

UNIVERSITE MOHAMMED V DE RABAT

ECOLE MOHAMMADIA D'INGENIEURS



Filière : Génie Informatique

Option : Systèmes d'information

Mémoire de Projet de Fin d'Études

N° INF 28/2022

**Mise en place d'une solution décisionnelle pour la
plateforme ANIA assurance santé**

Réalisé par :

KABBAJ Samiha

KRIMCHI Moataz

Encadré par :

Mr. El GHANAMI Driss

M. BERRADA Sofia

Année 2021-2022

UNIVERSITE MOHAMMED V DE RABAT

ECOLE MOHAMMADIA D'INGENIEURS



Filière : Génie Informatique

Option : Systèmes d'information

Mémoire de Projet de Fin d'Études

N° INF 28/2022

**Mise en place d'une solution décisionnelle pour la
plateforme ANIA assurance santé**

Réalisé par :

KABBAJ Samiha

KRIMCHI Moataz

Membres de Jury :

Pr. BENHLIMA Laila

Présidente

Pr. BAH Slimane

Rapporteur

Pr. El GHANAMI Driss

Encadrant EMI

M. BERRADA Sofia

Encadrante InsurBT

Dédicaces

Je dédie ce travail

À ma chère mère,

Qui a tous sacrifié pour moi, qui m'a comblé de son amour, qui m'a appris que les bonnes choses dans la vie n'arrivent pas facilement, qui était toujours présente, m'assistait et conseillait dans les bons moments comme dans les mauvais, reçois à travers ce travail aussi modeste soit-il, l'expression de mon éternelle gratitude.

À mon cher père

À cet homme qui a toujours été mon exemple, qui a sacrifié tout ce qu'il pouvait pour nous voir heureux, je te remercie pour les valeurs nobles et pour ton éducation.

À ma chère sœur Imane et mon cher frère Amine

Qui n'ont pas cessé de me conseiller, encourager et soutenir tout au long de mes études. Que Dieu les protège et leurs donne la chance et le bonheur.

À Mon binôme Moataz

Pour son encouragement, sa patience et pour tous les moments qu'on a vécus ensemble pour accomplir notre projet.

À tous mes amis, ma famille et les gens qui ont cru en moi et qui me donnent l'envie d'aller en avant.

KABBAJ Samiha

Je dédie ce travail du profond de mon cœur à tous ceux qui me sont chers

À ma chère mère

Aucune dédicace ne saurait exprimer mon très grand amour envers toi, mon respect pour tous les sacrifices que vous avez réservés pour moi. Cette femme qui n'a jamais dit non à mes exigences et qui n'a épargné aucun effort pour me rendre heureux.

À mon cher père

Ce travail est dédié à mon père qui m'a toujours poussé et motivé, a très bien travaillé dans mes études. L'homme qui doit ma vie, ma réussite et mon respect total puisqu'il a toujours accordé en témoignage de ma reconnaissance envers sa confiance, ses sacrifices et sa tendresse.

À mes frères Mohammed et Nadir

Je vous remercie pour l'encouragement que vous m'avez donné. Je vous souhaite tout le bonheur que vous méritez et un très brillant avenir sur vos parcours.

À ma Binôme Samiha

Merci énormément pour ton soutien moral plus que précieux et ta compréhension tout au long de notre travail ensemble.

À tous mes amis

Pour l'amitié sincère qui nous a liées et de bons moments que nous avons passés.

KRIMCHI Moataz

Remerciements

Au terme de ce travail, Il m'est agréable de remercier toutes les personnes dont leurs interventions ont permis d'accomplir ce projet de fin d'études.

Tout d'abord, nous remercions toute l'équipe de la société InsurBT pour leur accueil chaleureux et leur soutien.

Plus précisément, nous remercions : Notre encadrante **M. BERRADA Sofia** de la société InsurBT qui nous a soutenu par ses conseils enrichissants et son accompagnement tout au long de notre stage.

Nous tenons de remercier **Mr. AKANSOUS Rachid**, directeur de InsurBT pour sa disponibilité et ses conseils qui nous ont permis de toujours poser de nouvelles questions et ainsi avancer dans notre projet.

Notre encadrant interne à l'EMI **Mr. EL GHANAMI DRISS** pour sa bonne volonté d'accepter de nous encadrer ainsi que l'aide et les conseils concernant les missions évoquées dans ce rapport et la bonne démarche à suivre pour mener à bien notre travail.

Nos sincères remerciements aussi à **M. EL MOUSSAOUI Khadija** et à toute l'équipe d'InsurBT pour leur serviabilité, leurs conseils pertinents et leur soutien.

Nos vifs remerciements aussi pour tous nos professeurs et le corps administratif de l'EMI pour ces trois années de collaboration et pour la qualité de l'enseignement qui nous a été dispensé.

Nos remerciements vont également à Monsieur **Slimane BAH** et **Madame Laila BENHLIMA**, pour l'honneur qu'ils nous ont fait en acceptant de présider notre jury et d'évaluer notre modeste travail.

Résumé

L'industrie de l'assurance fournit un bon exemple de données exploitées pour aider à prendre des décisions commerciales bénéfiques étant donné que cette industrie est basée sur le risque, et c'est pour cette raison-là que les assurances choisissent l'Informatique Décisionnelle (BI) afin d'explorer des tableaux de bord interactifs et touchants le besoin des assureurs. InsurBT, en tant qu'entreprise de conseil et technologies d'assurance, s'intéresse à ce type de projet BI pour toutes les compagnies d'assurance & de réassurance.

Pour notre projet, afin d'assurer la bonne gestion, nous avons suivi une démarche agile grâce à la méthodologie Scrum qui a pour but d'améliorer la productivité des équipes, tout en permettant une optimisation du produit grâce à des feedbacks réguliers des clients.

Notre projet a passé par plusieurs étapes, tout d'abord on a commencé par analyser les besoins exprimés dans le cahier de charge et déterminer les différents KPIs répondants à ce dernier, ensuite l'élaboration de l'entrepôt de donnée a suivi une méthode propre à l'informatique décisionnelle : l'approche Ralph Kimball qui nous a permis à partir des Data Marts extraites par SSIS (ETL) de créer un modèle du Datawarehouse en flocon de neige, cette structuration de données se fait à l'aide de l'outil d'analyse SSAS qui s'appuie sur le langage DAX. Enfin une fois ces données sont structurées sous forme d'un modèle tabulaire, ils sont dirigés vers Power BI afin de créer les tableaux de bord et les intégrer sur la plateforme ANIA.

Mots clés : Assurance, Datawarehouse, BI, ETL, DataMarts, Visualisation

Abstract

The insurance industry provides a good example of data being leveraged to help make beneficial business decisions, as this is a risk-based industry, insurance companies are choosing Business Intelligence (BI) to explore interactive dashboards that address the needs of insurers. InsurBT, as an insurance technology and consulting company, is interested in this type of BI project for all insurance & reinsurance companies.

For our project to ensure proper management we followed an agile approach thanks to the Scrum methodology which aims to improve the productivity of the teams, while allowing an optimization of the product thanks to regular feedback from customers.

Our project went through several steps, first we started by analyzing the needs expressed in the specifications and determine the different KPIs responding to it, then the development of the data warehouse followed a method specific to business intelligence: the Ralph Kimball approach that allowed us from the Data Marts extracted by SSIS (ETL) to create a model of the Datawarehouse in snowflake, this structuring of data is done using the analysis tool SSAS which is based on the DAX language. Finally, once these data are structured in the form of a tabular model, they are directed to Power BI in order to create the dashboards and integrate them on the ANIA platform.

Keywords: Insurance, Datawarehouse, BI, ETL, DataMarts, Visualization

تقدم صناعة التأمين مثالا جيدا على البيانات التي يتم الاستفادة منها للمساعدة في اتخاذ قرارات تجارية مفيدة، نظرا لأن هذه الصناعة تعتمد على المخاطر، ولهذا السبب تختار شركات التأمين ذكاء الأعمال لاستكشاف لوحات المعلومات التفاعلية التي تلبي احتياجات شركات التأمين. إن InsurBT، بصفتها شركة استشارات وتكنولوجيا في مجال التأمين، مهتمة بهذا النوع من مشاريع BI لجميع شركات التأمين وإعادة التأمين.

بالنسبة لمشروعنا، لضمان الإدارة السليمة، اتبعنا منهجية Scrum، والتي تهدف إلى تحسين إنتاجية الفريق، مع السماح بتحسين المنتج من خلال ملاحظات العملاء المنتظمة.

مر مشروعنا بعدة مراحل، أولاً بدأنا بتحليل الاحتياجات التي تم التعبير عنها في المواصفات وتحديد مؤشرات الأداء الرئيسية المختلفة التي تستجيب لها، ثم اتبع تطوير مستودع البيانات أسلوباً محدداً (Ralph Kimball) الذي مكّننا من استخراج البيانات من DataMarts بواسطة SSIS (ETL) لإنشاء نموذج snowflake لمستودع البيانات، تتم هيكلة البيانات هذه باستخدام أداة تحليل SSAS استناداً إلى لغة DAX. أخيراً، بمجرد هيكلة هذه البيانات في شكل نموذج جدولي، يتم توجيهها إلى Power BI لإنشاء لوحات معلومات ودمجها في النظام الأساسي ANIA.

الكلمات الرئيسية: التأمين ، مستودع البيانات ، BI ، ETL ، DataMarts ، إعداد التقارير

Table des matières

Liste des figures	xi
Liste des tableaux	xiii
Glossaire.....	xiv
Introduction Générale.....	1
Chapitre 1 : Contexte général du projet	2
1 Présentation de l'organisme d'accueil.....	3
1.1 Présentation d'InsurBT	3
1.2 Présentation de la hiérarchie	3
2 Conseil et technologies des assurances	4
3 Cadrage du projet	5
3.1 Contexte du projet.....	5
3.2 Problématique	7
3.3 Objectifs	7
4 Méthodologie SCRUM	8
5 Planning du projet (Diagramme de Gantt)	9
Chapitre 2 : Spécification et Analyse des besoins.....	11
1 Etude des besoins	12
1.1 Besoins fonctionnels	12
1.1.1 Besoins décisionnels.....	12
1.1.2 Besoins applicatifs.....	14
1.2 Besoins non fonctionnels	15
2 Détermination des indicateurs	16
Chapitre 3 : Analyse et Conception	22
1 Architecture	23
1.1 Comparatif des approches BI.....	23
1.2 Descriptif de l'architecture	24
2 Modèle multidimensionnel.....	26
2.1 Base de données Oracle	26
2.2 Modélisation des Datamarts.....	28

2.3	Modélisation du Datawarehouse	32
Chapitre 4 : Etude technique et Réalisation		36
1	Outils et technologies	37
1.1	Suites BI.....	37
1.2	Serveur de gestion de base de données	38
1.3	Outil ETL (SSIS)	38
1.4	Outil de modélisation sémantique.....	39
1.5	Outil de Reporting.....	41
2	Réalisation du système décisionnel.....	44
2.1	Processus d'intégration des données.....	44
2.1.1	Création du package SSIS	45
2.1.2	Création des Jobs	48
2.2	Ecriture de la requête SQL.....	49
2.3	Conception du modèle tabulaire	52
2.3.1	Création des mesures (DAX).....	52
2.3.2	Implémentation du modèle	53
2.4	Réalisation des tableaux de bord.....	55
2.4.1	Connexion avec la source de données	55
2.4.2	Présentation des tableaux de bord	56
2.4.3	Test et publication sur Power BI Service	62
3	Intégration des Dashboards dans la solution ANIA	63
3.1	Technologies	63
3.2	Interface graphique	64
Conclusion.....		66
Références		67
Annexes		68

Liste des figures

Figure 1. 1: L’organigramme hiérarchique d’InsurBT.....	3
Figure 1. 2: Les composants de la suite ANIA	6
Figure 1. 3: Roadmap cible ANIA	6
Figure 1. 4: Workflow de la méthodologie SCRUM	8
Figure 1. 5: Diagramme GANTT du projet.....	10
Figure 2. 1: Catégorisation des Dashboards	12
Figure 2. 2: Diagramme de cas d'utilisation de la solution	15
Figure 3. 1: Architecture BI adoptée	25
Figure 3. 2: Modèle de la base de données	27
Figure 3. 3: Répartition des Datamarts.....	29
Figure 3. 4: Les modèles de données créés pour le Datamart Contrats	31
Figure 3. 5: Modèle en étoile	30
Figure 3. 6: Modèle en flocon de neige.....	30
Figure 3. 7: Modèle final du Data Warehouse	33
Figure 4. 1 : Magic Quadrant des outils d'intégration de données 2021[6]	39
Figure 4. 2 : le Magic Quadrant des outils de reporting [6].	41
Figure 4. 3 : Schéma récapitulatif de la suite MSBI	44
Figure 4. 4 : Gestionnaires de connexions SSIS	45
Figure 4. 5 : Flux de contrôle (Data Mart Contrat)	46
Figure 4. 6 : Flux de données SSIS	47
Figure 4. 7 : Package d'alimentation du Datawarehouse.....	48
Figure 4. 8 : Déploiement du Package d'alimentation du Datawarehouse	48
Figure 4. 9 : Configuration du job.....	49

Figure 4. 10 : Ecriture algébrique pour l'état "Récapitulatif des dépenses par tranches d'âge"	50
Figure 4. 11 : Requête SQL pour l'état "Récapitulatif des dépenses par tranches d'âge"	51
Figure 4. 12 : Requête DAX pour la création de la mesure Ratio combiné	52
Figure 4. 13 : Requête DAX pour la création de la table du temps	53
Figure 4. 14 : modèle tabulaire SSAS	54
Figure 4. 15 : Connexion de données PowerBI Desktop	55
Figure 4. 16 : Dashboard de l'état "Récapitulatif des dépenses par tranches d'âge"	57
Figure 4. 17 : Premier état de synthèse	59
Figure 4. 18 : Deuxième état de synthèse	61
Figure 4. 19 : Liste des tableaux de bord sur PowerBI Service	63
Figure 4. 20 : Intégration de l'état « comparatif des dépenses des années N et N-1 » sur la solution web	65

Liste des tableaux

Tableau 1. 1: Les catégories des Assurances	4
Tableau 2. 1 : Descriptif du KPI (Ratio combiné).....	16
Tableau 2. 2 : Descriptif du KPI (Prime)	17
Tableau 2. 3 : Descriptif du KPI (Cout moyen)	17
Tableau 2. 4 : Descriptif du KPI (Rendement).....	18
Tableau 2. 5 : Descriptif du KPI (Taux de couverture).....	19
Tableau 2. 6 : Descriptif des autres KPIs	20
Tableau 3. 1:Comparatif des approches Kimball et Inmon [8].....	24
Tableau 3. 2: Descriptif des attributs de la table Personne	34
Tableau 4. 1 : Comparatif des suites BI MSBI et OBIEE.....	37
Tableau 4. 2 : Comparatif des deux types de modèles SSAS	40
Tableau 4. 3 : Comparatif des outils de visualisation [4]	42

Glossaire

BI: Business Intelligence

CNSS: La Caisse Nationale de Sécurité Sociale

DW: Data Warehouse

ESN: Entreprise de services du numérique

ETL: Extract, Transform, Load

KPI: Key Performance Indicator

OBIEE : Oracle Business Intelligence Enterprise Edition

OLAP: Online Analytical Processing

SSAS: SQL Server Analysis Services

SSIS: SQL Server Integration Services

SSMS: SQL Server Management Studio

Introduction Générale

Dans un secteur aussi concurrentiel, les compagnies d'assurance ont besoin de meilleures informations pour pouvoir obtenir une vision plus globale des performances de l'entreprise. Afin d'augmenter l'efficacité et la rentabilité, il est essentiel que tous les membres de l'entreprise prennent les décisions optimales et, pour ce faire, le défi ne réside pas dans les données elles-mêmes, mais dans la façon selon laquelle elles sont présentées.

La BI est utilisée par les compagnies d'assurance pour créer des rapports précis, notamment des rapports financiers qui suivent les dépenses internes, les sinistres, les revenus générés par les primes et les rapports opérationnels qui suivent la productivité et le flux de travail des travailleurs afin de faciliter la prise de décision en affichant visuellement des données interactives et compréhensibles.

Ainsi, la mise en place d'une solution décisionnelle est indispensable pour la gestion des assurances. En effet, à partir de l'analyse des données, l'informatique décisionnelle fournit aux dirigeants de quoi prendre des décisions éclairées. Pour l'assureur, la business intelligence permet d'évaluer la rentabilité, détecter et lutter contre les fraudes en plus de pouvoir analyser les risques.

Dans ce cadre-là, notre projet de fin d'études s'inscrit ; ce dernier a pour objectif de réaliser une solution décisionnelle à l'aide des outils BI permettant de traiter les différents aspects humains et financiers d'une compagnie d'assurance.

Ce rapport présente les étapes suivies afin de réaliser notre projet ; pour ce faire nous avons commencé, tout d'abord, par traiter son cadre général notamment la problématique et les objectifs ainsi que la méthodologie adoptée. Le deuxième chapitre s'intéresse aux spécifications et analyses des besoins, et dans le même chapitre on a choisi de présenter les tableaux de bord à réaliser ainsi que les KPIs à traiter. Ensuite, on a passé à la conception du modèle de notre Data Warehouse en suivant une approche bien déterminée pour, finalement, aborder le dernier chapitre présentant les choix des outils et des technologies à utiliser, ainsi que les étapes de la réalisation de la solution à l'aide d'un exemple illustratif.

Chapitre 1 : Contexte général du projet

Dans ce chapitre, nous allons aborder le cadre général de notre projet en commençant par une présentation de l'organisme dans lequel on a effectué notre stage et son domaine d'activité, pour passer par la suite à détailler le contexte de notre projet ainsi que sa problématique et ses objectifs et finir par préciser la méthodologie en plus du planning adoptés afin de réaliser notre solution.

1 Présentation de l'organisme d'accueil

Cette partie est consacrée à présenter l'organisme d'accueil, son hiérarchie et son domaine d'activité qui est le conseil des assurances.

1.1 Présentation d'InsurBT

InsurBT est une ESN panafricaine spécialisée dans le conseil et les technologies du monde des assurances, elle vise principalement les Compagnies d'assurance & de Réassurance, les Caisses de prévoyance sociale ainsi que les Caisses pour l'assurance maladie.

InsurBT offre à ses clients une valorisation de leurs projets stratégiques par une utilisation optimale des meilleures pratiques du Lean et des nouvelles technologies.

InsurBT est active sur le marché Africain, et compte parmi ses partenaires CNSS Maroc (8 Millions de personnes protégées) et SANLAM : premier groupe panafricain d'assurance.

1.2 Présentation de la hiérarchie

InsurBT adopte une structure hiérarchique qui repose sur l'unicité du commandement afin d'assurer le fonctionnement optimal de l'organisme.

Selon la figure 1.1, On a représenté cette hiérarchie par un organigramme qui montre la répartition des domaines d'activité et de supervision des différents agents et départements de l'entreprise.

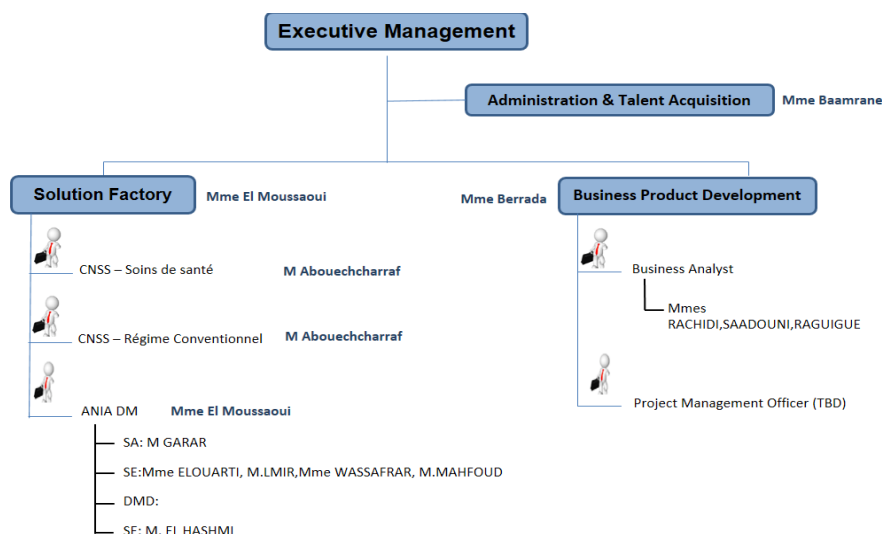


Figure 1. 1: L'organigramme hiérarchique d'InsurBT

2 Conseil et technologies des assurances

L'assurance est un secteur qui répond à un besoin impérieux des individus, il permet de se prémunir contre la survenance de certain événement affectant leur personne ou leur bien, il contribue à la sécurité de l'homme et ces activités. Ce qui oblige les compagnies d'assurance d'introduire des nouvelles méthodes et des outils modernes de gestion afin de pouvoir répondre aux exigences du marché et d'intégrer dans les objectifs et les priorités de développement.

Le marché de l'assurance est divisé en deux catégories principales : l'assurance de dommage et l'assurance de personne. Cette distinction nous l'avons récapitulée dans le tableau ci-dessous :

Tableau 1. 1: Les catégories des Assurances

Assurance non vie			Assurance vie
Assurance de biens (appartenant à l'assuré)	Assurance responsabilité (de l'assuré envers les tiers)	Assurance santé (accidents, maladie, invalidité, incapacité, frais médicaux)	Assurance vie (vie, décès, épargne et retraite)
Assurance de dommages			Assurance de personnes

La société InsurBT s'intéresse aux technologies des assurances afin d'améliorer leur métier et automatiser leurs processus. Dans le cadre de notre projet, nous nous focaliserons sur la partie assurance santé. L'opération d'assurance santé fait appel à plusieurs notions à savoir :

Police d'assurance : Document matérialisant le contrat d'assurance. Il indique les conditions générales et particulières.

Prime : Somme due par le souscripteur d'un contrat d'assurance en contrepartie des garanties accordées par l'assureur. La prime d'assurance peut être périodique (annuelle, semestrielle, trimestrielle ou mensuelle) ou unique (versement effectué une seule fois à la souscription du contrat).

Sinistre : C'est le risque qui se réalise : l'incendie, le vol, l'accident.... C'est pour vous prémunir de leurs conséquences que vous souscrivez une assurance.

Prestation : L'engagement de l'assureur vis-à-vis de l'assuré.

Bénéficiaire : Personne physique ou morale au profit de laquelle l'assurance a été contractée. C'est elle qui bénéficiera du versement de la prestation au moment de la matérialisation du sinistre.

Produit : Un produit d'assurance est un produit commercialisé par un assureur en agence ou en ligne à destination des personnes physiques ou des personnes morales dans le but de protéger et de garantir financièrement et juridiquement contre des risques inhérents à l'utilisation d'un bien ou d'un service.

Affection : représente la maladie pour laquelle la prestation est remboursée.

3 Cadrage du projet

Dans cette partie, nous allons examiner le contexte, la problématique et les objectifs du projet. Tous ces éléments vont sur le cadrage du projet.

3.1 Contexte du projet

ANIA est une suite progicielle full web et Cloud Native développée avec les dernières technologies de pointe. C'est une suite complètement intégrée et modulable qui offre :

- Un déploiement adapté au besoin de la compagnie grâce au fonctionnement des modules en stand alone ;
- Un atelier produit puissant (paramétrabilité, time to market, etc.) ;
- Un interfaçage intuitif ;
- Un déploiement à chaud ;
- Une vue 360 de l'activité ;
- Un haut niveau de paramétrage grâce, entre autres à un moteur de calcul, un flux, de Workflow et un archivage électronique des documents.

La figure 1.2 représente les différents composants de la suite progicielle ANIA :



Figure 1. 2: Les composants de la suite ANIA

Parmi les différentes parties de la suite progicielle ANIA, nous nous focaliserons sur la partie Reporting & BI qui est au cœur de notre projet.

La cible de la suite ANIA est de couvrir toutes les lignes métier du monde de l'assurance et inclure toutes les briques applicatives pour compléter les processus de bout en bout.

La roadmap cible ANIA est représentée sur la figure 1.3 suivante :

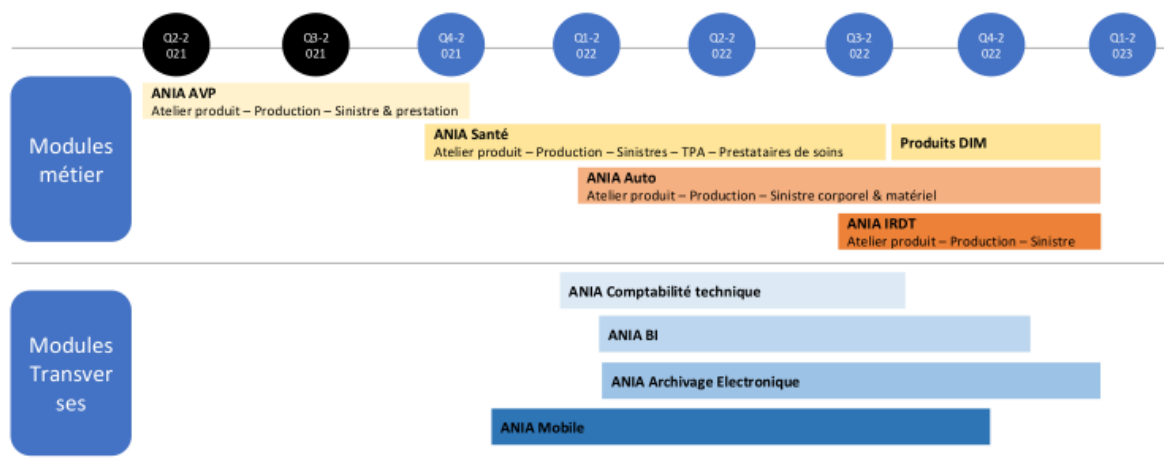


Figure 1. 3: Roadmap cible ANIA

On a commencé le développement des états de Reporting ANIA santé au début du premier trimestre de cette année, conformément au plan stratégique décrivant le développement probable de la solution ANIA.

3.2 Problématique

À l'heure du tout numérique, les entreprises collectent une quantité astronomique de données sur leurs clients, leur marché ou même en interne. Ces précieuses informations permettent à ceux qui sont capables de les analyser de mieux connaître leur marché et leur potentiel de croissance. L'assurance fournit un bon exemple de données exploitées pour aider à prendre des décisions commerciales bénéfiques, afin d'assurer la rentabilité tout en traitant équitablement ses clients en ce qui concerne les sinistres et les primes. Chaque compagnie d'assurance devra utiliser la richesse des données qu'elle génère quotidiennement pour optimiser leurs processus commerciaux, développer des produits plus intelligents et mieux satisfaire leurs clients, c'est bien grâce à la Business Intelligence qu'elle pourra exploiter la puissance de ces données.

Afin de combler ce besoin, InsurBT, organisme spécialisée dans les technologies et le conseil des assurances, a lancé le projet ANIA pour proposer une solution décisionnelle qui fournit un aperçu complet sur l'activité d'assurance, ceci afin que cette dernière soit en mesure de répondre aux interrogations suivantes : Quel est le nombre de sinistres reçues au cours d'un mois/d'une année ainsi que le montant ? Quelles sont les primes facturées vues des polices existantes ? Ce produit représente-t-il un gain ou une perte pour l'assurance ?

Cette solution va aider les assurances à prendre des meilleures décisions en affichant les données actuelles et historiques sur leur contexte commercial, grâce à des tableaux de bord présentant de nouvelles informations qui seront la base de prise des décisions.

3.3 Objectifs

Les tableaux de bord BI sont de plus en plus considérés comme des outils essentiels pour les entreprises, leur permettant de tirer des informations instantanées et à long terme de leurs stocks croissants de données et de prendre des décisions et d'analyser des données en meilleure connaissance de cause.

Notre PFE a pour objectif de mettre en place une solution décisionnelle pour le développement des états de Reporting BI de ANIA assurance santé, ces derniers vont être intégrés sur la suite progiciel ANIA afin de :

- Lutter contre les fraudes ;
- Renforcer l'efficacité de la gestion des sinistres ;

- Identifier les opportunités rentables ;
- Mieux activer l'équipe de vente ;
- Optimiser les processus de souscription et de vente ;
- Découvrir de nouvelles opportunités à l'aide de sources de données externes.

4 Méthodologie SCRUM

InsurBT applique une méthode agile pour assurer la bonne gestion de ses projets. L'ensemble de l'équipe InsurBT est répartie en multiple squads composés d'un scrum master, un business analyst et une équipe.

- Les sprints sont planifiés en fonction des priorités définies en amont. Il s'agit souvent de sprints de deux à trois semaines. Lors du sprint planning, les tâches ou user stories sont ensuite réparties entre l'équipe, et estimées en heures de travail ;
- Un stand-up meeting de 15 minutes est fait quotidiennement tous les matins pour faire le point sur la progression journalière du sprint ;
- Chacun doit gérer les tâches qui lui ont été affectées et ce en actualisant le statut de chaque tâche et en mettant à jour le suivi temporel quotidiennement ;
- Chaque tâche doit respecter le workflow établi ci-dessous sur la figure 1.4 :

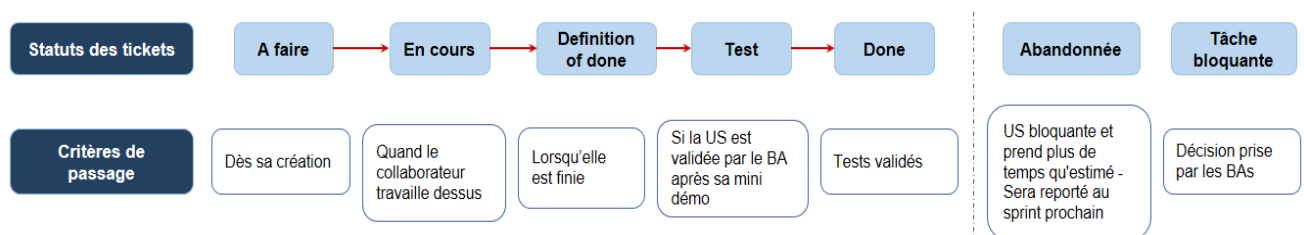


Figure 1. 4: Workflow de la méthodologie SCRUM

L'objectif phare de cette méthode est d'améliorer la productivité des équipes, tout en permettant une optimisation du produit grâce à des feedbacks réguliers du marché.

En parallèle, la méthode SCRUM permet d'avoir une vue d'ensemble du projet pour chacune des parties prenantes et elle permet aussi la réduction des bugs et une mise à jour régulière des priorités.

Dans SCRUM, une meilleure qualité de produit est garantie en vérifiant auprès des parties prenantes l'évolution des exigences lors des révisions de sprint, ainsi on aura la possibilité de pivoter et de changer de direction beaucoup plus tôt et beaucoup plus rapidement que dans un projet en cascade [3].

De plus, les projets sont plus contrôlés parce qu'ils sont chronométrés et que chacun sait quelle tâche il doit accomplir. Les équipes communiquent et collaborent quotidiennement afin de réduire les risques.

Ainsi Scrum a été conçu non seulement pour améliorer les projets mais aussi pour en accélérer la réalisation ce qui est un défi réel pour tout projet BI, en fait, le risque d'un projet de Business Intelligence conduit avec une méthode de type cycle en V est de produire tardivement des livrables qui ne sont plus en phase avec les besoins des utilisateurs.

Parmi les causes d'échecs fréquemment citées dans les projets Business Intelligence on retrouve un manque d'alignement avec la stratégie de l'entreprise et un manque d'adéquation avec les besoins des utilisateurs.

Les projets de Business Intelligence vont donc pouvoir tirer pleinement d'une approche de gestion de projet agile. Cette approche va permettre de s'adapter à l'évolution des exigences des utilisateurs et donc de développer leur engagement vis-à-vis du projet avec des livraisons opportunes et plus fréquentes.

Les sprints sont représentés par les phases de l'architecture qui va être représenté par la suite.

5 Planning du projet (Diagramme de Gantt)

Notre planning comporte quatre parties essentielles.

- Préparation : Etude de besoin et documentation sur les différents outils.
- Conception : création du modèle du datawarehouse.
- Réalisation : chaque tableau de bord suit le même processus de réalisation, à savoir :

- Choix des indicateurs ;
 - Préparation des données ;
 - Réalisation du tableau de bord ;
 - Tests fonctionnels ;
 - Tests de régressions et partage sur PowerBI Service.
- Intégration des tableaux de bord sur la solution web ANIA.

La figure 1.5 représente la distribution des taches faite lors de notre projet de fin d'étude (la partie réalisation ne contient que les deux premiers tableaux de bord réalisés, le même processus est suivi pour les autres Dashboards).

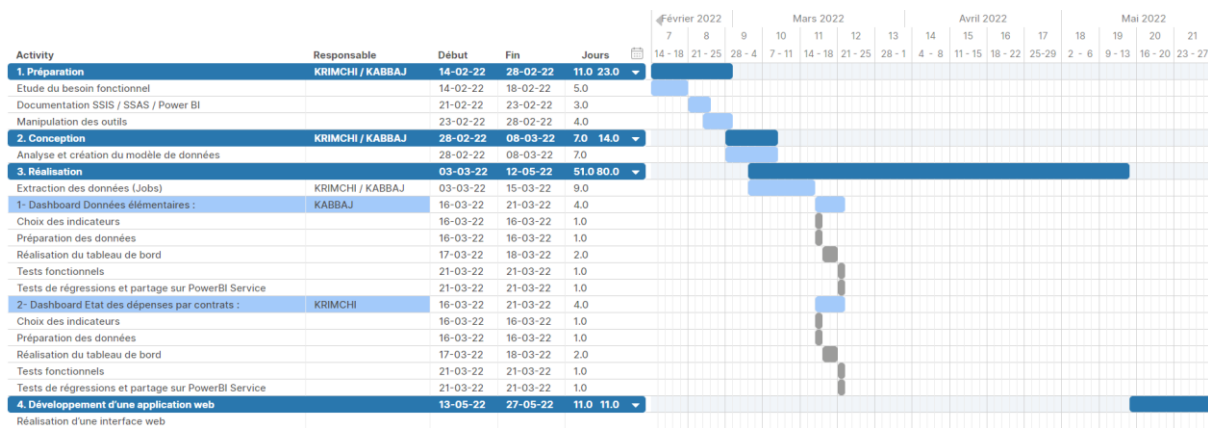


Figure 1. 5: Diagramme GANTT du projet

Chapitre 2 : Spécification et Analyse des besoins

Ce deuxième chapitre offre une vue sur les besoins fonctionnels (décisionnels et applicatifs) et non fonctionnels de notre projet ainsi que l'analyse des indicateurs clés de performance.

1 Etude des besoins

Cette partie est dédiée à l'analyse du besoin, pour ce faire, nous allons présenter en premier lieu les besoins fonctionnels (décisionnels, applicatifs) de notre solution décisionnelle, ensuite nous allons citer et expliquer les besoins non fonctionnels caractérisant le système.

1.1 Besoins fonctionnels

Pour notre cas, les besoins fonctionnels sont divisés en deux parties : des besoins décisionnels exprimant les attentes du client, en plus des besoins applicatifs concernant les fonctionnalités du système et qui seront détaillés à l'aide d'un diagramme de cas d'utilisation.

1.1.1 Besoins décisionnels

Le tableau de bord est un outil d'aide à la gestion composé d'un ensemble de rapports dont le but est de fournir une vision claire de la situation actuelle de son système par rapport aux objectifs fixés. C'est en théorie un ensemble d'indicateurs clés (KPIs) qui jouent un rôle d'alarme et se déclenchent automatiquement pour signaler une anomalie ou une performance dans le fonctionnement du système.

Pour notre besoin décisionnel, on a opté pour une catégorisation des différents tableaux de bord à réaliser où chaque tableau de bord appartient à une catégorie bien précise selon la figure 2.1 :

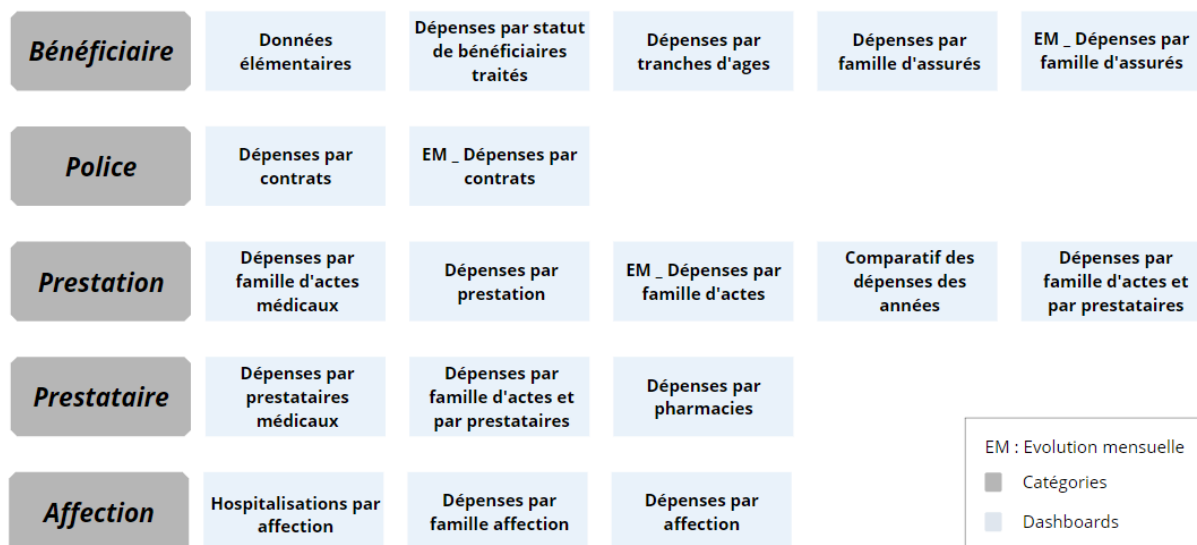


Figure 2. 1: Catégorisation des Dashboards

Nous détaillons chaque catégorie comme suit :

➤ Bénéficiaire

Cette catégorie regroupe les tableaux de bord qui s'intéressent à une analyse selon la vue des bénéficiaires afin de visualiser les informations suivantes :

- Les sinistres par tranches d'âge et par lien de parenté.
- Nombre d'assurés et de bénéficiaires traités.
- La population active et la population traitée par familles d'assuré.
- Evolution mensuelle des dépenses par familles d'assuré.

➤ Police

En se focalisant sur les contrats des clients, ce regroupement représente :

- Les montants remboursés pour chaque contrat dans une période sélectionnée.
- L'évolution mensuelle par année des différents contrats des produits.

➤ Prestation

Résumant la distribution des dépenses pour un contrat et une période déterminée en point de vue familles de prestation et prestations (familles d'acte médical et actes médicaux). Les tableaux de bord de cette catégorie s'intéressent à déterminer :

- Un comparatif des dépenses des années par familles de prestation dans une période considérée.
- Evolution mensuelle des sinistres par familles d'acte.
- Pour chaque prestataire médical, les sinistres remboursés par familles de prestation.

➤ Prestataire

Ces tableaux de bord traitent des données du point de vue prestataires (clinique, cabinet, pharmacie, ...) en visualisant la distribution des dépenses par contrat et par période, à savoir :

- Pour chaque prestataire médical, il s'agit d'afficher les montants des sinistres remboursés dans la période considérée et pour le contrat sélectionné.
- Pour chaque famille de prestation et pour chaque prestataire, il s'agit d'afficher le montant total des sinistres remboursés dans la période considérée et le contrat sélectionné.

- Pour chaque pharmacie, il s'agit de déterminer le montant global des sinistres remboursés dans la période considérée et pour le contrat sélectionné.

➤ Affection

Présentant la distribution des dépenses pour un contrat et une période déterminée en point de vue familles d'affection et affections, pour ce faire, on visualise :

- Pour chaque hospitalisation, le montant global des sinistres remboursés relatifs aux codes d'affection dans la période considérée et pour le contrat sélectionné.
- Pour chaque famille d'affection, le montant global des sinistres remboursés dans la période considérée et pour le contrat sélectionné.
- Pour chaque affection, le montant global de sinistres remboursés dans la période considérée et pour le contrat sélectionné.

Pour finir, on a choisi d'ajouter des tableaux de bord de synthèse n'appartenant à aucune catégorie car ils viennent pour synthétiser et rassembler plusieurs axes afin d'élargir l'espace d'analyse.

1.1.2 Besoins applicatifs

En nous appuyant sur le cahier de charge défini par le client, notre besoin fonctionnel exige un système offrant la possibilité de choisir parmi les tableaux de bord celui qui répond le plus à notre besoin en visualisant les KPIs, en plus de donner la main au responsable pour filtrer sur de différents critères et donc pouvoir tirer plus d'analyses, ainsi il peut générer les rapports fournis sous format PDF.

Dans ce sens, notre besoin fonctionnel est représenté à l'aide du diagramme de cas d'utilisation système sur la figure 2.2 suivante :

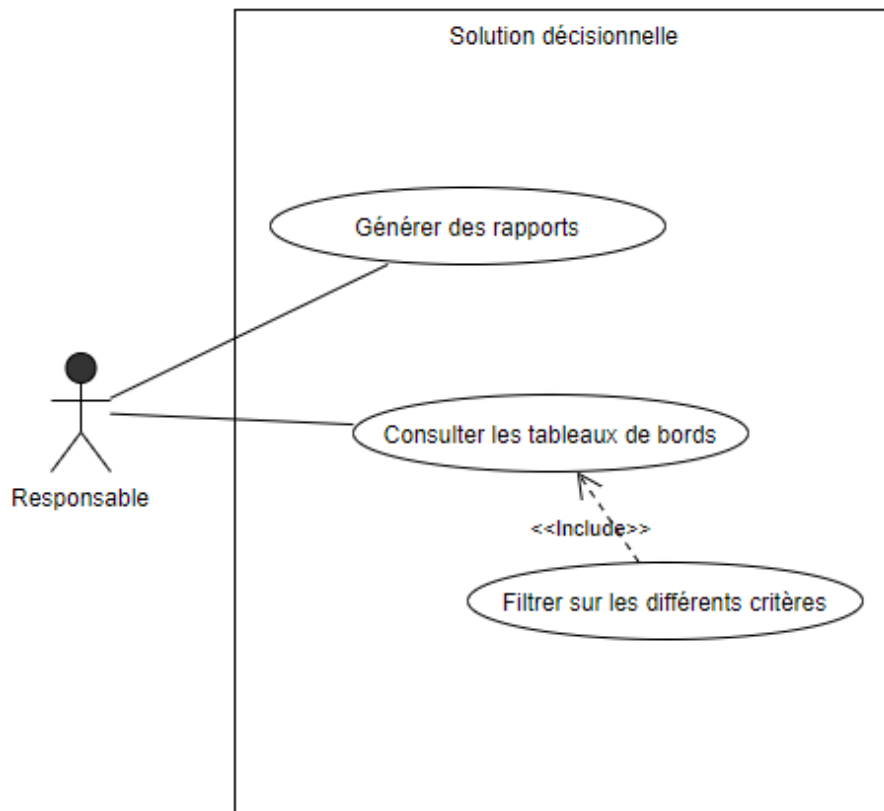


Figure 2. 2: Diagramme de cas d'utilisation de la solution

1.2 Besoins non fonctionnels

En ce qui concerne les besoins non fonctionnels du système on peut les résumer comme suit :

- **Qualité de l'information** : L'information présentée au niveau du tableau de bord doit être fiable vu qu'elle sera la base de la prise de décision.
- **L'ergonomie** : la densité des éléments sur les écrans, la disposition et les couleurs doivent être conformes aux règles ergonomiques.
- **Maintenabilité** : Facilité d'analyse, facilité de modification, stabilité et testabilité.
- **Portabilité** : Le déploiement doit être effectué facilement et rapidement.
- **Facilité d'utilisation** : Facilité de compréhension et d'exploitation ou d'utilisation.
- **Disponibilité** : Le système doit être disponible à tout moment et en tout emplacement.

2 Détermination des indicateurs

En nous basant sur les informations fournies auprès du métier, les différents ateliers tenus avec le manager ainsi que les informations exprimées dans le cahier de charge, nous avons pu déterminer les KPIs suivants : Ratio combiné, Revenue par période, Cout moyen, Rendement, Taux de couverture...

Ces indicateurs ont été choisis afin de répondre au mieux aux différents besoins stratégiques de l'assurance.

➤ Ratio combiné

Tableau 2. 1 : Descriptif du KPI (Ratio combiné)

Signification	Table source	Formule de calcul
Il s'agit du rapport entre le montant des sinistres et celui des cotisations encaissées sur un même contrat d'assurance. Pour connaître le pourcentage de leurs dépenses par rapport à leurs recettes, les assureurs calculent leur ratio de sinistres/primes (ou ratio combiné).	QTC_QUITT SIN_GAR_PREST	Montant des sinistres / Montant des cotisations Montant des cotisations = Prime nette + Taxe des primes nette + Taxe des accessoires Attributs : MT_REMB, PRIM_TOT (PRIM_TOT = PRIM_NET + TAX_PRIM_NET + TAX_ACCESSOIR + ACCESSOIR)

Ce ratio permet à l'assureur de mesurer la rentabilité de l'assurance délivrée sur une période donnée. Le ratio sinistres / primes doit rester inférieur à 1 pour que la compagnie reste viable. Ce ratio peut être calculé globalement, ou pour chaque assuré, en fonction des risques qu'il représente.

➤ Primes

Tableau 2. 2 : Descriptif du KPI (Prime)

Signification	Table source	Formule de calcul
Le revenu par titulaire de police est un indicateur de rendement clé (IRC) simple qui mesure le montant des revenus générés par la compagnie d'assurance, par titulaire de police desservi.	QTC_QUITT	<p>Dans la période considérée, il s'agit ici d'afficher la cotisation des bénéficiaires.</p> <p>Cette cotisation est une régularisation prenant en compte les quittances émises relatives au contrat sélectionné</p> <p>Attributs : PRIM_TOT</p>

Une valeur faible de ce KPI pourrait être due à plusieurs facteurs : une mauvaise agence, un service à la clientèle inférieur à la moyenne (ce qui conduit à un faible taux de rétention) ou un manque de bonnes pratiques d'investissement pourraient tous contribuer à une génération de revenus inférieure à la moyenne [1].

➤ Cout moyen

Tableau 2. 3 : Descriptif du KPI (Cout moyen)

Signification	Table source	Formule de calcul
L'indicateur de performance clé Coût moyen par sinistre mesure le montant que l'assurance paie pour chaque police. Avec ce KPI (comme avec d'autres KPIs d'assurance), il est important de catégoriser en fonction du produit, car pour chaque produit il y'aura une différence en coût.	SIN_GAR_PREST SIN_SANTE	<p>Montant des sinistres / Nombre des sinistres relative à un contrat</p> <p>C'est la moyenne des sinistres remboursés du contrat sélectionné dans la période considérée</p> <p>Attributs : MT_REMB, ID_SIN_SANTE</p>

L'objectif de cet indicateur est d'aider l'assurance à évaluer correctement le risque associé à chaque police et à ajuster les prix de la police en conséquence [1].

➤ Rendement

Tableau 2. 4 : Descriptif du KPI (Rendement)

Signification	Table source	Formule de calcul
Le rendement de l'excédent des assurés montre le montant des bénéfices qu'une compagnie d'assurance peut générer par rapport au montant des revenus qu'elle génère de la souscription de polices d'assurance et du produit d'investissement, l'excédent du preneur d'assurance représentant le montant des actifs d'un assureur dépassant ses passifs.	QTC_QUITT SIN_GAR_PREST	Prime totale – Montant des sinistres / Prime totale Attributs : MT_REMB, PRIM_TOT

Le rendement moyen de l'excédent dans toutes les branches d'assurance est de 8,8 %. Ce KPI surveille la performance bénéficiaire d'une assurance [1].

➤ Taux de couverture

Tableau 2. 5 : Descriptif du KPI (Taux de couverture)

Signification	Table source	Formule de calcul
Le taux de couverture est la proportion de la population cible qui participe au produit d'une assurance. Chaque produit cible un groupe de personnes à couvrir. Le taux de couverture correspond à un certain moment à la proportion des personnes qui sont activement couvertes par l'assurance.	SIN_SANTE POL_OBJ_ASSUR_PERSON	Nombre d'assuré active / Population actives Nombre d'assuré active : Nombre d'assurés (Lien de parenté = Lui-même) inscrit sur un contrat. Populations actives : Nombre de bénéficiaires inscrits sur un contrat. Attributs: POL_OBJ_ASSUR_PERSON_ID, REF_LIEN_PAR.DESIGN

Le taux de couverture peut être considéré comme un indicateur clé de l'efficacité de la commercialisation [2].

➤ Autre KPIs

Les indicateurs suivants permettent aussi de répondre aux besoins cités auparavant et interviennent sur l'axe d'analyse : Taux de bénéficiaires traités, Taux de bénéficiaires, Nombre d'actes par familles de prestation, Nombre d'actes par prestations, Nombre d'actes par prestataires médicaux, Nombre d'actes par affections, Nombre d'actes par familles d'affection.

Tableau 2. 6 : Descriptif des autres KPIs

Indicateur	Table source	Formule de calcul
Taux de bénéficiaires traités	SIN_SANTE POL_OBJ_ASSUR_PERSON	Population active / Population traités Populations actives : Nombre de bénéficiaires inscrits sur un contrat. Populations traitées : Nombre de bénéficiaires inscrits sur un contrat et qui ont déjà un sinistre. Attributs: POL_OBJ_ASSUR_PERSON_ID, FK_POL_OBJ_ASSUR_PERSON
Taux de bénéficiaires	POL_OBJ_ASSUR_PERSON	Le pourcentage du nombre des bénéficiaires du contrat sélectionné dans la période considérée. Attributs: POL_OBJ_ASSUR_PERSON_ID
Nombre d'actes par famille de prestation	SIN_SANTE REF_FAMILLE_PRESTA_SANT	Nombre d'actes remboursés par familles de prestation dans la période considérée et pour le contrat sélectionné. Attributs : ID_FAMILLE_PRESTA
Nombre d'actes par prestation	SIN_SANTE REF_PRESTA_SANT	Pour chaque famille de prestation, il s'agit du nombre de prestations remboursées pour le contrat sélectionné et la période considérée. Attributs : ID_PRESTA

Nombre d'actes par prestataires médicaux	SIN_SANTE TRS_ETABLIS_SOIN SIN_GAR_PREST	Nombre de sinistres remboursés dans la période considérée et pour le contrat sélectionné pour chaque type de prestataire médical. Attributs : FK_PS, FK_ETABLIS_SOIN, ID_TRS_ETABLIS_SOIN
Nombre d'actes par famille d'affection	SIN_SANTE REF_MALADIE	Pour chaque famille d'affection, il s'agit du nombre de sinistres remboursés dans la période considérée et pour le contrat sélectionné. Attributs : ID_FAMILLE_MALADIE
Nombre d'actes par affection	SIN_SANTE REF_FAMILLE_MALADIE	Pour chaque affection, il s'agit du nombre de sinistres remboursés dans la période considérée et pour le contrat sélectionné. Attributs : ID_MALADIE

Chapitre 3 : Analyse et Conception

Dans ce chapitre, nous allons migrer vers le volet d'analyse et conception dont lequel, nous allons concevoir le modèle de notre Datawarehouse à l'aide de la création des modèles des Datamarts selon une approche et une architecture bien choisie.

1 Architecture

Cette section va faire l'objet d'une étude comparative entre les deux approches BI Kimball et Inmon afin de choisir l'une parmi eux et définir l'architecture à suivre pour réaliser notre solution décisionnelle.

1.1 Comparatif des approches BI

La conception d'un entrepôt de données est une partie essentielle du développement d'une entreprise. Pour la conception, il existe deux architectures les plus courantes nommées Kimball et Inmon.

➤ Inmon

Dans son approche, il faut d'abord créer l'entrepôt qui comprend l'ensemble des données de l'entreprise, pour ensuite alimenter de mini-entrepôts de données départementales. Chaque mini-entrepôt vise un secteur particulier de l'entreprise et sera la source des requêtes et rapports utilisés par les décideurs de l'entreprise.

Ce modèle sert à la création d'un modèle logique détaillé qui est utilisé pour les opérations majeures, il est ensuite utilisé pour développer un modèle physique. Ce modèle est normalisé et réduit la redondance des données. Il s'agit d'un modèle complexe difficilement utilisable à des fins commerciales pour lesquelles des Datamarts sont créés et chaque département est capable de l'utiliser à ses fins [8].

➤ Kimball

Ralph Kimball prend une approche tout à fait différente d'Inmon pour la création d'entrepôts. Premièrement, il a développé une technique pour modéliser un entrepôt selon un modèle dimensionnel. Au lieu d'être basé sur le modèle entités-association, le modèle dimensionnel repose sur des tables représentant des faits et des dimensions. Cette approche commence par la reconnaissance des processus métier et des questions auxquelles le Data Warehouse doit répondre. Ces ensembles d'informations sont analysés puis bien documentés. Le processus Extract Transform Load (ETL) rassemble toutes les données de plusieurs sources de données appelées Datamarts, puis les charge dans une zone commune. Ensuite, cela est transformé en cube OLAP ou modèle tabulaire [8].

Le tableau 3.1 représente un comparatif des deux approches BI :

Tableau 3. 1:Comparatif des approches Kimball et Inmon [8]

Kimball	Inmon
Une approche ascendante pour la mise en œuvre.	Une approche descendante pour la mise en œuvre.
Commence par la conception du modèle dimensionnel pour les datamarts	Commence par la conception du modèle du DW
Efficace et prend moins de temps.	Complexe et prend beaucoup de temps.
Il comporte des étapes itératives et rentables.	Le coût initial est énorme et le coût de développement est faible.

En nous basant sur les besoins de notre solution décisionnelle exigeant la réalisation d'un seul modèle final pour le DataWarehouse, nous avons opté pour l'approche Kimball.

1.2 Descriptif de l'architecture

Avant de passer aux détails des phases de notre architecture, commençons tout d'abord par la définition de quelques notions à savoir :

- ETL (Extract, Transform, Load) : est un processus d'intégration de données utilisé pour combiner des données provenant de sources multiples. Dans un processus ETL, les données sont extraites de la source, transformées pour être agrégées ou pour mettre en œuvre des règles de gestion, puis chargées dans le système cible. Ces données sont utilisées pour l'analyse. Le processus ETL est utilisé pour construire un entrepôt de données.
- Data Warehouse (ou entrepôt de données) : est un processus de gestion de grandes quantités de données dans une organisation. Il est conçu pour faciliter les tâches de BI, en particulier l'analyse. Compte tenu d'une grande quantité de données historiques, un entrepôt de données permet une interrogation et une analyse plus rapides des données.

La figure 3.1 représente l'architecture de notre solution décisionnelle basée sur l'approche Kimball :

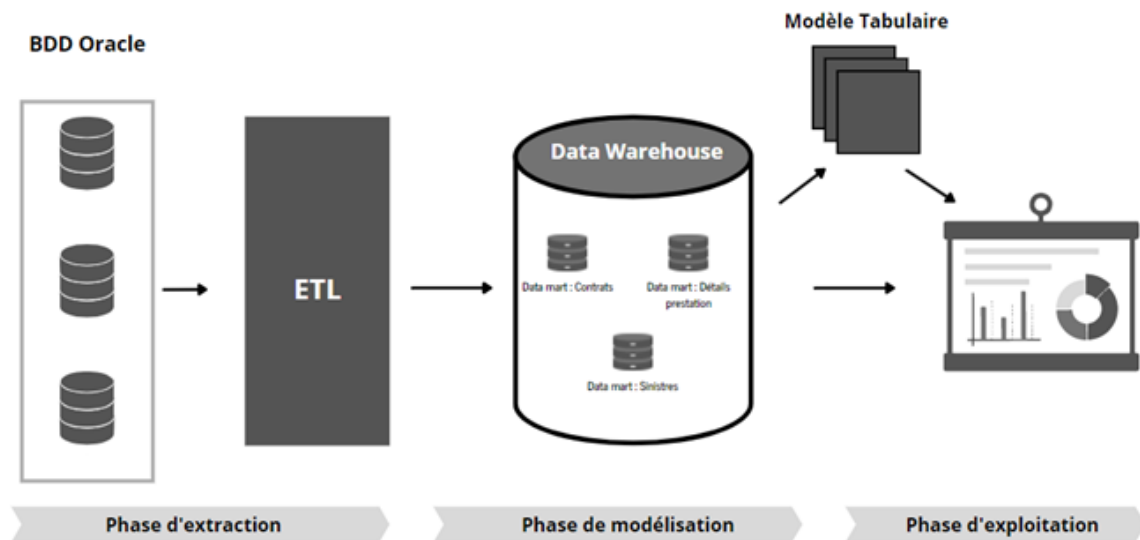


Figure 3. 1: Architecture BI adoptée

Dans ce qui suit, nous allons présenter les différentes phases de la transformation des données en informations exploitables.

- Phase d'extraction des données :

Pour notre solution, la collecte des données va être faite à partir d'une source de base de données oracle via les outils d'ETL qui permettent d'automatiser les traitements à savoir le nettoyage et la consolidation puis de gérer les flux de données alimentant les bases de stockage (Datawarehouse ou Datamart).

- Phase de modélisation :

Dans cette étape, les données, au préalable nettoyées et consolidées, seront stockées dans le Datawarehouse en créant les modèles des Datamarts selon la distribution logique de la source de données en plus, on procède pour la structuration à travers des requêtes effectuant des transformations adéquates aux besoins de la solution afin de créer des vues construisant le modèle final de notre Datawarehouse.

- Phase d'exploitation :

Les utilisateurs finaux interviennent dans cette dernière étape pour exploiter et analyser les données qui leur sont fournies. Selon les besoins, différents types d'outils d'exploitation sont disponibles tels que les outils OLAP et les modèles tabulaires pour les analyses multidimensionnelles et les outils de Reporting pour communiquer sur la performance.

2 Modèle multidimensionnel

Il s'agit dans cette section de décrire la base de données, spécifier les Datamarts ainsi que notre schéma du Datawarehouse en flocon de neige et ses tables de fait et de dimension.

2.1 Base de données Oracle

Dans un premier temps, les données des assurances sont stockées dans plusieurs applicatifs suivant un modèle relationnel. Dans notre cas, celle qui nous intéresse est nommée ANIA. Afin de réaliser les tableaux de bord cités auparavant on doit récupérer les informations nécessaires depuis notre source de données Oracle.

Le modèle de notre base de données pour le module ANIA Santé est représenté comme suit sur la figure 3.2 :

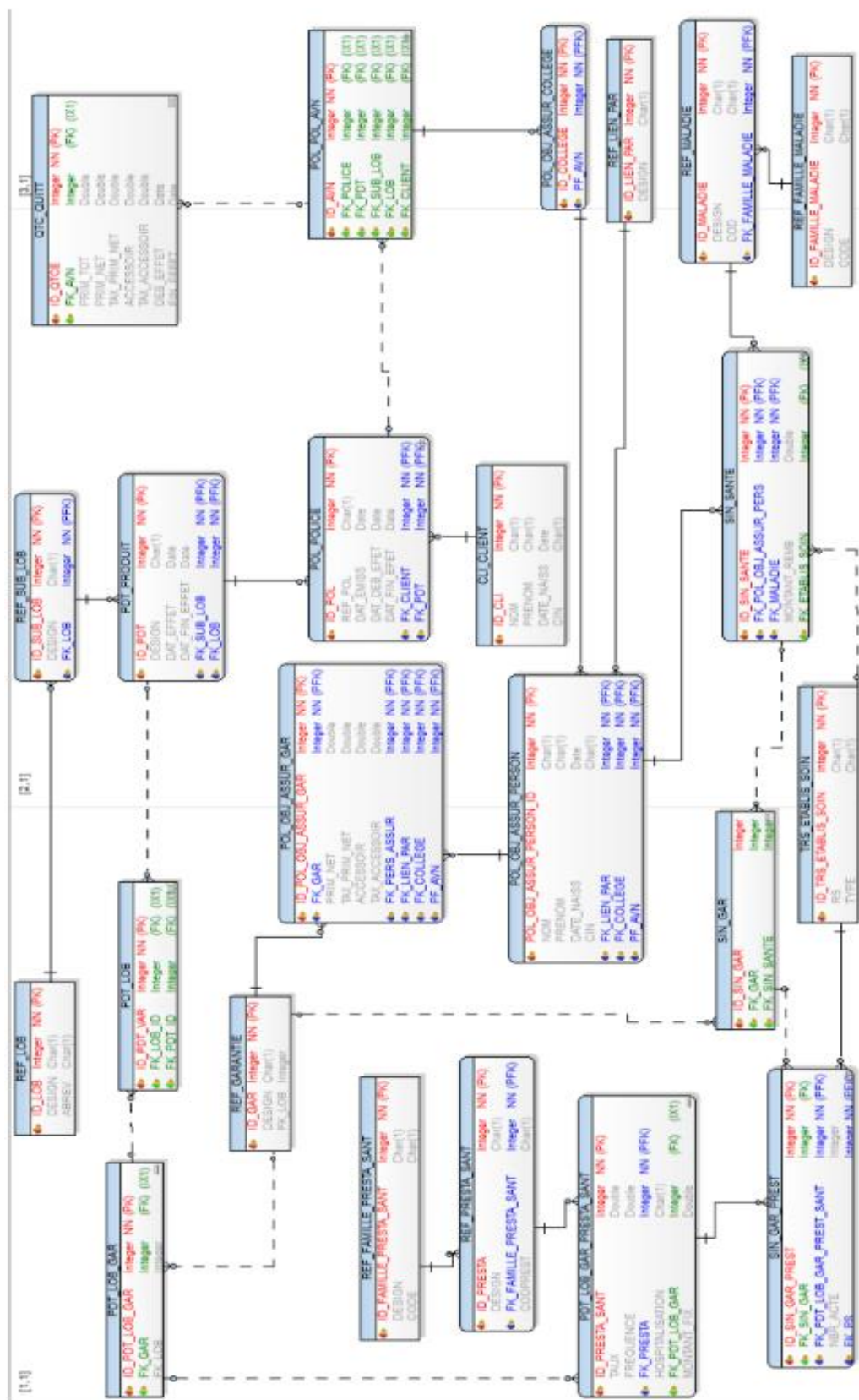


Figure 3. 2: Modèle de la base de données

Le premier traitement à faire est la collecte de ces données. En effet, les données sont extraites de la base de données ANIA, transformées selon nos besoins puis chargées dans la base de destination (l'entrepôt de données).

Cette phase d'alimentation consiste à réorganiser les données dans nos magasins de données (Datamarts), à savoir Contrats, Détails prestation et Sinistres, afin de supporter efficacement les processus d'analyse et d'interrogation, et de nous fournir les vues appropriées à nos besoins.

2.2 Modélisation des Datamarts

Un Datamart est une base de données orientée sujet conçue pour rendre les données organisationnelles spécifiques et facilement disponibles. Un magasin de données est une version condensée d'un entrepôt de données, qui stocke toutes les données générées par les départements d'une organisation [7].

Avec les Datamarts, les utilisateurs peuvent accéder rapidement aux données pertinentes et obtenir des informations sans avoir à parcourir tout un entrepôt de données.

Avec sa petite taille et sa conception ciblée, les Datamarts offre plusieurs avantages à l'utilisateur final, notamment :

- Contient des données précieuses pour des groupes spécifiques au sein d'une organisation.
- Plus rentable à construire qu'un entrepôt de données.

Permet un accès simplifié aux données :

- Accès rapide aux informations sur les données.
- Le magasin de données nécessite moins de temps de mise en œuvre que l'entrepôt de données, car il ne faut se concentrer que sur un petit sous-ensemble de données.

Selon cette méthode, les Datamarts sont d'abord créés pour fournir la capacité de Reporting et d'analyse pour un processus métier spécifique, plus tard, avec ces Datamarts, un entrepôt de données d'entreprise est créé.

La figure 3.3 suivante représente les trois Datamarts construisant notre entrepôt de données :

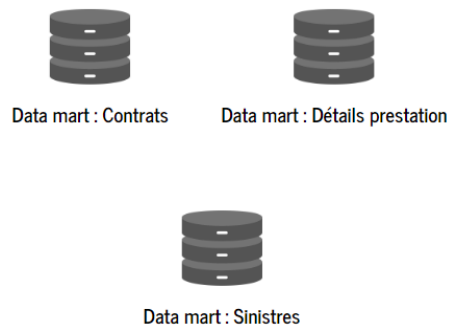


Figure 3. 3: Répartition des Datamarts

Avant d’entamer notre schéma multidimensionnel, nous allons tout d’abord définir les principaux ingrédients de toute mise en œuvre de la Business Intelligence.

- Table de dimension : contient les attributs descriptifs ou qualitatifs des données. Cette table est construite à partir de la base de données relationnelle et représente l’axe en relation avec la mesure dans la table de fait.
- Table de fait : contient les clés étrangères de toutes les tables de dimension. Elle stocke l’attribut mesurable ou quantitatif des données.

Ces tables sont utilisées pour former les schémas en étoile ou en flocon, qui sont conçus dans le cadre de la construction d’un entrepôt de données ou d’un Data Mart.

- Modèle en Etoile : Il se compose d’une table de fait entourée de plusieurs tables de dimension. Les requêtes exécutées par le schéma en étoile sont plus rapides en raison du nombre réduit de jointures utilisées pour interroger les données [7].

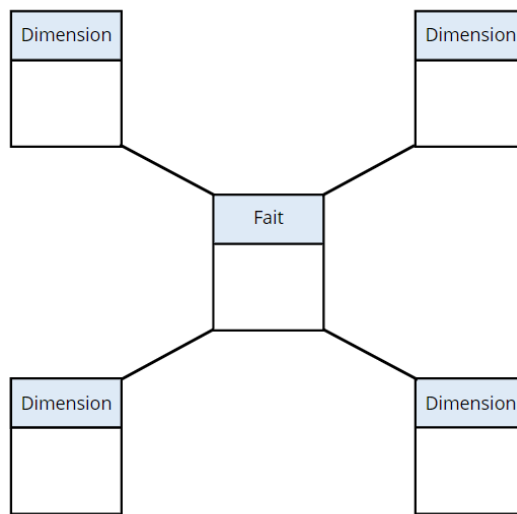


Figure 3. 4: Modèle en étoile

- **Modèle en Flocon de neige** : Il s'agit d'une extension du schéma en étoile. Dans un schéma en étoile chaque dimension est stockée dans une seule table de dimension, alors que dans un schéma en flocon de neige, une dimension explose ou possède une sous dimension [7].

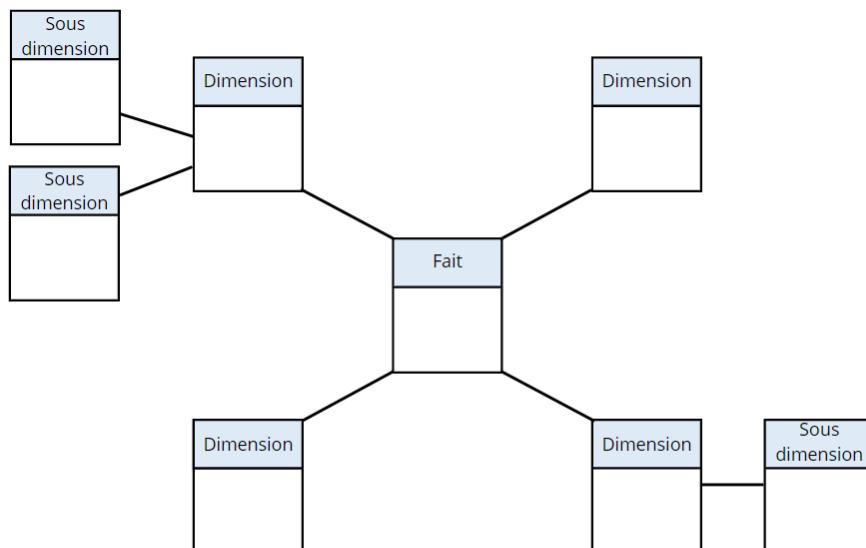


Figure 3. 5: Modèle en flocon de neige

En détails voici les schémas de données réalisés pour le Datamart Contrats sur la figure 3.4 :

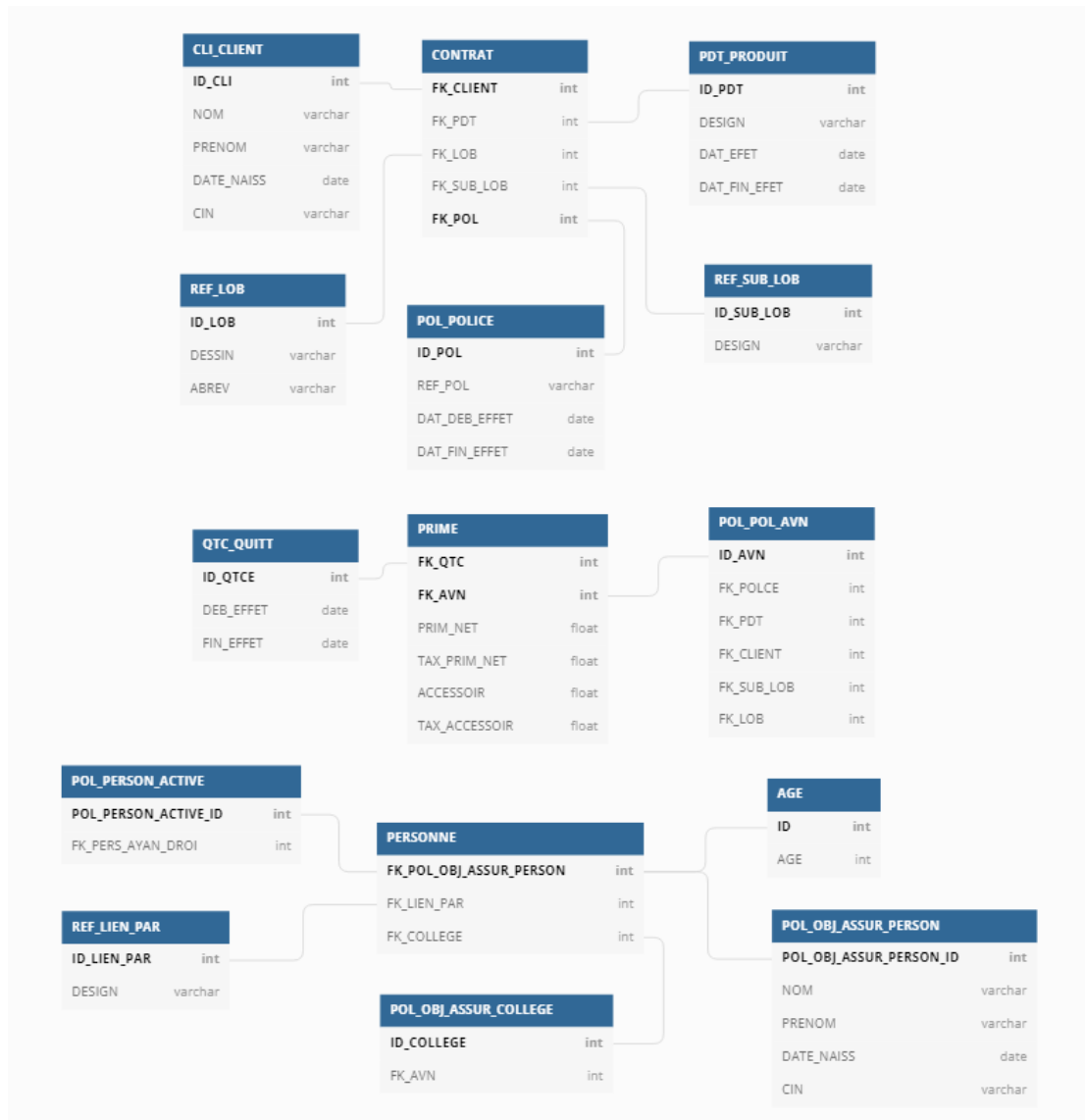


Figure 3. 6: Les modèles de données créés pour le Datamart Contrats

Notre Datamart relatif aux Contrats est composé de trois schémas en étoile, où chacun d'eux s'intéresse à un axe d'analyse.

- **CONTRAT** : rassemble toutes les informations liées à la police et son souscripteur ainsi que son produit et son type (individuel ou groupe).
- **PRIME** : compose les éléments nécessaires au calcul de la Prime totale notamment taxe de la prime nette, prime nette, taxe des accessoires et accessoires.
- **PERSONNE** : regroupe les données relatives à chaque bénéficiaire, afin d'avoir une vision d'analyse sur les liens de parentés, les tranches d'âge et même la population active dans notre assurance.

Pour les deux autres Datamarts (Détails prestation et sinistres), on a opté pour présenter leurs modèles sur l'**Annexe**.

2.3 Modélisation du Datawarehouse

Pour notre cas, afin de formaliser une hiérarchie au sein d'une dimension et de réduire la taille des dimensions, on a travaillé par un schéma en flocon de neige représenté sur la figure 3.7.



Figure 3. 7: Modèle final du Data Warehouse

Le schéma en flocon précédent représente des mesures qu'on peut projeter sur les axes d'analyse et qu'on peut présenter par des tables de dimension.

Les principales dimensions sur lesquelles nous nous appuyerons afin de réaliser nos Dashboards sont :

- **PERSONNE** : Cette dimension englobe toutes les informations liées aux bénéficiaires de l'assurance et leurs ayants droit (parent, enfant, conjoint).

Tableau 3. 2: Descriptif des attributs de la table Personne

Attributs	Type	Description de l'attribut
POL_OBJ_ASSUR_PERSON_ID	INT	Identifiant du bénéficiaire (Clé primaire)
NOM	VARCHAR	Nom du bénéficiaire
PRENOM	VARCHAR	Prénom du bénéficiaire
TRANCHE	VARCHAR	La tranche d'âge à laquelle le bénéficiaire appartient
DESIGN_LIEN	VARCHAR	Lien de parenté avec l'assuré principal (parent, conjoint, enfant, lui-même)
ASSURE_PRINCIPAL	INT	L'identifiant de l'assuré principal
NOM_ASSURE_PRINCIPAL	VARCHAR	Nom de l'assuré principal
PRENOM_ASSURE_PRINCIPAL	VARCHAR	Prénom de l'assuré principal
AGE	INT	L'âge du bénéficiaire
FK_POL	INT	L'identifiant de la police dont il appartient

- **CONTRAT** : Représente les données liées à la police telle que la référence, le type, le souscripteur et le produit dont elle appartient.

- **PRIME** : Contient les informations relatives aux cotisations à savoir la date d'échéance et la police à laquelle elle appartient.
- **AFFECTION** : C'est une dimension permettant l'analyse par rapport aux maladies et les familles de maladie.
- **FAMILLE D'ACTES** : S'intéresse aux prestations et aux familles de prestation de chaque police de l'assurance santé.
- **ETABLISSEMENT** : Comporte les différents établissements (prestataires médicaux) ainsi que leurs types (pharmacie, clinique dentaire, hôpital, ...).
- **DATE** : C'est une dimension très essentielle dont le but est de faire des analyses par rapport à l'axe du temps afin de visualiser des rapports périodiques. Cette dernière contient les dates de début et fin des sinistres.

Chapitre 4 : Etude technique et Réalisation

Ce chapitre fait l'objet de la dernière partie de notre rapport, il englobe les différentes étapes de la réalisation des tableaux de bord en commençant par le choix des outils à l'aide de quelques études comparatives, et en passant par la phase de la réalisation où on va détailler les aspects de l'implémentation de notre solution décisionnelle à savoir le processus d'intégration des données, l'écriture de la requête SQL, la conception du modèle tabulaire et la partie Reporting, et finir par l'intégration de notre projet dans la solution ANIA.



1 Outils et technologies

Dans cette partie, nous allons aborder le choix technologique à l'aide de quelques études comparatives pour finir par élaborer l'acheminement par lequel les données vont passer afin d'accomplir notre solution décisionnelle.

1.1 Suites BI

Avant d'entamer tout projet BI, il est indispensable de choisir la suite ou les outils dont on aura besoin pour accomplir la solution, pour ce faire, commençant par comparer les deux suites BI les plus répandues :

Tableau 4. 1 : Comparatif des suites BI MSBI et OBIEE

	
<ul style="list-style-type: none">• Exploration et visualisation faciles des données.• MSBI est un outil en libre-service, il est très facile à utiliser par rapport aux outils conventionnels.• MSBI prend en charge des services Web• Solutions d'entreprise de bout en bout : Divers outils MSBI sont utilisés pour diverses solutions commerciales telles que le traitement des transactions en ligne, l'entrepôt de données, le magasin de données, le modèle sémantique BI et l'exploration de données.• SSIS permet d'obtenir l'entrepôt de données à partir de sources de données différentes.	<ul style="list-style-type: none">• OBIEE est conçu spécifiquement pour les environnements de bases de données et d'applications Oracle.• La solution OBIEE a une portée large et fonctionnelle et les différents composants sont liés les uns aux autres.• OBIEE est très utile pour une entreprise qui utilise déjà Oracle et qui ne dispose pas encore d'une solution de BI d'un autre fournisseur.• Extensif : OBIEE est très convivial et extrêmement intuitif, et l'interface Web apporte une BI robuste et dynamique à partir de diverses sources.

Cette étude comparative nous a menée à confirmer le choix de l'entreprise concernant la suite MSBI qui permet d'offrir les meilleures solutions pour les requêtes de Business Intelligence et d'exploration de données. Cette suite utilise Visual Studio avec SQL Server et propose des outils spéciaux pour les procédures distinctes requises dans les solutions de Business Intelligence (BI).

1.2 Serveur de gestion de base de données

SQL Server Management Studio (SSMS) est un environnement intégré pour la gestion des infrastructures SQL. Il fournit des outils pour configurer et surveiller les instances de SQL Server et des bases de données. SSMS est utilisé pour mettre à niveau les composants de la couche de donnée utilisés par les applications, et créer des requêtes et des scripts [5].



1.3 Outil ETL (SSIS)

SQL Server Integration Services est une plate-forme permettant de créer des solutions d'intégration et de transformation de données au niveau de l'entreprise.

SSIS peut extraire et transformer des données à partir d'une grande variété de sources telles que des fichiers de données XML, des fichiers plats et des sources de données relationnelles, et enfin charger les données dans une ou plusieurs destinations [5].



Les services d'intégration comprennent :

- Un ensemble de tâches et de transformations intégrées ;
- Outils graphiques pour créer des packages ;
- Une base de données de catalogue SSIS pour stocker, exécuter et gérer les packages.

D'après Gartner, SSIS de Microsoft est parmi les leaders dans le Magic Quadrant des outils d'intégration de données selon la figure 4.1.

Figure 1: Magic Quadrant for Data Integration Tools



Source: Gartner (August 2021)

Figure 4. 1 : Magic Quadrant des outils d'intégration de données 2021[6]

1.4 Outil de modélisation sémantique

Les données décisionnelles, une fois traitées et stockées, doivent être structurées pour répondre aux besoins d'analyse. Pour ce faire, SQL Server Analysis Services fournit une plateforme d'analyse rapide et intuitive permettant de structurer les données et de mettre en place des modèles sémantiques multidimensionnels ou tabulaires pour les rapports d'entreprises et les applications clientes [5].



Les raisons typiques de l'utilisation de SSAS sont les suivantes :

- Couche sémantique simple à mettre en place ;
- Performances élevées et accès intuitifs aux données ;
- Outil de développement complètement intégré dans Visual Studio ;
- Accès par requêtes DAX (Data Analysis Expressions) pour les cubes TABULAR ;
- Augmentation de la vitesse de requête.

SSAS permet de travailler avec deux types de modèles différents suivant les besoins.

Tableau 4. 2 : Comparatif des deux types de modèles SSAS

Modèle Multidimensionnel	Modèle Tabulaire
<ul style="list-style-type: none">- Analyser des données déjà agrégées et stockées selon les besoins de l'utilisateur.- Modéliser les données sous forme de dimensions et de mesures.- Capacité à manipuler de gros volumes de données agrégés en croisant différentes dimensions.- Un cube OLAP utilise les fonctions classiques d'agrégation : Min, Max, Count, Sum, mais peut être enrichi avec des calculs spécifiques complexes.	<ul style="list-style-type: none">- Analyser des données « in-memory » ce qui permet d'assurer des temps de réponse très réduits.- Possibilité de ne pas charger les données en mémoire et configurer le mode « Direct Query ».- Modéliser les données sous forme de « tableaux » liés avec des relations.- Les cubes « TABULAR » peuvent être enrichis avec des calculs spécifiques.

En s'appuyant sur l'architecture déjà définie selon l'approche de Kimball et le modèle du Datawarehouse réalisé, nous avons travaillé selon un modèle tabulaire afin d'avoir une organisation d'une manière équivalente aux bases de données relationnelles et même adéquate avec l'outil de visualisation choisi (Power BI).

1.5 Outil de Reporting

Voyant le nombre important des solutions BI existantes actuellement, il est difficile de décider pour quel outil BI opter. Dans cette partie nous allons faire un comparatif de quelques produits qui représentent les solutions leader du marché. Cette liste n'a aucune prétention d'exhaustivité, et bien entendu, le choix d'un logiciel dépend toujours de la nature du projet et surtout du budget alloué, il est donc préférable de construire une base de critères pour bien évaluer les produits susceptibles de répondre au mieux aux besoins.

Nous nous sommes basés sur Gartner qui désigne Microsoft, Tableau et Qlik comme leaders du dernier Gartner Magic Quadrant for Business Intelligence and Analytics Platforms 2022.






Source: Gartner (March 2022)

Figure 4. 2 : le Magic Quadrant des outils de reporting [6].

Sur le marché, les trois outils sont des marques bien connues. Cependant, il existe encore de nombreuses similitudes et différenciations entre ces trois produits, donc dans ce tableau comparatif, nous nous sommes basés sur les critères suivants :

- Versions de base ;
- Versions avancées ;
- Analytique et interprétation ;
- Capture et stockage des données ;
- Communauté de clients.

Tableau 4. 3 : Comparatif des outils de visualisation [4]

Critère	 Power BI	 tableau	 QlikView
Versions de base	Power BI Desktop est gratuit, les utilisateurs profitent des différentes fonctionnalités	Une version de base de Tableau est également gratuite mais dispose de fonctionnalités très limitées.	La version de base de Qlikview est également gratuite, mais comme Tableau, les fonctionnalités avec lesquelles travailler sont très limitées.
Versions avancées	Power BI est livré avec deux versions avancées. Ces forfaits coûtent 10 \$ par utilisateur et par mois, semblent moins chers que Tableau et Qlikview.	La version avancée de Tableau coûte 100 \$ par utilisateur, ce qui semble cher pour les industries de taille moyenne.	30 \$ par mois et par utilisateur est le coût d'acquisition d'une version avancée de Qlikview.

Analytique et interprétation	Power BI a beaucoup de fonctionnalités avancées aide à l'analyse et l'interprétation des données.	Tableau est le meilleur ici. En raison des options d'exploration et de filtrage.	Qlikview manque d'analyses éprouvées.
Capture et stockage des données	Power BI est capable de récupérer des données de n'importe quelle source de données telles que : le Web, le cloud SQL, Azure, etc.	Tableau est également capable d'établir des bonnes connexions aux sources de données, probablement à égalité avec Power BI.	L'élément clé de QlikView est la transformation des données et il peut également récupérer des données de n'importe où.
Communauté de clients	Power BI est fourni avec le package Office 365, ceux qui achètent d'autres produits Microsoft peuvent également être exposés à cet outil.	Tableau a une histoire en soi et a donc une énorme communauté de clients.	Qlikview ne peut être utilisé que par des personnes connaissant la programmation, de sorte que peu de personnes sont issues de la programmation en termes de communauté de clients.

Après cette étude comparative des outils de Reporting, on a opté pour Power BI comme générateur des rapports pour notre cycle BI en se basant sur nos besoins et notre environnement de travail, puisque la création d'un tableau de bord sur Power BI nécessite une bonne connaissance du langage DAX. Ce dernier est un langage de requête sur les bases de données

qui permet d'explorer les cubes tabulaires sur SSAS, interroger les données qui y sont contenues et les manipuler. Ceci est résumé sur la figure 4.3 :

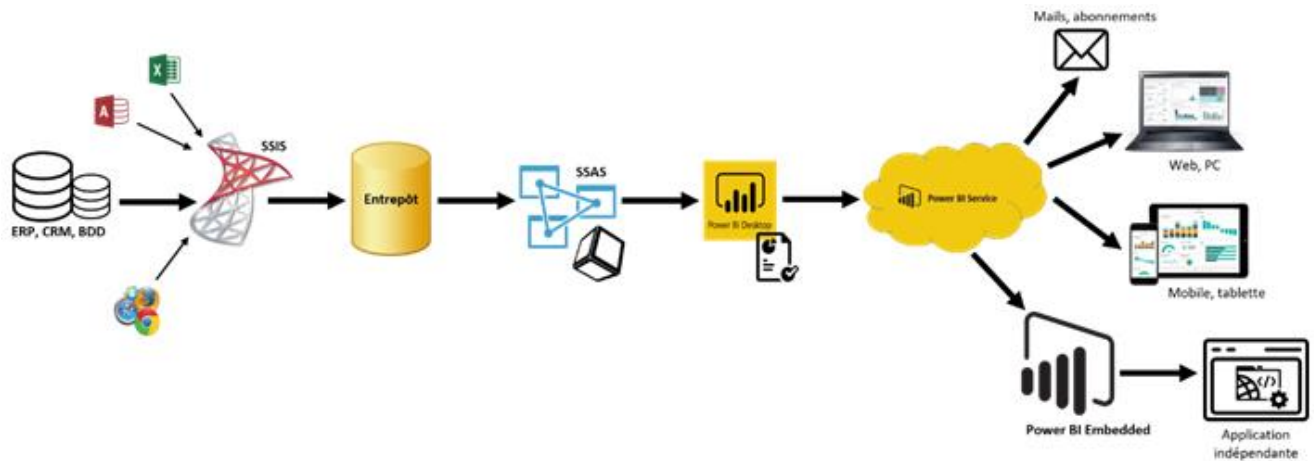


Figure 4. 3 : Schéma récapitulatif de la suite MSBI

Pour conclure, la figure 4.3 récapitule l'acheminement suivi par les données afin de les rendre des informations décisionnelles grâce à des visuels interactifs. Ceci commence par une extraction depuis une base de données Oracle vers notre entrepôt de données à l'aide de l'outil SSIS, puis la création du modèle tabulaire avec SSAS, ensuite introduire ce modèle de données sur PowerBI Desktop afin de générer des tableaux de bord et les partager sur la solution web créée par l'entreprise.

2 Réalisation du système décisionnel

Dans cette partie, notre système décisionnel sera concrétisé ; pour ce faire, nous allons commencer par présenter le processus d'intégration des données contenant la création du package SSIS ainsi que la création des Jobs, ensuite on passe à l'écriture de la requête SQL pour la transformation des données, puis la création du modèle tabulaire, pour finalement présenter la réalisation d'un tableau de bord à l'aide d'un exemple illustratif.

2.1 Processus d'intégration des données

Cette section s'intéresse à la réalisation du processus ETL à l'aide des outils d'intégration choisis précédemment ; ce processus va passer par la création de deux éléments essentiels qui sont le package SSIS et la configuration des Jobs.

2.1.1 Création du package SSIS

Un package SSIS est une collection organisée de connexions, d'éléments de flux de contrôle, d'éléments de flux de données, de gestionnaires d'événements, de variables, de paramètres et de configurations assemblés à l'aide des outils de conception graphique fournis par SQL Server Integration Services.

➤ Gestionnaire de Connexion

L'ETL, comme l'on a précisé dans ce qui précède, est un processus automatisé prenant en entrée des données brutes, extrait les informations nécessaires à l'analyse, les transforme en un format qui peut répondre aux besoins opérationnels et les charge dans un Data Warehouse, et afin de réaliser ce processus on aura besoin de créer des connexions avec une base de données source et une autre de destination.

Pour notre cas, on a configuré deux connexions ; la première est une connexion Oracle vu que notre source est une base de données Oracle, et la deuxième est une connexion SQL Server pour notre Data Warehouse de destination.

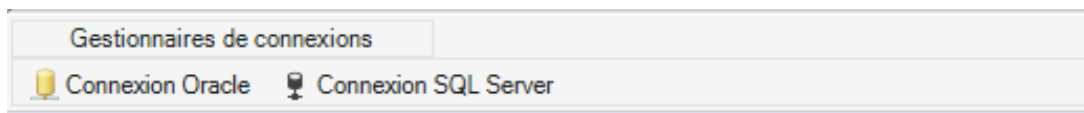


Figure 4. 4 : Gestionnaires de connexions SSIS

➤ Flux de contrôle

Un package se compose d'un flux de contrôle et, éventuellement, d'un ou plusieurs flux de données. SSIS fournit trois types d'éléments de flux de contrôle :

- Des conteneurs qui fournissent des structures dans des packages : ils implémentent des flux de travail répétitifs dans un package ou divisent un flux de contrôle en sous-ensembles ;
- Des tâches qui fournissent des fonctionnalités ;
- Contraintes de priorité qui connectent les exécutable.

Prenant comme exemple la création du flux de contrôle du datamart Contrats, on a choisi de structurer ce dernier à l'aide des conteneurs d'exécution parallèle des taches à savoir :

- Tache de début ;

- Tache de flux de données : cette tache contient le flux de données qui sera configuré vers la suite ;
- Taches d'exécution de script SQL : ces taches donnent la main pour écrire des requêtes SQL afin d'effectuer des transformations et créer des vues pour finalement obtenir les tables (Primes, Bénéficiaires, Contrats) composant notre Datamart Contrats ;
- Tache de fin.

La figure 4.5 représente le flux de contrôle correspondant au Datamart Contrats.

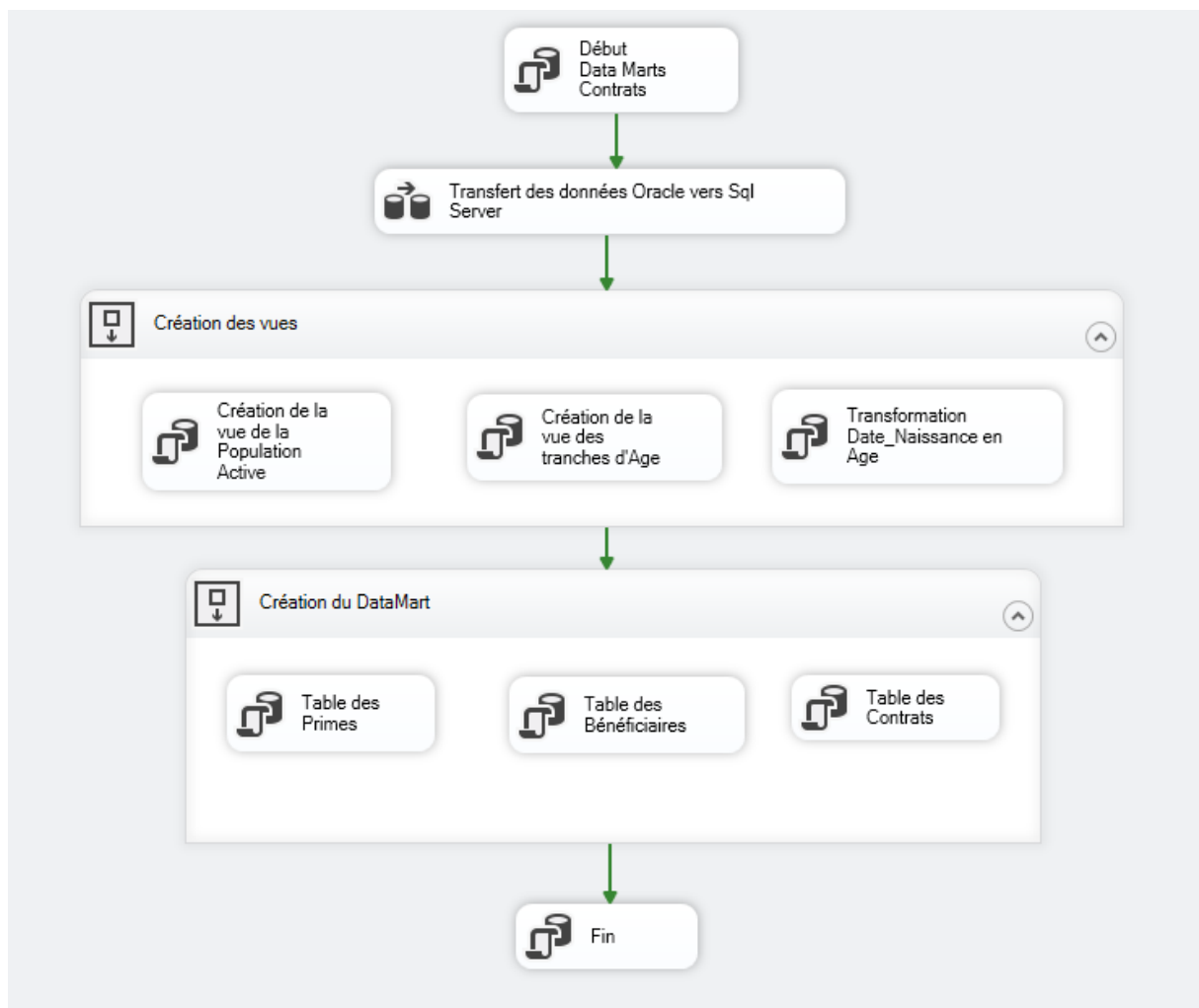


Figure 4. 5 : Flux de contrôle (Data Mart Contrat)

➤ Flux de données

SSIS fournit trois types de composants de flux de données : sources, transformations et destinations. De plus, il offre des chemins qui connectent la sortie d'un composant à l'entrée d'un autre composant.

La connexion des composants de flux de données se fait en liant la sortie des sources et des destinations à l'entrée des transformations et des destinations. Lors de la construction d'un flux de données, une fois le composant connecté, les colonnes d'entrée sont disponibles pour être utilisées dans la configuration du composant.

Pour notre cas, les flux de données se composaient de quatre composants à savoir :

- Une source Oracle à l'aide d'une extension Devart ;
- Un convertisseur de données pour éliminer les anomalies en relation avec les types d'attribut de la source et de la destination ;
- Un transformateur de recherche afin d'insérer uniquement les nouveaux enregistrements dans la table de destination ;
- Une destination SQL SERVER pour alimenter notre Datawarehouse.

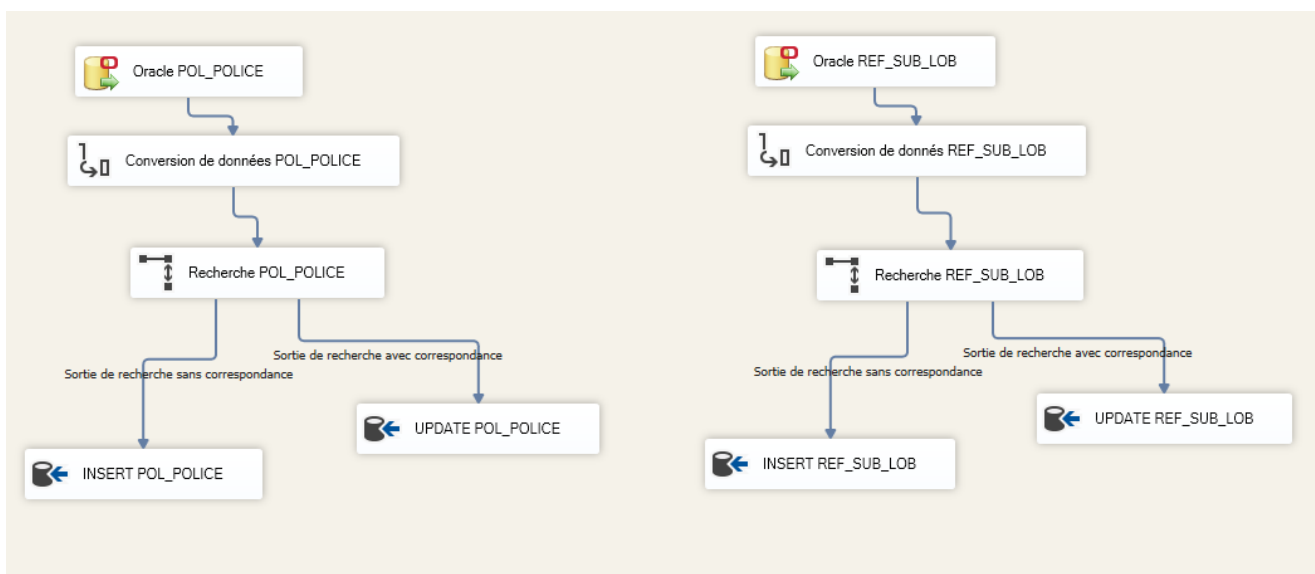


Figure 4. 6 : Flux de données SSIS

Après avoir créé les packages correspondants à l'alimentation des Datamarts, on passe maintenant à configurer le package global qui fera appel à tous les packages qui sont créés précédemment en utilisant le composant « Tâche d'exécution du package », ces derniers seront exécutés en parallèle et vers la fin la tâche « Envoyer un message » permet de notifier le manager de l'alimentation du Data Warehouse.

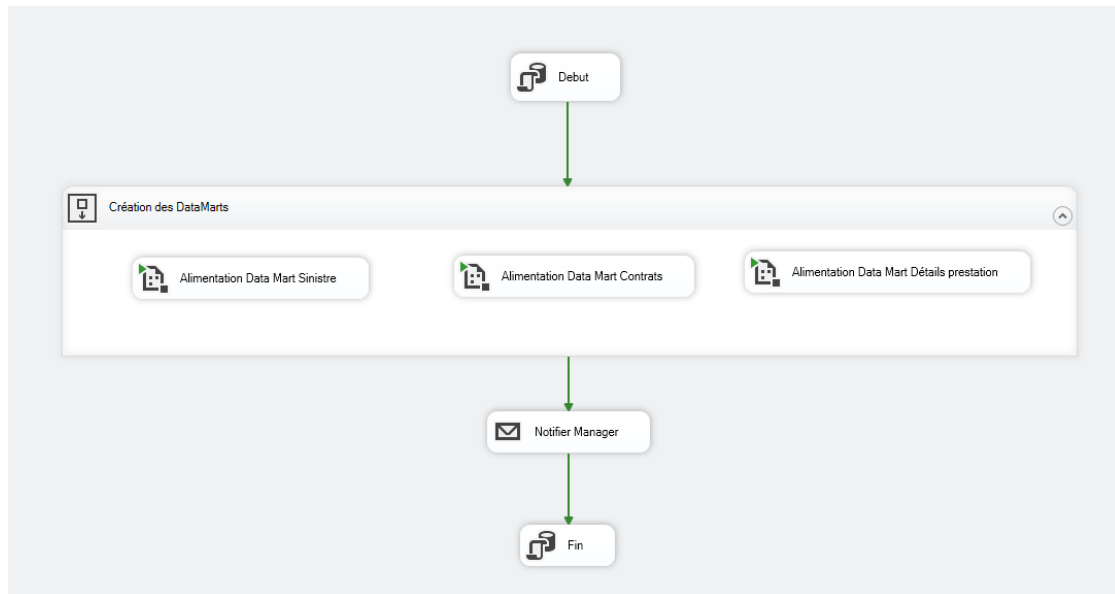


Figure 4. 7 : Package d'alimentation du Datawarehouse

Dans la partie suivante, on va voir comment exécuter ce package régulièrement afin de mettre à jour notre Datawarehouse.

2.1.2 Création des Jobs

L'automatisation et la planification de l'exécution du package global d'alimentation SQL Server Integration Services se fait à l'aide de l'Agent SQL Server. Dans un premier lieu, le déploiement se fait sur le catalogue d'Integration Services et se stocke dans SQL Server.

La figure suivante représente le déploiement de notre package global « All-Etat » dans SQL Server.

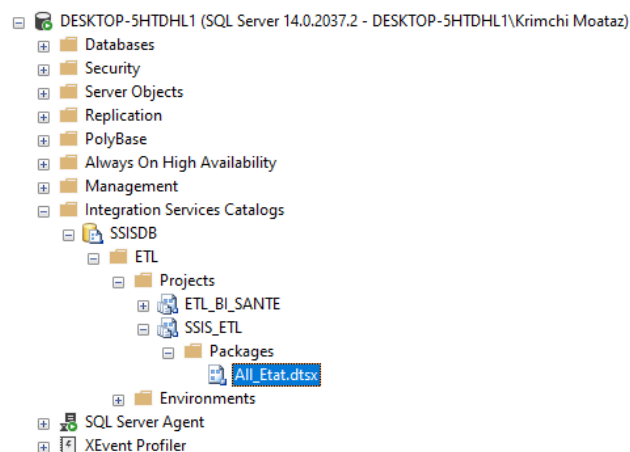


Figure 4. 8 : Déploiement du Package d'alimentation du Datawarehouse

Il faut planifier les packages afin que notre ETL puisse s'exécuter dans des fenêtres horaires spécifiques (probablement la nuit), afin d'éviter une coïncidence avec les heures de travail en base de données. On peut même définir une ou plusieurs planifications pour exécuter le package à des moments prédéfinis.

SSMS offre le choix entre plusieurs types de plannings : quotidien, hebdomadaire ou mensuel. Pour notre cas nous avons programmé le job pour qu'il s'exécute tous les jours à 3h30 du matin, à partir du 12/04.

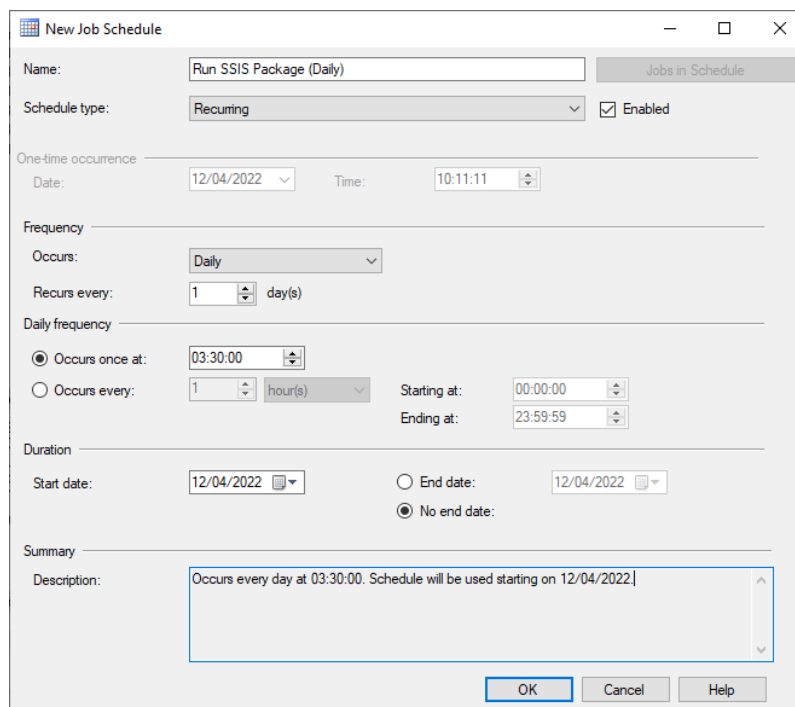


Figure 4. 9 : Configuration du job

2.2 Ecriture de la requête SQL

Avant de commencer l'écriture de notre requête SQL, il faut tout d'abord définir l'objectif du tableau de bord par une lecture du cahier de charge correspondant à ce dernier afin d'en tirer les ambiguïtés et finir par programmer une réunion avec le manager dans le but d'éclaircir et discuter les attentes du client.

Dans ce sens, prenant comme exemple l'état « Récapitulatif des dépenses par tranches d'âge » après avoir défini les objectifs, on passe à préciser parmi les KPIs cités précédemment ceux qui répondent aux besoins tracés. En fait, Pour chacun des objectifs fixés, on définit l'indicateur principal qui doit présenter un résultat quantitatif ou une mesure précise. Pour notre

cas, le nombre de bénéficiaires par tranches d'âge, le cout moyen par tranches d'âge et les montants remboursés par tranches d'âge sont les principaux indicateurs de performance répondant à notre besoin.

Par la suite, on passe à déterminer les données nécessaires afin d'extraire les informations de la base de données, pour ce faire, on aura besoin d'effectuer des transformations et créer des vues via des requêtes SQL bien précises.

Continuant sur l'exemple précédent ; afin de tirer les informations des montants remboursés par les tranches d'âge des bénéficiaires on procède selon les jointures et les unions suivantes :

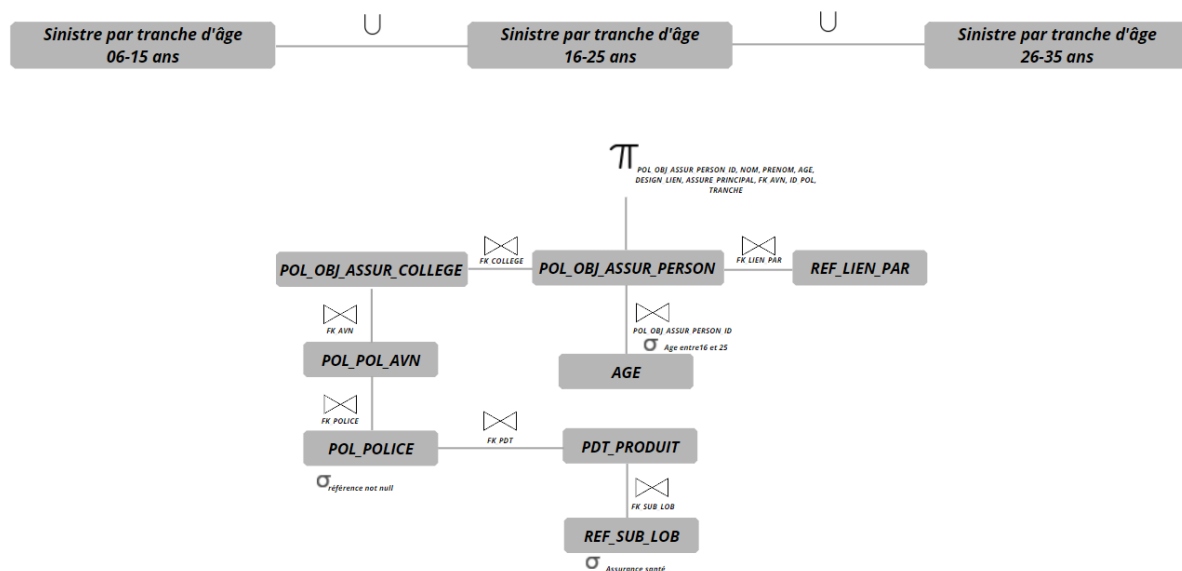


Figure 4. 10 : Ecriture algébrique pour l'état "Récapitulatif des dépenses par tranches d'âge"

Cette requête permet de donner les données concernant chaque bénéficiaire à savoir :

- **La tranche d'âge** : cette information a été générée à l'aide d'une transformation des dates de naissance en âges à travers la vue AGE.
- **Le lien de parenté** (parent, conjoint, enfant ou lui-même) : une jointure avec la table REF-LIEN-PAR permet de préciser le lien de parenté pour chacun des bénéficiaires.
- **Le montant des sinistres** : en effectuant une jointure avec la vue SINISTRE, on peut extraire les informations sur les sinistres.

Ainsi, on effectue des filtres afin d'extraire que les informations concernant les produits d'assurance santé en plus d'éliminer les références des polices nulles.

En se basant sur l'écriture algébrique il est vraiment très simple d'écrire la requête SQL correspondante à la création de la vue PERSON qui englobe toutes les informations liées à chaque bénéficiaire notamment sa tranche d'âge, son lien de parenté et d'autres colonnes supplémentaires.

```

/* TABLE PERSON */
create view PERSON
(TRANCHE, POL_OBJ_ASSUR_PERSON_ID, NOM, AGE, PRENOM, DESIGN_LIEN, ASSURE_PRINCIPAL, FK_AVN, ID_POL)
AS
SELECT 'No age' AS tranche, POL_OBJ_ASSUR_PERSON.POL_OBJ_ASSUR_PERSON_ID, POL_OBJ_ASSUR_PERSON.NOM, age, PRENOM, REF_LIEN_PAR.DESIGN,
isnull([POL_PERSON_ACTIVE].[FK_PERS_AYAN_DROI],[POL_PERSON_ACTIVE].POL_OBJ_ASSUR_PERSON_ID), FK_AVN, ID_POL
FROM POL_OBJ_ASSUR_PERSON
JOIN POL_OBJ_ASSUR_COLLEGE ON POL_OBJ_ASSUR_PERSON.FK_COLLEGE = POL_OBJ_ASSUR_COLLEGE.ID_COLLEGE
JOIN POL_POL_AVN ON POL_OBJ_ASSUR_COLLEGE.FK_AVN = POL_POL_AVN.ID_AVN
JOIN POL_POLICE ON POL_POL_AVN.FK_POLICE = POL_POLICE.ID_POL
JOIN PDT_PRODUIT ON POL_POLICE.FK_PDT = PDT_PRODUIT.ID_PDT
join REF_SUB_LOB ON REF_SUB_LOB.ID_SUB_LOB = PDT_PRODUIT.FK_SUB_LOB
JOIN REF_LIEN_PAR ON REF_LIEN_PAR.ID_LIEN_PAR = POL_OBJ_ASSUR_PERSON.FK_LIEN_PAR
JOIN Age ON Age.id = POL_OBJ_ASSUR_PERSON.POL_OBJ_ASSUR_PERSON_ID
JOIN POL_PERSON_ACTIVE ON POL_OBJ_ASSUR_PERSON.POL_OBJ_ASSUR_PERSON_ID = POL_PERSON_ACTIVE.POL_OBJ_ASSUR_PERSON_ID
WHERE FK_SUB_LOB IN (183,184) and REF_POL is not null AND (Age.age IS NULL )

UNION ALL SELECT '0-5 ans' AS tranche, POL_OBJ_ASSUR_PERSON.POL_OBJ_ASSUR_PERSON_ID, POL_OBJ_ASSUR_PERSON.NOM, age, PRENOM, REF_LIEN_PAR.DESIGN,
isnull([POL_PERSON_ACTIVE].[FK_PERS_AYAN_DROI],[POL_PERSON_ACTIVE].POL_OBJ_ASSUR_PERSON_ID), FK_AVN, ID_POL
FROM POL_OBJ_ASSUR_PERSON
JOIN POL_OBJ_ASSUR_COLLEGE ON POL_OBJ_ASSUR_PERSON.FK_COLLEGE = POL_OBJ_ASSUR_COLLEGE.ID_COLLEGE
JOIN POL_POL_AVN ON POL_OBJ_ASSUR_COLLEGE.FK_AVN = POL_POL_AVN.ID_AVN
JOIN POL_POLICE ON POL_POL_AVN.FK_POLICE = POL_POLICE.ID_POL
JOIN PDT_PRODUIT ON POL_POLICE.FK_PDT = PDT_PRODUIT.ID_PDT
join REF_SUB_LOB ON REF_SUB_LOB.ID_SUB_LOB = PDT_PRODUIT.FK_SUB_LOB
JOIN REF_LIEN_PAR ON REF_LIEN_PAR.ID_LIEN_PAR = POL_OBJ_ASSUR_PERSON.FK_LIEN_PAR
JOIN Age ON Age.id = POL_OBJ_ASSUR_PERSON.POL_OBJ_ASSUR_PERSON_ID
JOIN POL_PERSON_ACTIVE ON POL_OBJ_ASSUR_PERSON.POL_OBJ_ASSUR_PERSON_ID = POL_PERSON_ACTIVE.POL_OBJ_ASSUR_PERSON_ID
WHERE FK_SUB_LOB IN (183,184) and REF_POL is not null AND (CAST(LEFT(Age.age, 4) AS INTEGER) BETWEEN 0 AND 5)

UNION ALL SELECT '06-15 ans' AS tranche, POL_OBJ_ASSUR_PERSON.POL_OBJ_ASSUR_PERSON_ID, POL_OBJ_ASSUR_PERSON.NOM, age, PRENOM, REF_LIEN_PAR.DESIGN,
isnull([POL_PERSON_ACTIVE].[FK_PERS_AYAN_DROI],[POL_PERSON_ACTIVE].POL_OBJ_ASSUR_PERSON_ID), FK_AVN, ID_POL
FROM POL_OBJ_ASSUR_PERSON
JOIN POL_OBJ_ASSUR_COLLEGE ON POL_OBJ_ASSUR_PERSON.FK_COLLEGE = POL_OBJ_ASSUR_COLLEGE.ID_COLLEGE
JOIN POL_POL_AVN ON POL_OBJ_ASSUR_COLLEGE.FK_AVN = POL_POL_AVN.ID_AVN
JOIN POL_POLICE ON POL_POL_AVN.FK_POLICE = POL_POLICE.ID_POL
JOIN PDT_PRODUIT ON POL_POLICE.FK_PDT = PDT_PRODUIT.ID_PDT
join REF_SUB_LOB ON REF_SUB_LOB.ID_SUB_LOB = PDT_PRODUIT.FK_SUB_LOB
JOIN REF_LIEN_PAR ON REF_LIEN_PAR.ID_LIEN_PAR = POL_OBJ_ASSUR_PERSON.FK_LIEN_PAR
JOIN Age ON Age.id = POL_OBJ_ASSUR_PERSON.POL_OBJ_ASSUR_PERSON_ID
JOIN POL_PERSON_ACTIVE ON POL_OBJ_ASSUR_PERSON.POL_OBJ_ASSUR_PERSON_ID = POL_PERSON_ACTIVE.POL_OBJ_ASSUR_PERSON_ID
WHERE FK_SUB_LOB IN (183,184) and REF_POL is not null AND (CAST(LEFT(Age.age, 4) AS INTEGER) BETWEEN 6 AND 15)

UNION ALL SELECT '16-25 ans' AS tranche, POL_OBJ_ASSUR_PERSON.POL_OBJ_ASSUR_PERSON_ID, POL_OBJ_ASSUR_PERSON.NOM, age, PRENOM, REF_LIEN_PAR.DESIGN,
isnull([POL_PERSON_ACTIVE].[FK_PERS_AYAN_DROI],[POL_PERSON_ACTIVE].POL_OBJ_ASSUR_PERSON_ID), FK_AVN, ID_POL
FROM POL_OBJ_ASSUR_PERSON
JOIN POL_OBJ_ASSUR_COLLEGE ON POL_OBJ_ASSUR_PERSON.FK_COLLEGE = POL_OBJ_ASSUR_COLLEGE.ID_COLLEGE
JOIN POL_POL_AVN ON POL_OBJ_ASSUR_COLLEGE.FK_AVN = POL_POL_AVN.ID_AVN
JOIN POL_POLICE ON POL_POL_AVN.FK_POLICE = POL_POLICE.ID_POL
JOIN PDT_PRODUIT ON POL_POLICE.FK_PDT = PDT_PRODUIT.ID_PDT
join REF_SUB_LOB ON REF_SUB_LOB.ID_SUB_LOB = PDT_PRODUIT.FK_SUB_LOB
JOIN REF_LIEN_PAR ON REF_LIEN_PAR.ID_LIEN_PAR = POL_OBJ_ASSUR_PERSON.FK_LIEN_PAR
JOIN Age ON Age.id = POL_OBJ_ASSUR_PERSON.POL_OBJ_ASSUR_PERSON_ID
JOIN POL_PERSON_ACTIVE ON POL_OBJ_ASSUR_PERSON.POL_OBJ_ASSUR_PERSON_ID = POL_PERSON_ACTIVE.POL_OBJ_ASSUR_PERSON_ID
WHERE FK_SUB_LOB IN (183,184) and REF_POL is not null AND (CAST(LEFT(Age.age, 4) AS INTEGER) BETWEEN 16 AND 25)

```

Figure 4. 11 : Requête SQL pour l'état "Récapitulatif des dépenses par tranches d'âge"

Ce regroupement d'informations au sein d'une entité (vue) va nous aider d'une part de travailler qu'avec les données nécessaires à la création des rapports et d'une autre part de créer un Datawarehouse ouvert à n'importe quelle source de données.

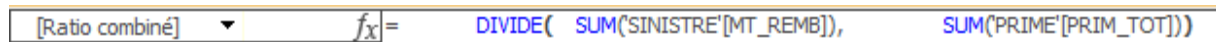
2.3 Conception du modèle tabulaire

Après avoir fini le processus ETL à l'aide de l'outil SSIS afin de créer les différents Datamarts de notre Data Warehouse, on passe dans ce qui suit à fixer notre modèle de données à utiliser pour la réalisation de nos Dashboards ; ce modèle est modélisé grâce à l'outil SSAS qui permet de créer un modèle tabulaire basé sur des mesures programmées en langage DAX.

2.3.1 Création des mesures (DAX)

Data Analysis Expressions (DAX) est un langage de formule utilisé pour créer des calculs personnalisés dans Analysis Services. Les formules DAX incluent des fonctions, des opérateurs et des valeurs pour effectuer des calculs avancés sur les données des tables et des colonnes [9].

SSAS offre la possibilité de créer des mesures et les ajouter aux tables du Data Warehouse avant de fixer le modèle ; ces mesures permettent de préparer les données exactes de visualisation, prenons comme exemple la création de la mesure correspondante au KPI Ratio combiné qui se fait à travers la formule DAX suivante :



[Ratio combiné] f_x = `DIVIDE(SUM('SINISTRE'[MT_REMB]), SUM('PRIME'[PRIM_TOT]))`

Figure 4. 12 : Requête DAX pour la création de la mesure Ratio combiné

De même, le langage DAX permet de créer non seulement les mesures mais aussi les tables ; dans notre cas, on a ajouté une table de temps appelée « Datum » afin de régler des problèmes temporels à savoir la comparaison entre deux années et les cumulés mensuels qui exigent la continuité du temps.

[Date]	f_x	=
		<pre> VAR DebDate = DATE(2012,11,21) VAR FinDate = TODAY() RETURN ADDCOLUMNS(CALENDAR(DebDate,FinDate), "Annee", FORMAT([Date],"YYYY"), "Trimestre", "T"&FORMAT([Date],"Q"), "Numero du mois", MONTH([Date]), "Mois", "M"&FORMAT([Date],"M"), "Annee mois", FORMAT([Date],"YYYYMM"), "Nom du mois", FORMAT([Date],"MMMM YY"), "Nom court du mois", FORMAT([Date],"MMM YY"), "Semaine", "S" & FORMAT(WEEKNUM([Date],2),"00"), -- mise en place de la semaine ISO 8601 (WEEKNUM([Date],21) "Semaine ISO", "W" & FORMAT(WEEKNUM([Date],21),"00"), -- mise en place de la colonne année-semaine ISO (pour le cumul hebdomadaire) "Annee_Sem_ISO", VAR An = YEAR([Date]) VAR Sem = WEEKNUM([Date],2) VAR SemISO = WEEKNUM([Date],21) RETURN IF(AND (SemISO < 5, Sem > 50), An+1, IF(AND (SemISO > 50, Sem < 5), An-1, An)) & "-W" & FORMAT(SemISO,"00"), "Numero du jour", WEEKDAY([Date],2), "Nom du jour", FORMAT([Date],"DDDD"), "Jour ouvre", IF(WEEKDAY([Date],2)<6,"Ouvre","Week-end")) </pre>

Figure 4. 13 : Requête DAX pour la création de la table du temps

2.3.2 Implémentation du modèle

Une fois les mesures et les transformations DAX sont finalisées, la dernière étape est la génération du modèle final de notre Data Warehouse. Cette étape s'intéresse à créer les liaisons entre les tables selon le principe des tables de fait et de dimension ; dans notre cas, la table Sinistre est notre table de Fait et les autres tables sont des tables de dimension et sous dimension.

La figure suivante présente notre modèle tabulaire final, ce dernier va être importé par la suite depuis Power BI afin de réaliser nos visuels :

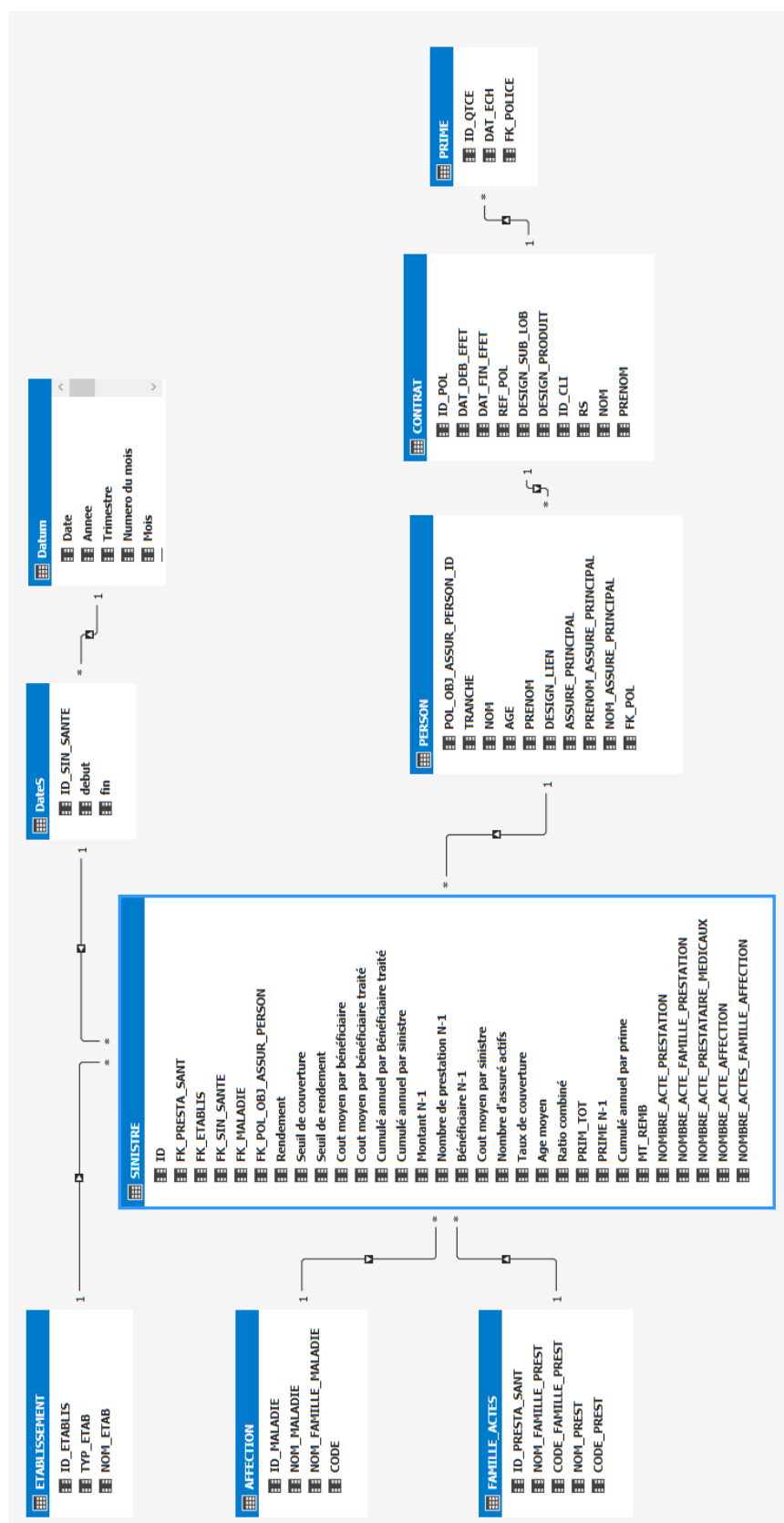


Figure 4. 14 : modèle tabulaire SSAS

2.4 Réalisation des tableaux de bord

Dans cette partie, nous allons voir les différentes étapes de la réalisation de nos tableaux de bord, commençant par la connexion avec la source de données, puis la mise en place des différents visuels et enfin les tests effectués afin de partager les Dashboards sur PowerBI service et les rendre accessibles.

2.4.1 Connexion avec la source de données

Sur PowerBI Desktop, on commence par la connexion avec la source de données à travers l'onglet « Obtenir les données », ce dernier offre la possibilité de se connecter à plusieurs sources distinctes. Pour notre cas, on a effectué une connexion avec la source SQL Server Analysis Services (SSAS) où nous avons déjà créé notre modèle tabulaire des données.

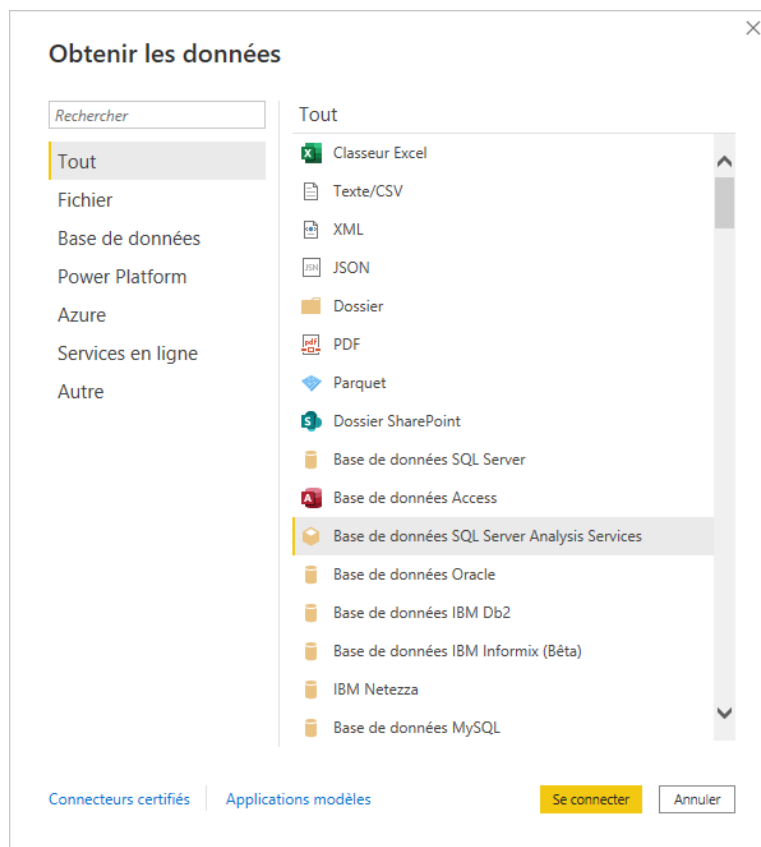


Figure 4. 15 : Connexion de données PowerBI Desktop

La connexion avec SSAS permet d'apporter le modèle tabulaire qui nous servira de créer les tableaux de bord qui seront présentées par la suite.

2.4.2 Présentation des tableaux de bord

Afin d'assurer à l'entreprise une supervision continue de leur activité, la visualisation des tableaux de bord est la fonctionnalité principale pour toute plateforme décisionnelle.

Nous réalisons ces Dashboards de manière à suivre tous les indicateurs présentés précédemment qui permettent de répondre à un besoin précis.

Pour nos Dashboards, il existe quatre filtres essentiels à savoir :

- Type de contrat : individuel ou groupe ;
- Produit : qui désigne le pack vendu par la compagnie d'assurance ;
- Police : les différentes références des contrats ;
- Date : date de début et fin des sinistres.

On va présenter par la suite quelques Tableaux de bord qu'on a pu réaliser lors de notre stage de fin d'études.

➤ Dépenses par tranches d'Age

Ce tableau de bord contient quatre figures d'analyse :

Le premier représente les montants remboursés par tranche d'âge où on peut observer les tranches avec les bénéficiaires qui ont profité d'un montant élevé des sinistres, puis le deuxième visuel s'intéresse à l'effectif des bénéficiaires par tranches d'âge afin d'extraire les tranches les plus actives pour l'assurance. En plus d'un visuel qui représente les couts moyens par tranches d'âge qui montre à quel point les montants par bénéficiaire sont élevés pour chaque tranche d'âge.



➤ Etat de synthèse

Ces deux tableaux de bord englobent toutes les informations essentielles sur les rapports qui se trouvent sur L'**Annexe**

Ce premier état de synthèse contient six visuels d'analyse :

Le premier est sous la forme d'une arborescence de décomposition qui représente les montants remboursés par familles de maladies, de même ces derniers sont décomposés afin de les détailler en montants remboursés par maladies. Puis le deuxième représente les tranches d'âge et les statuts de bénéficiaires les plus sinistrés. Ce tableau de bord comporte aussi un graphe en anneau qui présente les trois types d'établissements avec un pourcentage du montant remboursé élevé et deux graphes mensuels en courbe où l'un montre le taux de couverture des bénéficiaires par rapport au seuil (75%) et l'autre représente le taux des bénéficiaires traités. Enfin le dernier visuel sous forme d'un histogramme représente les montants remboursés et le ratio combiné afin de vérifier la rentabilité d'une police d'assurance par mois.

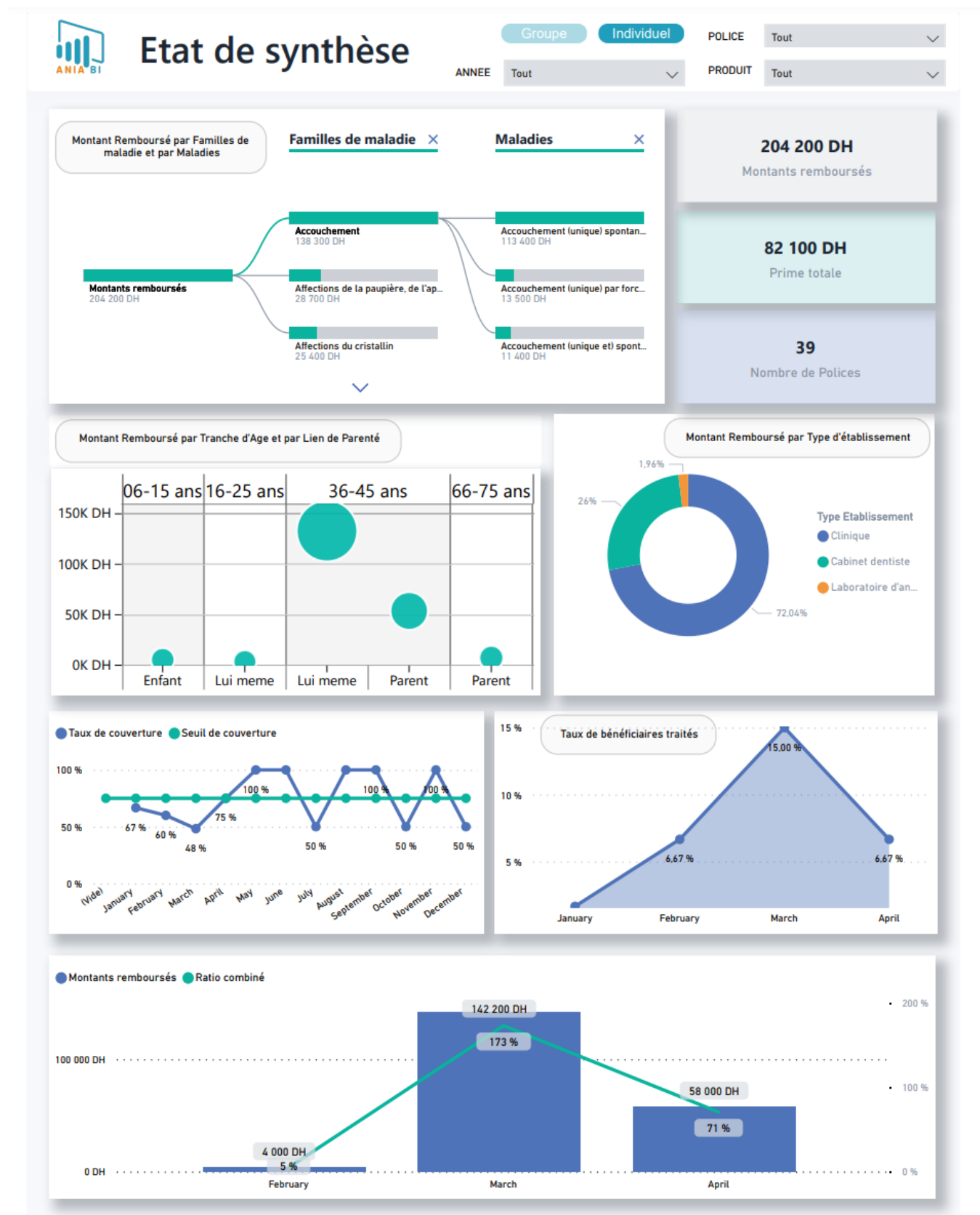


Figure 4. 17 : Premier état de synthèse

Ce deuxième état de synthèse contient aussi six visuels où chacun d'eux se focalise sur un axe d'analyse différents :

Pour ce tableau de bord, il présente un premier visuel qui montre les primes et le nombre d'assurés actifs pour chaque mois. Ensuite un diagramme en courbe où on observe le rendement et son seuil afin de surveiller la performance bénéficiaire de l'assurance. Puis un tableau qui joue un rôle déterminant dans le développement de la stratégie de la compagnie d'assurance grâce à l'observation des nombres des contrats par produit. Et un quatrième visuel qui présente un nuage des points de familles de prestation où la taille des points montre le nombre des sinistres pour cette famille, ainsi un TREEMAP qui représente les quatre familles d'assuré les plus sinistrés dans le but de surveiller les fraudes. Et finalement, une carte radar observe les primes par produit afin de déterminer les packs d'assurance les plus commercialisés.



Figure 4. 18 : Deuxième état de synthèse

2.4.3 Test et publication sur Power BI Service

L'un des facteurs clés de la réussite des projets BI est d'une part le niveau de confiance sur les données affichées dans les rapports. Le manque de confiance dans les données réduit l'adoption par les utilisateurs et entraîne souvent des échecs de projets BI, d'une autre part la performance des rapports est un autre problème courant dans les projets BI. La raison des problèmes de performances peut être une mauvaise conception ou des limitations de capacité du système.

Pour notre PFE, on a effectué deux types de test afin d'assurer le bon fonctionnement de nos rapports et la fiabilité de ces données.

➤ Tests fonctionnels BI :

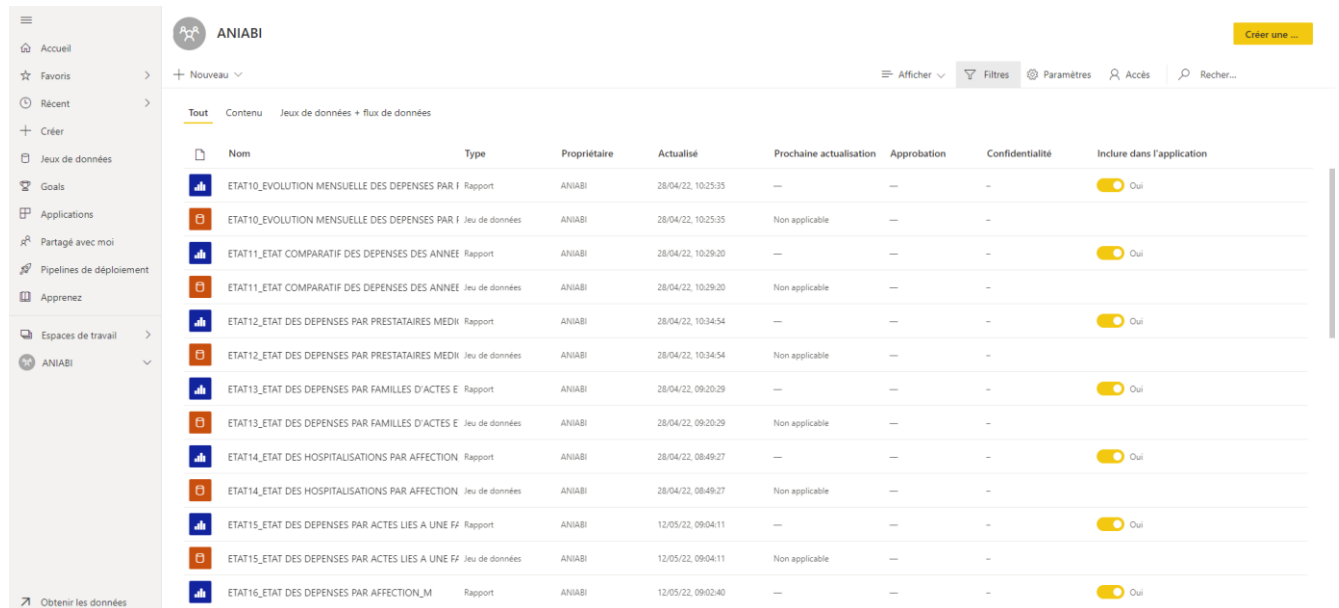
- Vérifier que le nouveau rapport ou tableau de bord est conforme aux exigences de rapport/spécifications de conception ;
- Vérifier les filtres utilisés ;
- Vérifier que les données affichées dans le rapport sont exactes ;
- Vérifier que les délais d'affichage des rapports et des pages de tableau de bord respectent les exigences.

➤ Tests de régression BI :

- L'objectif principal de ce test est de vérifier que les données affichées sur une page de tableau de bord sont les mêmes avant et après la modification (ou la mise à niveau) ;
- Il est utile de comparer la requête SQL de la base de données avec le résultat affiché sur le tableau de bord ;
- Vérifier que la liste des valeurs affichées dans les filtres correspond aux valeurs attendues.

Après la vérification du tableau de bord, on le partage sur PowerBI Service afin qu'il soit exploitable et utilisable pour la solution ANIA.

La figure suivante représente les différents tableaux de bord partagés sur l'espace de travail ANIABI, ce dernier est accessible par le manager afin de vérifier et confirmer les tableaux de bord réalisés.



The screenshot shows the ANIABI dashboard interface. On the left is a sidebar with navigation options: Accueil, Favoris, Récent, Créer, Jeux de données, Goals, Applications, Partagé avec moi, Pipelines de déploiement, Apprenez, Espaces de travail, and ANIABI. The main area displays a table of reports and datasets. The table has columns for Nom, Type, Propriétaire, Actualisé, Prochaine actualisation, Approbation, Confidentialité, and Inclure dans l'application. The data is organized into two tabs: 'Tout' and 'Contenu'. The 'Contenu' tab is active, showing a list of reports and datasets. The reports are categorized by 'Type' (Rapport or Jeu de données) and 'Actualisé' (28/04/22, 10:25:35 or 12/05/22, 09:04:11). The 'Inclure dans l'application' column has a toggle switch for each item, all of which are currently turned on (Oui).

Nom	Type	Propriétaire	Actualisé	Prochaine actualisation	Approbation	Confidentialité	Inclure dans l'application
ETAT10_EVOLUTION MENSUELLE DES DEPENSES PAR F	Rapport	ANIABI	28/04/22, 10:25:35	—	—	—	Oui
ETAT10_EVOLUTION MENSUELLE DES DEPENSES PAR F	Jeu de données	ANIABI	28/04/22, 10:25:35	Non applicable	—	—	Oui
ETAT11_ETAT COMPARATIF DES DEPENSES DES ANNEE	Rapport	ANIABI	28/04/22, 10:29:20	—	—	—	Oui
ETAT11_ETAT COMPARATIF DES DEPENSES DES ANNEE	Jeu de données	ANIABI	28/04/22, 10:29:20	Non applicable	—	—	Oui
ETAT12_ETAT DES DEPENSES PAR PRESTATAIRES MEDI	Rapport	ANIABI	28/04/22, 10:34:54	—	—	—	Oui
ETAT12_ETAT DES DEPENSES PAR PRESTATAIRES MEDI	Jeu de données	ANIABI	28/04/22, 10:34:54	Non applicable	—	—	Oui
ETAT13_ETAT DES DEPENSES PAR FAMILLES D'ACTES E	Rapport	ANIABI	28/04/22, 09:20:29	—	—	—	Oui
ETAT13_ETAT DES DEPENSES PAR FAMILLES D'ACTES E	Jeu de données	ANIABI	28/04/22, 09:20:29	Non applicable	—	—	Oui
ETAT14_ETAT DES HOSPITALISATIONS PAR AFFECTION	Rapport	ANIABI	28/04/22, 08:49:27	—	—	—	Oui
ETAT14_ETAT DES HOSPITALISATIONS PAR AFFECTION	Jeu de données	ANIABI	28/04/22, 08:49:27	Non applicable	—	—	Oui
ETAT15_ETAT DES DEPENSES PAR ACTES LIES A UNE F	Rapport	ANIABI	12/05/22, 09:04:11	—	—	—	Oui
ETAT15_ETAT DES DEPENSES PAR ACTES LIES A UNE F	Jeu de données	ANIABI	12/05/22, 09:04:11	Non applicable	—	—	Oui
ETAT16_ETAT DES DEPENSES PAR AFFECTION_M	Rapport	ANIABI	12/05/22, 09:02:40	—	—	—	Oui

Figure 4. 19 : Liste des tableaux de bord sur PowerBI Service

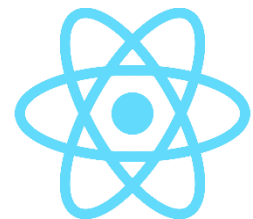
3 Intégration des Dashboards dans la solution ANIA

La dernière étape de notre projet consiste à intégrer les tableaux de bord déjà partagés sur PowerBI Service sur la solution web ANIA afin de les rendre utilisables par les responsables des assurances.

3.1 Technologies

Comme mentionné sur la première partie de notre rapport, la solution ANIA est une suite progicielle full web développée avec React. Donc afin d'intégrer ces différents tableaux de bord, on a besoin de travailler avec les technologies suivantes :

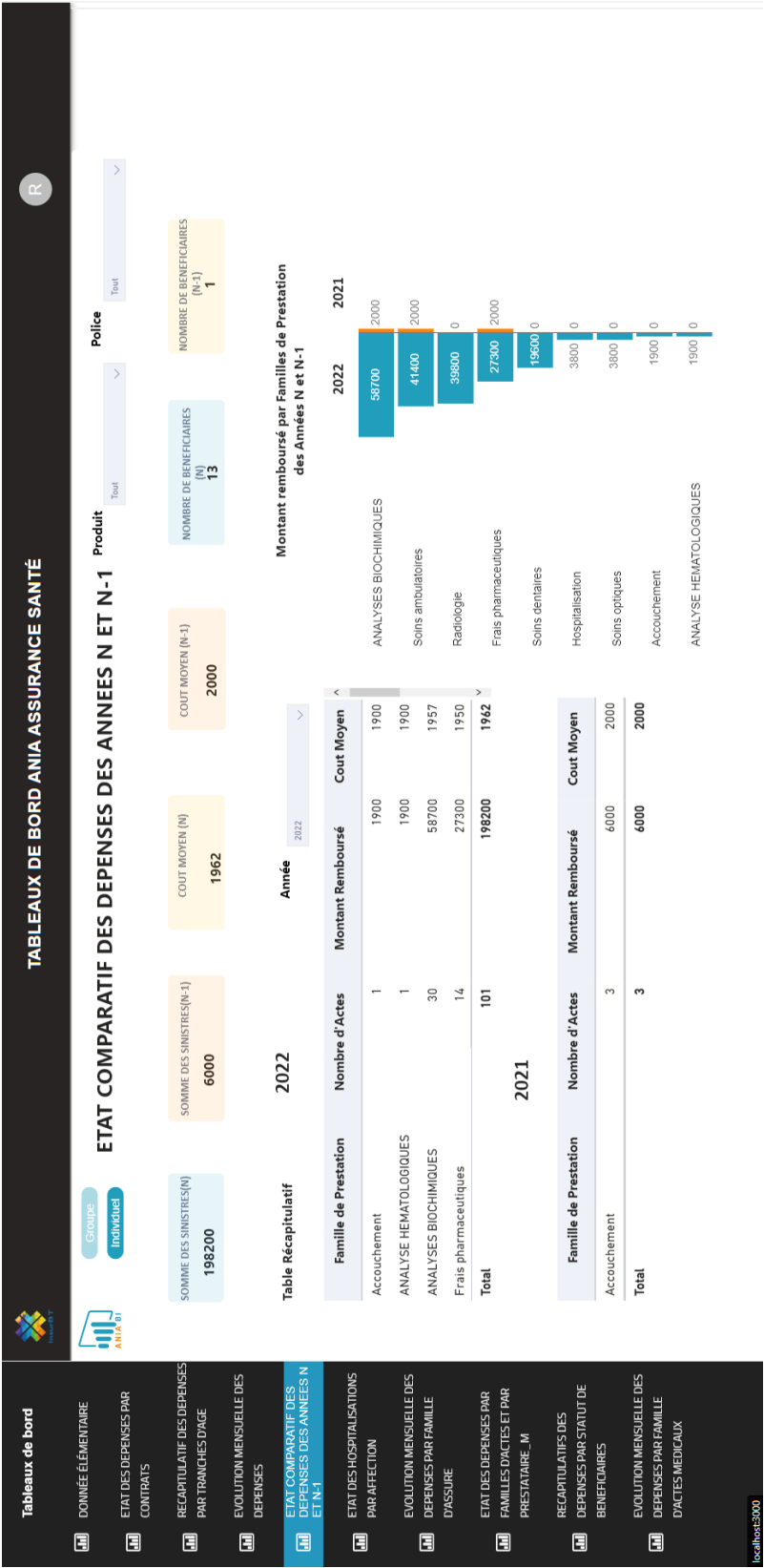
- ReactJS : est la bibliothèque JavaScript open source utilisée pour créer des interfaces utilisateur spécifiquement pour les applications Web et mobiles. Cette Framework est basée sur JSX qui permet de mélanger HTML avec Javascript.
- Power BI Embedded : est un service Azure qui permet aux développeurs d'intégrer du contenu Power BI (tableaux de bord, rapports, vignettes et visuels) dans une application web.



3.2 Interface graphique

Afin d'intégrer nos rapports sur la solution ANIA, on a utilisé un Wrapper open source qui permet aux utilisateurs d'intégrer facilement du contenu Power BI dans des applications basées sur React. Ce dernier consiste en une bibliothèque de composants appelée powerbi-client-react, qui élimine le besoin d'écrire de grandes quantités de code.

La figure suivante représente l'intégration de « l'état comparatif des dépenses des années N et N-1 » sur l'application web :



Conclusion

Au terme de notre projet de fin d'études au sein de la société InsurBT, on a pu atteindre tous les objectifs fixés par le cahier de charge qui nous a été accordé en réalisant une solution décisionnelle pour la plateforme ANIA Assurance Santé afin de garantir aux dirigeants des assurances une meilleure visibilité sur leurs activités ainsi ils pourront aligner leurs objectifs sur la vision stratégique de l'entreprise.

Notre mission consistait en l'étude, l'analyse, la conception ainsi que la mise en place de la solution décisionnelle ; pour ce faire nous avons adopté une méthodologie agile SCRUM. Cette dernière même si elle trouve ses origines dans le développement des logiciels, elle est, aujourd'hui, en train de faire ses preuves dans d'autres types de gestion de projet à savoir les projets BI. En fait, la répartition des tâches sur de courtes phases de travail avec une réévaluation fréquente et régulière du travail nous a aidé énormément à avancer correctement sur le projet.

De même, ce travail nous a permis de mettre en pratique nos connaissances en matière de la Business Intelligence en suivant sa chaîne complète et plus précisément l'approche Kimball. Dans ce sens, nous avons eu l'occasion de maîtriser les outils de la suite MSBI notamment SSIS pour la réalisation du processus ETL, SSAS pour la génération des modèles tabulaires et finalement Power BI pour la phase de la visualisation et l'exploitation des données traitées.

Également, ce projet était, pour nous, une découverte du métier des assurances qui reste un secteur assez particulier en termes de fonctionnement de ses entreprises, puisque les compagnies d'assurance ont des méthodes de gestion spécifiques adaptées à la nature de leur activité.

Finalement, notre projet de fin d'études s'est intéressé par la partie assurance santé, alors en perspectives, nous proposons d'élargir l'espace pour inclure les différents types d'assurance à savoir l'assurance de vie et de prévoyance et l'assurance automobile, techniquement, répartir les Datamarts selon ces trois types (secteurs) d'assurance.

Références

➤ Webographies

- [1] : <https://guidingmetrics.com/content/insurance-industrys-18-most-critical-metrics/>
- [2] : [https://www.shareweb.ch/site/EI/Documents/FSD/Topics/Micro-Insurance/PERFORMANCE%20INDICATORS%20-%20A%20Handbook%20for%20Microinsurance%20Practitioners%20-%202010\(fr\).pdf](https://www.shareweb.ch/site/EI/Documents/FSD/Topics/Micro-Insurance/PERFORMANCE%20INDICATORS%20-%20A%20Handbook%20for%20Microinsurance%20Practitioners%20-%202010(fr).pdf)
- [3] : <https://www.blogdumoderateur.com/role-scrum-master-product-owner-equipe-developpement/>
- [4] : <https://www.wallstreetmojo.com/power-bi-vs-tableau-vs-qlikview/>
- [5] : <https://docs.microsoft.com/en-us/documentation/>
- [6] : <https://www.gartner.com/en/research/magic-quadrant>

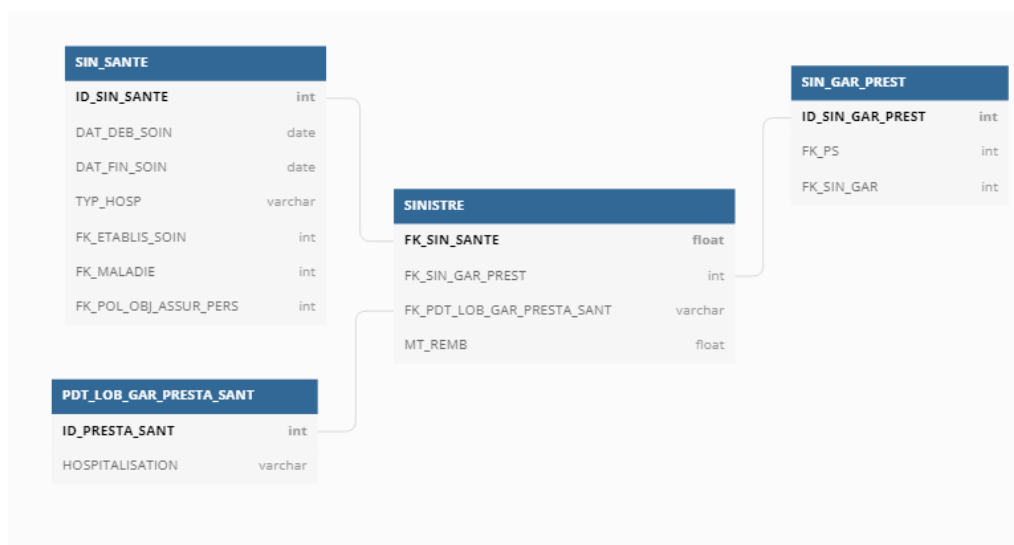
➤ Bibliographies

- [7] : Mastering Power BI Build Business Intelligence Applications Powered with DAX Calculations, Insightful Visualisation, Advanced BI Techniques, and Loads of Data Sources
- [8] : Pro Power BI Architecture Sharing, Security, and Deployment Options for Microsoft Power BI Solutions
- [9] : Beginning DAX with Power BI The SQL Pro's Guide to Better Business Intelligence

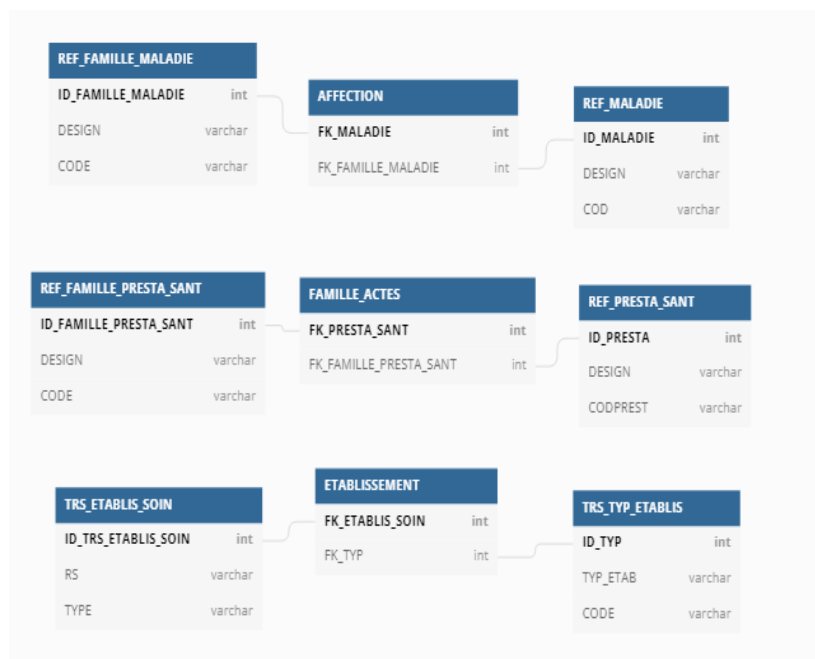
Annexes

Annexe 1

Modélisation du Datamart « Sinistres » :



Modélisation du Datamart « Détails prestation » :



Lors de notre projet de fin d'études, on a réalisé 17 Tableaux de bord en plus des Dashboards de synthèse ; on a présenté dans ce rapport 3 parmi eux et cette annexe présentera 6 autres Dashboards.

<div> <div>ANAF</div> <div> <div>Groupe</div> <div>Individuel</div> </div> </div>	ETAT COMPARATIF DES DEPENSES DES ANNEES N ET N-1			Produit	Police
				Tout	Tout

SOMME DES SINISTRES(N)

198200

SOMME DES SINISTRES(N-1)

6000

COUT MOYEN (N)

1962

COUT MOYEN (N-1)

2000

NOMBRE DE BENEFICIAIRES (N)

13

NOMBRE DE BENEFICIAIRES (N-1)

1

Table Récapitulatif

2022

Année

2022

Famille de Prestation	Nombre d'Actes	Montant Remboursé	Cout Moyen
Accouchement	1	1900	1900
ANALYSE HEMATOLOGIQUES	1	1900	1900
ANALYSES BIOCHIMIQUES	30	58700	1957
Frais pharmaceutiques	14	27300	1950
Total	101	198200	1962

2021

Famille de Prestation	Nombre d'Actes	Montant Remboursé	Cout Moyen
Accouchement	3	6000	2000
Total	3	6000	2000

Montant remboursé par Familles de Prestation des Années N et N-1

2022

2021

ANALYSES BIOCHIMIQUES	58700	2000
Soins ambulatoires	41400	2000
Radiologie	39800	0
Frais pharmaceutiques	27300	2000
Soins dentaires	19600	0
Hospitalisation	3800	0
Soins optiques	3800	0
Accouchement	1900	0
ANALYSE HEMATOLOGIQUES	1900	0

—



- EVOLUTION MENSUELLE DES DEPENSES PAR FAMILLE D'ASSURE



- ETAT DES DEPENSES PAR FAMILLES D'ACTES ET PAR PRESTATATAIRE



Montant Remboursé par Prestataires et par Familles de Prestation



Famille de prestation

- Accouchement
- ANALYSE HE...
- ANALYSES...
- Frais pharmac...
- Hospitalisation
- Radiologie
- Soins ambul...

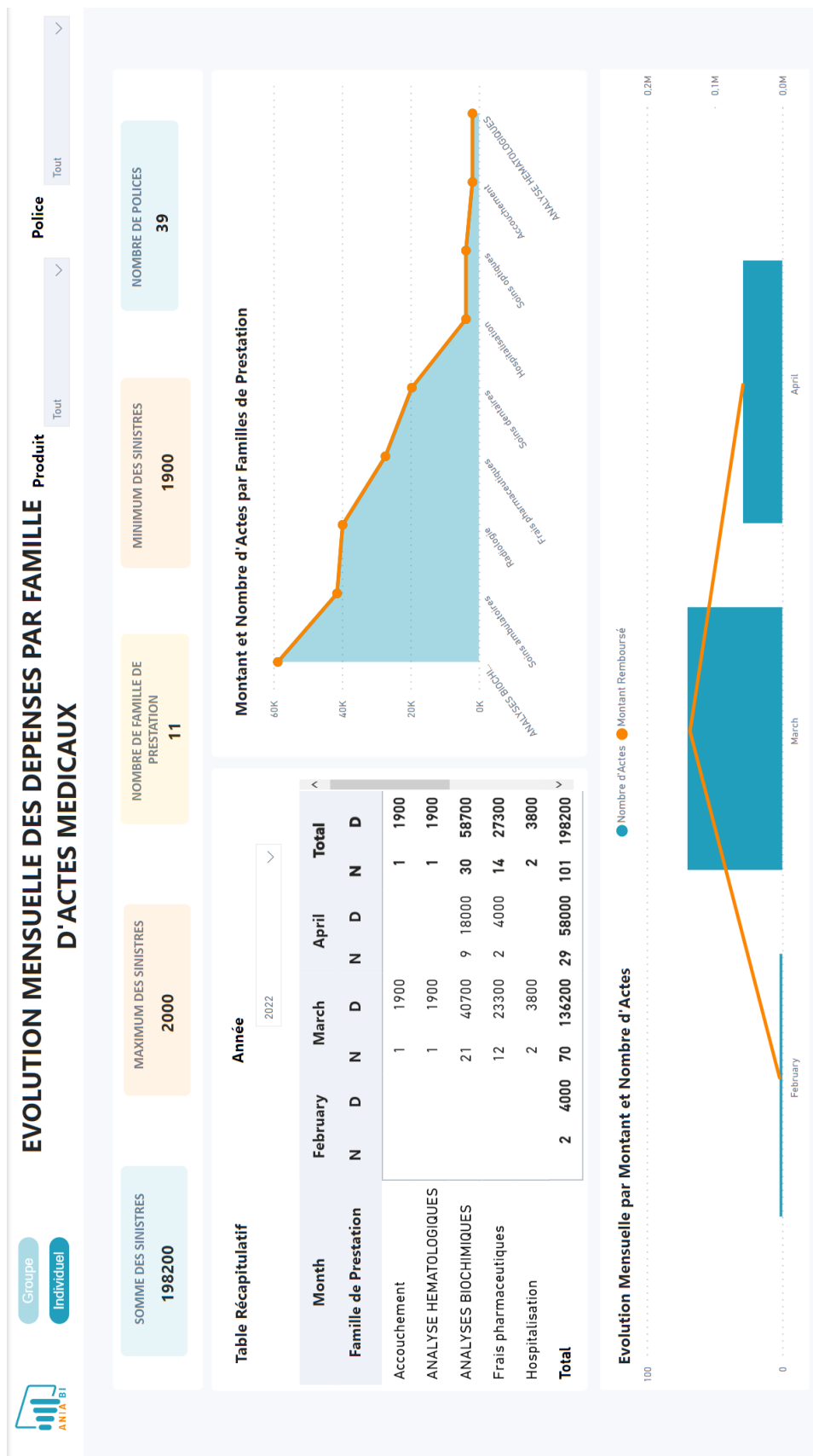
Les 3 Prestataires les plus sinistrés



POLYCLINIQUE BENJELLOUN: 147100

OMFAM CABINET D...: 53100

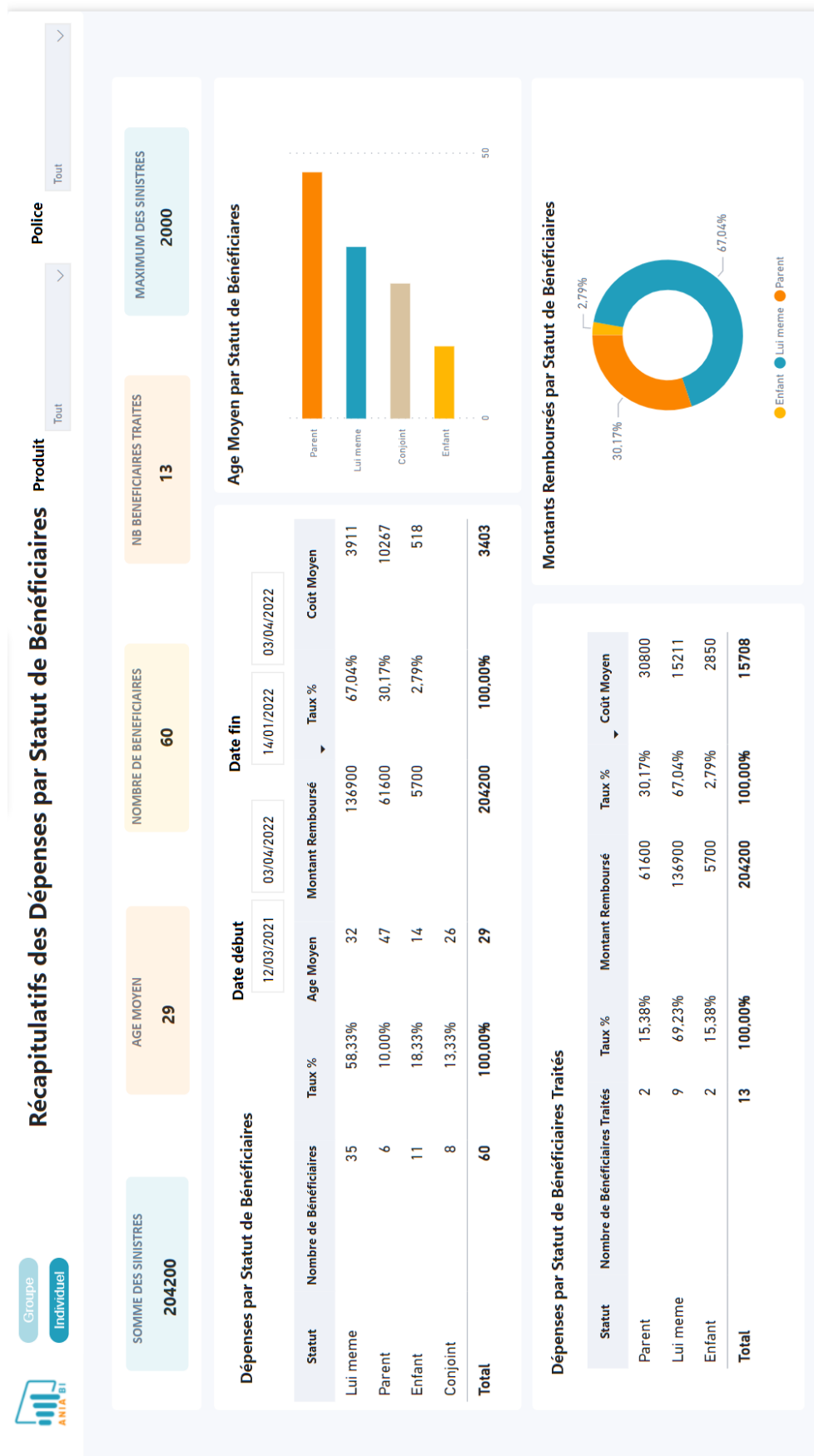
- EVOLUTION MENSUELLE DES DEPENSES PAR FAMILLE D'ACTES MEDICAUX



- ETAT DES HOSPITALISATIONS PAR AFFECTION



- RECAPITULATIFS DES DEPENSES PAR STATUT DE BENEFICIAIRES



- EVOLUTION MENSUELLE DES DEPENSE ET DES PRIMES

