

CAS DE SYNTHESE



Enseignants:
LINARD Claire
DOHET Nicolas

Etudiants :
DIOP Sidy
KAMARA Emmanuella
MAMOUDOU KAKA Abdoul-Kader
ABDOU DANGALADIMA Nana Fatoumatou

Table des matières

I. Introduction.....	2
II. Présentation de l'équipe.....	3
III. Mission et problématique.....	4
IV. Modélisation.....	5
A. Modélisation des tables SAS.....	5
B. Modélisation des tables Data Warehouse.....	6
V. Intégration des données.....	7
1. Intégration des fichiers sources avec la table TABLE_FILE.....	7
2. Alimentation des tables SAS.....	8
3. Alimentation des tables DWH.....	8
a. DWH_COMPTE.....	8
b. DWH_CLIENT.....	9
c. Tables de contact : DWH_ADRESSE, DWH_TELEPHONE, DWH_EMAIL.....	10
4. Ordonnancement et automatisation.....	10
VI. Restitution.....	12
VII. Retour d'expériences.....	14
A. Difficultés rencontrés :.....	14
B. Axes d'amélioration :.....	14
VIII. Conclusion.....	15
IX. Remerciements.....	16

I. Introduction

Un groupe dentaire situé dans le nord de la France souhaite améliorer sa connaissance client. Dans cette optique, nous avons été contactés pour prendre en charge ce projet. Notre mission consiste à leur fournir des visualisations répondant à leurs besoins et à identifier les axes d'amélioration, notamment en ce qui concerne la qualité des données client.

Pour ce projet, il nous a été mis à disposition des données non structurées sous forme de fichiers plats. Notre rôle sera d'assurer toutes les étapes, allant de la transformation et du nettoyage des données à leur analyse et visualisation. Nous veillerons également à anticiper et à gérer les contraintes susceptibles de survenir dans le cadre de l'activité.

Nous avons mobilisé nos compétences en data, de la conception d'une base de données à la restitution des informations, en utilisant des outils adaptés tels que l'ETL Semarchy pour effectuer les différents traitements et transformations des données et l'outil de restitution Power BI pour les visualisations.

Consciente du caractère transversal d'un projet data et de l'importance de la prise de recul à chaque étape, notre équipe a été constituée de quatre experts spécialisés dans différentes branches de la data. Nous avons ainsi un data analyst, un data engineer, un tech lead, et un scrum master. Cette organisation nous a permise de prendre des décisions plus éclairées, d'être mieux organisés et d'identifier les problèmes le plus tôt possible.

Afin d'assurer une progression cohérente, nous avons structuré notre démarche autour d'un fil conducteur clair. Ce rapport est organisé en plusieurs volets distincts :

1. Une présentation détaillée de notre équipe.
2. Une description de nos missions et de la problématique abordée.
3. Une explication des étapes de modélisation, d'intégration et de restitution des données.
4. Un retour d'expérience sur les défis rencontrés et les solutions mises en place.

Nous espérons que ce rapport offrira une vision claire et approfondie du travail accompli et des résultats obtenus. Cette réalisation, dont nous sommes persuadés, apportera des ressources non négligeables qui seront bénéfiques pour l'entreprise.

II. Présentation de l'équipe

Pour mener à bien ce projet, il était essentiel de constituer une équipe équilibrée, composée de profils complémentaires couvrant presque toute la chaîne de la data. Chaque membre de l'équipe a été sélectionné pour son expertise et son rôle spécifique, contribuant ainsi à une répartition efficace des tâches, un élément clé de la réussite du projet. L'équipe, composée de quatre personnes, était organisée comme suit :

1. Tech Lead : Sidy Diop

Sidy, notre Tech Lead, est un développeur expérimenté possédant une expertise technique approfondie. Il a joué un rôle central dans les décisions techniques du projet, guidant l'équipe dans la résolution de problèmes complexes et veillant au respect des normes de codage et des bonnes pratiques. Son travail de planification technique a permis de déterminer les meilleures approches pour atteindre les objectifs définis. Il a également collaboré étroitement avec les autres membres pour assurer la cohérence technique du projet.

2. Data Analyst : Emmanuella Kamara

Emmanuella, en tant que Data Analyst, a joué un rôle clé dans la restitution des résultats. Elle a analysé les données collectées et produit des visualisations telles que des graphiques, des tableaux de bord et des rapports. Ces livrables permettent de rendre les données plus accessibles et compréhensibles, facilitant ainsi la prise de décision. Ses compétences ont été particulièrement précieuses dans la création d'un reporting de qualité, contribuant directement à la valeur ajoutée du projet pour le cabinet dentaire.

3. Data Engineer : Abdoul Kader Mamoudou

Abdoul Kader a occupé le rôle de Data Engineer, avec pour mission principale la gestion technique des flux de données à l'aide de l'outil ETL Semarchy. Il a conçu et implémenté les pipelines de traitement des données, veillant à leur intégration correcte dans la base de données. En plus de s'assurer de l'automatisation et de l'optimisation des processus, Abdoul Kader a défini les contrôles nécessaires pour garantir la qualité et la traçabilité des données tout au long du projet. Son expertise technique a été un atout majeur pour répondre aux exigences complexes du projet.

4. Scrum Master : Nana Fatoumatou Abdou

Nana, notre Scrum Master, a été la facilitatrice de l'équipe. Elle a assuré la coordination du projet en appliquant les principes de la méthodologie Agile. Elle a organisé les réunions de suivi, aidé l'équipe à surmonter les obstacles, et veillé à ce que les besoins du Product Owner (PO) soient compris et pris en compte. En collaboration étroite avec le PO, Nana a clarifié et priorisé les éléments du backlog. Son rôle a été déterminant pour maintenir une dynamique positive au sein de l'équipe et garantir l'avancement fluide du projet.

III. Mission et problématique

Les données sont une ressource précieuse, voire essentielle, à la bonne gestion d'une organisation. Les exploiter efficacement est désormais indispensable pour améliorer la connaissance client, les fidéliser, mieux comprendre leurs besoins, optimiser les produits et, in fine, atteindre les objectifs stratégiques.

Notre client, un cabinet bancaire, n'a pas les mêmes besoins de fidélisation de la clientèle qu'une organisation vendant des produits spécifiques, comme nous l'avons étudié au cours de ces deux années de master.

La problématique ici est différente : notre rôle consiste à permettre au cabinet de visualiser la qualité de ses données client. Dans ce contexte, il est primordial, par exemple, de vérifier si l'adresse email d'un client est obsolète, car elle est essentielle pour l'envoi des résultats, l'identification des clients par région, etc.

C'est précisément pour cela que nous menons ce projet : l'objectif est d'améliorer les activités du cabinet, de fluidifier les opérations et d'assurer une organisation optimale.

Une base de données bien conçue constitue le socle indispensable pour transformer les données en informations pertinentes. Notre mission consiste donc à concevoir et réaliser cette base de données à partir des données fournies, tout en suivant un ensemble d'étapes cruciales : modélisation, prise en compte des contraintes spécifiques à l'activité du client, intégration, gestion de la qualité des données et restitution.

Chacune de ces étapes nécessite rigueur et expertise pour garantir un résultat à la hauteur des attentes. En complément de la base de données, notre objectif est également de produire un reporting clair et synthétique. Celui-ci permettra au groupe dentaire de visualiser les résultats obtenus, tout en mettant en lumière les données clés de manière compréhensible et exploitable.

En résumé, notre mission principale est de doter le cabinet dentaire d'un outil fiable et performant pour optimiser la gestion de son activité. À travers ce projet, nous visons à transformer les données en un levier stratégique, en combinant une infrastructure robuste et une restitution efficace et informative.

IV. Modélisation

La conceptualisation et la modélisation jouent un rôle central dans la mise en place d'une base de données clients fiable et efficace. Première étape cruciale du projet, elles conditionnent la qualité des phases ultérieures, notamment l'intégration et la restitution des données. Une modélisation rigoureuse garantit une structure claire et cohérente, facilitant ainsi le traitement des données et les analyses futures.

A. Modélisation des tables SAS

Dans le cadre de ce projet, nous avons créé des **tables SAS** afin de préparer les données avant de les intégrer dans le **Data Warehouse**. Cette étape intermédiaire nous permet de structurer et de nettoyer les données, garantissant ainsi leur qualité et leur cohérence avant leur stockage final.

Pour cela, nous avons d'abord conceptualisé des tables SAS qui reflètent les fichiers sources reçus. Ces tables (SAS_COMPTE,SAS_CLIENT,SAS_EMAIL,SAS_ADRESSE, SAS_TELEPHONE) servent d'image intermédiaire des données avant leur intégration dans le Data Warehouse.

Pour répondre aux exigences de conformité et de traçabilité, une table supplémentaire, appelée **FILE_SOURCE**, a été créée. Cette table permet d'enregistrer la date d'intégration et d'identifier précisément le fichier source à l'origine des données. Ce mécanisme assure une transparence totale dans le suivi des données et prévient tout risque d'incohérence.

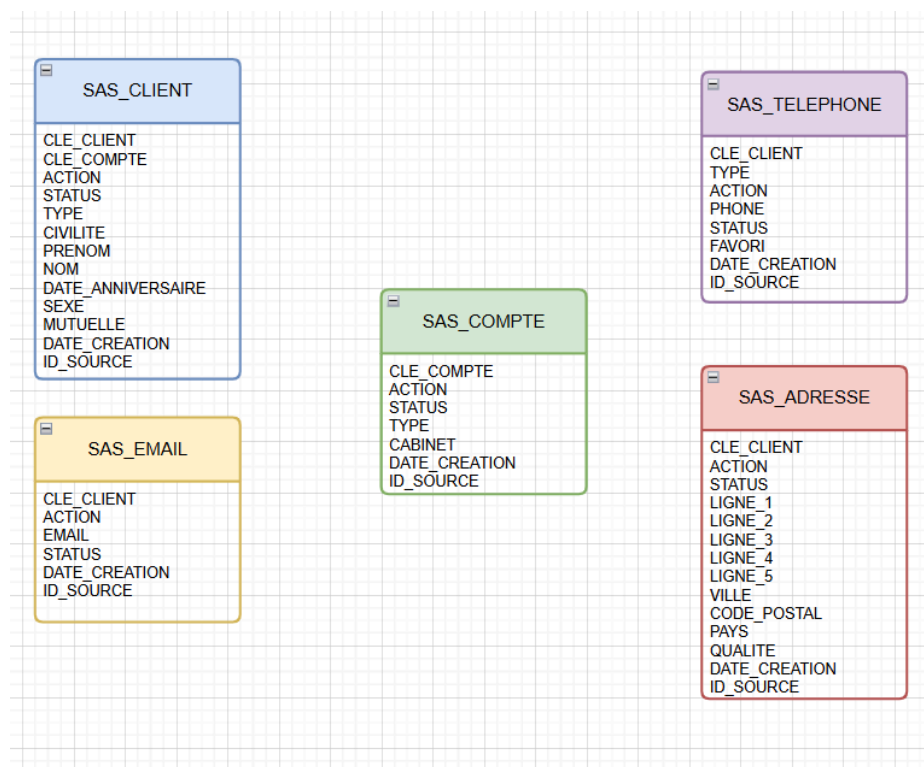


Figure 1 : Modélisation des tables SAS

B. Modélisation des tables Data Warehouse

Les tables SAS servent de base pour alimenter les tables finales du Data Warehouse (DWH). Ces dernières sont structurées de manière à garantir une restitution claire et exploitable des données. Avant l'alimentation, des contrôles rigoureux sont effectués (comme défini dans le cahier de test) pour assurer l'intégrité et la cohérence des données.

La structure du Data Warehouse repose sur une table centrale, **DWH_CLIENT**, qui consolide les informations principales relatives aux clients. Cette table est liée aux tables annexes suivantes :

- **DWH_COMPTE** : Informations liées aux comptes des clients.
- **DWH_ADRESSE** : Adresses postales des clients.
- **DWH_TELEPHONE** : Numéros de téléphone associés.
- **DWH_EMAIL** : Adresses email des clients.

Cette architecture en étoile simplifie les analyses tout en assurant une organisation claire des données. Une modélisation soigneusement conçue garantit ainsi la fiabilité des informations extraites du Data Warehouse et facilite leur exploitation à des fins décisionnelles.

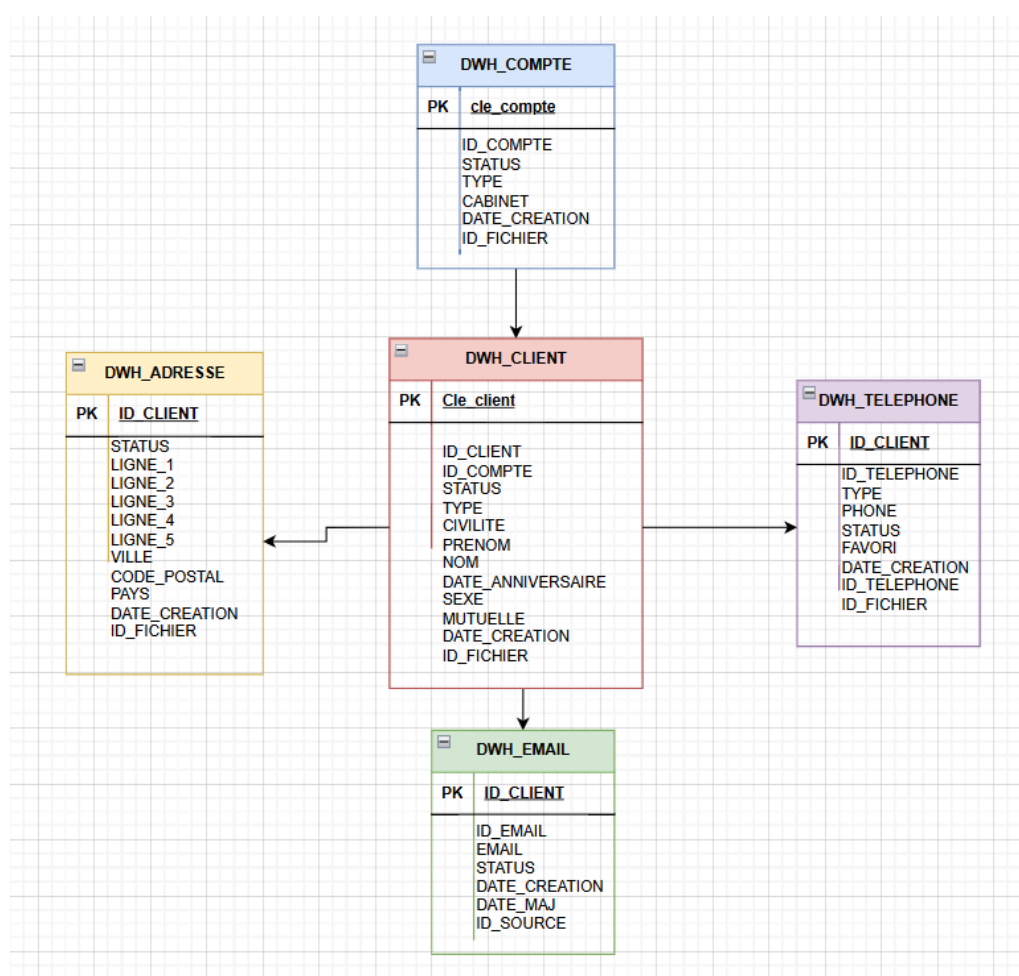


Figure 2 : Modèle Logique de Données

Technologie utilisée:

Toutes ces tables ont été créées sur une base de données **ORACLE**, choisie pour ses performances et sa robustesse dans le traitement de gros volumes de données.

En résumé, cette modélisation, tant au niveau des tables SAS que des tables DWH, constitue le socle du projet et assure une gestion optimale des données tout au long du processus.

V. Intégration des données

L'intégration des données est une étape clé de notre projet, et nous avons choisi d'utiliser **Stambia**, une plateforme d'intégration de données puissante et polyvalente, pour alimenter les différentes tables du Data Warehouse. Cette section décrit les étapes suivies pour intégrer les données, des fichiers sources jusqu'aux tables finales du Data Warehouse.

1. Intégration des fichiers sources avec la table **TABLE_FILE**

La première étape de l'intégration consistait à importer les fichiers fournis par le métier en utilisant la technologie de rétro-ingénierie de Stambia. Ces fichiers sont ensuite intégrés dans une table nommée **TABLE_FILE**, où sont effectués les contrôles nécessaires pour valider leur conformité. Ces contrôles garantissent la qualité et la fiabilité des données avant leur utilisation.

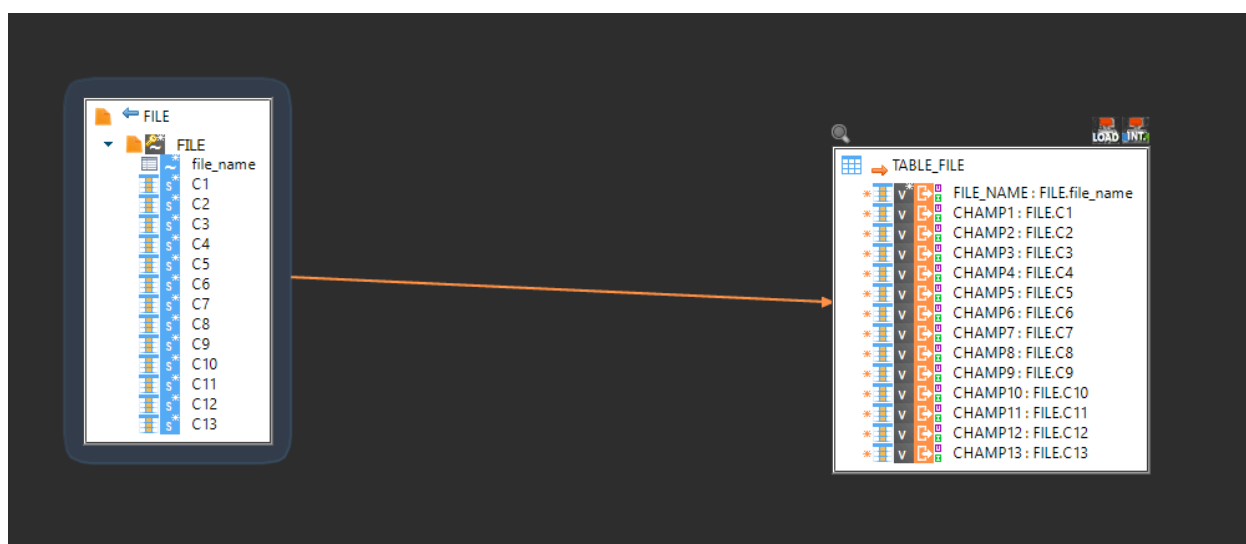


Figure 3 : Alimentation de la table **TABLE_FILE**

2. Alimentation des tables SAS

Une fois les fichiers validés dans TABLE_FILE, ils sont intégrés dans les tables SAS (SAS_COMPTE, SAS_CLIENT, SAS_EMAIL, SAS_ADRESSE, SAS_TELEPHONE) sans transformation.

Le processus se déroule comme suit :

1. **TABLE_FILE** alimente la table **FILE_SOURCE** pour tracer l'origine des données.
2. Une table de staging est utilisée pour effectuer les jointures nécessaires avec la table **FILE_SOURCE**.
3. Les données sont ensuite filtrées selon le champ 2 qui correspond au type de la ligne pour alimenter les tables SAS correspondantes.

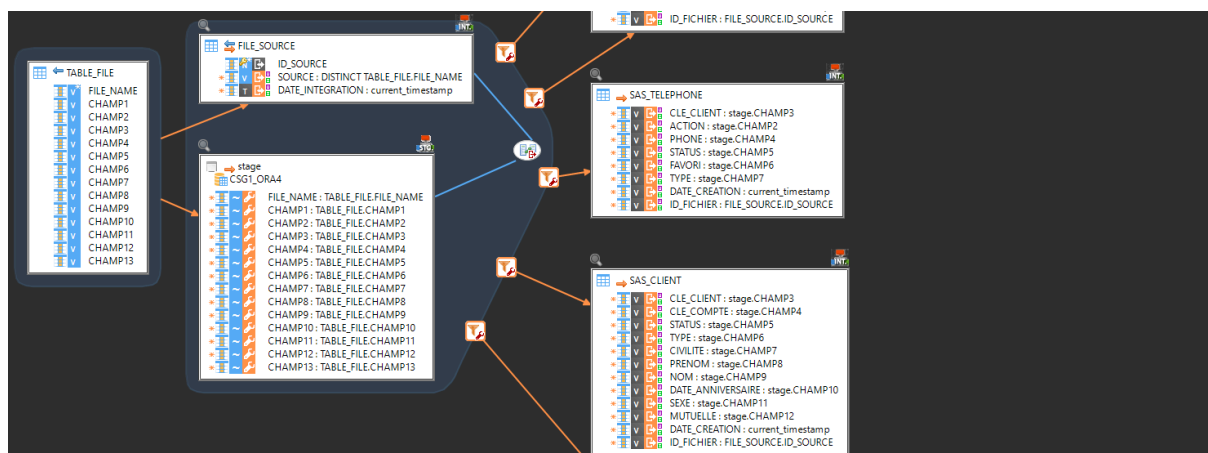


figure4 : Alimentation des tables SAS

3. Alimentation des tables DWH

Les tables DWH sont alimentées à partir des tables SAS selon un ordre prédéfini pour garantir la cohérence des données.

a. DWH_COMPTE

La première table à être alimentée est **DWH_COMPTE**, basée sur **SAS_COMPTE**. Une jointure avec la table **TRANSCO** permet de transcoder le champ **CABINET**, et **TYPE** essentiel pour le reporting. Un contrôle supplémentaire rejette les cabinets non conformes à la liste préétablie.

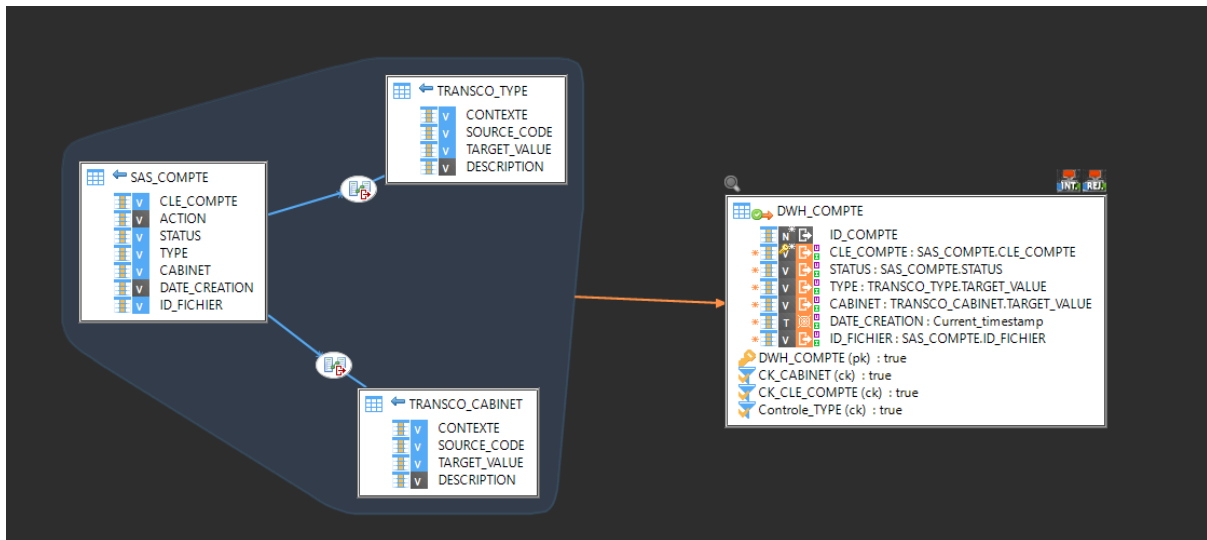


Figure5 : Alimentation de DWH_COMPTE

b. DWH_CLIENT

Ensuite, la table **DWH_CLIENT** est alimentée à partir de **SAS_CLIENT** et de **DWH_COMPTE**. Cette étape assure que seuls les clients ayant un compte valide sont intégrés. Des transcodifications supplémentaires sont réalisées pour les champs **SEXE**, **CIVILITÉ**, **TYPE**

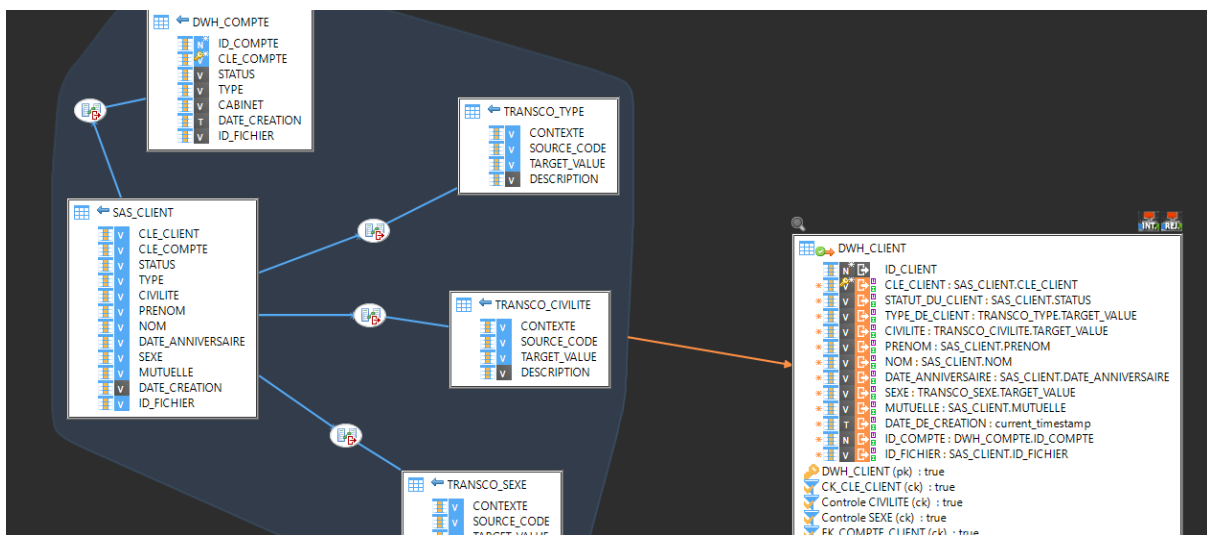


Figure6 : Alimentation de DWH_CLIENT

c. Tables de contact : DWH_ADRESSE, DWH_TELEPHONE, DWH_EMAIL

Les tables **DWH_ADRESSE**, **DWH_TELEPHONE**, et **DWH_EMAIL** sont alimentées à partir des tables SAS respectives, avec une jointure sur **DWH_CLIENT**. Ce processus garantit que seules les informations de contact associées à des clients valides sont intégrées.

- Des contrôles sont appliqués pour valider la structure des adresses e-mail et des numéros de téléphone.
- Des transcodifications sont également réalisées pour certains champs, tels que le pays dans les adresses.

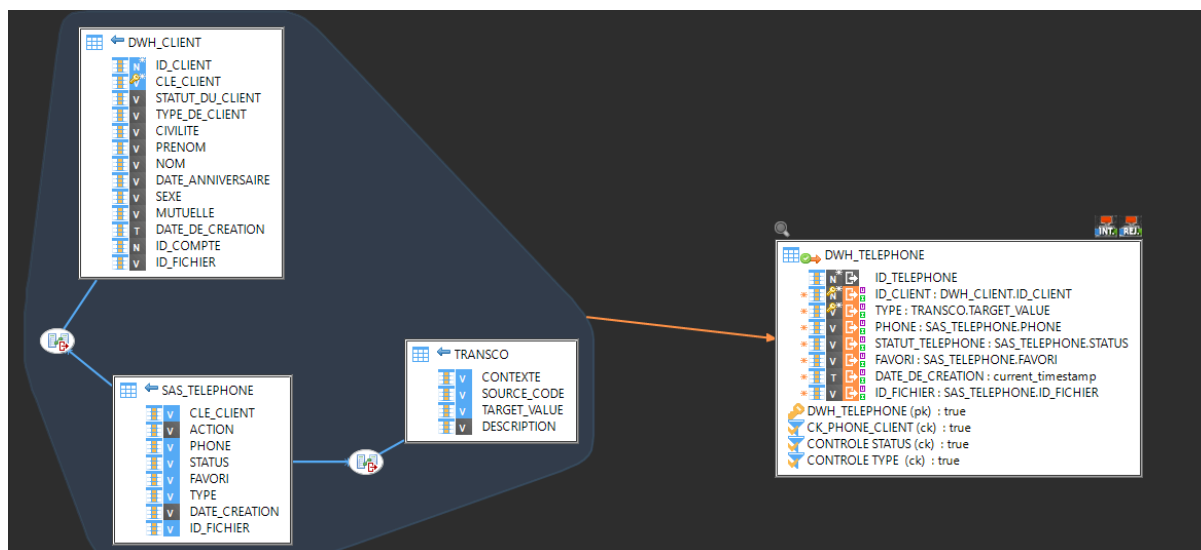


Figure7 : Alimentation de DWH_TELEPHONE

4. Ordonnancement et automatisatisation

L'ensemble des étapes d'intégration a été orchestré pour garantir un fonctionnement fluide et structuré. Le processus suit un ordre logique et cohérent, permettant une intégration fiable des données tout au long de la chaîne.

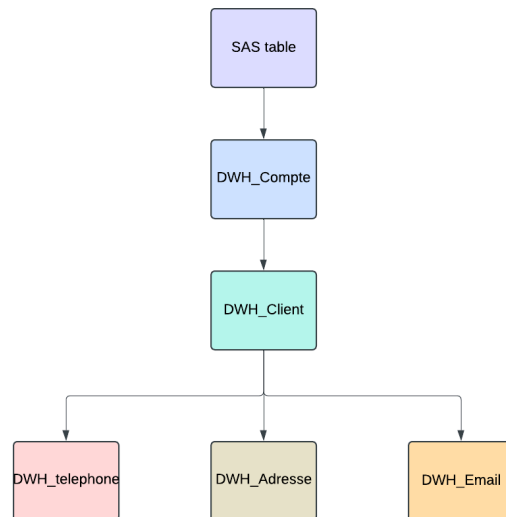


Figure8 : Processus global d'intégration

Cette approche méthodique, associée aux outils et contrôles mis en place, garantit la qualité et la traçabilité des données tout au long du processus.

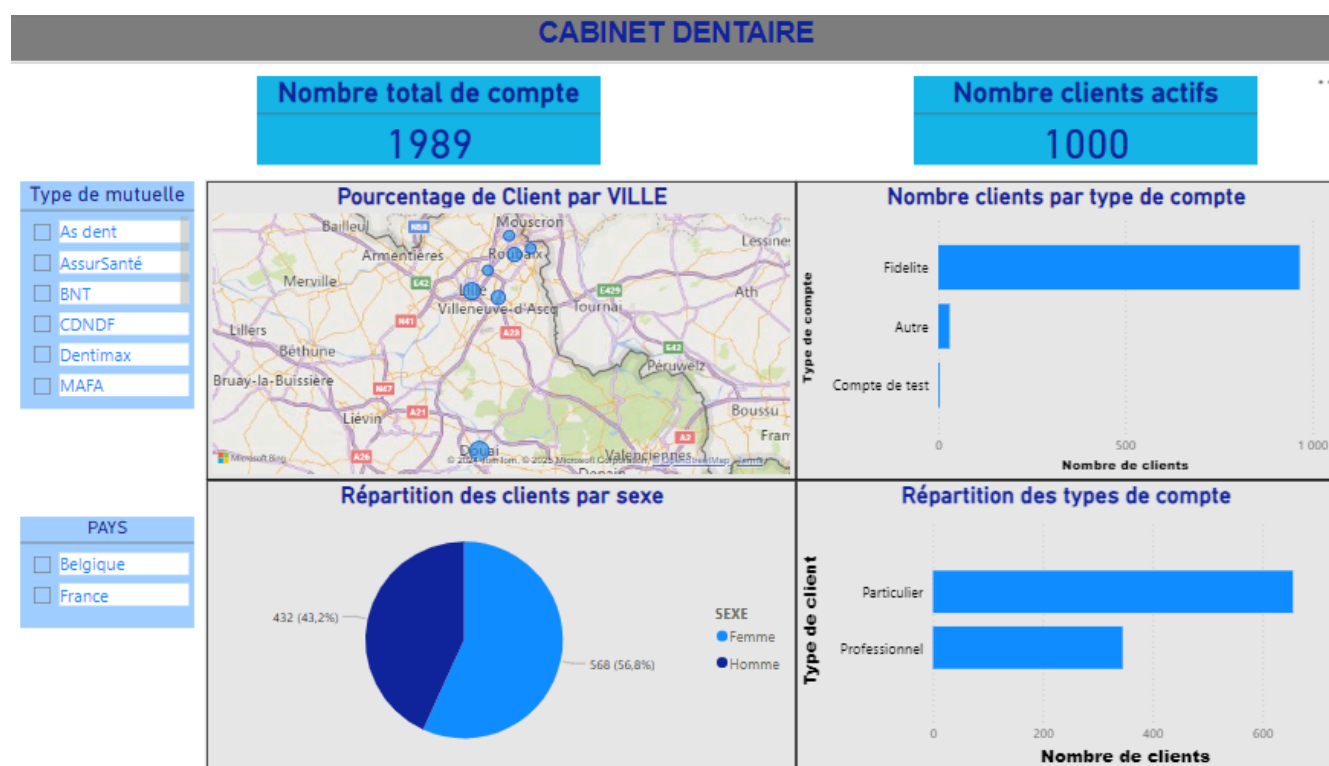
VI. Restitution

Après l'intégration complète des données dans la base, nous avons entamé la conception du tableau de bord. Grâce au travail effectué en amont, la création de cet outil a été grandement simplifiée : les relations entre les tables ont été générées automatiquement, éliminant ainsi la nécessité de traitements supplémentaires après la connexion à la base de données.

Nous avons structuré notre analyse autour de la **connaissance client**, en menant une exploration approfondie des différents types de comptes, des profils clients, et de leur répartition géographique. Une attention particulière a été portée à l'étude des affiliations aux mutuelles, permettant d'identifier les tendances et les spécificités liées à ces données.

Cette restitution offre une vision globale et précise de la base client, facilitant des prises de décisions stratégiques éclairées. Elle permet également de mieux comprendre les comportements et les besoins des clients, ouvrant la voie à des actions ciblées et à une amélioration continue des services.

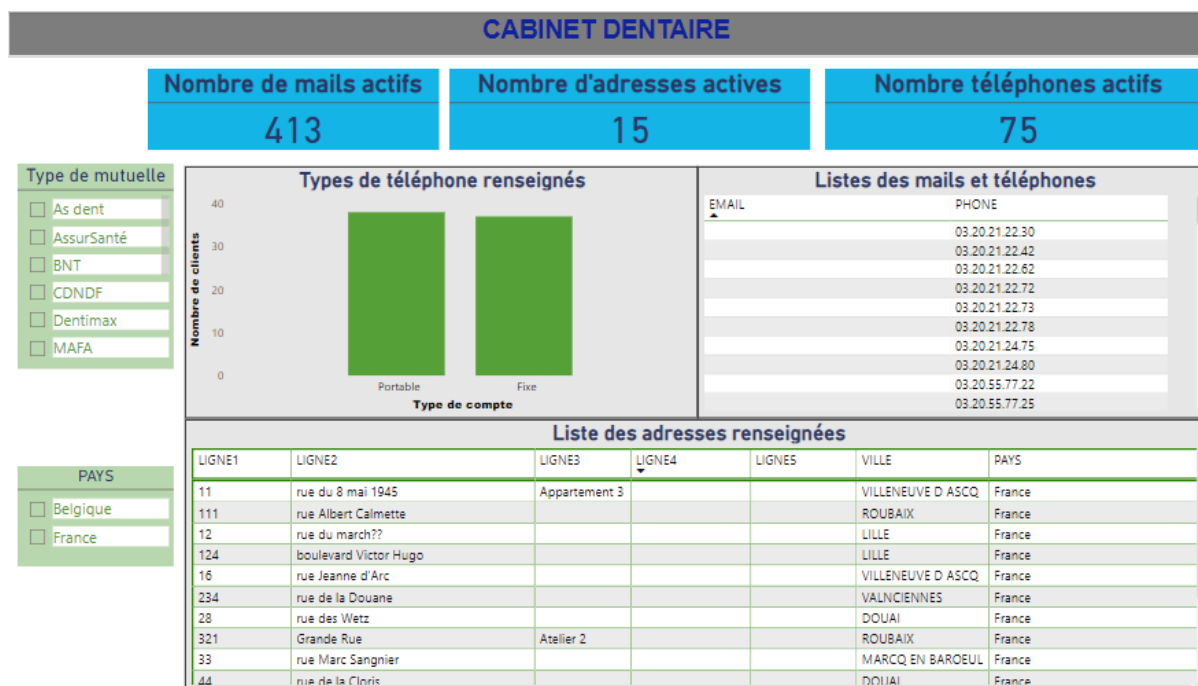
Ci-dessous une image du tableau de bord :



Nous avons également conçu un tableau de bord dédié à l'analyse des coordonnées des clients, incluant les adresses e-mail, les numéros de téléphone et les adresses postales. Ce tableau de bord permet une exploration approfondie de la validité et de la disponibilité de ces données, offrant ainsi une vue claire sur leur fiabilité.

Cette analyse permet d'optimiser le ciblage des campagnes marketing en se concentrant sur des informations précises et à jour. Cela contribue à améliorer l'efficacité des actions marketing tout en réduisant les coûts liés aux envois inefficaces ou aux contacts obsolètes.

Ci dessous une image du tableau de bord :



Ces tableaux de bord, conçus pour être à la fois simples et fonctionnels, regroupent toutes les coordonnées des clients, y compris leurs adresses postales, emails, et numéros de téléphone. Cette centralisation facilite grandement la recherche de clients spécifiques, permettant aux gestionnaires de gagner du temps et d'améliorer leur efficacité au quotidien.

Un des points forts de ces tableaux de bord est l'intégration de filtres interactifs sur la gauche, permettant de trier les données selon différents critères, comme les types de mutuelles ou les pays de résidence. En cochant simplement les cases, les gestionnaires peuvent rapidement accéder aux informations recherchées, sans avoir à naviguer dans des volumes de données inutiles.

Ces filtres permettent aussi de combiner plusieurs critères pour des recherches encore plus précises. Par exemple, il devient facile d'identifier les clients affiliés à une mutuelle spécifique ou localisés dans une région donnée, ce qui est particulièrement utile pour mener des actions ciblées, telles que l'envoi de communications personnalisées ou la mise en place de campagnes marketing adaptées.

VII. Retour d'expériences

A. Difficultés rencontrés :

Nous avons initialement rencontré des difficultés à comprendre les relations complexes entre certaines tables, ce qui a freiné notre progression au début du projet. La nécessité de réaliser un cahier des charges n'était pas immédiatement évidente. Cependant, à mesure que nous avançons, plusieurs concepts expliqués par Mme Linard en classe ont pris tout leur sens et nous ont permis de progresser. Les différentes contraintes du cahier des charges, indispensables pour garantir la qualité et la cohérence des données, ainsi que le concept des tables SAS, jouant le rôle de tables intermédiaires où nous pouvions effectuer les modifications nécessaires avant leur intégration dans le Data Warehouse, se sont révélés cruciaux.

Sur la partie ETL, nous avons investi beaucoup de temps pour apprendre et maîtriser Semarchy, un outil avec lequel nous n'avions que peu d'expérience. Grâce au soutien de M. Dohet et à nos recherches approfondies, nous avons pu surmonter ces obstacles et renforcer considérablement nos compétences sur cet outil.

Par ailleurs, des problèmes d'accès à Oracle via une connexion à distance ont causé des retards. Ces contraintes techniques nous ont contraints à vérifier scrupuleusement que les fonctionnalités implémentées fonctionnaient correctement, ce qui a prolongé certaines étapes du projet. Cette rigueur, bien que chronophage, a été essentielle pour assurer la fiabilité des solutions mises en place.

B. Axes d'amélioration :

Avec le temps dont nous disposions, nous avons réussi à concevoir une solution fonctionnelle répondant aux objectifs initiaux. Cependant, avec davantage de temps, nous aurions pu intégrer des fonctionnalités supplémentaires pour enrichir le projet, optimiser son utilisation et améliorer son impact global.

Il convient également de noter que la majorité des étudiants ne disposent pas des prérequis nécessaires pour maîtriser efficacement l'outil d'intégration Semarchy. Pour pallier cette lacune, il serait judicieux de renforcer les formations sur cet outil en proposant des sessions pratiques plus approfondies ou en organisant des ateliers spécialisés. Cela permettrait aux futurs étudiants de mieux appréhender l'outil et d'être plus autonomes dans leur utilisation lors de projets similaires.

Au vu du secteur d'activité du cabinet, il aurait été particulièrement pertinent d'intégrer des contraintes spécifiques sur la validité des adresses e-mail et des numéros de téléphone. En effet, dans un cabinet médical ou dentaire, les résultats d'analyses, les rappels de rendez-vous, ou toute autre communication importante sont souvent envoyés par e-mail.

En intégrant ces contraintes, les gestionnaires pourraient calculer le ratio d'e-mails et de numéros de téléphone invalides, leur permettant d'identifier rapidement les données incorrectes ou obsolètes. Cela garantirait non seulement une meilleure traçabilité des informations, mais aussi une amélioration significative de l'expérience client, en minimisant les risques de communication manquée.

VIII. Conclusion

Ce projet de mise en place d'une base de données pour un groupe dentaire a été une expérience riche et formatrice. De la modélisation des données à la restitution via des tableaux de bord, chaque étape a permis de répondre aux besoins du client.

Malgré les défis rencontrés, notamment liés à l'apprentissage d'outils nouveaux comme Semarchy et aux contraintes techniques d'accès à distance, nous avons su nous adapter grâce à une approche collaborative et à des efforts continus. Le projet a non seulement permis de renforcer notre maîtrise des technologies utilisées, mais également d'approfondir notre compréhension des bonnes pratiques en gestion de données.

Ce projet met également en lumière l'importance de la planification et de la structuration des tâches. Il a permis aussi de révéler des axes d'amélioration, comme le renforcement des formations sur les outils spécifiques.

IX. Remerciements

Globalement, ce projet a été extrêmement enrichissant. Chaque membre du master SIAD a un parcours différent, des aspirations variées et des ambitions qui évoluent au fil du temps. Certains souhaitent devenir data analyst, d'autres data scientist ou encore data engineer, et nous avons souvent tendance à nous concentrer sur les matières directement liées à la profession que nous envisageons d'exercer. Cependant, il est essentiel de comprendre qu'une approche transversale est primordiale, surtout dans un domaine aussi dynamique et en constante évolution, dans lequel nous voulons évoluer et faire carrière.

Ce cas de synthèse a appris à chacun d'entre nous à adopter une vision globale de la chaîne d'exploitation des données, de la prise en charge des données brutes jusqu'à leur restitution. Nous avons réfléchi aux différentes contraintes liées à la problématique, à l'identification des analyses pertinentes et surtout à la manière de détecter des solutions capables d'améliorer concrètement le quotidien des gestionnaires. Cette expérience a renforcé notre compréhension du travail d'équipe et de l'importance de la collaboration interdisciplinaire pour réussir un projet de cette envergure.

Nous tenons à remercier tous ceux qui ont contribué à ce projet, en particulier nos enseignants et nos collègues, pour leur soutien, leurs conseils et leur expertise.