle et présenté le schéma en étoile de l'entrepôt de données.

Chapitre 4 : La mise en place d'un processus ETL

1 Introduction

Le processus ETL est une phase indispensable dans les projets de prise de décision, son objectif principal est de faciliter l'accès aux données et de produire des données standardisées.

Dans ce quatrième chapitre, nous nous discuterons de la mise en place du processus ETL et détaillerons le processus de transformation des données

2 Le deuxième sprint : Mise en place d'un processus ETL

2.1 Backlog du sprint

L'objectif principal de ce sprint est le déploiement de l'ETL afin d'obtenir une base de données propre et des données homogènes.

User Story

L'utilisateur veut que les différents formats des données soient standardisés afin qu'elles aient la même apparence.

2.2 Processus ETL

Puisque les données peuvent provenir de plusieurs sources, parfois hétérogènes, on doit passer par une phase qui est très importante pour rapprocher les données des sources hétérogènes et pour résoudre le problème de manque de cohérence des données [W26].

Au cours de cette phase, nous avons utilisé l'outil **Talend open studio for Data Integration**.

Talend Data Integration fournit les outils unifiés nécessaires pour intégrer, nettoyer, masquer et profiler toutes les données.

Il offre :

- Un développement et déploiement 10 fois plus rapide.
- Une confiance accrue dans les données.
- Un coût total de possession minimisé.

Les étapes du processus ETL :

1. L'extraction des Données

La première étape de l'ETL est l'extraction, elle permet de sélectionner les données à partir des systèmes de stockage des données afin de les récupérer.

Par conséquent, cette étape nous permet de lire sélectivement les données source, filtrant ainsi les données afin que seules les données pertinentes soient extraites pour construire notre magasin de données (Datamart).

2. Transformation des données :

La deuxième étape de l'ETL est la transformation, au cours de cette étape, les données doivent être filtrées, nettoyées, reformatées et surtout homogènes.

Les données doivent être synchronisées afin que les valeurs obtenues avant la phase de chargement soient cohérentes.

Cette étape est divisée en deux parties :

Standardisation des formats : Cela inclut la détermination des données à traiter, leur format et leur stockage, ainsi que d'autres considérations de base pour suivre ces étapes.

Nettoyage des données : Cette étape gère les données manquantes, élimine les doublons, valide les contraintes d'intégrité et gère les valeurs erronées ou incohérentes.

3. Chargement des données :

La phase de chargement des données est la dernière étape, elle permet d'insérer les données nettoyées, filtrées et préparées dans l'entrepôt de données en temps réel ou périodiquement qu'elles soient ensuite disponibles pour les différents outils d'analyse et de présentation.

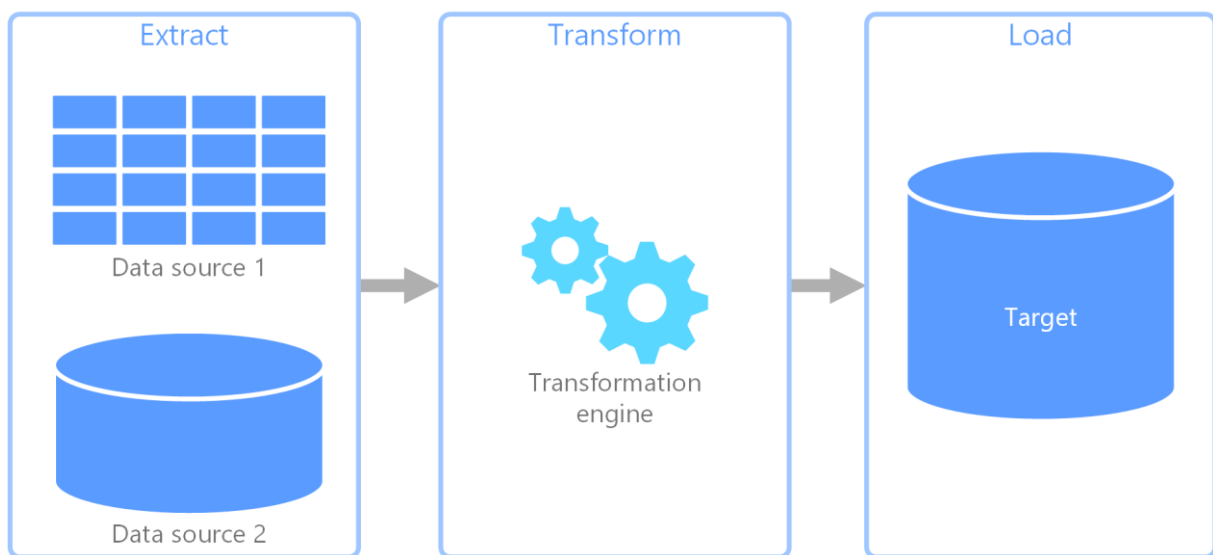


Figure 17: Processus ETL

Source : [W27]

2.3 Réalisation

Nous avons commencé le processus ETL, tout d'abord nous avons créé une base de données cible appelée "datawarehouse " dans MySQL Workbench et puis nous avons créé les tables cibles.

La figure ci-dessous indique la base de données destination créée dans MySQL Workbench :

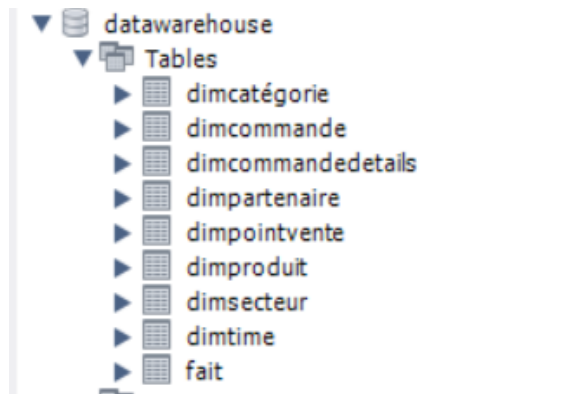


Figure 18: Les tables de l'entrepôt de données

2.3.1 Extraction des données

Tout d'abord, nous avons extrait la base de données source "Bprice" nommée "ad_bi" et la base de données destination nommée "datawarehouse". Ensuite, nous avons créé les Jobs dont nous parlerons dans la deuxième étape.

Cette figure illustre la connexion vers les bases de données :

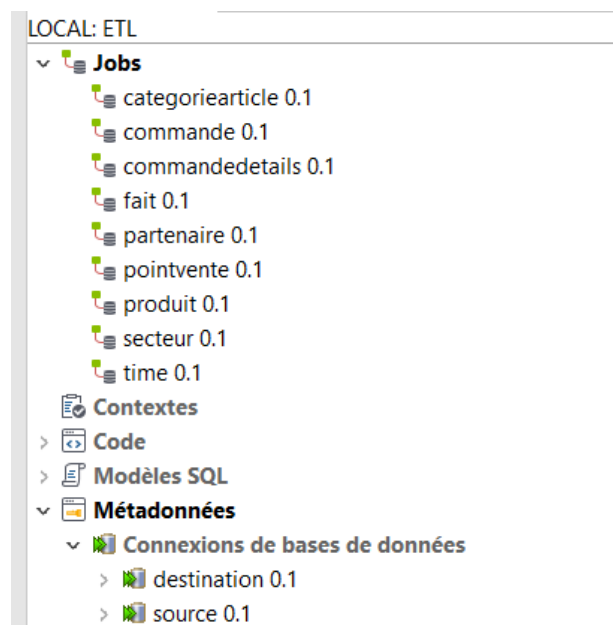


Figure 19: Connexions de bases de données

2.3.2 Transformation des données et chargement

2.3.2.1 Nettoyage des données avec un script

Le volume des données ne cesse de grandir. Pour bien exploiter ces connaissances et les rendre plus accessibles, il faut alors transformer ces données, les convertir d'un format à un autre.

Ce processus est exécuté manuellement en utilisant des requêtes différentes.

Les données sources dans la base de données Bprice sont brutes, elles sont donc inutilisables et on ne peut pas créer un Dashboard qui satisfait notre client.

Notre cas d'utilisation nous nécessite à changer le format des données :

Explication :

La société TWINS DIGITAL LABS collabore avec plusieurs partenaires qui travaillent en plusieurs secteurs, on peut prendre l'exemple le cas du secteur : La restauration.

On note par exemple les partenaires suivant : **BATBOUT, THE HOOD, LES MORISQUES**, chacun d'entre eux a plusieurs produits dans la carte.

L'inconvénient est que chaque partenaire nomme ces produits de sa manière, notre tâche est d'unifier noms des produits similaires en un seul nom.

- **La table produit :**

Pour transformer nos données, nous avons implémenté la fonction SOUNDEX

Définition : La fonction SOUNDEX dans MYSQL prend une chaîne de caractère source et il la convertit en son Xcode, c'est-à-dire la façon d'un string sonnera [W28].

Nous avons utilisé la requête suivante pour regrouper les produits qui partagent un son commun :


```

8
9 • SELECT SOUNDEX(code) as a ,count(soundex(code)) as b from bprice.produit
10 group by a
11 order by b desc;

```

a	b
W45424245414254	107
L645264	60
L645	55
I542414254	44
S134	38
D423	26
K5353246	22
L645264241624635265313	19
I500	19
R262	17
L640	15
W454261612	14
L64362	13
Z542	13
L64526424216	10
T500	10

Result 1 x

Output

Action Output

Figure 20: Fonction SOUNDEX appliquée sur la table produit (1)

Ensuite, nous avons utilisé la requête suivante pour connaître le nom du produit qui a le son dont nous l'avons trouvé dans le tableau :

```

7
8 • SELECT soundex(code),code,prixTtc,idPartenaire FROM bprice.produit
9 where soundex(code) = 'S4326';
10

```

soundex(code)	code	prixTtc	idPartenaire
S4326	SALADE CESAR	12	60d9d5c033c27601f47328d6
S4326	SALADE CESAR	8	60dc527e33c27601f46f2f07
S4326	SALADE CÉSAR	18	613881a433c27601f463535d
S4326	salade cesare	14.5	616896b633c27601f411d751

Figure 21: Fonction SOUNDEX appliquée sur la table produit (2)

Après nous avons utilisé la requête "update" pour regrouper tous les produits et standardiser les noms et les formats :

```
21
22 • update bprice.produit
23   set code ='SALADE CÉSAR '
24   where soundex(code) ='S4326';
25
26
27
```

Figure 22: Fonction UPDATE

On a eu le résultat suivant :

```
14 • select soundex(code),code,prixTtc,idPartenaire from bprice.produit
15   where soundex(code) = 'S4326' ;
16
17
```

soundex(code)	code	prixTtc	idPartenaire
S4326	SALADE CÉSAR	12	60d9d5c033c27601f47328d6
S4326	SALADE CÉSAR	8	60dc527e33c27601f46f2f07
S4326	SALADE CÉSAR	18	613881a433c27601f463535d
S4326	SALADE CÉSAR	14.5	616896b633c27601f411d751

Figure 23: Résultat des transformations

On a eu un problème dans la colonne prixTtc, certaines valeurs sont mises en MILLIMES, nous avons alors les converti tous en DINAR.

- **La table PointVente**

Dans cette table, la colonne "adresse" est l'une des colonnes les plus importantes pour la création du tableau de bord, il faut donc unifier le format des adresses et les rendre tous en minuscule.

```
12
13 • update bprice.pointVente
14   set adresse = 'enasser'
15   where adresse like '%nasser%';
16
```

Figure 24: Transformation des adresses

- **La table CategorieArticle**

Comme c'est indiqué dans la partie conception d'une base de données, chaque catégorie contient plusieurs produits.

Exemple : Une catégorie Pizza possède les produits suivants : Pizza Neptune, Pizza Salami, Pizza mini, Pizza de l'ambassadeur, etc...

C'est pour cela, on va faire le même travail que dans la table produit. Chaque partenaire nomme ces catégories de sa manière, notre tâche est de transformer les noms des catégories similaires en un seul.

- **La table PartenaireBprice**

Nous avons constaté que les données sont manquantes. Pour corriger ce problème, nous avons ajouté des valeurs dans les colonnes nulles.

Interprétation

Les données de notre entreprise sont brutes et inutilisables. Pour leurs exploiter, on a investie du temps et des méthodes techniques pour les transformer.

2.3.2.2 Transformation et chargement des données avec l'outil Talend

Dans cette section, nous avons créé des jobs de chargement pour alimenter les tables de dimensions et la table des faits où nous avons effectués des jointures et des transformations. Ensuite nous avons utilisé le mapping pour obtenir uniquement les colonnes dont nous avons besoin et nous avons adopté des outils pour assurer le contrôle de la qualité des données.

Tout d'abord, on a créé les Jobs suivants :

- **Job produit : table produit**

Dans ce Job, nous avons extrait la table "produit " de la base de données source qui contient des données nettoyées, suivie du Mapping afin d'alimenter la table "dimproduit ".

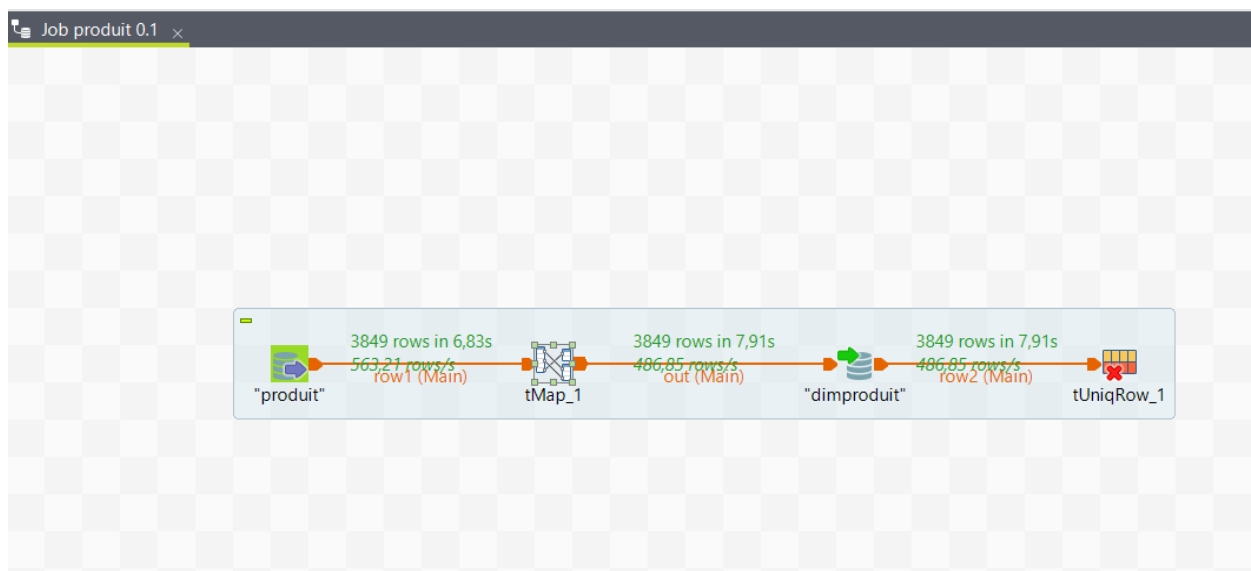


Figure 25: Job produit

- **Job pointVente : table pointVente :**

Dans le pointVente, nous avons extrait la table "pointVente" de la base de données source, on a effectué le Mapping afin d'alimenter la table "dimpointVente".

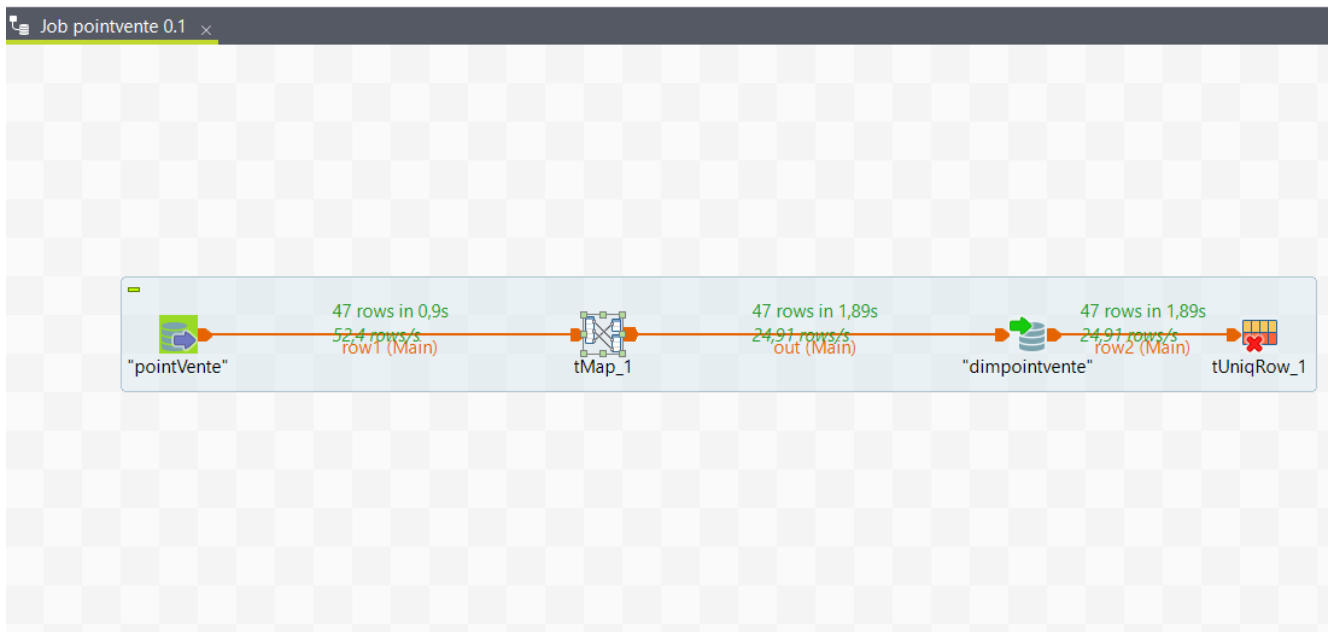


Figure 26: Job pointVente

- **Job partenaire : table partenaire**

Dans le Job suivant, nous avons extrait la table "partenaireBprice" de la base de données source. Comme les autres jobs, on a utilisé les deux outils TUniqRow pour la qualité de données et Tmap pour le mapping.

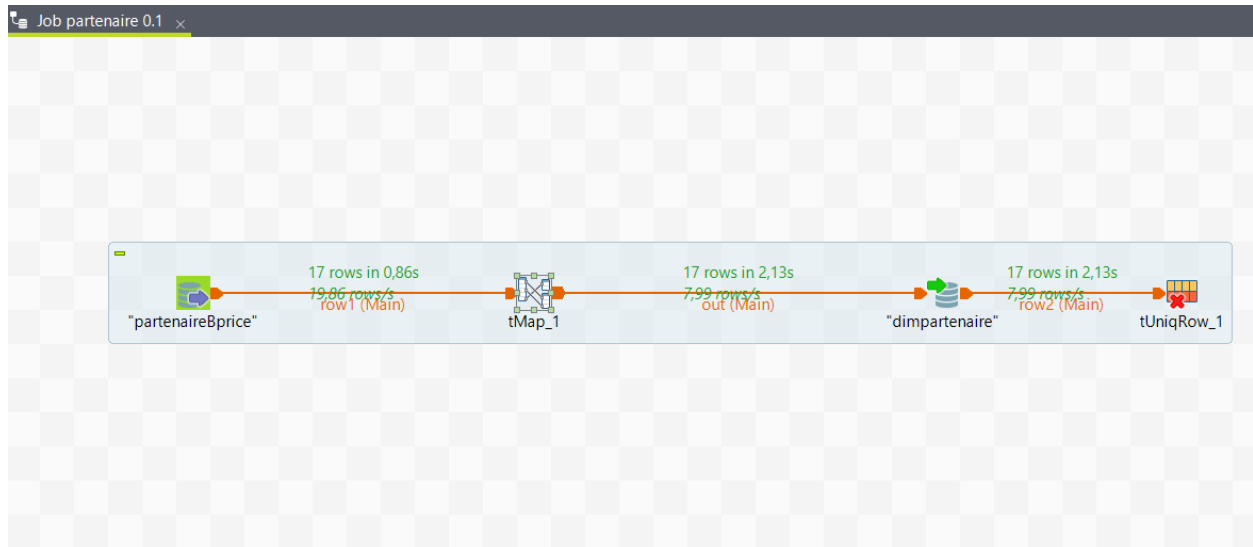


Figure 27: Job partenaire

- **Job commandetails : table commandeDetails**

Dans le Job commandetails, nous avons extrait la table "commandeDetails" de la base de données source qui passe par l'outil Tmap pour alimenter la table "dimcommandeDetails".

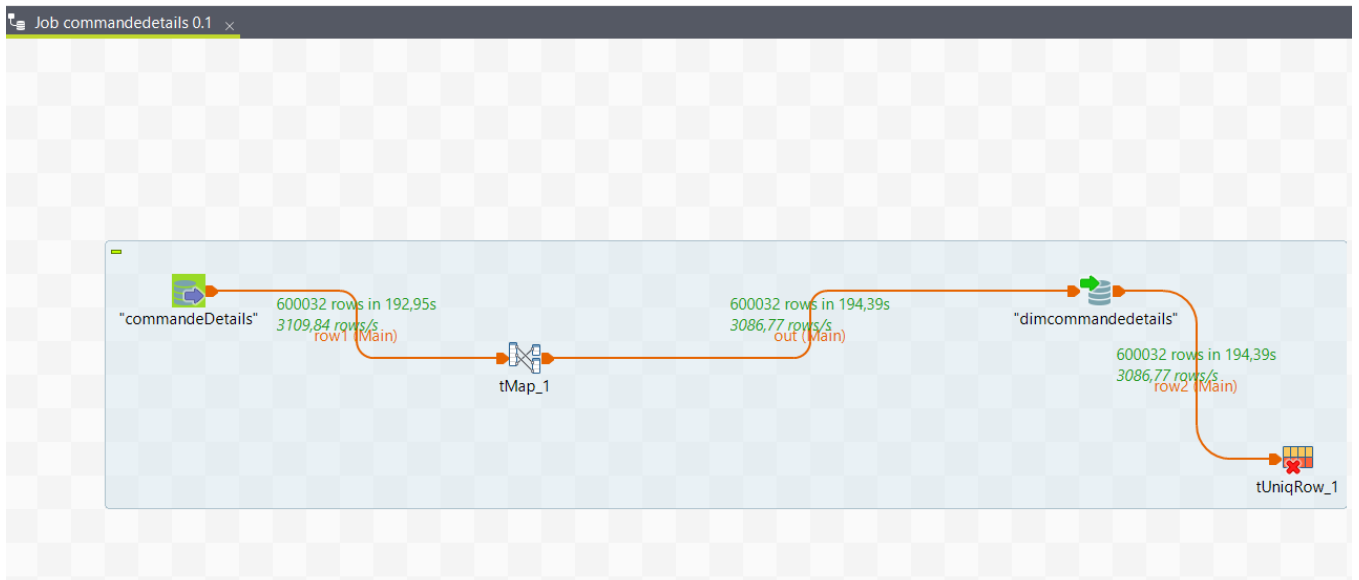


Figure 28: Job commandeDetails

- **Job categorieArticle : table categorie**

Dans le Job suivant, nous avons extrait la table "categorieArticle" de la base de données source, puis nous avons utilisé le Mapping pour alimenter la table "dimCategorieArticle".

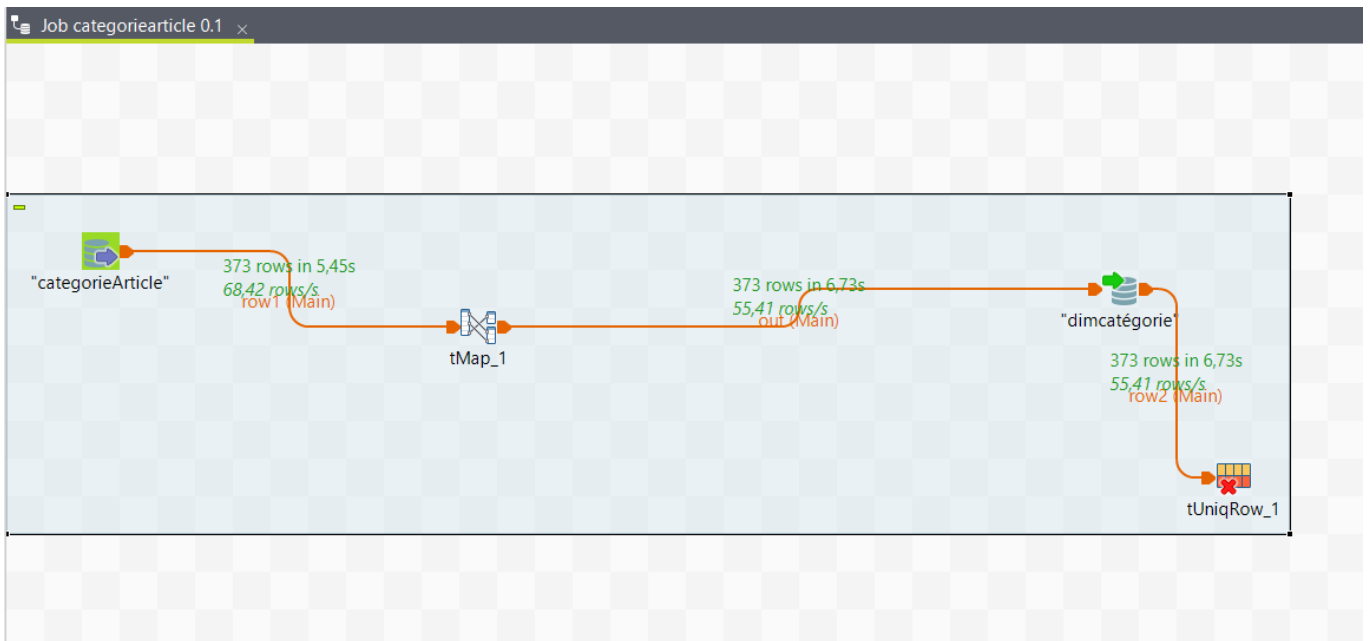


Figure 29: Job categorieArticle

- **Job secteur : table sector**

Dans le Job secteur, nous avons récupéré la table "sector" de la base de données source, suivie du Mapping afin d'alimenter la table "dimSecteur".

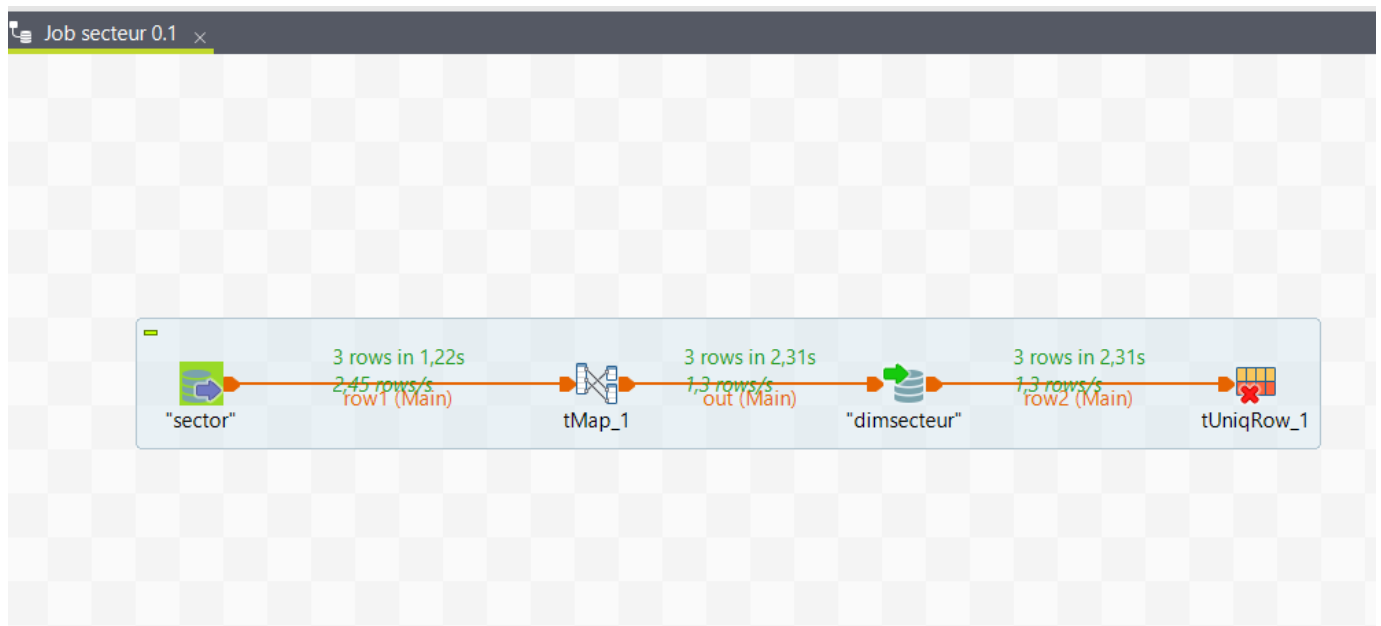


Figure 30: Job Secteur

- **Job time**

Dans le Job suivant, nous avons extrait la colonne dateCreation de la table "commande" de la base de données source et converti cette colonne en quatre colonnes : année, mois, jour et heure en utilisant l'outil tMap et la fonction suivante :

`TalendDate.formatDate("yyyy",row1.dateCreation) = Année`

`TalendDate.formatDate("mm",row1.dateCreation) = Mois`

`TalendDate.formatDate("dd",row1.dateCreation) = Jour`

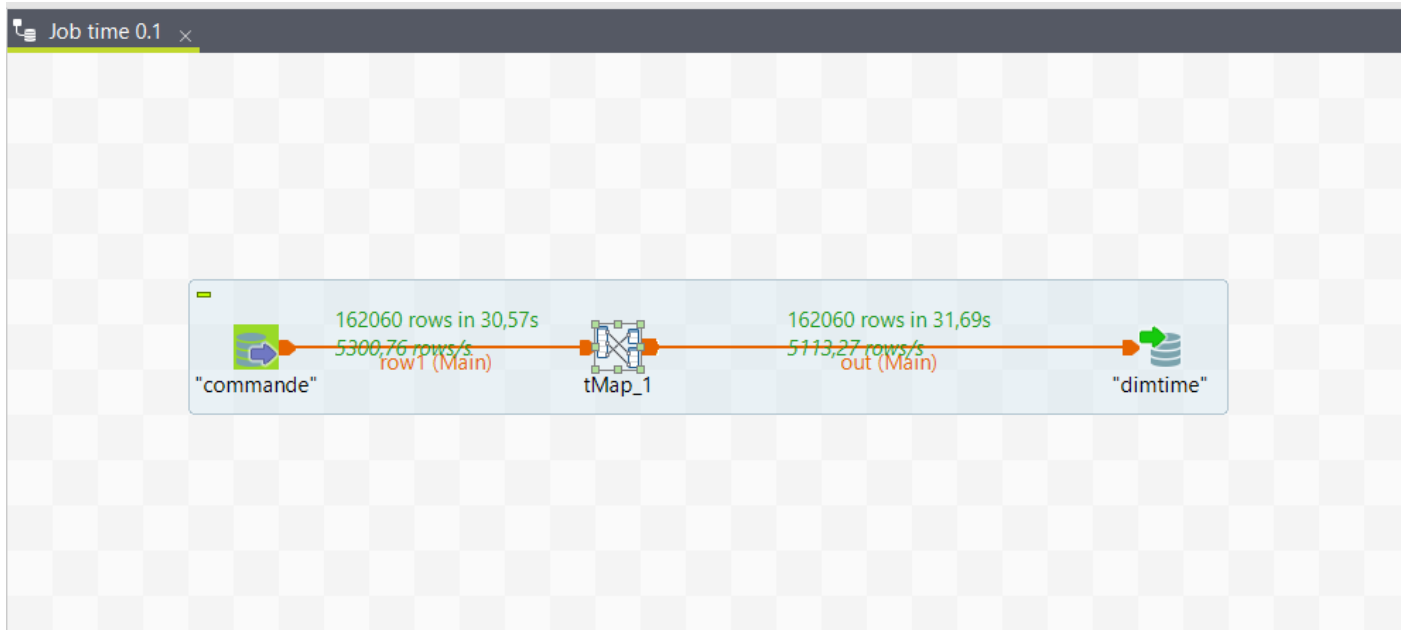


Figure 31: Job time

- **Job commande : table commande**

Dans le Job suivant, nous avons extrait la table commande de la base de données source, puis nous avons effectué les transformations nécessaires pour alimenter la table "dimcommande" du datawarehouse.

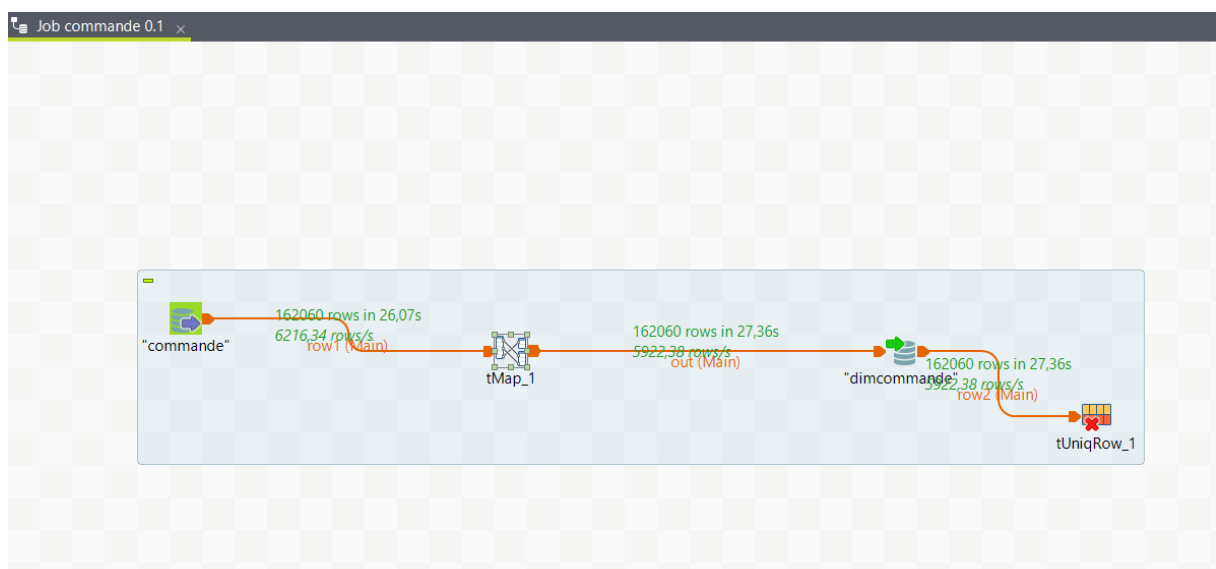


Figure 32: job commande

- **Job Vente : table des faits Vente**

Dans ce job, on a alimenté la table fait a partir de tous les dimensions, nous avons choisi la table dimCommandeDetails comme row Main car elle contient les mesures prix et quantité :

Nous avons alors 600032 lignes dans cette table

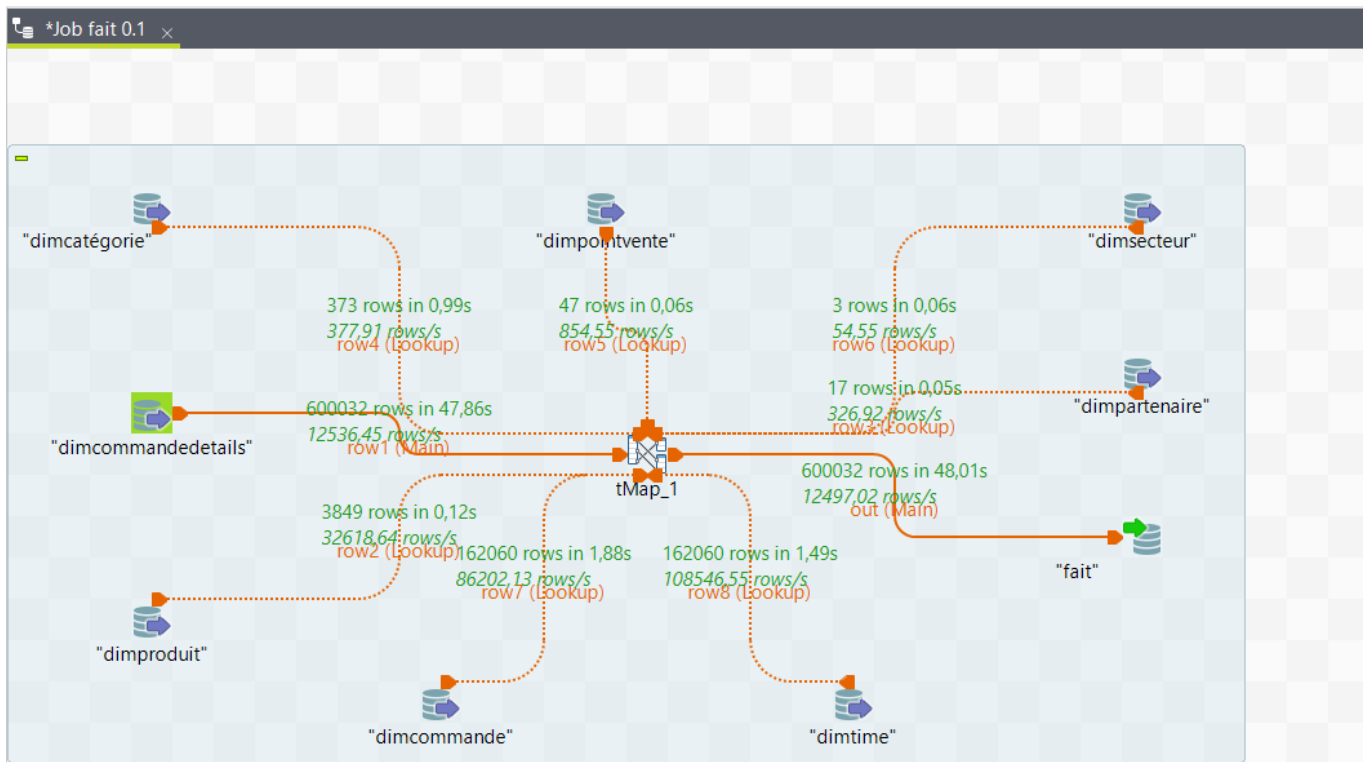


Figure 33: Job fait

Dans les jobs ci-dessus, on a utilisé les deux outils le tMap et le tUniqRow dont on va parler de leurs utilité dans la section suivante:

- **Le tMap** : cet outil permet d'effectuer des jointures, des filtres de colonnes ou de lignes, des transformations et de multiples sortis.
- **Le tUniqRow** : cet outil permet de crée un flux de données unique à partir de la clé définie dans le schéma, et il permet d'assurer une bonne qualité de données.

La figure suivante présente la configuration dans le composant tMap :

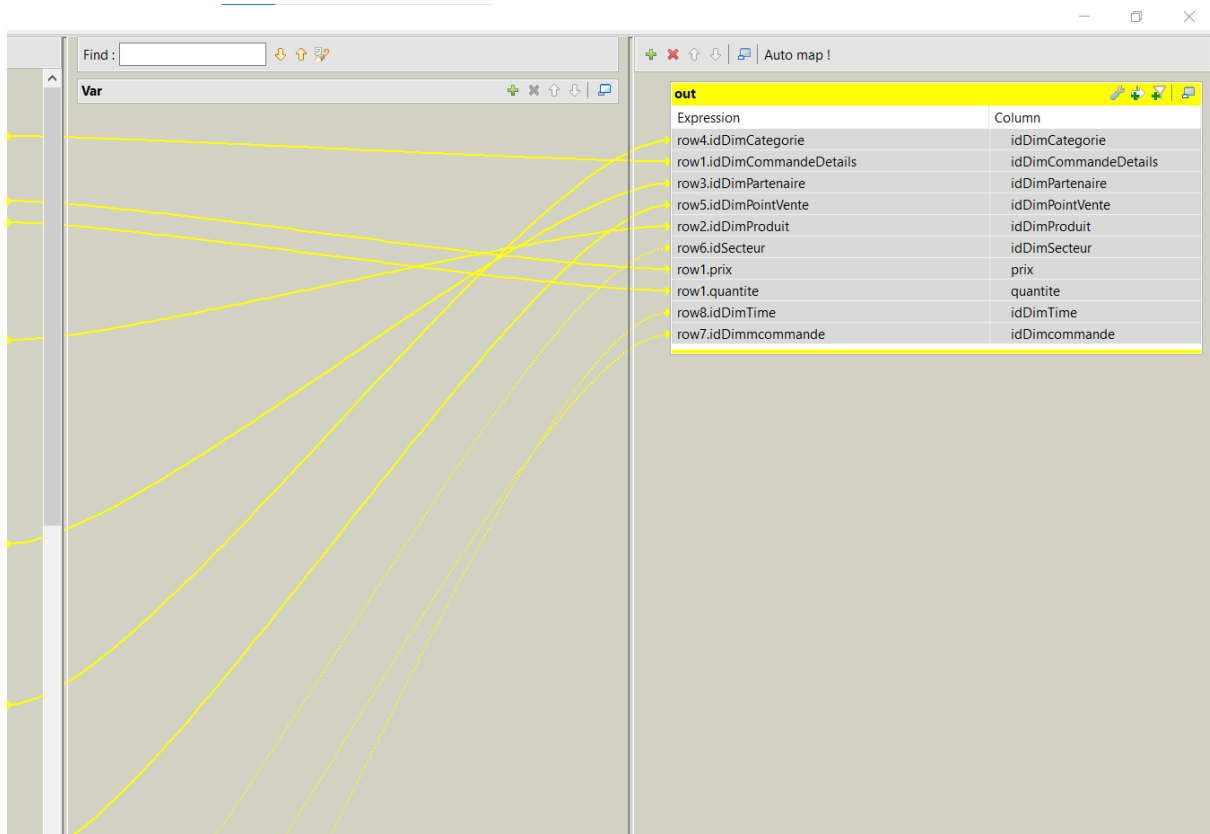


Figure 34: L'outil tMap

3 Conclusion

Durant ce chapitre, nous avons présenté l'étape la plus importante dans un projet de prise de décision.

En premier lieu, nous avons montré le nettoyage des données à travers un script dans l'outil MySQL Workbench, ensuite, nous avons montré les transformations aboutit dans l'outil Talend Open Studio, et enfin l'alimentation des tables du datawarehouse.

Chapitre 5 : Réalisation du tableau de bord

1 Introduction

Dans ce dernier chapitre, nous allons présenter la démarche à suivre afin de réaliser un tableau de bord qui regroupe les indicateurs de performance nécessaire afin d’avoir une vision claire des données et de faciliter les tâches aux décideurs de la société.

2 Backlog du sprint

Ce Sprint a pour objectif d’effectuer des analyses et des reporting et de livrer un tableau de bord qui identifie les indicateurs de performances afin d’exporter les données.

3 Définition des indicateurs

Dans cette partie nous allons identifier les différents indicateurs réalisés dans ce projet.

3.1 Définition d’un indicateur clé de performance

Un indicateur clé de performance ou KPI en anglais est une mesure quantitative ou un ensemble de mesures qui permet à une société de suivre son progrès afin d’obtenir une vision globale de ses objectifs stratégiques.

Les KPIs servent à établir clairement les objectifs que l'équipe veut atteindre et l'échéance visée, ainsi qu'à préciser la façon dont le Scrum master mesure leur accomplissement.

Les indicateurs de performance ne laisseront jamais les décideurs indifférents. Quand les décideurs n'agissent pas, c'est en pleine conscience. [W29].

Un indicateur clé de performance donne l'occasion de répondre à ces objectifs :

- La communication
- Le diagnostic
- L'évaluation
- Le progrès et la motivation

3.2 Les indicateurs choisis

Tout d'abord, notre Scrum master nous a proposé de deviser le travail en plusieurs modules dont chacun contient des indicateurs :

Module Général

- Nombre total des produits.
- Nombre total des catégories articles.
- Nombre total des points de vente.
- Nombre total des partenaires.
- Classement des produits les plus vendus.
- Classement des partenaires selon leurs chiffres d'affaire (CA).
- Classement des produits les plus chers.
- Nombre des produits stockables ou services.
- Les produits primaires / non primaires.

Module produit

- Comparaison du prix des produits par une zone géographique.
- Consommation d'un produit sélectionné.

- Chiffre d'affaire d'un produit par une zone géographique.
- Classement des produits primaires les plus consommés.
- Consommation d'un produit par mois.

Module Partenaire

- Classement des produits les plus vendus pour un partenaire donnée.
- Les produits avec le chiffre d'affaires le plus élevé d'un partenaire donnée.
- Nombre total des commandes annulés.
- Nombre des points de vente pour chaque partenaire.

Module Secteur

- Les produits le plus vendus du secteur d'activité sélectionné.
- Produits les plus chers pour un secteur sélectionné

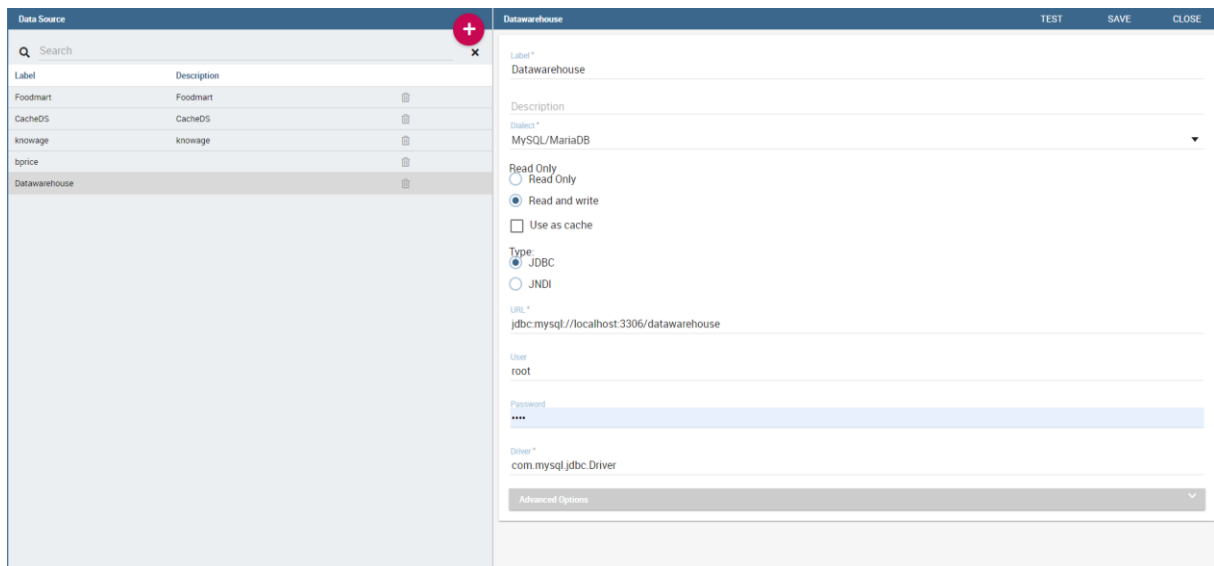
4 Configuration de l'outil de création des tableaux de bord

Tout d'abord, et comme nous l'avons indiqué dans le chapitre 2, l'outil de reporting choisi est Knowage.

4.1 Connexion

La première étape à faire est d'établir une connexion vers le data warehouse. Il s'agit de remplir les champs vides et de préciser l'URL.

La figure ci-dessous représente l'interface de connexion vers le data warehouse nommé « Data Source »:



Label	Description
Foodmart	Foodmart
CacheDS	CacheDS
knowage	knowage
bprice	bprice
Datawarehouse	

Label *
Datawarehouse

Description

Driver *
MySQL/MariaDB

Read Only
☐ Read Only
☒ Read and write

☐ Use as cache

Type
☒ JDBC
☐ JNDI

URL *
jdbc:mysql://localhost:3306/datawarehouse

User
root

Password

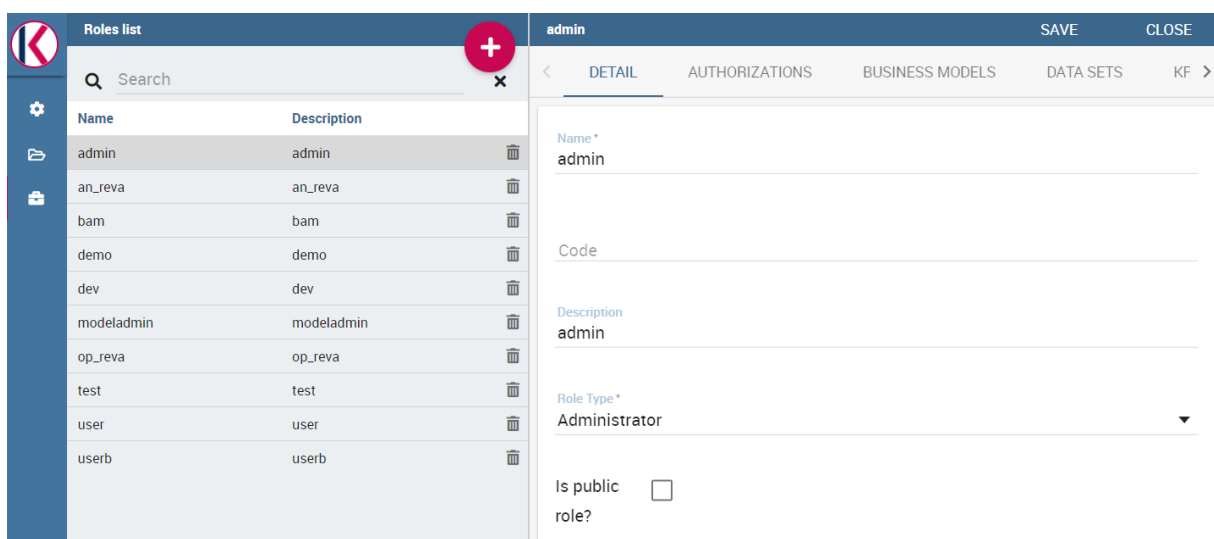
Driver *
com.mysql.jdbc.Driver

Advanced Options

Figure 35: Interface de connexion vers le datawarehouse

4.2 Gestion des utilisateurs et des rôles

La deuxième étape consiste à gérer les utilisateurs, la figure suivante représente l'interface de la gestion des autorisations. Tout d'abord, nous avons précisé les caractéristiques du rôle « admin » comme l'indique la figure ci-dessous



Name	Description
admin	admin
an_reva	an_reva
bam	bam
demo	demo
dev	dev
modeladmin	modeladmin
op_reva	op_reva
test	test
user	user
userb	userb

Name *
admin

Code

Description
admin

Role Type *
Administrator

Is public role? ☐

Figure 36: Gestion des rôles

La figure ci-dessous représente l'interface de la gestion des utilisateurs et plus précisément l'utilisateur "biadmin" auquel on attribue un mot de passe.

The screenshot shows a web application interface for user management. On the left, a sidebar contains a 'Users list' section with a search bar and a table of users. The table has columns 'User ID' and 'Full Name'. The 'biadmin' user is highlighted. On the right, a form titled 'biadmin' is open, showing the 'DETAIL' tab. The form contains fields for 'User ID *' (filled with 'biadmin'), 'Full Name *' (filled with 'Knowage Administrator'), 'Password' (masked with dots), and 'Confirm Password' (masked with dots). Buttons for 'SAVE' and 'CLOSE' are visible at the top right of the form.

User ID	Full Name
biadmin	Knowage Administrator
bidemo	Knowage Demo User
bidev	Knowage Developer
bitest	Knowage Test User
biuser	Knowage User

Figure 37: Gestion des utilisateurs

Après avoir saisi les informations de l'utilisateur, on doit attribuer à l'utilisateur "biadmin" un rôle bien précis, comme le montre la figure ci-dessous :

The screenshot shows the same web application interface, but now the 'ROLES' tab is selected for the 'biadmin' user. The form displays a table of roles with columns 'Name' and 'Value'. The 'admin' role is selected with a checkbox. Other roles listed include 'an_reva', 'bam', 'demo', 'dev', 'modeladmin', 'op_reva', 'test', 'user', and 'userb'. Buttons for 'SAVE' and 'CLOSE' are visible at the top right of the form.

Name	Value
<input checked="" type="checkbox"/> admin	admin
<input type="checkbox"/> an_reva	an_reva
<input type="checkbox"/> bam	bam
<input type="checkbox"/> demo	demo
<input type="checkbox"/> dev	dev
<input type="checkbox"/> modeladmin	modeladmin
<input type="checkbox"/> op_reva	op_reva
<input type="checkbox"/> test	test
<input type="checkbox"/> user	user
<input type="checkbox"/> userb	userb

Figure 38: Attribution d'un rôle à l'utilisateur

5 Création des indicateurs par l'outil du reporting

5.1 Création des Data Sets

Notre outil de visualisation Knowage exige de passer par plusieurs étapes afin de créer un tableau de bord.

Tout d'abord, il faut écrire des requêtes SQL pour chaque indicateur, prenons l'exemple ci-dessous : pour l'indicateur « Consommation du produit par mois », nous avons écrit la requête suivante :

DETAIL TYPE ADVANCED

DataSet Type *

Query

Data Source *

Datawarehouse

Query

```
1 select nomProduit,c.prix,nomPartenaire,round(sum(quantité),0),d.adresse,month(b.dateCreation) as mois,round(sum(quantité)*c.prix,2) as CA,
2 CASE
3   WHEN month(b.dateCreation)=1 THEN 'Janvier'
4   WHEN month(b.dateCreation)=2 THEN 'Février'
5   WHEN month(b.dateCreation)=3 THEN 'Mars'
6   WHEN month(b.dateCreation)=4 THEN 'Avril'
7   WHEN month(b.dateCreation)=5 THEN 'Mai'
8   WHEN month(b.dateCreation)=6 THEN 'Juin'
9   WHEN month(b.dateCreation)=7 THEN 'Juillet'
10  WHEN month(b.dateCreation)=8 THEN 'Août'
11  WHEN month(b.dateCreation)=9 THEN 'Septembre'
12  WHEN month(b.dateCreation)=10 THEN 'Octobre'
13  WHEN month(b.dateCreation)=11 THEN 'Novembre'
14  ELSE 'Décembre'
15 END
16 as nomMois
17 from datawarehouse.dimcommandeDetails a
18 join datawarehouse.dimcommande b
19 on b.idDimcommande = a.idCommande
20 join datawarehouse.dimproduit c
21 on a.idProduit = c.idDimProduit
22 join datawarehouse.dimpointvente d
23 on b.idPointVente = d.idDimPointVente
24 join datawarehouse.dimpartenaire e
25 on e.idDimPartenaire = d.idPartenaire
26 group by d.idDimPointVente,mois,nomProduit;
```

EDIT SCRIPT

Figure 39: Création des datasets

Ensuite, nous allons créer une charte pour cet indicateur dans la section myWorkspace de Knowage, on doit donc choisir le chart qui le convient

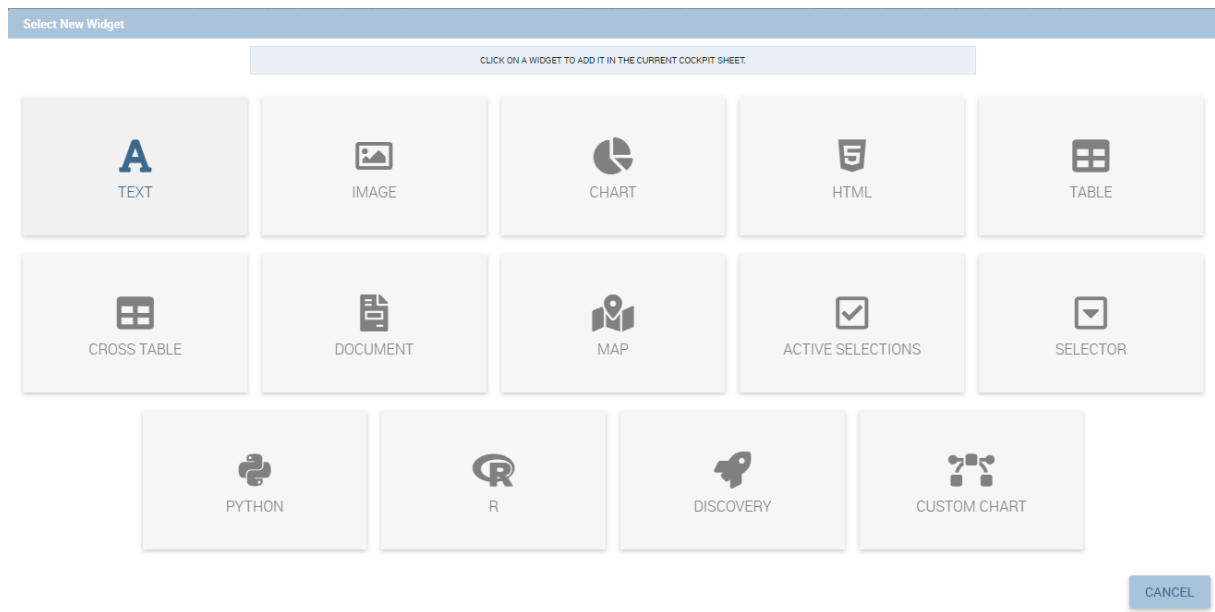


Figure 40: les différents widgets

5.2 Création des chartes

5.2.1 Tableau de bord du module Général

Les utilisateurs peuvent enfin voir les tableaux de bord, la figure suivante représente la partie « Générale ». Ce module apporte des informations générales sur l'application AD CAISSE

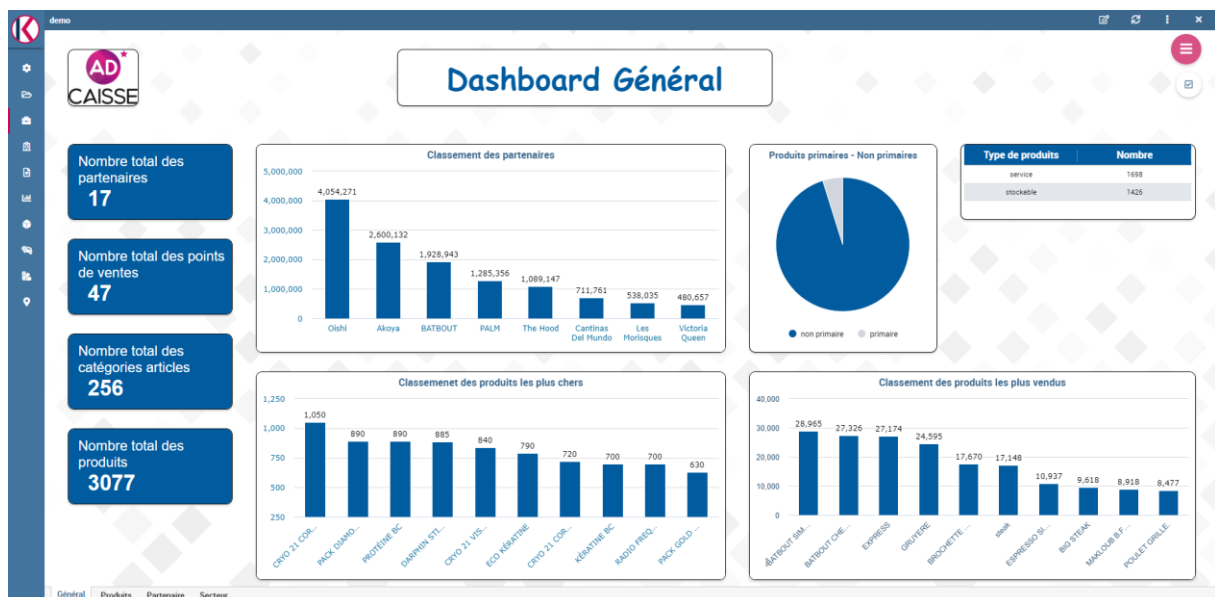


Figure 41: Le Module Générale

La figure suivante nous montre le classement des partenaires selon leurs chiffres d'affaires, l'utilisateur peut alors constater que les repas asiatiques du partenaire « OISHI » et les produits du secteur esthétique comme « AKOYA » sont chers

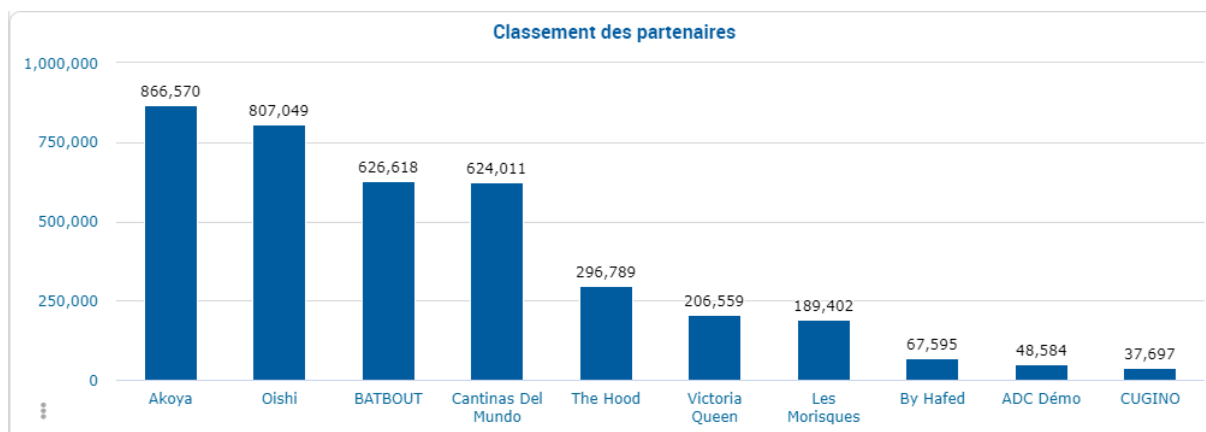


Figure 42: Classement des partenaires

Le rapport suivant est utile pour comprendre les produits les plus chers pour tous les partenaires. Nous avons choisi de les illustrer à travers des bar charts :

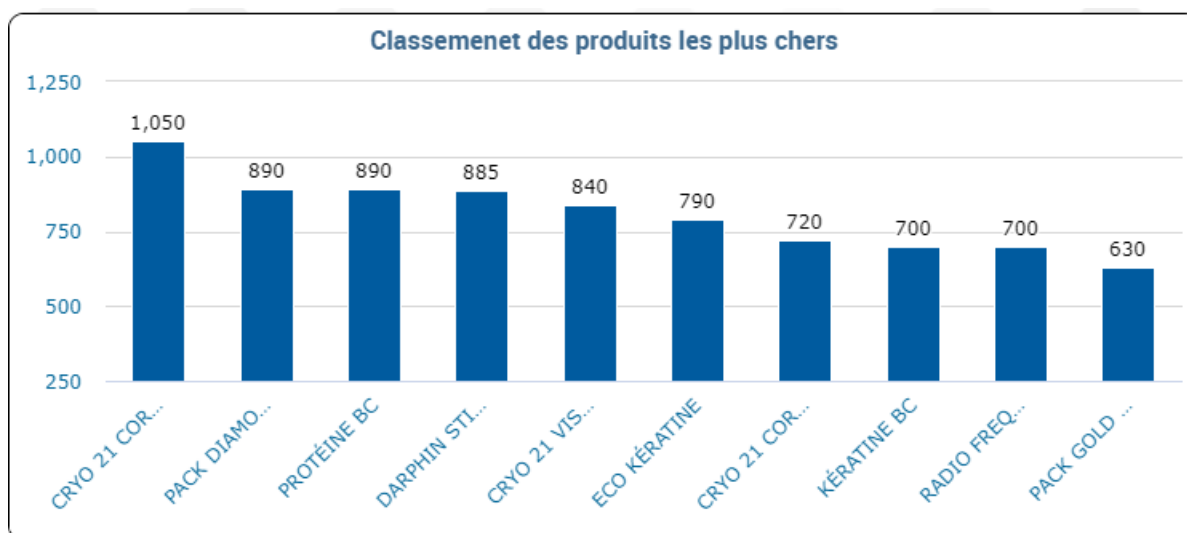


Figure 43: Rapport des produits les plus vendus

Le rapport ci-dessous représente des informations générales sur l'application AD-Caisse tel que le nombre total des produits, des partenaires, des points de ventes et des catégories des produits :



Figure 44: Informations d'ADCAISSE

5.2.2 Tableau de bord du module Produit

La figure suivante représente le module « Produit » dans lequel nous avons mis les rapports qui sont en relation avec les produits :

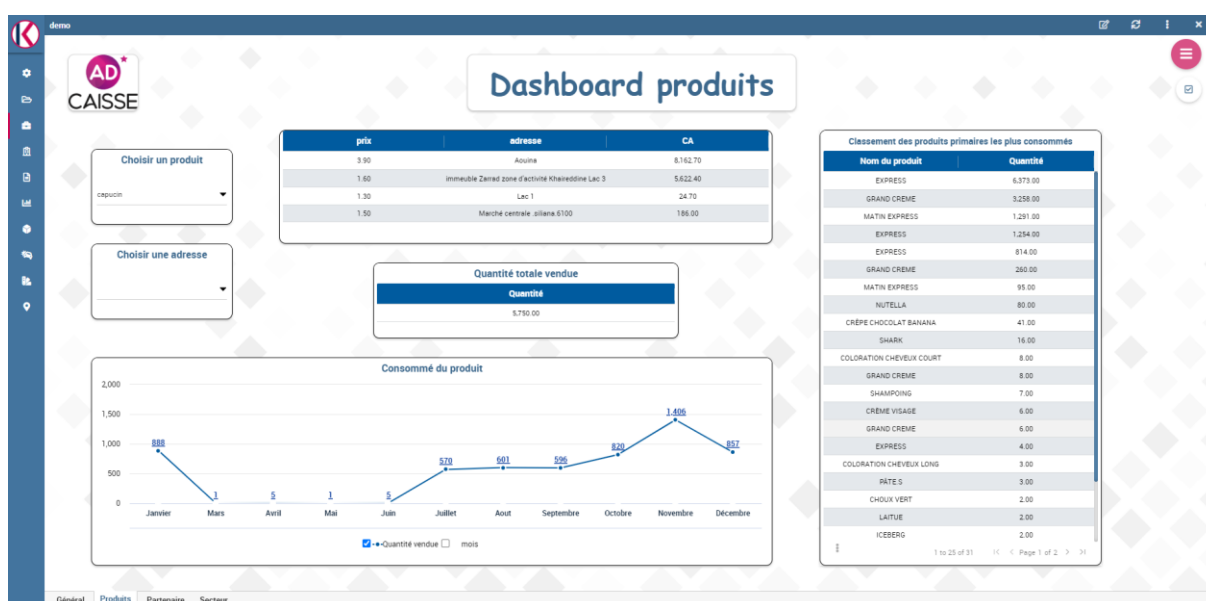


Figure 45: Module produit

Le rapport suivant nous montre la différence des prix et le chiffre d'affaire pour un produit sélectionné. En utilisant ce graphe, l'utilisateur distinguer le prix d'un produit pour chaque adresse.

prix	adresse	CA
3.90	Aouina	8,162.70
1.60	immeuble Zarrad zone d'activité Khaireddine Lac 3	5,622.40
1.30	Lac 1	24.70
1.50	Marché centrale .siliana.6100	186.00

Figure 46: comparaison des prix par une zone géographique

Dans cette partie, nous avons créé un filtre afin de sélectionner un produit et pour monter l'interactivité entre les graphiques dans le module Produit

Figure 47: Interface de création de sélection

Dans la figure suivante nous montre l'évolution de la consommation d'un produit sélectionné pour chaque mois, on a pris l'exemple du produit « eau 0.5L » dans ce cas :

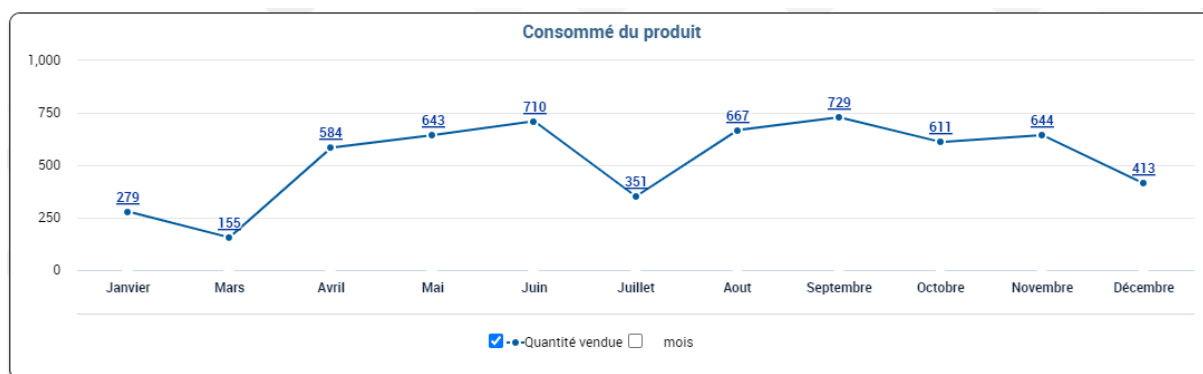


Figure 48: Consommé du produit

5.2.3 Tableau de bord du module Partenaire

La figure suivante représente le module « Partenaire » contenant toutes les informations par rapport aux partenaires.

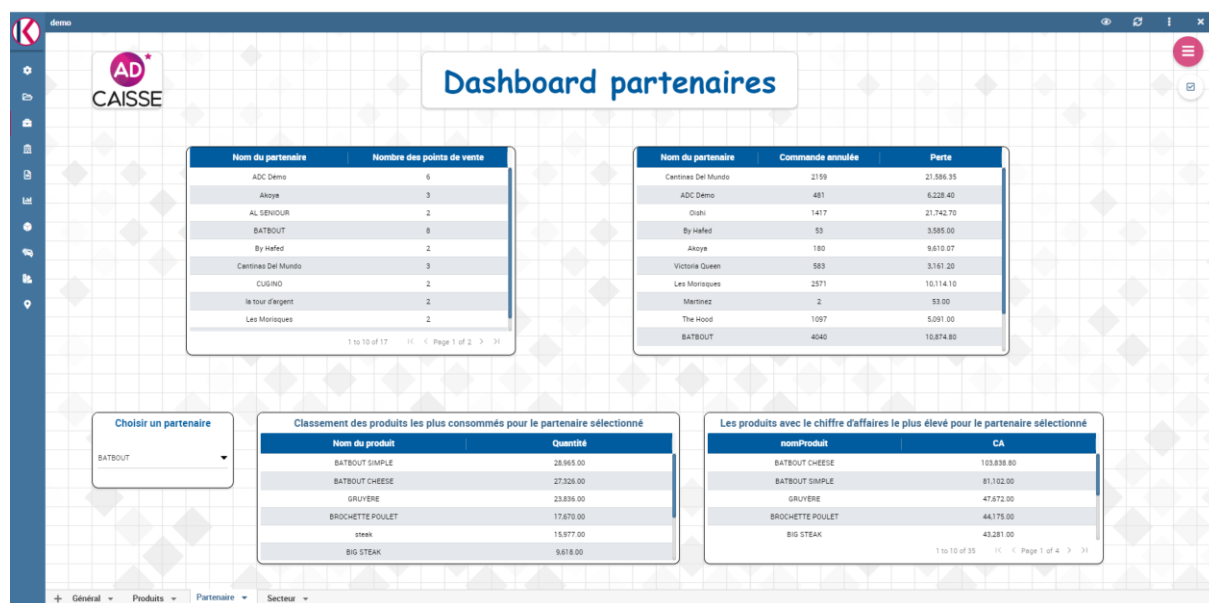


Figure 49: Rapport du module partenaire

Nous avons mis sous forme d'un tableau les produits les plus vendus et leurs chiffres d'affaires pour un partenaire donné. On peut alors voir que le produit fromage est trop cher par rapport aux autres produits.

Prenons l'exemple de la quantité vendue de « Batbout Cheese » dans le premier tableau qui est inférieure à la quantité vendue de « Batbout Simple » alors que dans le deuxième tableau le chiffre d'affaire de « Batbout Cheese » est très élevé par rapport à « Batbout simple » alors on peut conclure que le prix du supplément est plus élevé.

Nom du produit	Quantité
BATBOUT SIMPLE	28.965.00
BATBOUT CHEESE	27.326.00
GRUYERE	23.836.00
BROCHETTE POULET	17.670.00
steak	15.977.00
BIG STEAK	9.618.00

nomProduit	CA
BATBOUT CHEESE	103.838.80
BATBOUT SIMPLE	81.102.00
GRUYERE	47.672.00
BROCHETTE POULET	44.175.00
BIG STEAK	43.281.00

Figure 50: Classement des produits pour un partenaire

5.2.4 Tableau de bord du module Secteur

L'application AD-Caisse travaille sur les deux secteurs suivants : Esthétique et Restauration. La figure suivante représente le module « Secteur »



Figure 51: rapport du module secteur

La figure ci-dessus montre qu'après la sélection d'un secteur d'activité, on obtient une classification des produits les plus chers et des produits les plus vendus selon le secteur donné.

L'exemple ci-dessous représente les produits les plus vendus pour le secteur « Restauration » :

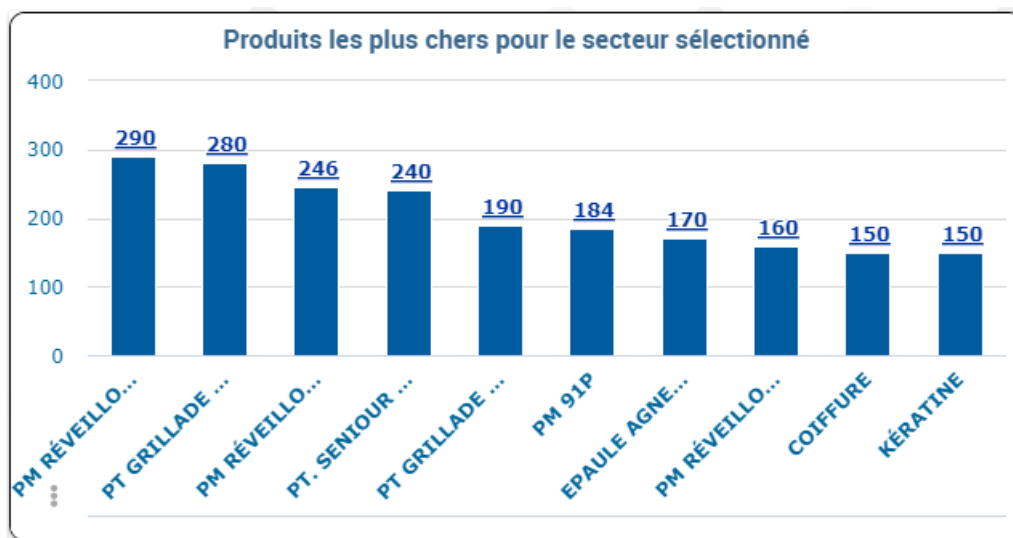


Figure 52: Classement des produits les plus chers pour un secteur

6 Conclusion

Durant ce chapitre, nous avons présenté le dernier sprint où on a défini tous les indicateurs et on a affiché les résultats du tableau de bord en passant par plusieurs étapes, nous avons établi une connexion entre MySQL et l'outil de visualisation KNOWAGE, suivie de création des datasets afin de créer des Dashboard interactives.

Conclusion et perspectives

Toutes les entreprises doivent tirer les meilleures informations de leurs flux de données afin de faciliter l'accès aux données et les analyser. Dans ce contexte, les responsables de TWINS DIGITAL LABS ont défini comme objectif principal « Mettre en place une solution décisionnelle au sein de la société pour les différentes thématiques ». La solution proposée s'est élaborée en suivant la méthode agile Scrum.

Durant ce travail, nous avons passé par des plusieurs étapes. Après avoir étudié l'existant, défini les problèmes et mettre un objectif, nous avons défini l'outil de gestion de base de données MySQL, un outil permettant d'assurer un mécanisme d'ETL par rapport à l'ensemble des données TALEND et un outil de visualisation KNOWAGE. Nous avons défini les indicateurs de performances nécessaires pour fournir des résultats à fin d'être visualisés par les personnes qui souhaitent avoir du consulting par rapport à la phase de pré lancement d'un projet basé sur une activité commerciale. Le concept clé pour la construction d'un système BI efficace c'est la conception d'un entrepôt de données. Pour cela nous avons commencé par la modélisation d'un data warehouse qui présente tous les dimensions et la table de fait, ensuite nous avons passé par le processus de l'intégration de données après avoir les transformer, suivie de créer des rapports dynamiques et des tableaux de bords.

Quant aux perspectives de ce projet, nous proposons parvenir à génération d'un SDK qui sera intégré dans le back office central, ainsi le processus de data mining qui va nous aider à prédire et fournir une aide décisionnelle par rapport à une étude ou bien une idée de projet (Quelles sont les produits les plus vendus, avec combien d'argent on peut ouvrir un restaurant dans certaines adresses ...).

Ce projet est le fruit d'une agréable opportunité d'intégration du monde professionnelle qui nous a permis de mettre en œuvre nos connaissances académiques et les enrichir.

WEBOGRAPHIE ET BIBLIOGRAPHIE

- [W1] <http://twinsdigitallabs.tech/agency>
- [W2] <https://www2.stardust-testing.com/blog-fr/cascade-scrum-testing-qa>
- [W3] <https://www2.stardust-testing.com/blog-fr/cascade-scrum-testing-qa>
- [W4] <https://www.piloter.org/projet/methode/methode-agile.htm>
- [W5] <https://www.softfluent.fr/blog/12-principes-du-manifeste-agile/>
- [W6] <https://www.piloter.org/projet/methode/scrum.htm#:~:text=D%C3%A9finition%20de%20Scrum,au%20fil%20de%20son%20avancement>
- [W7] <https://www.planzone.fr/blog/quest-ce-que-la-methodologie-scrum>
- [W8] [https://fr.wikipedia.org/wiki/Scrum_\(d%C3%A9veloppement\)](https://fr.wikipedia.org/wiki/Scrum_(d%C3%A9veloppement))
- [W9] <https://www.piloter.org/projet/methode/scrum.htm#:~:text=D%C3%A9finition%20de%20Scrum,au%20fil%20de%20son%20avancement>
- [W10] <https://www.toolsqa.com/agile/scrum/sprint/>
- [W11] <https://www.tuleap.org/fr/agile/comprendre-methode-agile-scrum-10-minutes>
- [W12] <https://www.edureka.co/blog/agile-scrum-tutorial/>
- [W13] <https://docs.microsoft.com/>
- [W14] <https://www.data-bird.co/sql/mysql>
- [W15] <https://www.oracle.com/fr/database/definition-postgresql.html>
- [W16] <https://www.techno-science.net/definition/7708.html>
- [W17] <https://www.mysql.com/fr/>

[W18] [https://www.next-decision.fr/editeurs-bi/etl/talend-open-studio#:~:text=Talend%20Open%20Studio%20\(TOS\)%20est,'un%20syst%C3%A8me%20d'information.](https://www.next-decision.fr/editeurs-bi/etl/talend-open-studio#:~:text=Talend%20Open%20Studio%20(TOS)%20est,'un%20syst%C3%A8me%20d'information.)

[W19] <https://www.next-decision.fr/editeurs-bi/etl/pentaho-pdi>

[W20] <https://fr.wikipedia.org/wiki/Fichier:Logo-talend.jpg>

[W21] <https://www.appvizer.fr/analytique/analyse-de-donnees/power-bi>

[W22] <https://www.next-decision.fr/editeurs-bi/restitution/knowledge>

[W23] <https://www.synaltic.fr/webinars/knowledge-les-donnees-ouvertes-sante-publique>

[W24] <https://www.knowledge-suite.com/site/>

[W25] <https://www.oracle.com/fr/database/data-warehouse-definition.html>

[B1] [Inmon, 1996] INMON, Bill. « Building the Data Warehouse », QED Technical Publishing Group, Wellesley, Massachusetts, U.S.A., 1996

[W26] http://formations.imtatlantique.fr/bi/bi_atelier_easymorph_partie1.html

[W27] <https://www.astera.com/fr/type/Blog/donn%C3%A9es-ETL/>

[W28] <https://fr.acervolima.com/fonction-soundex-dans-mysql/>

[W29] <https://www.piloter.org/mesurer/tableaudebord/indicateur-performance.htm>