

MEMOIRE

Présenté à

**L’Institut Supérieur d’Informatique
Et de Multimédia de Sfax**

En vue de l’obtention du diplôme de

**LICENCE
EN SCIENCES INFORMATIQUES**

Intitulé

**Mise en place d’une application décisionnelle
pour la gestion des cliniques de CliniSys**

Par

Sirine BOUZIDI & Taha DAOUD

Soutenu le 09/06/2023, devant le jury composé de :

M. Ahmed KHARRAT

Président

Mme Mouna KTARI

Membre

Mme Ines ZOUARI

Encadrante

Mme Saida DHIB

Invitée Entreprise



DÉDICACE 1

Avec l'expression de ma reconnaissance et mon amour, je dédie ce travail

À l'homme, ma précieuse offre de Dieu, mon cher papa **Ahmed**.

À la femme, qui a souffert sans me laisser souffrir, la plus douce des mamans **Saida**.

Les fruits de l'arbre que vous avez planté et n'avez jamais cessé d'arroser, commencent à mûrir, j'espère que ce mémoire constitue pour vous un signe de fierté. Que Dieu, le tout puissant vous garde, vous procure santé et bonheur et vous accorde une longue vie.

À mes chères frères **Achraf** et **Ayoub**, à tous les moments d'enfance passés avec vous, en gage de ma profonde estime pour l'aide que vous m'avez apporté. Puissent nos liens fraternels se consolider et se pérenniser encore plus.

À mon encadrante professionnel **Mme Saida dhib**, je ne peux pas trouver les mots justes et sincères pour vous exprimer mon affection, merci pour votre aide et votre soutien.

À tous ceux que j'aime et à toutes les personnes qui m'ont prodigué des encouragements et ont donné la peine de me soutenir durant cette formation, **Emna, Ghalya, Melek, Achraf, Rakia, Ghada, Amira, Sirine et Hanen** Aucun hommage ne pourrait être à la hauteur de ses sacrifices.

Mille Mercis...



DÉDICACE 2

Je dédie ce modeste travail,

À mes chers parents,Sameh Sellami et Habib Daoud

pour la reconnaissance que je leur dois, pour leur présence, leurs conseils et leur aide inestimable dans tout ce que j'entreprends. Ce travail n'est que le fruit de leurs énormes sacrifices, qu'ils puissent trouver dans ce travail l'expression de ma grande gratitude et affection.

A mon Frère Yessin et ma binôme Sirine Bouzidi, mes amis Mariam,ma chère Nada pour leurs supports, leurs présences et leurs conseils tout au long de mon parcours. A toute ma famille et tous mes amis qui en aucun moment, n'ont cessé de m'épauler et m'encourager à toujours donner le meilleur de moi- même.

Aux personnels de Clinisys, Mme Saida dhib qui grâce à eux j'ai été toujours motivé et déterminé à m'améliorer et me battre pour ma réussite académique

REMERCIEMENTS

Après avoir rendu grâce à Dieu le tout puissant et le miséricordieux, qui nous a donné la force et la persévérence d'accomplir ce travail.

Nous tenons à exprimer nos gratitude et nos vifs remerciements à tous ceux qui nous ont aidé à réaliser ce stage dans des conditions favorables, à tous ceux qui n'ont cessé de nous prodiguer leurs conseils et encouragements à travers ces lignes : Ce travail n'aurait pu aboutir à un résultat sans une réelle collaboration et un échange d'idées entre tous ceux qui y ont participé.

Notre grande gratitude s'adresse à notre encadrante universitaire **Mme Ines ZOUARI** qui a accepté de nous encadrer durant cette période de stage. Sa disponibilité, ses orientations, ses conseils et son cil critique nous ont été très précieux pour structurer le travail.

Nous tenons à remercier vivement le personnel de la société **CliniSys**, pour l'opportunité et la confiance qu'il nous ont accordé pour être parmi eux, notre encadrante professionnelle **Mme Saida DHIB** pour ses efforts, le temps qu'elle nous a consacré et pour les précieuses informations qu'elle nous a accomplies avec intérêt et compréhension. Ainsi que toute l'équipe de Clinisys sans oublier l'équipe de développement BI **Mme Saida, Mme Mariem, M. Ayman** pour la chaleureuse ambiance de travail, le temps qu'ils nous ont consacré, leurs directives, et pour la qualité de leurs suivis durant toute la période de notre stage.
nos vifs remerciements vont également aux membres du jury **Mr Ahmed kharrat** et **Mme Mouna ktari** pour le grand honneur qu'ils nous ont fait. Nos remerciements vont enfin à toute personne qui a contribué de près ou de loin à l'élaboration de ce projet.

■ TABLE DES MATIÈRES

INTRODUCTION GÉNÉRALE	1
1 Cadre général du projet	3
1.1 Présentation de l'organisme d'accueil	4
1.2 Présentation de l'application	5
1.2.1 Définition de la mission	5
1.2.2 Objectifs à atteindre	6
1.3 Etude de l'existant	6
1.4 Analyse de l'existant	6
1.4.1 Critique de l'existant	8
1.4.2 Solution proposée	8
1.5 Définition des besoins fonctionnels et non fonctionnels	9
1.5.1 Besoins fonctionnels	9
1.5.2 Besoins non fonctionnels	9
1.6 Cycle de vie d'un projet décisionnel	10
1.7 Plan du déroulement du projet	11
2 Informatique décisionnelle	13
2.1 Définition de l'informatique décisionnelle	14
2.2 L'architecture du décisionnel	14
2.3 Entrepôt de données et magasins de données	16
2.3.1 Définition de l'entrepôt de données	16
2.3.2 Caractéristique de l'entrepôt de donnée	16
2.3.3 Magasins de données (Data Mart)	17
2.4 Modélisation des entrepôts de données	17
2.4.1 Définition de la modélisation multidimensionnelle	18
2.4.2 Concepts de base de la modélisation multidimensionnelle	18
2.4.3 Les schémas multidimensionnels	19
2.5 L'alimentation de l'entrepôt de données	21

TABLE DES MATIÈRES

2.6 La technologie OLAP (Online Analytical Processing)	21
2.7 Le modèle tabulaire	22
2.8 Langage DAX	23
3 Modélisation conceptuelle	24
3.1 Méthodes de conception d'un entrepôt de données	25
3.1.1 Méthode descendante	25
3.1.2 Méthode ascendante	25
3.1.3 Méthode mixte	26
3.2 Modélisation conceptuelle	26
3.2.1 Méthode descendante	26
3.2.1.1 Collecte des données	27
3.2.1.2 Spécification des besoins	28
3.2.1.3 Formalisation des besoins	30
3.2.2 Méthode ascendante	35
3.2.2.1 Définition de faits	35
3.2.2.2 Définition des dimensions	36
3.2.2.3 Définition des hiérarchies	38
3.2.3 Méthode mixte	40
4 Construction de l'entrepôt de données	43
4.1 Processus ETL	44
4.2 Modélisation logique et physique	45
4.2.1 Modèle logique	45
4.2.2 Modèle physique	46
4.2.2.1 Optimisation physique	46
4.2.2.2 Schéma physique	47
4.3 Conception détaillé de l'ETL	47
4.4 Développement de l'ETL	49
4.4.1 Outils utilisées	49
4.4.2 Présentation du travail réalisé	51
5 Restitution des données de l'entrepôt	67
5.1 Modélisation de l'application BI	68
5.1.1 Types d'analyse	68
5.1.2 Types d'application BI	69
5.1.3 Types d'utilisateurs BI	70

TABLE DES MATIÈRES

5.1.4	Modèle conceptuel de l'application BI	70
5.1.4.1	Elaboration du modèle de cas d'utilisation	70
5.1.4.2	Description textuelle des cas d'utilisation	71
5.2	Développement de l'application BI	74
5.2.1	Cube tabulaire de l'entrepôt de données	74
5.2.2	Projet Power BI	77
5.2.2.1	Importation de la base de données	77
5.2.2.2	Présentation du tableau de bord	78
	CONCLUSION GÉNÉRALE	85
	BIBLIOGRAPHIE	86

LISTE DES FIGURES

1.1	Logo de CliniSys	4
1.2	Organigramme général de CliniSYS	5
1.3	Cycle de vie d'un projet décisionnel	10
1.4	Planning prévisionnel	11
2.1	Architecture du système d'information décisionnel	15
2.2	Exemple d'un fait	18
2.3	Exemple d'un modèle en étoile	19
2.4	Exemple d'un modèle en constellation	20
2.5	Exemple d'un modèle en flocon de neige	20
2.6	Cube-OLAP	22
3.1	Les étapes de la modélisation descendante	26
3.2	La matrice des besoins	29
3.3	Simplification de la matrice des besoins	30
3.4	Hiérarchies des dimensions p1 (démarche descendante)	32
3.5	Hiérarchies des dimensions p2 (démarche descendante)	33
3.6	Schéma multidimensionnel en constellation (démarche descendante)	34
3.7	Diagramme de classe de la source de données	35
3.8	Schéma multidimensionnel en constellation (Démarche ascendante)	39
3.9	Schéma multidimensionnel en constellation (démarche mixte)	41
4.1	Le processus ETL	45
4.2	Schéma physique de l'entrepôt de données	47
4.3	Diagramme d'activités pour l'alimentation des tables de dimensions et de faits	48
4.4	Logo SQL Server Management Studio	49
4.5	Logo Microsoft SQL Server Integration Services	50
4.6	Logo Visual Studio 2019	50
4.7	Logo Microsoft Power BI	51
4.8	Logo Enterprise Architect	51

LISTE DES FIGURES

4.9	L'entrepôt de données	52
4.10	Le projet SSIS	53
4.11	Base de données des cliniques	53
4.12	Configuration des variables	54
4.13	Configuration de la partie générale de la tâche d'exécution de requêtes SQL . .	54
4.14	Configuration de la partie jeu de résultats de la tâche d'exécution des requêtes SQL	55
4.15	Configuration de la collection du conteneur de boucle Foreach	56
4.16	Configuration du mappage de variables de conteneur de boucle Foreach	56
4.17	Configuration de la test connexion	57
4.18	Configuration de la test connexion	57
4.19	Flux de contrôle	58
4.20	Projet SSIS validé	58
4.21	Flux de données relative à la dimension "nationalité"	59
4.22	Editeur de transformation de colonne dérivé	60
4.23	Editeur de transformation de la variation lente	61
4.24	Flux de données « F_Ca »	62
4.25	Editeur de transformation de la tache suppression	62
4.26	Editeur de transformation de la tache suppression	63
4.27	Les requêtes SQL de la tache suppression	63
4.28	Résultats du déploiement validé	64
4.29	Création du job dans SQL	65
4.30	La planification du job créé	65
5.1	Diagramme de cas d'utilisation	71
5.2	préparation de la clé primaire dans la table " DMedecin"	74
5.3	Modèle tabulaire	75
5.4	Hiérarchie de la table "DEmployee"	75
5.5	Déploiement du cube tabulaire	76
5.6	Connexion directe de la base de données	77
5.7	Chargement des tables	77
5.8	Le projet power BI	78
5.9	Page d'accueil	79
5.10	Tableau de bord « CA »	79
5.11	Histogramme de chiffre d'affaires par spécialité médecin	80
5.12	Tableau du chiffre d'affaires par service	80

LISTE DES FIGURES

5.13 Tableau du chiffre d'affaires par service	81
5.14 Chiffre d'affaires par type de payement	81
5.15 Tableau de bord « patientèle »	82
5.16 Tableau de bord « Employée »	82
5.17 Graphique en anneau pour le nombre de jour d'absence par trimestre	83
5.18 Tableau de bord " Charge Personnel "	83
5.19 Les filtres de temps et de cliniques	84

Liste des tableaux

5.1	Description du cas d'utilisation « Développer le processus ETL »	72
5.2	Description du cas d'utilisation « Créer les Dashboard »	73
5.3	Description du cas d'utilisation « Consulter les Dashboard »	73

INTRODUCTION GÉNÉRALE

Avec l'évolution rapide de la technologie, les entreprises disposent aujourd'hui de quantités massives de données provenant de sources multiples et variées, telles que les transactions commerciales, les médias sociaux, les enquêtes, les informations sur les clients, les opérations internes, etc. Cependant, cette abondance de données crée également des défis pour les entreprises qui cherchent à comprendre et à tirer parti de ces informations pour prendre des décisions stratégiques informées. Les tableaux de bord permettent de visualiser et d'interpréter les données de manière intuitive et facilement compréhensible, ce qui facilite la prise de décisions basées sur des faits. toutefois, la création de tableaux de bord efficaces peut être complexe et fastidieuse, nécessitant des compétences en analyse de données, en visualisation, en informatique et en gestion de projets. Dans ce sens, l'informatique décisionnelle offre aux entreprises la possibilité d'homogénéiser leurs données et d'y ajouter une dimension intelligente, afin de rendre leurs décisions plus adaptées. Les nombreux avantages de cette technologie incitent un grand nombre d'entreprises à l'adopter et à en faire la base de leurs systèmes d'information. C'est dans ce contexte que s'inscrit notre projet de fin d'études, dont l'objectif est de concevoir et de réaliser un outil d'analyse et de reporting pour le suivi de la gestion financière, la gestion des ressources humaines et la gestion de capacité de trois cliniques. À travers cet outil, nous souhaitons fournir aux décideurs les informations nécessaires pour prendre des décisions éclairées et optimiser leurs activités, en centralisant les données pertinentes, en les analysant de manière approfondie et en les présentant de manière claire et intuitive.

Le présent rapport est structuré en 5 chapitres décrits comme suit :

INTRODUCTION GÉNÉRALE

- Le chapitre 1 décrit une présentation de l'organisme d'accueil, établissant ainsi le cadre général de notre projet, en exposant le contexte et les objectifs à atteindre.
- Le chapitre 2 aborde les fondamentaux de l'informatique décisionnelle.
- Dans le chapitre 3, nous avons traité la modélisation conceptuelle de l'entrepôt de données en suivant une approche mixte.
- Le chapitre 4 se consacre au développement du processus ETL (Extract-Transform-Load) des données. Il présente les sources de données ainsi que les méthodes d'extraction et de chargement des données.
- Le dernier chapitre se focalise sur la restitution des données de notre entrepôt de données. Il met en évidence les fonctionnalités essentielles des outils de reporting que nous avons développés, ainsi que l'architecture technique adoptée. Notre travail est synthétisé par une conclusion générale qui récapitule les grandes lignes de notre solution accompagnée par quelques perspectives.

Cadre général du projet

Sommaire

1.1	Présentation de l'organisme d'accueil	4
1.2	Présentation de l'application	5
1.2.1	Définition de la mission	5
1.2.2	Objectifs à atteindre	6
1.3	Etude de l'existant	6
1.4	Analyse de l'existant	6
1.4.1	Critique de l'existant	8
1.4.2	Solution proposée	8
1.5	Définition des besoins fonctionnels et non fonctionnels	9
1.5.1	Besoins fonctionnels	9
1.5.2	Besoins non fonctionnels	9
1.6	Cycle de vie d'un projet décisionnel	10
1.7	Plan du déroulement du projet	11

Introduction

Le présent chapitre porte sur l'étude préliminaire de notre projet. Nous commencerons par la présentation de l'organisme d'accueil, CLINISYS. Nous présentons ensuite les principales caractéristiques de notre application et nous fixons les principaux objectifs à atteindre. La troisième partie de ce chapitre concerne l'étude de l'existant, les exigences du projet et son cadre.

1.1 Présentation de l'organisme d'accueil

CLINSYS est une société de services et d'ingénierie en informatique SSII. Elle a été créée en Tunisie depuis 1994. Sa filiale à Sfax est située à route manzel chaker km 4,5. CLINISYS ERP est une solution de gestion hospitalière totalement intégrée installée dans 90% des cliniques en Tunisie. Elle est spécialement conçue afin de couvrir l'intégralité des fonctions des établissements de santé. Cet ERP gravite autour du dossier médical informatisé qui est conçu de façon à centraliser, en temps réel, l'ensemble des informations relatives à la prise en charge des patients ce qui permet d'améliorer la qualité des soins. CLINSYS est actuellement le leader en Afrique du Nord dans le développement de systèmes d'information hospitaliers intégrés. Depuis plus de 25 ans, elle a réussi à moderniser la gestion hospitalière dans des environnements durs et très hostiles. Elle a intégré avec succès tous les aspects administratifs, financiers, organisationnels, médicaux et décisionnels avec un produit unique dans plus de 110 hôpitaux en Tunisie, en Égypte, en Libye et au Maroc. Elle migre vers les architectures les plus modernes et vers le Cloud. Elle a innové dans les domaines de l'analyse et l'intelligence artificielle.



FIGURE 1.1 – Logo de CliniSys

Organigramme de CLINISYS La figure 1.2 présente l'organigramme générale de la société CLINISYS

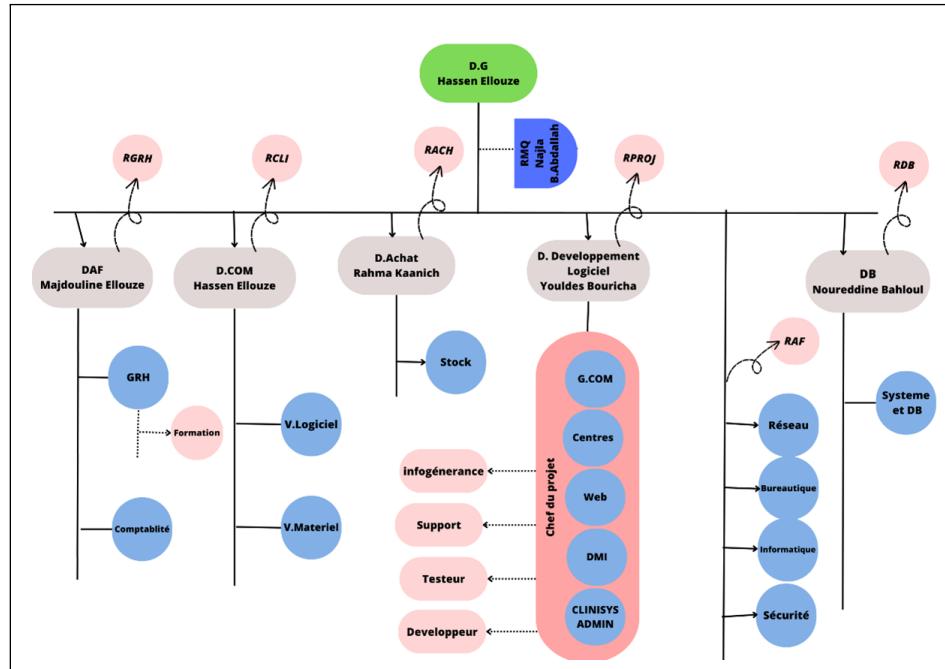


FIGURE 1.2 – Organigramme général de Clinisys

1.2 Présentation de l'application

1.2.1 Définition de la mission

Dans le cadre de notre projet de fin d'études, nous avons été appelés à concevoir et à mettre en place une application décisionnelle autour de la gestion financière, gestion ressources humaines et gestion de la capacité de trois cliniques tunisiens (CliniqueF, CliniqueM, CliniqueP). Il s'agit tout d'abord de construire un entrepôt de données permettant d'analyser pour chaque clinique :

- Le chiffre d'affaires selon le type de prestation, service, société ...
- L'effectif patientèle selon, la nationalité, par type de société ...
- La capacité d'hébergement par service
- Le nombre d'employés selon, la qualification, le type de contrat, l'âge...
- Les salaires et le nombre d'absences des employés.

Ensuite, il s'agit d'exploiter les données de cet entrepôt à travers la création de tableau de bord

1.2.2 Objectifs à atteindre

Les principaux objectifs que nous visons atteindre par notre application sont les suivants :

- Simplifier la collecte fastidieuse des données dispersées dans les bases de données des trois cliniques.
- Créer une seule source de données stable et centralisée qui regroupe les indicateurs d'analyse (chiffre d'affaires, nombre de patients, ...)
- Assurer une vision optimale des données grâce :
 - La suppression des champs superflus provenant des sources de données.
 - L'exclusion des champs qui ne semblent pas pertinents pour l'analyse.
- Faciliter l'analyse de données selon différents critères et à travers plusieurs visuels regroupés dans des tableaux de bord interactifs.
- Permettre aux utilisateurs de filtrer les données pour une analyse plus approfondie.

1.3 Etude de l'existant

L'étude de l'existant vise à comprendre le fonctionnement du système en place, à identifier ses forces et ses faiblesses, ainsi qu'à identifier les opportunités d'amélioration et les problèmes à résoudre. Elle permet de collecter les informations nécessaires pour déterminer les exigences du projet, les besoins des utilisateurs et les attentes des parties prenantes.

1.4 Analyse de l'existant

Il existe certainement des projets de création de tableaux de bord de gestion financière des cliniques. Ces projets peuvent être initiés par des cliniques elles-mêmes, des cabinets de

CADRE GÉNÉRAL DU PROJET

consultants en gestion, des sociétés de développement de logiciels ou des organisations de santé. Les projets de création de tableaux de bord de gestion financière peuvent varier en taille et en complexité, en fonction des besoins de la clinique et de la portée du projet. Les outils utilisés pour créer ces tableaux de bord peuvent varier en fonction des besoins et des préférences de chaque clinique, mais voici quelques exemples courants :

- **Microsoft Excel** : Excel est un logiciel de tableur populaire qui peut être utilisé pour créer des tableaux de bord financiers en utilisant des graphiques, des tableaux croisés dynamiques et des formules.
- **Logiciels de business intelligence (BI)** : les logiciels de BI tels que Power BI, Tableau ou QlikView peuvent être utilisés pour créer des tableaux de bord interactifs et visuels, en intégrant des données provenant de différentes sources
- **Logiciels de comptabilité et de gestion financière** : les logiciels de comptabilité tels que QuickBooks, Sage ou Xero peuvent être utilisés pour extraire des données financières de la clinique et les intégrer dans des tableaux de bord

Avant de mettre en place un système de tableau de bord efficace, il est important de réaliser une étude approfondie de l'existant. Notre société a déjà travaillé sur ce projet par le passé, où l'analyse des données était réalisée à l'aide d'Excel. Cependant, au fil du temps et en raison de la croissance des données et de la complexité croissante des besoins des clients, ils nous ont sollicités pour mettre en place une solution plus robuste et efficace. Ils nous ont expressément demandé d'utiliser un processus ETL (Extraction, Transformation et Chargement) pour intégrer les données provenant de différentes sources et de les rendre cohérentes et exploitables. De plus, ils ont souhaité utiliser Power BI, une plateforme de visualisation de données, pour créer des rapports interactifs et des tableaux de bord dynamiques. Cette transition vers l'ETL et Power BI leur permettra de bénéficier d'une meilleure gestion des données, d'une visualisation claire et d'une analyse approfondie pour prendre des décisions plus éclairées et réactives.

1.4.1 Critique de l'existant

L'objectif du Critique de l'existant dans notre cas est d'identifier les forces et les faiblesses du système existant pour en faire ressortir les aspects à améliorer et les éléments à valoriser.

- L'étude de l'existant a révélé que les données des cliniques sont mal organisées, souvent redondantes, incomplètes et incohérentes, vu qu'elles proviennent de sources différentes.
- En raison de la redondance et des valeurs superflues dans les bases de données, l'exécution des requêtes analytiques est très lente en raison de leur taille importante.

1.4.2 Solution proposée

Une fois les problèmes et les causes identifiés, il est important de proposer des solutions pour améliorer l'organisation des données. Suite à notre critique de l'existant, Il a été observé que la construction d'un entrepôt de données était essentielle pour :

- Récupérer des données provenant des bases de données, des trois cliniques.
- Appliquer les transformations nécessaires sur ces données sources afin de les nettoyer, les intégrer et les rendre cohérentes
- Collecter les données relatives aux chiffres d'affaires, l'étude patientèle, l'étude de capacité d'hébergement et l'étude des ressources humaines dans une seule base (entrepôt de données).
- Créer un tableau de bord en utilisant différents types de graphiques tels que des histogrammes, des graphiques en secteurs et des tableaux pour faciliter l'analyse des chiffres d'affaires du nombre de patients.

1.5 Définition des besoins fonctionnels et non fonctionnels

1.5.1 Besoins fonctionnels

Les besoins fonctionnels de la construction d'un entrepôt de données comprennent la collecte, le stockage, la transformation, l'analyse, la visualisation, la sécurité et l'évolutivité des données, afin de répondre aux besoins de l'entreprise en matière de prise de décision et de gestion des données.

- Collecter des données provenant de sources diverses
- Stocker de grandes quantités de données de manière efficace dans l'entrepôt
- Nettoyer et transformer les données pour les rendre cohérentes et exploitables
- Assurer le rafraîchissement périodique des données de l'entrepôt
- Analyser les données dans des cubes tabulaires
- Visualiser des données en créant des tableaux de bord

1.5.2 Besoins non fonctionnels

Les besoins non fonctionnels incluent la performance, la fiabilité, la facilité d'utilisation, la sécurité, l'extensibilité, l'accessibilité et la maintenabilité, afin de garantir une expérience utilisateur optimale et une utilisation à long terme.

- **Performance** : L'entrepôt de données doit être capable de gérer de grandes quantités de données et de les traiter efficacement pour garantir des performances optimales lors des requêtes et des analyses.
- **Extensibilité** : L'entrepôt de données doit être conçu de manière à pouvoir évoluer facilement en fonction des besoins futurs, notamment en termes de volume de données, de nouvelles sources de données et de besoins d'analyse supplémentaires.
- **Fiabilité** : Il est essentiel que l'entrepôt de données soit fiable et disponible en permanence, car il constitue une source d'informations essentielle pour la prise

de décision. Il doit être capable de gérer les pannes, les erreurs et de récupérer rapidement

- Le tableau de bord doit être performant et réactif pour offrir une expérience utilisateur fluide et sans délai
- Le tableau de bord doit être facile à utiliser pour les utilisateurs, quel que soit leur niveau de compétence en informatique.
- **Maintenabilité :** Le tableau de bord doit être facile à maintenir et à mettre à jour pour garantir sa longévité.

1.6 Cycle de vie d'un projet décisionnel

Le cycle de vie d'un projet décisionnel comprend plusieurs étapes clés, allant de la modélisation conceptuelle des données au développement des applications BI. Comme le montre la figure 1.3.

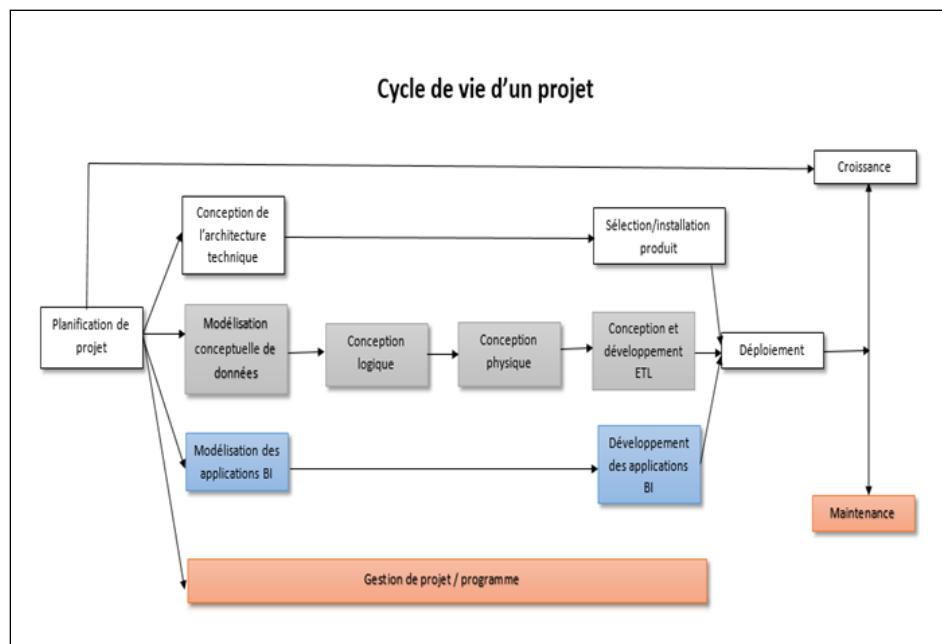


FIGURE 1.3 – Cycle de vie d'un projet décisionnel

Voici une brève description de quelques étapes :

La planification du projet : la planification du projet joue un rôle essentiel. Cette étape permet de définir les objectifs, les ressources, les délais et les livrables du projet, tout en identifiant les risques et en établissant une feuille de route claire pour sa réalisation.

CADRE GÉNÉRAL DU PROJET

Modélisation conceptuelle des données : Cette étape consiste à comprendre les besoins et les objectifs de l'entreprise et à définir les concepts et les relations entre les différentes entités métier. Elle aboutit à la création d'un schéma conceptuel qui représente la structure globale des données.

Conception logique : Cette étape implique la normalisation des données, la définition des clés primaires et étrangères, ainsi que la modélisation de données détaillées pour chaque entité.

Conception et développement ETL (Extraction, Transformation et Chargement) : Une fois la conception physique en place, l'étape suivante consiste à concevoir et développer le processus ETL qui permettra d'extraire les données des sources, de les transformer en un format adapté à l'entrepôt de données et de les charger dans celui-ci.

Modélisation des applications BI : Une fois que les données sont disponibles dans l'entrepôt de données, la modélisation des applications BI intervient pour définir les structures et les relations nécessaires pour fournir des informations significatives aux utilisateurs finaux.

Développement des applications BI : Cela comprend la création de rapports, de tableaux de bord, d'analyses et d'autres fonctionnalités permettant aux utilisateurs finaux d'explorer et d'analyser les données de manière intuitive. Les applications BI sont développées en utilisant les outils appropriés tels que Power BI, Tableau, etc.

1.7 Plan du déroulement du projet

La figure 1.4 présente le planning prévisionnel de notre projet.

Semaine	Mars				Avril				Mai		
	1	2	3	4	1	2	3	4	1	2	3
Spécification des besoins	X	X	X	X							
Modélisation conceptuelle					X	X	X	X			
Conception et développement de l'ETL							X	X	X		
Modélisation et développement de l'application BI									X	X	X
Rédaction du rapport					X	X	X	X	X	X	X

FIGURE 1.4 – Planning prévisionnel

Conclusion

Le premier chapitre de ce rapport a débuté par une présentation de CliniSys, notre organisme d'accueil. Nous avons ensuite expliqué le contexte de notre stage avant de poser la problématique de notre projet. Nous avons mené une analyse de l'existant puis nous avons évoqué des critiques liées aux traitement des données de l'entreprise. Nous avons proposé les solutions envisageables.

Informatique décisionnelle

Sommaire

2.1	Définition de l'informatique décisionnelle	14
2.2	L'architecture du décisionnel	14
2.3	Entrepôt de données et magasins de données	16
2.3.1	Définition de l'entrepôt de données	16
2.3.2	Caractéristique de l'entrepôt de donnée	16
2.3.3	Magasins de données (Data Mart)	17
2.4	Modélisation des entrepôts de données	17
2.4.1	Définition de la modélisation multidimensionnelle	18
2.4.2	Concepts de base de la modélisation multidimensionnelle	18
2.4.3	Les schémas multidimensionnels	19
2.5	L'alimentation de l'entrepôt de données	21
2.6	La technologie OLAP (Online Analytical Processing)	21
2.7	Le modèle tabulaire	22
2.8	Langage DAX	23

Introduction

Dans ce chapitre, nous allons explorer les fondements de l'informatique décisionnelle, qui permet de transformer des données brutes en informations exploitables. Nous commencerons par présenter l'entrepôt de données et ses diverses caractéristiques. Ensuite, nous discuterons des concepts clés de la modélisation multidimensionnelle, qui permet de représenter les données sous forme de tableaux de bord interactifs et facilement compréhensibles. Enfin, nous aborderons les outils ETL (Extract-Transform-Load).

2.1 Définition de l'informatique décisionnelle

Le terme Business Intelligence (BI), ou informatique décisionnelle, désigne les applications, les infrastructures, les outils et les pratiques offrant l'accès à l'information. La BI permet d'analyser l'information pour améliorer et optimiser les décisions et les performances d'une entreprise. En d'autres termes, la Business Intelligence constitue le processus d'analyse des données piloté par la technologie pour découvrir des informations exploitables. Celles-ci aident les chefs d'entreprise et les autres utilisateurs finaux à prendre des décisions plus éclairées. En tant que telle, la BI englobe une grande variété d'outils, d'applications et de méthodologies. Elle permet de collecter des données à partir de systèmes internes et de sources externes, de les préparer à l'analyse, de les développer et d'exécuter des requêtes. Ces outils créent ensuite des rapports, des tableaux de bord et des visualisations de données pour mettre les résultats de l'analyse à la disposition des décideurs [1].

2.2 L'architecture du décisionnel

L'architecture de système d'information décisionnel se compose généralement de plusieurs couches ou composants, chacun ayant une fonction spécifique dans le processus de collecte, d'analyse et de présentation des données. Les principales composantes de l'architecture de l'informatique décisionnelle sont les suivants :

- 1. La source de données :** Cette couche comprend toutes les sources de données de l'entreprise, telles que les bases de données transactionnelles, les fichiers plats, les systèmes ERP (Enterprise Resource Planning), CRM (Customer Relationship Management) ou encore les réseaux sociaux.
- 2. L'ETL (Extract-Transform-Load) :** Cette couche est responsable de l'extraction, de la transformation et du chargement des données de la source vers l'entrepôt de données. Les données sont nettoyées, transformées et structurées pour qu'elles soient cohérentes et exploitable.
- 3. L'entrepôt de données :** Cette couche est la base de l'informatique décisionnelle. Elle stocke les données collectées et organisées pour permettre des analyses multidimensionnelles et des requêtes rapides et efficaces.
- 4. Les outils de traitement des données :** Cette couche comprend des outils pour l'analyse des données, tels que les OLAP (Online Analytical Processing), les data mining, les tableaux de bord et les rapports. Ces outils permettent de transformer les données en informations exploitables et de les présenter de manière conviviale pour l'utilisateur final.
- 5. La couche de présentation :** Cette couche comprend les interfaces utilisateur, les tableaux de bord et les rapports destinés aux utilisateurs finaux. Ces interfaces permettent aux décideurs de visualiser les données et de prendre des décisions éclairées.

La figure 2.1 nous montre les quatre composantes essentielles de l'architecture du système d'information décisionnel.

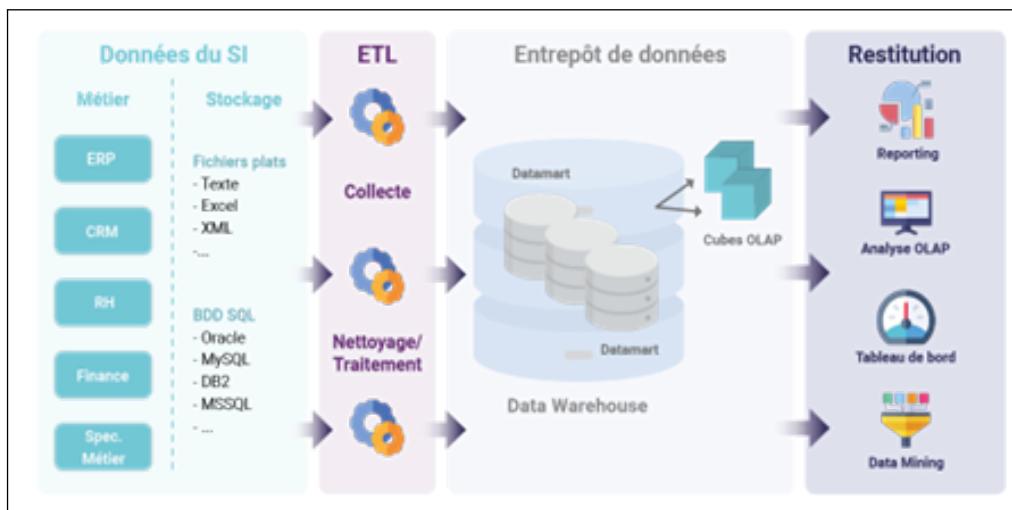


FIGURE 2.1 – Architecture du système d'information décisionnel

2.3 Entrepôt de données et magasins de données

2.3.1 Définition de l'entrepôt de données

Un entrepôt de données est un référentiel centralisé qui stocke des données structurées (tables de base de données, feuilles Excel) et des données semi-structurées (fichiers XML, pages web) à des fins de création de rapports et d'analyse. Les données circulent à partir de diverses sources, telles que les systèmes de point de vente, les applications métier et les bases de données relationnelles, et elles sont généralement nettoyées et normalisées avant d'arriver dans l'entrepôt. Étant donné qu'un entrepôt de données peut stocker de grandes quantités d'informations, il permet aux utilisateurs d'accéder facilement à une multitude de données historiques, qui peuvent être utilisées pour l'exploration de données, la visualisation des données et d'autres formes de rapports décisionnels [2].

2.3.2 Caractéristique de l'entrepôt de donnée

Voici quelques caractéristiques clés d'un entrepôt de données :

1. **Orientation vers les sujets** : L'entrepôt de données est organisé autour de sujets métier, tels que les ventes, les clients ou les produits, plutôt que d'être organisé en fonction de l'application ou du service qui les a générés.
2. **Intégration des données** : L'entrepôt de données intègre des données provenant de diverses sources et les organise de manière cohérente, en supprimant les redondances et en assurant la qualité des données.
3. **Données historiques** : L'entrepôt de données stocke des données historiques qui permettent de suivre l'évolution des données dans le temps, facilitant ainsi l'analyse des tendances et des modèles.
4. **Architecture spécifique** : L'entrepôt de données est conçu selon une architecture spécifique, comme l'architecture en étoile ou en flocon, qui permet de faciliter l'analyse de données à grande échelle.

5. **Accès rapide aux données :** L'entrepôt de données est optimisé pour fournir un accès rapide et efficace aux données, permettant ainsi aux utilisateurs de trouver rapidement les informations dont ils ont besoin.
6. **Séparation des données et des applications :** L'entrepôt de données est conçu pour séparer les données des applications qui les ont générées, permettant ainsi aux données d'être utilisées par de multiples applications et de multiples utilisateurs.
7. **Support des outils d'analyse :** L'entrepôt de données est conçu pour prendre en charge une variété d'outils d'analyse et de visualisation, facilitant ainsi l'analyse des données par les utilisateurs.

2.3.3 Magasins de données (Data Mart)

Une fonction opérationnelle particulière d'une organisation. Contrairement à un entrepôt de données, qui peut contenir des données de l'ensemble de l'organisation, un data Mart est généralement plus petit et plus ciblé, contenant des données spécifiques à un groupe d'utilisateurs ou à une unité métier particulière.

Un data Mart est souvent utilisé lorsque les besoins en données d'un groupe d'utilisateurs sont spécifiques et limités à un domaine particulier, et que la création d'un entrepôt de données complet serait trop coûteuse ou trop complexe. Les data marts peuvent être construits de manière autonome ou en tant que sous-ensembles d'un entrepôt de données plus large, et ils peuvent être utilisés pour l'analyse des données, la prise de décision, la gestion des performances ou d'autres tâches liées aux données dans l'entreprise.

2.4 Modélisation des entrepôts de données

La modélisation de l'entrepôt de données est le processus de conception et de structuration d'un entrepôt de données, qui est une base de données centralisée et intégrée destinée à stocker des données provenant de diverses sources. Cette modélisation implique l'identification et la définition des différents types de données qui doivent être stockés, ainsi que la façon dont ils sont liés les uns aux autres.

2.4.1 Définition de la modélisation multidimensionnelle

La modélisation multidimensionnelle est une technique de modélisation de l'entrepôt de données qui permet de représenter les données sous forme de dimensions et de mesures, plutôt que sous forme de tables relationnelles traditionnelles. Elle est utilisée pour organiser et analyser des données complexes et multidimensionnelles.

2.4.2 Concepts de base de la modélisation multidimensionnelle

Les concepts de base de la modélisation multidimensionnelle sont :

- **Le fait** : Un fait est un centre d'intérêt décisionnel. Il regroupe un ensemble d'attributs numériques représentant les mesures d'activité, telles que les ventes, les coûts, les quantités, les temps, les distances, etc.
- **La mesure** : Une mesure est un indicateur d'analyse de type numérique et cumulable. Une mesure est accompagnée d'un ensemble de fonctions d'agrégation qui permettent de l'agréger en fonction des axes d'analyse.

La figure 2.2 nous montre un exemple d'un fait avec ses mesures.

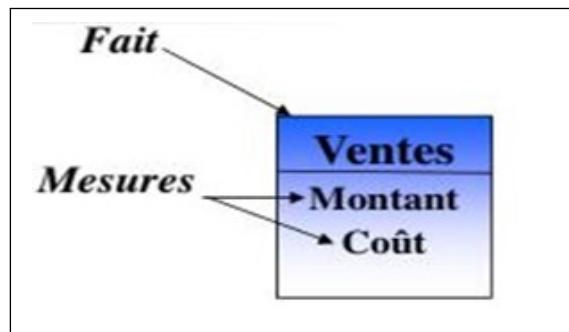


FIGURE 2.2 – Exemple d'un fait

- **La dimension** : Une dimension est un axe d'analyse selon lequel sont visualisées les mesures d'activité d'un sujet d'analyse, tels que le temps, le produit, le lieu, le client, le fournisseur, etc..
- **un paramètre** : Un paramètre est un attribut appartenant à une dimension. Il représente un niveau de détail selon lequel sont visualisées les mesures d'activité d'un sujet d'analyse.

- **L'attribut faible :** Un attribut faible est un descripteur de paramètre. Cet attribut n'est pas utilisé dans les calculs de regroupement lors des opérations d'agrégation ; il a un rôle informationnel permettant de faciliter les analyses.
- **Les hiérarchies :** Une hiérarchie est une perspective d'analyse définie dans une dimension. Elle regroupe un ensemble de paramètres organisés de la granularité la plus fine vers la granularité la plus générale, tels qu'année > mois > jour pour la dimension du temps.

2.4.3 Les schémas multidimensionnels

Les schémas multidimensionnels sont des représentations graphiques utilisées pour modéliser les données dans un entrepôt de données selon une approche multidimensionnelle. Il existe plusieurs types de schémas multidimensionnels, parmi lesquels on peut citer :

- **Le schéma en étoile :** Il s'agit d'un schéma simple et facile à comprendre, dans lequel les mesures sont regroupées dans un seul fait relié à plusieurs dimensions regroupant les paramètres de l'analyse.

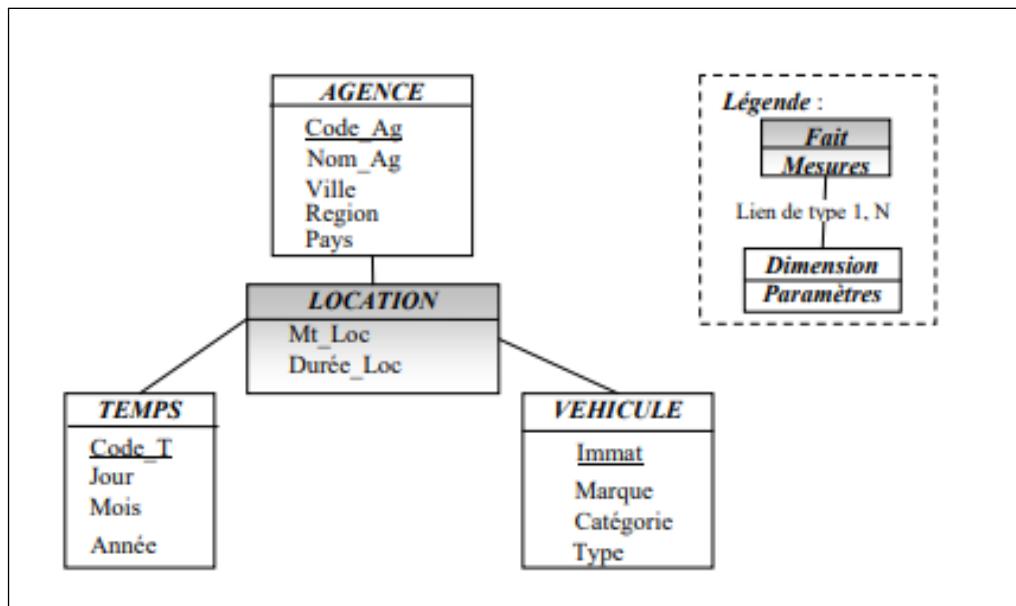


FIGURE 2.3 – Exemple d'un modèle en étoile

- **Le schéma en constellation :** Ce schéma est une extension du schéma en étoile. Il consiste à fusionner plusieurs schémas en étoile qui utilisent des dimensions communes. Un schéma en constellation comprend donc plusieurs faits reliés à un ensemble de

dimensions qui peuvent être partagées. Ce schéma présente l'avantage de pouvoir corrélérer les sujets d'analyse tels que la comparaison des montants des locations réalisées dans les différentes agences par rapport aux chiffres d'affaires réalisés par son personnel. En outre, le partage des dimensions par plusieurs faits permet d'éviter de les définir plusieurs fois.

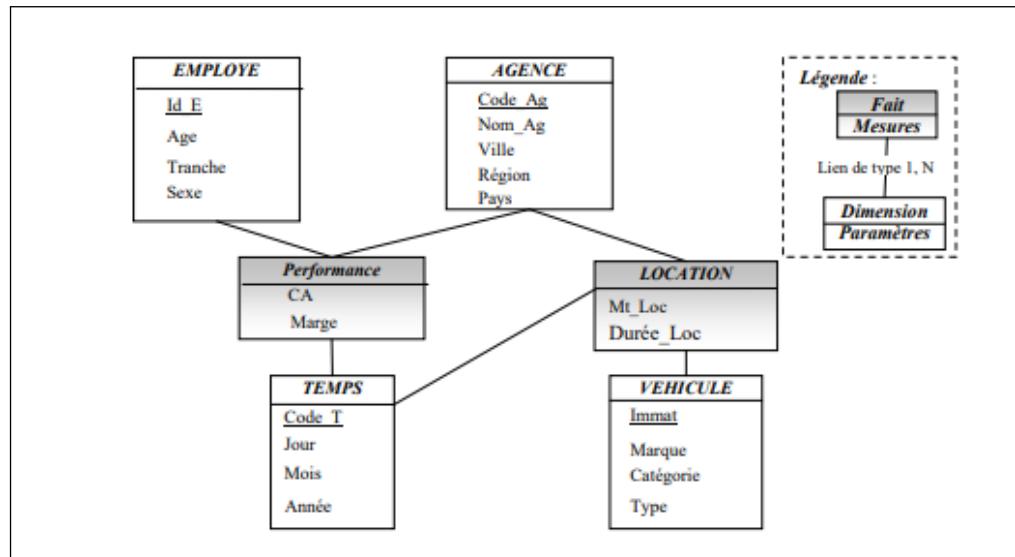


FIGURE 2.4 – Exemple d'un modèle en constellation

- **Le schéma en flocon de neige** : Il s'agit d'une variante du schéma en étoile, dans lequel les tables de dimensions sont normalisées, c'est-à-dire qu'elles sont divisées en sous-tables pour éviter les redondances. Cette approche permet de réduire la taille des tables et d'améliorer les performances de requête, mais peut rendre le schéma plus complexe à comprendre [3].

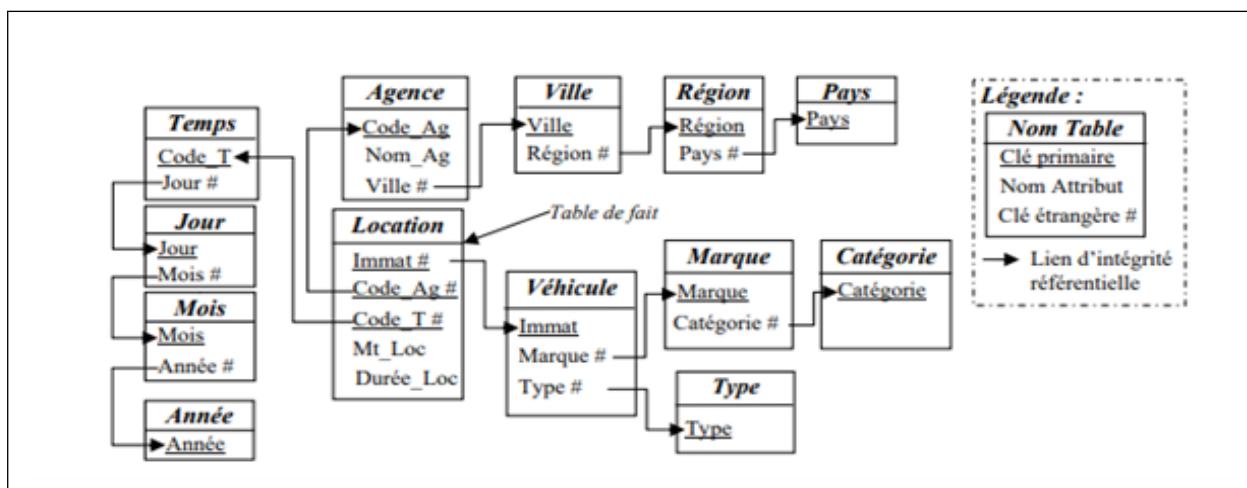


FIGURE 2.5 – Exemple d'un modèle en flocon de neige

2.5 L'alimentation de l'entrepôt de données

L'alimentation de l'entrepôt de données est le processus qui consiste à extraire, transformer et charger (ETL) les données à partir de sources hétérogènes vers l'entrepôt de données. Le processus ETL est une étape critique dans la création et la maintenance de l'entrepôt de données, car il permet de garantir l'intégrité, la qualité et la cohérence des données. Voici les étapes du processus ETL :

- **Extraction** : cette étape consiste à extraire les données à partir de sources externes, telles que des bases de données, des fichiers plats, des flux de données en temps réel, etc. L'extraction peut être effectuée manuellement ou automatiquement à l'aide de logiciels d'extraction de données.
- **Transformation** : cette étape consiste à transformer les données extraites en les nettoyant, en les enrichissant, en les structurant et en les conformant à la structure de l'entrepôt de données. Les transformations peuvent inclure des opérations telles que la fusion, la déduplication, la normalisation, l'enrichissement, la restructuration, etc.
- **Chargement** : cette étape consiste à charger les données transformées dans l'entrepôt de données. Le chargement peut être effectué à l'aide de plusieurs techniques, telles que le chargement en bloc, le chargement incrémental, le chargement en temps réel.

2.6 La technologie OLAP (Online Analytical Processing)

OLAP : pour online analytical processing est une technologie d'organisation de grandes bases de données commerciales qui prend en charge des analyses multidimensionnelles. Elle peut être utilisée pour effectuer des requêtes analytiques complexes sans affecter négativement les systèmes transactionnels. Par exemple, il est possible de rapidement obtenir le nombre de ventes puis de rajouter d'autres dimensions comme les dates, les régions géographiques ou d'autres caractéristiques de la vente ce qui prenait auparavant un temps conséquent. Cette forme de pré-tri des données permet de grandement réduire le temps entre une requête dans la base de

données et l'achèvement de cette dernière. La figure nous montre un exemple d'un cube OLAP [4].

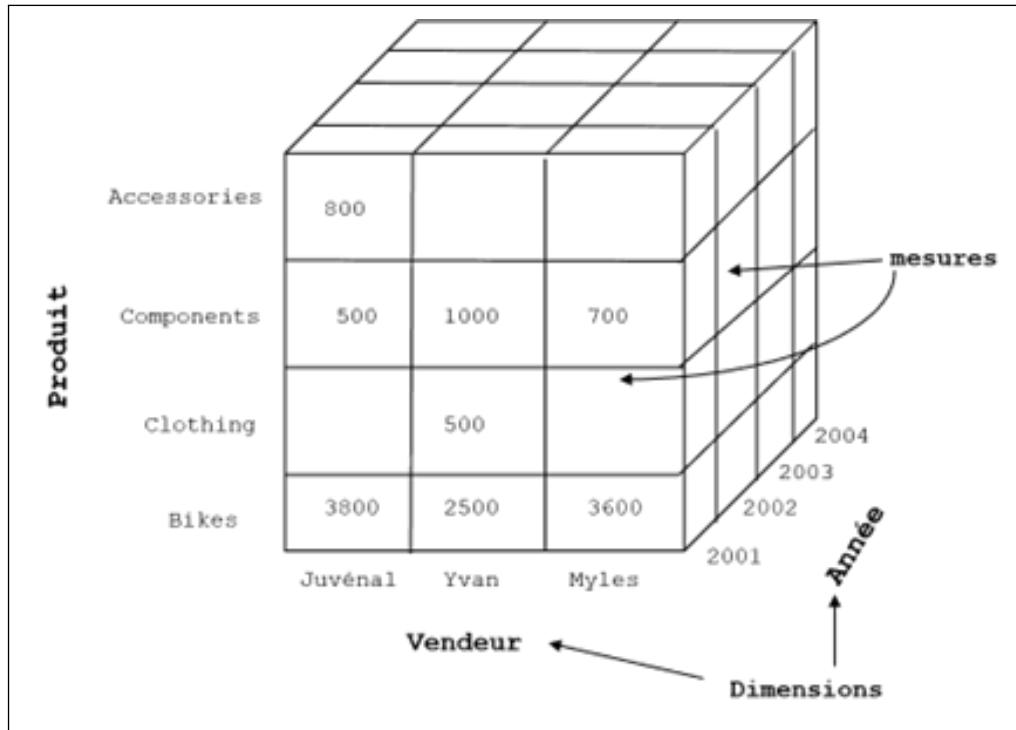


FIGURE 2.6 – Cube-OLAP

2.7 Le modèle tabulaire

Un cube tabulaire, également connu sous le nom de modèle tabulaire, est une composante clé de l'outil de traitement analytique SQL Server Analysis Services (SSAS) . Il s'agit d'une structure de données multidimensionnelle utilisée pour l'analyse des données et la création de rapports. il est conçu pour fournir une expérience d'exploration interactive et rapide des données. les autres cubes multidimensionnels sont basés sur un modèle de données en étoile ou en flocon, avec des agrégations précalculées, et sont adaptés aux scénarios d'analyse complexes. Les cubes tabulaires utilisent un modèle de données en colonnes, avec des données en mémoire et des calculs dynamiques, et sont adaptés aux scénarios d'analyse ad hoc et interactifs nécessitant des performances élevées et une flexibilité accrue. Le choix entre les deux dépend des besoins spécifiques de l'entreprise et des caractéristiques des données à analyser.

2.8 Langage DAX

Le langage DAX (Data Analysis Expression) est un langage d'analyse de données très puissant. Il est utilisé par Power Pivot dans Excel, Power BI Desktop et SSAS (SQL Server Analysis Services) tabulaire. Il s'agit d'un langage inspiré en partie des fonctions Excel mais adapté à la réalité d'analyse de bases de données. Il permet notamment la création d'indicateurs sur mesure répondant aux problématiques des utilisateurs métiers. Il est facile de créer un classeur et d'y importer des données. Vous pouvez même créer des tableaux croisés dynamiques ou des graphiques croisés dynamiques qui affichent des informations importantes sans utiliser de formules DAX [5].

Conclusion

En somme, le chapitre sur l'informatique décisionnelle met en lumière l'importance croissante des technologies et méthodes d'analyse de données pour aider les entreprises à prendre des décisions plus éclairées. En s'appuyant sur des outils tels que les entrepôts de données, les tableaux de bord et les modèles d'analyse prédictive, l'informatique décisionnelle aide les entreprises à comprendre leur environnement et à identifier les opportunités et les risques. Bien que l'adoption de ces technologies puisse poser des défis, les avantages qu'elles apportent sont cruciaux pour la prise de décision efficace et la réussite à long terme de l'entreprise.

Modélisation conceptuelle

Sommaire

3.1 Méthodes de conception d'un entrepôt de données	25
3.1.1 Méthode descendante	25
3.1.2 Méthode ascendante	25
3.1.3 Méthode mixte	26
3.2 Modélisation conceptuelle	26
3.2.1 Méthode descendante	26
3.2.2 Méthode ascendante	35
3.2.3 Méthode mixte	40

Introduction

Dans ce chapitre, nous allons aborder la conception d'un entrepôt de données. Nous présenterons les différentes étapes de sa conception afin de permettre une exploitation efficace des données et une amélioration de la qualité de la prise de décision.

3.1 Méthodes de conception d'un entrepôt de données

3.1.1 Méthode descendante

Dans (Kimball et al, 2002) différentes études de cas de bases dimensionnelles sont proposées. La modélisation dimensionnelle est basée sur le schéma en étoile et ses différentes variations (schéma en flocon et schéma en constellation). En outre, une méthode, appelée architecture en matrice de BUS, est proposée pour la construction d'un schéma dimensionnel à partir de la définition des besoins des utilisateurs (Kimball et al, 2002). Cette méthode permet de collecter les différents sujets d'intérêt dans l'entreprise et de les combiner avec les différents axes d'analyse pour former une matrice. Cette matrice globale englobe plusieurs sujets d'analyse et l'ensemble de leurs dimensions. Ces travaux ne proposent pas de méthode formelle de conception et de construction d'une base dimensionnelle. Ainsi, nous ne trouvons ni une démarche formelle de spécification des besoins décideurs, ni un outil de transformation de ces besoins dans la matrice proposée..

3.1.2 Méthode ascendante

(Golfarelli et al, 1998) propose le modèle dimensionnel des faits et une méthode semi-automatique de conception du schéma dimensionnel à partir d'un schéma EntitéAssociation décrivant les sources. Cette méthode se base sur le schéma logique des données décrivant les sources opérationnelles car souvent la documentation des schémas Entité-Association est incomplète (Golfarelli et al, 2002). Cette méthode est basée sur les trois niveaux d'abstraction : conceptuel, logique et physique. Au niveau conceptuel, la méthode propose un ensemble d'étapes pour la définition des faits, des dimensions et des hiérarchies à

partir du schéma de la source. Le fait représente les événements fréquents dans le monde de l'entreprise. Les dimensions sont formées à partir d'une portion du schéma source qui dépend de la relation représentant le fait. Les attributs de cette portion sont extraits puis réorganisés et épurés pour former les hiérarchies [6].

3.1.3 Méthode mixte

La méthode mixte est la fusion des deux approches : ascendante et descendante. Elle intègre à la fois les données des sources opérationnelles et les besoins des décideurs dans la définition du modèle dimensionnel. En effet, elle résout les problèmes rencontrés dans les méthodes précédentes pour satisfaire les différents besoins utilisateurs tout en respectant les données sources.

3.2 Modélisation conceptuelle

3.2.1 Méthode descendante

La figure 3.1 nous montre la démarche de la méthode descendante en se basant sur les besoins des décideurs et sur les règles de gestion relatives aux données décisionnelles. Cette démarche se base sur trois étapes :

- La collecte des données
- La spécification des besoins
- La formalisation des besoins

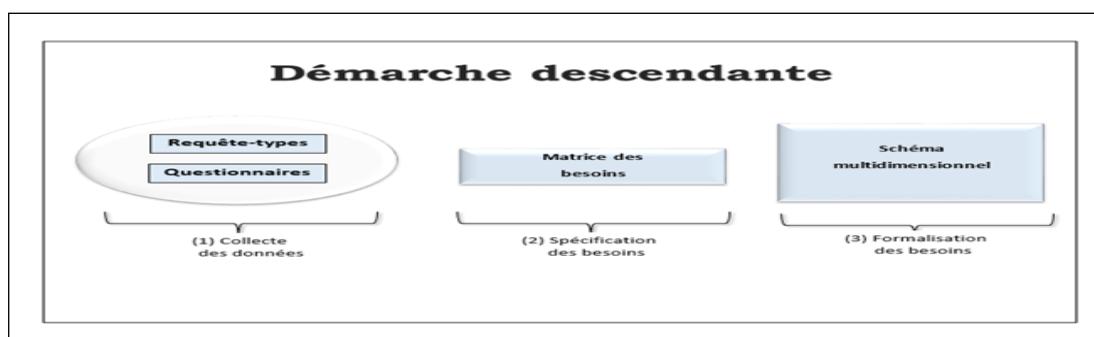


FIGURE 3.1 – Les étapes de la modélisation descendante

3.2.1.1 Collecte des données

À cette étape, nous cernons les besoins décisionnels initiaux en identifiant les types d'information qui peuvent intéresser chaque groupe de décideurs. On procède alors à :

- La collecte des requêtes-types pertinentes en interviewant les décideurs.
- La mise au point d'un questionnaire permettant de mieux caractériser et identifier les besoins des décideurs.
- La clause « Analyser » répond à la question Quoi ? - Elle définit les données que les décideurs souhaitent analyser.
- La clause « En fonction de » répond aux questions Qui ? Où ? Et quand ? - Elle indique les paramètres d'analyse des données décrites par la clause Analyser. Ainsi, nous avons déterminé les requêtes-types suivantes :

R1 : Analyser le chiffre d'affaires

En fonction de clinique, type de payeur, famille de prestation.

R2 : Analyser le chiffre d'affaires.

En fonction de la nationalité de la patientèle

R3 : Analyser le chiffre d'affaires.

En fonction de service.

R4 : Analyser le chiffre d'affaires.

En fonction des patients couverts par la CNAM..

R5 : Analyser le nombre des patients par nationalité.

R6 : Analyser le nombre des patients par clinique.

R7 : Analyser le nombre d'admission par service, par nationalité.

R8 : Analyser le nombre de patients par société, par nationalité, pour la société CNAM, nationalité tunisienne.

R9 : Analyser le nombre de lits par clinique, service.

R10 : Analyser le nombre de chambre par clinique, service.

R11 : Analyser le nombre d'employées par clinique.

R12 : Analyser le nombre d'employées par type de contrat.

R13 : Analyser le nombre d'employées par qualification.

R14 : Analyser le salaire brut d'employées par qualification.

R15 : Analyser l'âge moyen des employées par qualification.

R16 : Analyser le moyen de taux d'absences des employées par année.

3.2.1.2 Spécification des besoins

En sortie de l'étape précédente, nous avons obtenu une liste de requêtes-types formulées à l'aide de notre pseudo langage et un ensemble de règles de gestion. L'étape de spécification permet d'analyser les données collectées afin de spécifier les besoins des décideurs. Ces besoins seront organisés en termes de paramètres et de mesures afin de préparer la définition du schéma multidimensionnel.

Construction de la matrice des besoins La matrice des besoins est un outil important pour la planification et la gestion de projet, car elle permet d'identifier et de hiérarchiser les différents besoins du projet, en fonction de leur importance et de leur complexité. Cela permet de mieux allouer les ressources et de garantir que toutes les fonctionnalités requises sont fournies. Le figure constitue la matrice des besoins de notre application.

MODÉLISATION CONCEPTUELLE

	CA	Clinique	Type de payeur	Famille_prestation	Nationalité	Service	Nbre_patients	Nbre_Adm	Société	Nbre_lits	Nbre_chambres	Nbre_employée	Contrat	Qualification	Age_moyen	Salaires	Nbr_Absences	Jour	Mois	Année
CA	X	X	X		X				X										X	
Clinique		X																		
Type_payeur			X																	
Famille_prestation				X																
Nationalité					X															
Service						X														
Nbre_patients	X			X													X	X	X	
Nbre_Adm	X		X	X	X												X	X	X	
société										X										
Nbre_lits	X																X	X	X	
Nbre_chambres	X																X	X	X	
Nbre_employée	X												X	X						
Contrat													X							
Qualification														X						
Age_moyen														X						
Salaire_moyen														X						
Nbr_Absences															X					
Jour																X				
Mois																	X			
Année																		X		

FIGURE 3.2 – La matrice des besoins

La simplification de la matrice des besoins dans la figure 3.3 est réalisée en deux étapes :

- Chaque colonne vide est supprimée de la matrice : En effet, une colonne vide indique que la propriété correspondante ne décrit aucun indicateur d'analyse et donc qu'elle ne fait pas partie des paramètres de l'analyse.
- Chaque ligne vide est supprimée de la matrice : ceci permet d'enlever la propriété correspondante, de la liste des indicateurs. Cette propriété n'est analysée en fonction d'aucun paramètre (sa ligne est vide). Elle ne correspond pas à un indicateur dans notre analyse dimensionnelle

	Nationalité	société	Type_payer	Clinique	Service	Famille_prestation	Contrat	Qualification	Jour	Mois	Trimestre	Année
CA		X	X	X	X	X						X
Nbre_Adm	X			X	X	X			X	X	X	X
Nbre_patients	X	X		X					X	X	X	X
Nbre_chambres				X	X				X	X	X	X
Nbre_lits				X	X				X	X	X	X
Nbre_employée							X	X				
Age_moyen								X				
Salaire_moyen								X				
Nbr_Absences												X

FIGURE 3.3 – Simplification de la matrice des besoins

3.2.1.3 Formalisation des besoins

Après avoir collecté et spécifié les besoins des décideurs, nous réalisons dans cette étape la formalisation de ces besoins sous forme d'un schéma multidimensionnel. La conception de ce schéma est basée sur la matrice des besoins définie dans l'étape précédente.

Définition des faits

La phase de spécification des faits est une étape importante dans de nombreux processus décisionnels et d'analyse de données. Cette phase consiste à déterminer les faits pertinents qui seront utilisés pour prendre des décisions ou analyser une situation donnée. A partir de notre matrice des besoins, nous avons déduit les faits suivants :

- F_Chiffre_Affaires ((Chiffre_Affaire, SUM, MIN, MAX, AVG), (Nbre_Admission, SUM, MIN, MAX, AVG))
- F_Patientèle (Nbr_Patient, SUM, MIN, MAX, AVG)
- F_Capacité ((Nbr_Chambres, SUM, MIN, MAX, AVG), (Nbr_Lits, SUM, MIN, MAX, AVG))
- F_Employé ((Nbr_employé SUM, AVG), (Moyenne_AgeMIN,MAX,SUM,AVG))

MODÉLISATION CONCEPTUELLE

- F_Paie ((SalaireSUM, AVG), (Nbr_AbsenceSUM, MIN, MAX, AVG))

Définition des dimensions

Après l'analyse des dépendances fonctionnelles entre les paramètres nous avons déterminé les dimensions suivantes : Client (Nom)

Clinique (libelle clin)

Nationalité (libelle Nat)

Service (libelle Ser)

Temps (Mois, Année)

Nature Admission (Affiche principal)

Contrat (CodeCont)

Qualification (CodQual)

Enrichissement des dimensions

Une fois que nous avons intégré les nouveaux attributs obtenus à partir des questionnaires réalisés pour renforcer les dimensions, nous établissons la définition des dimensions de la manière suivante :

Client (Id_Client, Nom, Prénom, Type payeur, Nat, Id_Soc, libelle soc, âge)

Clinique (Id_Clinique, libelle clinique)

Nationalité (Id_Nat, libelle Nat)

Service (Id_service, libelle service, Id_famillePrest, libelle famillePrest)

Temps (Jour, Mois, Année)

Nature Admission (Code, Desc, Affiche principal)

Contrat (CodeCont, LibCont)

Qualification (CodQual, LibQual)

Définition des hiérarchies

L'objectif de cette étape est de structurer les paramètres de chaque dimension en utilisant le

MODÉLISATION CONCEPTUELLE

symbole " \rightarrow " pour indiquer l'agrégation et former ainsi des hiérarchies.

H^{temps}	$: Jour \rightarrow Mois \rightarrow Année \rightarrow All$
H_1^{client}	$: Id_Client \rightarrow Id_Soc \rightarrow All$
H_2^{client}	$: Id_Client \rightarrow TypePayeur \rightarrow All$
H_3^{client}	$: Id_Client \rightarrow Nat \rightarrow All$
H_4^{client}	$: Id_Client \rightarrow age \rightarrow All$
H_5^{client}	$: Id_Client \rightarrow Sexe \rightarrow All$
$H^{clinique}$	$: Id_Clinique \rightarrow All$
$H^{nationalit}$	$: Id_Nat \rightarrow All$
$H^{service}$	$: Id_Service \rightarrow Id_Familleprest \rightarrow All$
$H^{natureAdmission}$	$: Code \rightarrow AffichePrincipal \rightarrow All$
$H^{Qualification}$	$: CodQual \rightarrow All$
$H^{Contrat}$	$: CodeCont \rightarrow All$

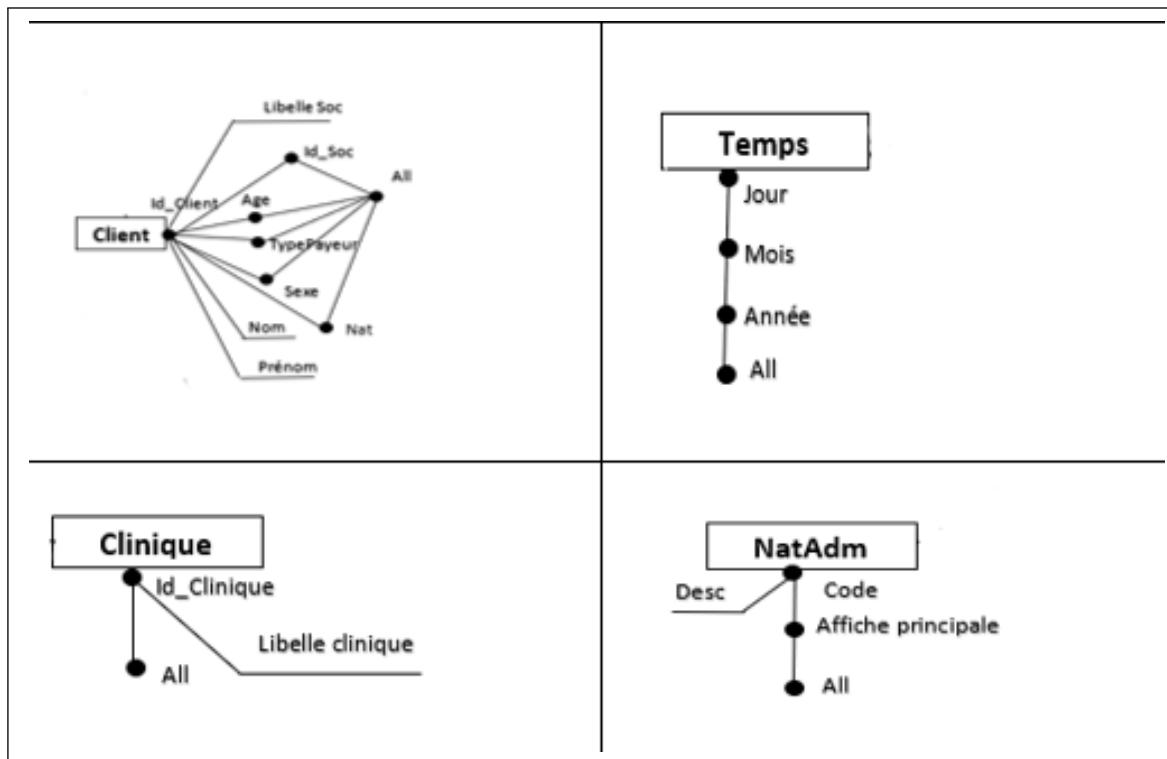


FIGURE 3.4 – Hiérarchies des dimensions p1 (démarche descendante)

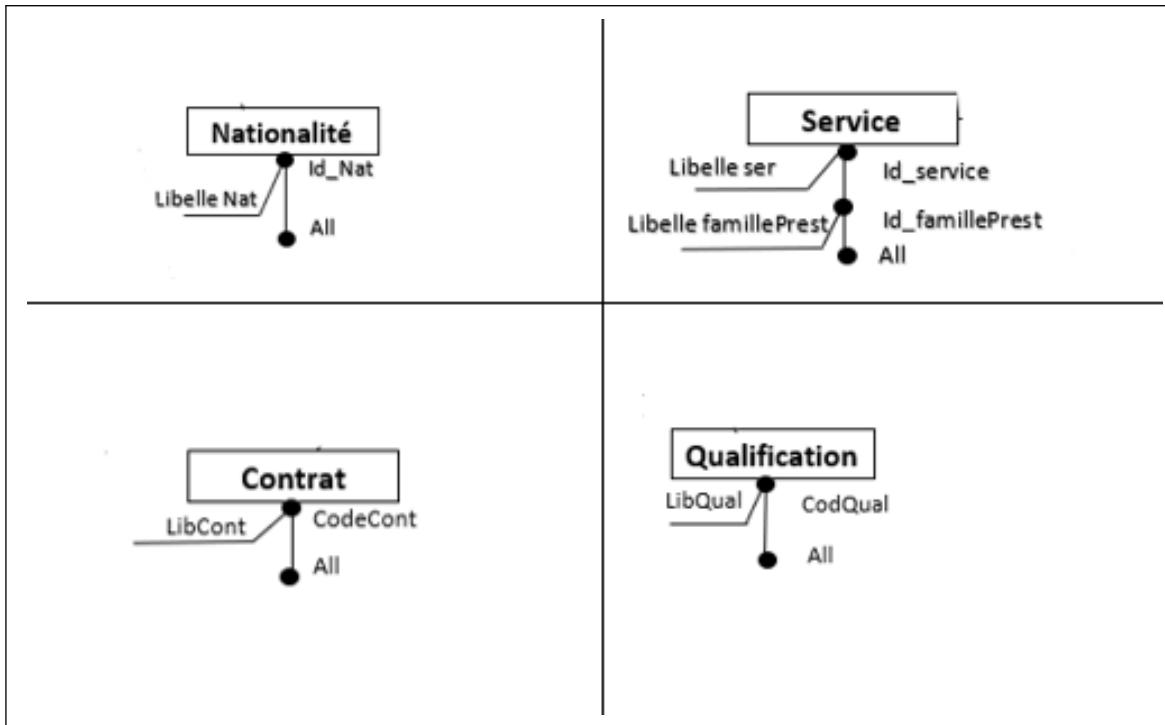


FIGURE 3.5 – Hiérarchies des dimensions p2 (démarche descendante)

Définition du schéma multidimensionnel

En utilisant la matrice des besoins comme guide, nous assignons des dimensions aux faits en examinant les paramètres de chaque dimension qui ont une corrélation avec une mesure donnée d'un fait. Ainsi, chaque fait est associé à ses dimensions pertinentes. La figure 3.6 illustre le schéma multidimensionnel qui est obtenu en utilisant cette approche descendante.

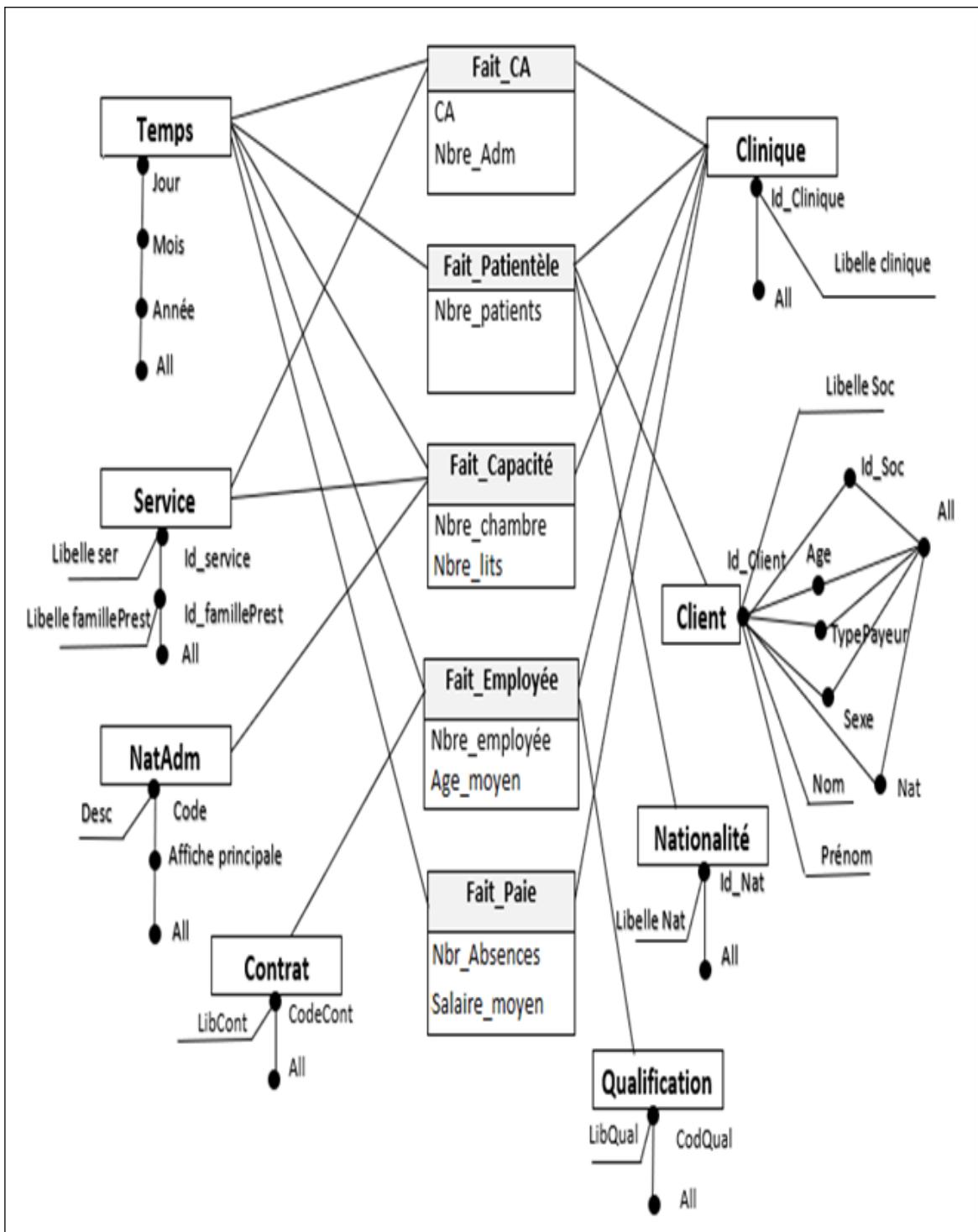


FIGURE 3.6 – Schéma multidimensionnel en constellation (démarche descendante)

3.2.2 Méthode ascendante

3.2.2.1 Définition de faits

Avant d'élaborer la conception de notre schéma multidimensionnel, nous introduisons le diagramme de classe pour nos sources de données. Dans notre contexte, ces sources sont constituées de plusieurs tables distinctes qui ne coexistent jamais pour une même série de données, engendrant ainsi des problèmes d'hétérogénéité. Suite à l'analyse de ces tables, nous avons créé le diagramme de classe suivant afin de mettre en œuvre notre approche de conception ascendante. Le diagramme de classe de la source de données est présenté dans la figure 3.7.

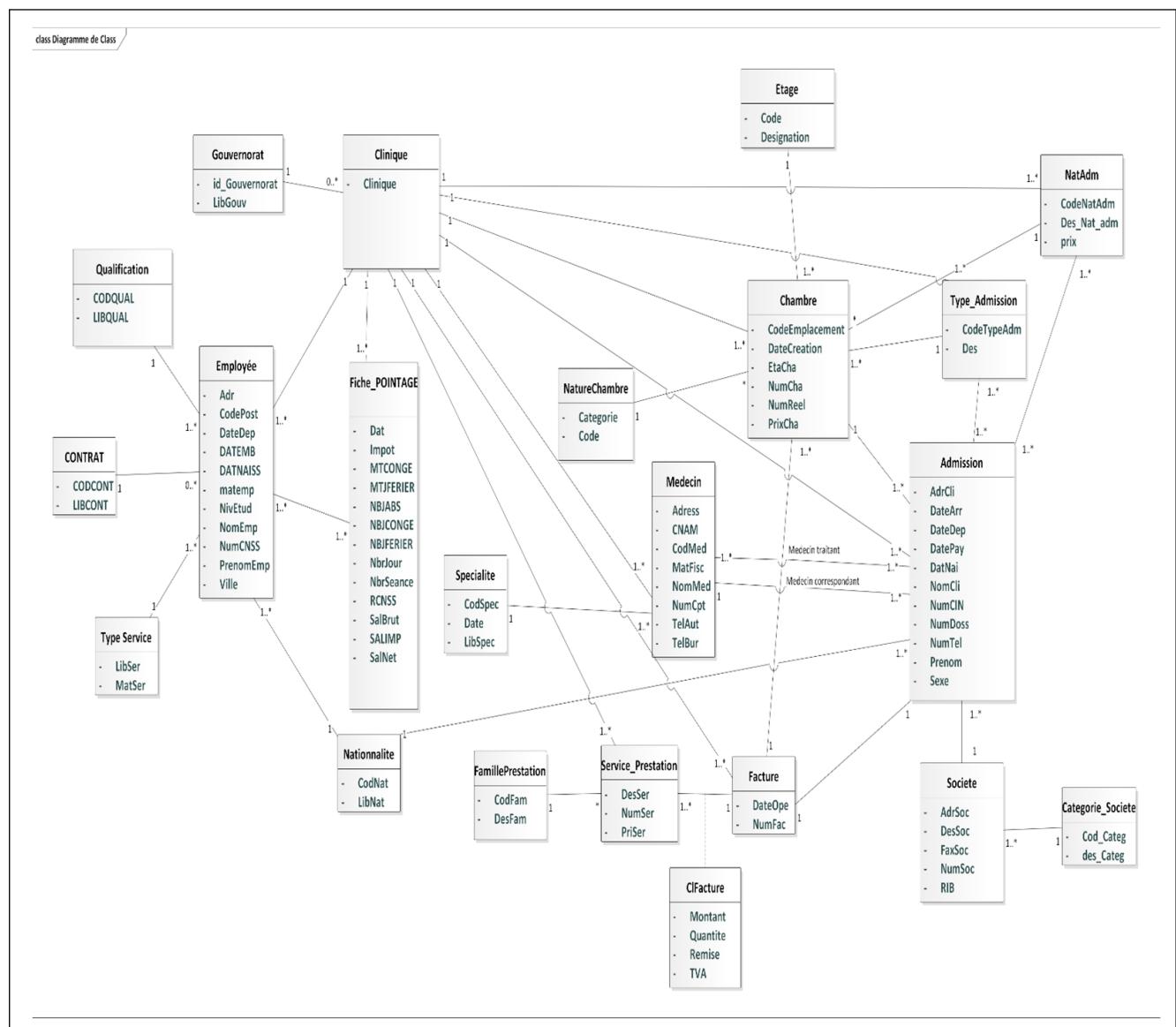


FIGURE 3.7 – Diagramme de classe de la source de données

Classes représentatives :

Une classe représentative (CR) décrit un événement qui se produit à un instant donné, elle contient les mesures d'analyse. Dans notre cas, nous avons détecté les classes représentatives suivantes : **Les classes représentatives = ClFacture, Admission, Chambre, Employé, Fpaie**

Définition des faits :

- **Fait_patientèle** = (NbreAdm=COUNT(S.Admission.NumDoss).(SUM,AVG))
- **Fait_CA** = Quantité = COUNT (S.ClFacture.NumSer) (SUM, AVG, MIN, MAX),
CA = S.ClFacture.Montant*(1-S.ClFacture.Remise) (SUM, AVG))
- **Fait_Capacité** =(NbreLits=COUNT(S.Chambre.All) (SUM, AVG),
NbreChambres=COUNT (DISTINCT(S.Chambre.NumReel)) (SUM, AVG))
- **Fait_Employée** =(NbreEmp = COUNT(S.Employée.Matemp) (SUM, AVG),
Moyenne_Age = (S.Employée.DATNAISS) (SUM, AVG))
- **Fait_Fpaie** =Taux_abscence = COUNT(S.Fiche_Pointage.NBJABS) (SUM,AVG) ,
Salaire = (S.Fiche_Pointage.SalBrut) (SUM,AVG))

3.2.2.2 Définition des dimensions

Les dimensions ci-dessous ont été identifiées suite à une analyse sémantique :

Medecin (

codMed = S.Admission.Medecin.CodMed,

Adress = S.Admission.Medecin.Adress,

NomMed = S.Admission.Medecin.NomMed,

Soc = S.Admission.Medecin.CNAM);

Spécialité (

id_specialite = S.Admission.Medecin.Specialite.CodSpec,

LibSpec = S.Admission.Medecin.Specialite.LibSpec);

Client (

id_client = S.Admission.NumDoss,

NomCli = S.Admission.NomCli,

MODÉLISATION CONCEPTUELLE

Prenom = S.Admission.Prenom,

Sexe = S.Admission.sex,

Nationalite = S.Admission.Nationalite,

Societe = S.Admission.Societe.NumSoc,

AdrCli = S.Admission.AdrCli);

Nationalite (

CodNat = S.Admission.Nationalite.CodNat,

LibNat = S.Admission.nationalite.LibNat);

Société (

NumSoc = S.Admission.Societe.NumSoc,

DesSoc = S.Admission.Societe.DesSoc);

Service (

CodeTypeAdm= S.Admission.Service.CodeTypeAdm,

Des = S.Admission.Service.Des);

NatAdm (

CodeNatAdm = S.Admission.NatAdm.CodNatAdm,

DesNatAdm = S.Admission.NatAdm.Des_Nat_Adm,

Des_Nat_Adm_Ar = S.Admission.NatAdm.Des_Nat_Adm_Ar,

Prix = S.Admission.NatAdm.Prix);

Service Prestation(

NumSer = S.ClFacture.Service Prestation.NumSer,

DesSer= S.ClFacture.Service Prestation. DesSer,

Prix = S.ClFacture.Service Prestation.Prix);

FamPres (

CodFam = S.ClFacture.Service Prestation.FamillePrestation.CodFam,

DesFam = S.ClFacture.Service Prestation.FamillePrestation. DesFam);

Qualification (

CODQUAL = S.Qualification.CODQUAL,

LIBQUAL = S.Qualification.LIBQUAL,

nbreHeureParSemaine = S.Qualification.NbreHeureParSemaine);

Contrat (

CODCONT = S.CONTRAT.CODCONT,

LIBCONT = S.CONTRAT.LIBCONT);

Type Service (

LibSer = S.Type Service.LibSer,

MatSer = S.Type Service.MatSer);

3.2.2.3 Définition des hiérarchies

Dans cette étape, nous définissons les hiérarchies complètes des dimensions en fonction du schéma de la source.

$$H^{prestation} : NumSer \rightarrow CodFam$$

$$H_1^{fact} : NumFac \rightarrow NumDoss \rightarrow CodNat$$

$$H_2^{fact} : NumFac \rightarrow NumDoss \rightarrow NumSoc$$

$$H_3^{fact} : NumFac \rightarrow NumDoss \rightarrow CodMed \rightarrow CodSpec$$

$$H_4^{fact} : NumFac \rightarrow NumDoss \rightarrow NumReel$$

$$H_5^{fact} : NumFac \rightarrow NumDoss \rightarrow CodeNatAdm \rightarrow NumReel$$

$$H_1^{fact} : NumFac \rightarrow NumDoss \rightarrow CodeTypeAdm \rightarrow NumReel$$

$$H_1^{Adm} : NumDoss \rightarrow NumReel$$

$$H_2^{Adm} : NumDoss \rightarrow CodeTypeAdm \rightarrow NumReel$$

$$H_3^{Adm} : NumDoss \rightarrow CodeNatAdm \rightarrow NumReel$$

$$H_4^{Adm} : NumDoss \rightarrow CodNat$$

$$H_5^{Adm} : NumDoss \rightarrow NumSoc$$

$$H_6^{Adm} : NumDoss \rightarrow CodMed \rightarrow CodSpec$$

$$H^{cap} : NumReel \rightarrow code$$

MODÉLISATION CONCEPTUELLE

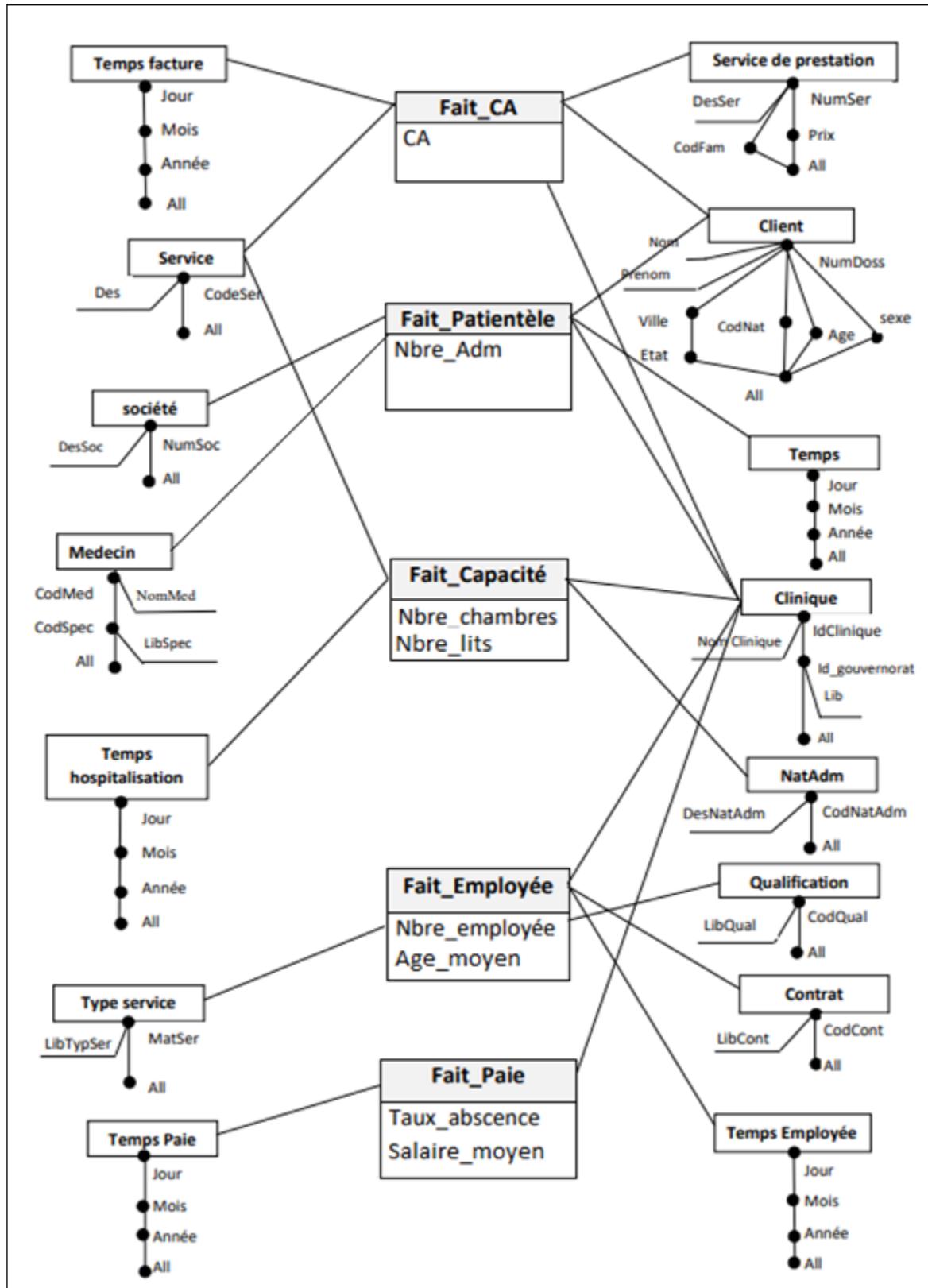


FIGURE 3.8 – Schéma multidimensionnel en constellation (Démarche ascendante)

3.2.3 Méthode mixte

Nous procédons maintenant à la comparaison des schémas créés à partir des deux approches, ascendante et descendante. Cette comparaison nous permet de créer un schéma commun qui intègre les données provenant des deux schémas initiaux. Ensuite, nous effectuons des ajustements tels que l'ajout d'un nouveau fait, la correction de la granularité de l'analyse ou encore la suppression d'une dimension. Ces ajustements sont illustrés dans la Figure 15. Une fois que nous avons défini les approches ascendante et descendante, nous avons pris la décision de supprimer juste la dimension « NatAdm ». Cette décision a été motivée par le fait que tout le reste des dimensions sont utilisées lors de l'analyse des sujets.

MODÉLISATION CONCEPTUELLE

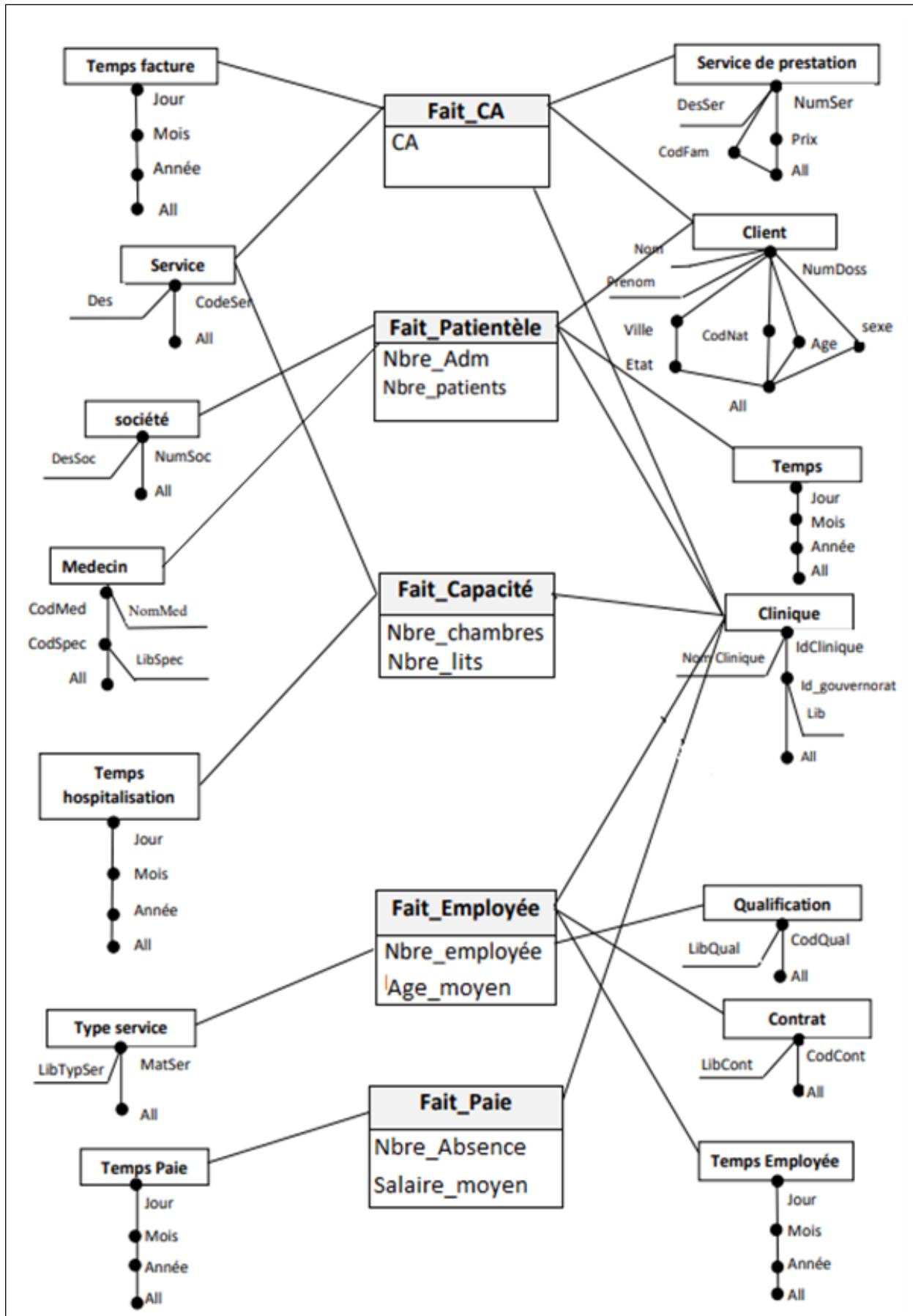


FIGURE 3.9 – Schéma multidimensionnel en constellation (démarche mixte)

Conclusion

En conclusion, ce chapitre a abordé la modélisation conceptuelle d'un entrepôt de données en utilisant différentes approches. Nous avons exploré les méthodes ascendante et descendante, puis nous avons confronté les schémas obtenus pour parvenir à un schéma conceptuel final. Le prochain chapitre se concentrera sur la construction de l'entrepôt de données et le développement des processus ETL, où nous mettrons en pratique les concepts présentés ici afin de concrétiser notre vision de l'entrepôt de données.

Construction de l'entrepôt de données

Sommaire

4.1	Processus ETL	44
4.2	Modélisation logique et physique	45
4.2.1	Modèle logique	45
4.2.2	Modèle physique	46
4.3	Conception détaillé de l'ETL	47
4.4	Développement de l'ETL	49
4.4.1	Outils utilisées	49
4.4.2	Présentation du travail réalisé	51

Introduction

Ce chapitre présente les différentes étapes de construction d'un entrepôt de données, en commençant par la modélisation logique et physique, pour aboutir à la phase d'intégration des données. Nous examinerons en détail la conception et le développement de l'ETL, qui est l'outil clé pour extraire, transformer et charger les données dans l'entrepôt. En somme, nous allons décrire les différentes phases de la construction de l'entrepôt de données et mettre l'accent sur l'importance de chaque étape pour garantir la qualité des données et la performance de l'entrepôt de données.

4.1 Processus ETL

Le processus ETL est un processus informatique qui permet de transférer des données depuis une source de données, de les transformer et de les charger dans une destination. Les lettres ETL signifient Extract, Transform and Load, qui se traduit en français par Extraction, Transformation et Chargement. Le processus ETL peut être divisé en trois étapes principales :

1. **Extraction** : Dans cette étape, les données sont extraites de différentes sources de données telles que des fichiers plats, des bases de données relationnelles, des sources de données cloud, etc.
2. **Transformation** : Dans cette étape, les données sont transformées et nettoyées pour assurer leur qualité. Cette étape implique souvent l'utilisation de logiciels de nettoyage de données et d'outils de transformation des données pour normaliser les formats de données, effectuer des agrégations, des jointures, etc.
3. **Changement** : Dans cette étape, les données transformées sont chargées dans la destination finale, qui peut être une base de données relationnelle, un data warehouse, un data lake ou tout autre système de stockage.

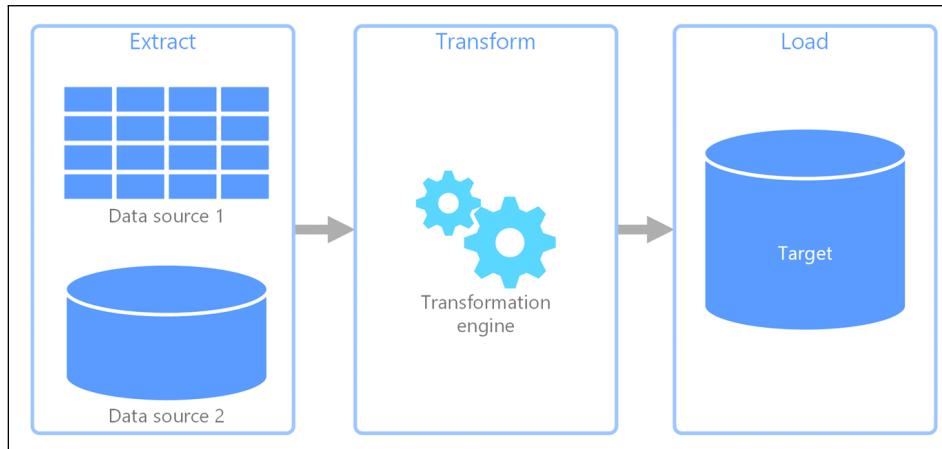


FIGURE 4.1 – Le processus ETL

4.2 Modélisation logique et physique

4.2.1 Modèle logique

Pour la modélisation logique de notre entrepot nous avons choisi le modèle ROLAP en étoile ou chaque fait se transforme en une table ayant comme identifiant les identifiants des dimensions correspondantes et chaque dimension se transforme en une table :

Fait_CA (FK_client#, FK_Societe#, FK_CodFam#, FK_NumSer#, FK_Date#, Montant, Remise, CodeTva, DatOpe, NumSer, DatFac, Clinique).

Fait_Capacité (FK_TypAdm#, FK_Etage#, FK_Service#, FK_Date#, Clinique, NatCha, Externe, Etage, NumSer, NumCha).

Fait_Patientèle (FK_Natclinique#, FK_soc_Clinique#, FK_Client_Clinique#, FK_Nation#, FK_NatAdm#, FK_Date#, NumDoss, NomCli, Prenom, DatNai, Nation, AdrCli, NumSoc, NatAdm, Clinique).

Fait_Employée (FK_Contrat#, FK_Qualification#, FK_Date#, MATEMP, Age)

Fait_POINTAGE (FK_Employee#, FK_MatSer#, NBJABS, SALBRUT)

Clinique (PK_clinique, Clinique)

Medecin (PK_CodMed, Clinique, NumCpt, MatFisc, TypMed, Adresse, NomMed, CodMed).

Societe (PK_numsoc, Clinique, RIB, NumCpt, FacSoc, AdrSoc, DesSoc, NumSoc)

NatAdm (PK_CodMed, CodeNatAdm, Clinique, AffichePrincipal, DesNatAdm)

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

Client (PK_Client, NumDoss, NumCha, NomCli, Prenom, DatNai, Nation, AdrCli, TypAdm, NatAdm, NumSoc, MedTrait, MedSpec, DatArr, NumFac, DatFac, Sexe, DatDep, Clinique)

Service (PK_Code, Code, Des, Clinique)

Service Pestation (PK_NumSer, NumSer, DesSer, CodFam, PriSer, DesSerAr)

Type Service (PK_MatSer, LibTypeSer)

Contrat (PK_CodCont, LibCont)

Qualification (PK_CodQual, LibQual)

4.2.2 Modèle physique

4.2.2.1 Optimisation physique

Dimension temporelle : Dans le contexte de notre entrepôt de données, il n'est pas nécessaire de créer une dimension temporelle pour enregistrer chaque minute ou chaque seconde sur une période donnée. Alors selon notre cas, la granularité jour est généralement considérée comme suffisante, c'est pourquoi nous l'avons adoptée comme la granularité analytique.

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

4.2.2.2 Schéma physique

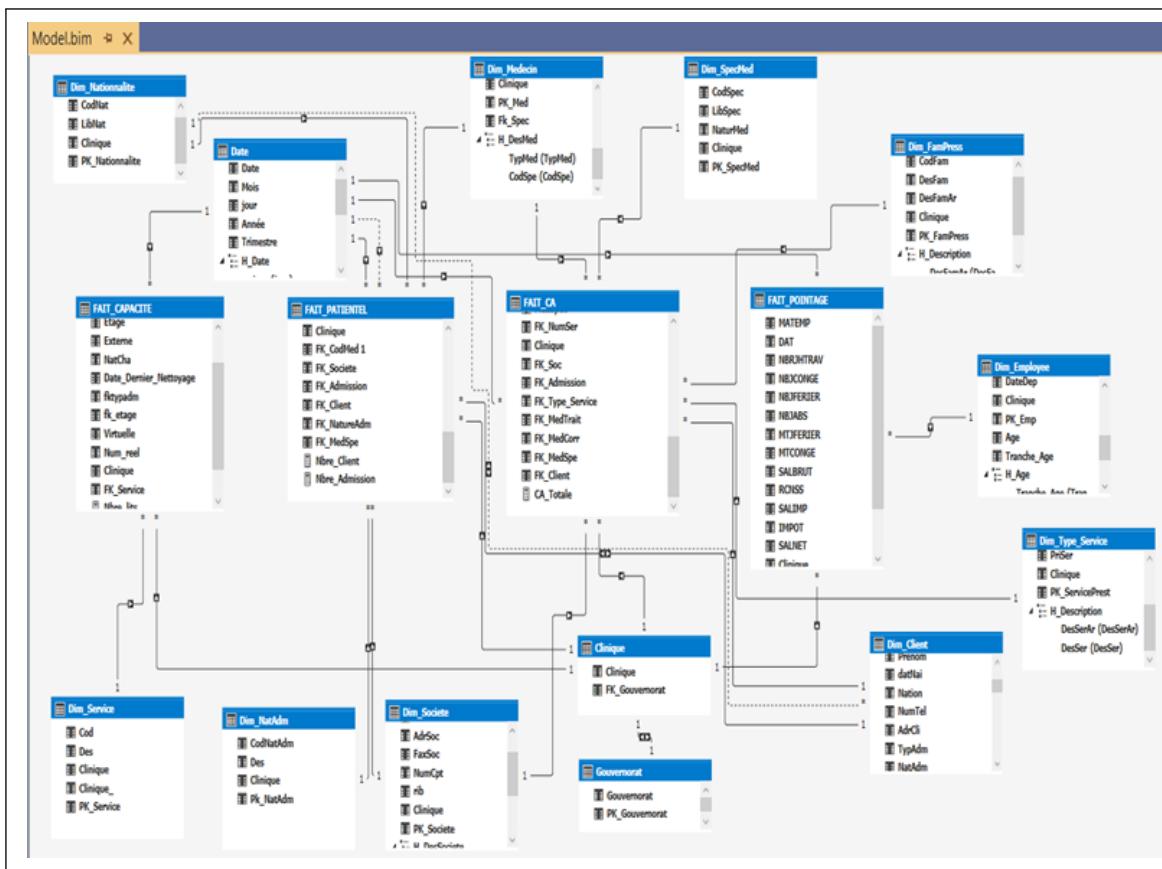


FIGURE 4.2 – Schéma physique de l'entrepôt de données

4.3 Conception détaillé de l'ETL

Les processus ETL sont une étape cruciale de la gestion des données dans un projet. Ils servent à collecter, nettoyer, transformer et intégrer les données provenant de différentes sources, dans le but de créer un entrepôt de données (data warehouse) ou un lac de données (data lake). Ces données seront ensuite utilisées pour la prise de décisions en fournissant une vision complète et fiable des informations concernant l'entreprise ou l'organisation.

Diagramme d'activité

Dans le contexte de l'ETL, les diagrammes d'activités peuvent être utilisés pour modéliser les étapes du processus ETL, qui comprennent l'extraction des données à partir de sources de

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

données externes, la transformation des données pour répondre aux exigences de la source de données cible, et le chargement des données transformées dans la source de données cible.

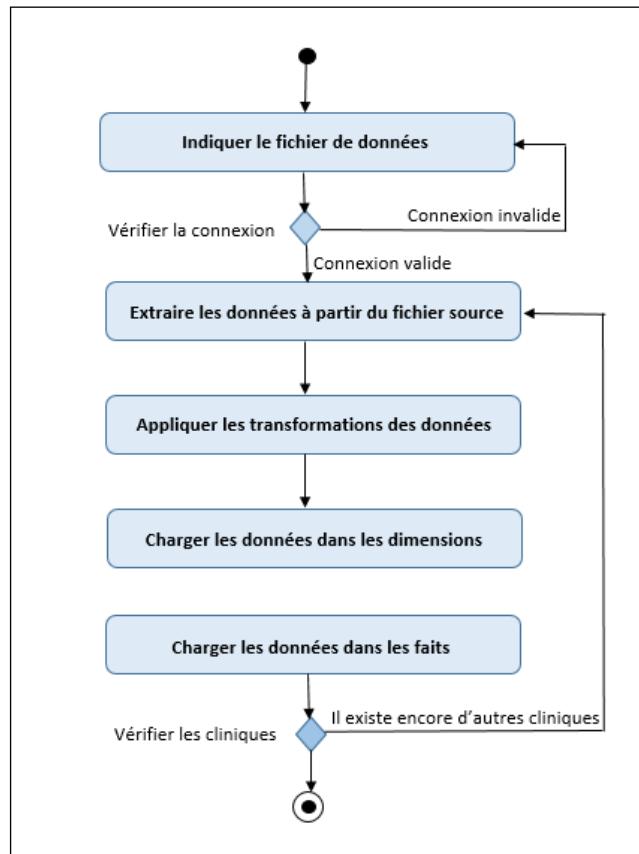


FIGURE 4.3 – Diagramme d'activités pour l'alimentation des tables de dimensions et de faits

Le diagramme représente la phase d'intégration globale qui a pour objectif de récupérer les données à partir de la source et de les charger dans la base de destination. Pour atteindre cet objectif, plusieurs activités doivent être effectuées, telles que celles mentionnées ci-dessous. Ces activités ont pour but de garantir que les données sont correctement intégrées et prêtes à être utilisées pour les besoins de l'entreprise ou de l'organisation :

- Vérification de la présence des données dans la source.
- Validation des données pour s'assurer de leur qualité et de leur intégrité.
- Transformation des données si nécessaire pour les préparer à l'intégration.
- Chargement des données dans l'entrepôt de données afin de les rendre accessibles pour les analyses et les rapports ultérieurs.

4.4 Développement de l'ETL

4.4.1 Outils utilisées

Ci-dessous, nous présentont une liste des outils de développement utilisés pour assurer l'intégration des données.

SQL Server Management Studio (SSMS) : SQL Server Management Studio (SSMS) est un environnement intégré pour la gestion de serveurs de bases de données SQL Server. Il est utilisé pour se connecter à des instances de SQL Server, pour effectuer des opérations de maintenance et de développement de base de données, pour écrire et exécuter des requêtes, pour déployer des bases de données et pour surveiller les performances de SQL Server. SSMS est une application gratuite développée par Microsoft, qui est largement utilisée par les administrateurs de bases de données, les développeurs de logiciels et les analystes de données pour gérer les bases de données SQL Server.



FIGURE 4.4 – Logo SQL Server Management Studio

Microsoft SQL Server Integration Services : SSIS est un outil puissant pour l'intégration de données dans un environnement Microsoft SQL Server, qui peut aider les entreprises à améliorer la qualité et la disponibilité de leurs données, ainsi qu'à réduire les coûts de gestion des données. SSIS permet de se connecter à différentes sources de données, de transformer les données pour les préparer à l'utilisation et de les charger dans une ou plusieurs destinations.



FIGURE 4.5 – Logo Microsoft SQL Server Integration Services

Visual Studio 2019 : Visual Studio 2019 est un environnement de développement intégré (IDE) développé par Microsoft pour les développeurs de logiciels. Il offre un éditeur de code sophistiqué, des outils de débogage avancés, des assistants pour la création d'interfaces utilisateur, des outils de test et de déploiement, ainsi que des intégrations avec des outils de gestion de code source. Visual Studio 2019 prend en charge une variété de langages de programmation et est utilisé pour créer des applications de qualité professionnelle pour une variété de platesformes, telles que Windows, Android, iOS, Web et Cloud.



FIGURE 4.6 – Logo Visual Studio 2019

Microsoft Power BI Microsoft Power BI est une plateforme d'analyse de données et de business intelligence qui permet aux utilisateurs de visualiser, analyser et partager les données de leur entreprise à partir d'une variété de sources, en créant des tableaux de bord interactifs et des visualisations de données.



FIGURE 4.7 – Logo Microsoft Power BI

Enterprise Architect Enterprise Architect est un outil de modélisation UML (Unified Modeling Language) et de gestion de cycle de vie des applications développé par Sparx Systems. Il est utilisé pour concevoir, documenter et maintenir des systèmes logiciels et des processus métier complexes. Enterprise Architect permet aux utilisateurs de créer des diagrammes UML pour représenter les différentes vues d'un système, y compris les diagrammes de classes, de séquence, d'activité, de composants, de déploiement, etc.



FIGURE 4.8 – Logo Enterprise Architect

4.4.2 Présentation du travail réalisé

La première étape consiste à créer notre entrepôt de données cible dans SSMS pour implémenter l'ETL. Ensuite, nous avons procédé à la création de la base de données et desmodèles de tous les tableaux, comme illustré dans la figure 4.9.

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

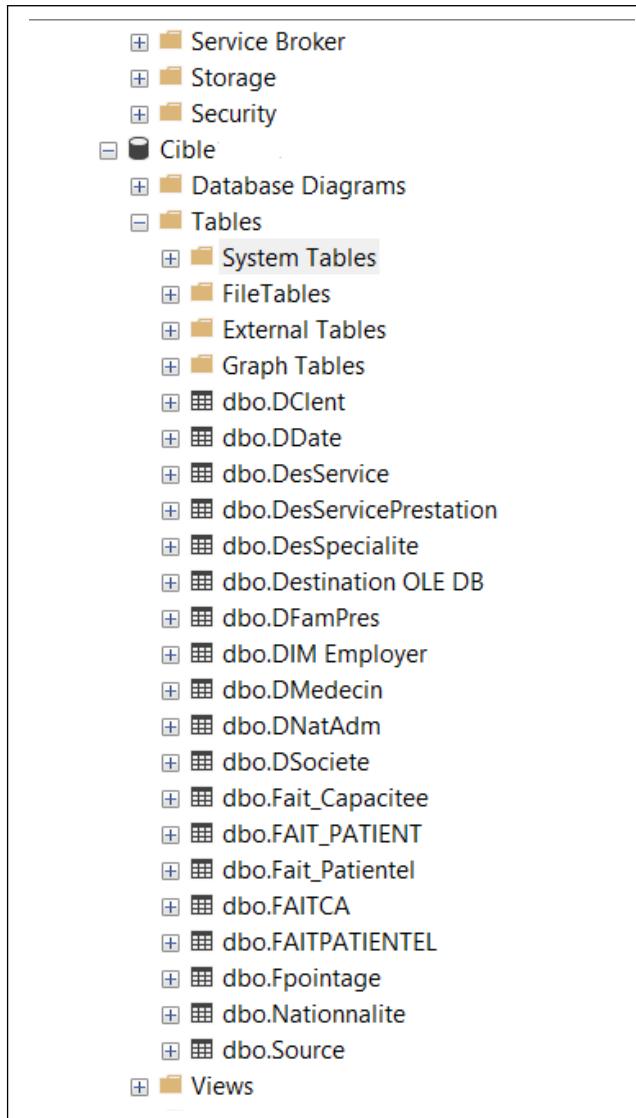


FIGURE 4.9 – L'entrepôt de données

Une fois la création de la base de données terminée, nous passerons à la deuxième étape, qui consiste à créer un nouveau projet dans Visual Studio appelé "PFE". Dans ce projet, nous avons créé un nouveau paquet SSIS nommé "CA_Cliniques.dtsx", tel qu'illustré dans la figure.

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

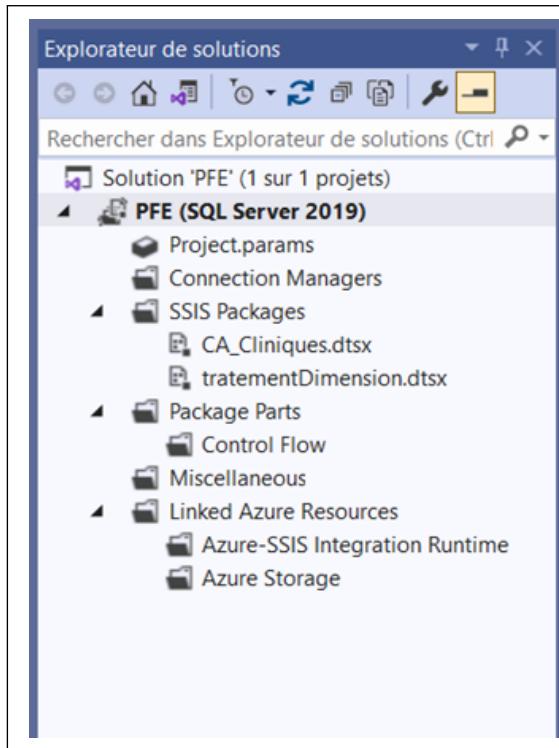


FIGURE 4.10 – Le projet SSIS

Nous avons ajouté une table appelée "Sourcelist" à notre entrepôt de données. Cette table contient les cliniques et leurs sources de données. La figure 4.11 représente la table des bases de données.

ID	ServerName	Clinique	ConnStringsource
1	LAPTOP-6JVFC03QIVE_SERVER	CliniqueF	Data Source="LAPTOP-6JVFC03QIVE_SERVER";User ID=...
2	LAPTOP-6JVFC03QIVE_SERVER	CliniqueM	Data Source="LAPTOP-6JVFC03QIVE_SERVER";User ID=...
3	LAPTOP-6JVFC03QIVE_SERVER	CliniqueP	Data Source="LAPTOP-6JVFC03QIVE_SERVER";User ID=...

ConnStringsource2	ConnStringsource3	ConnStringdestination
Data Source="LAPTOP-6JVFC03QIVE_SERVER";User ID=...	Data Source="LAPTOP-6JVFC03QIVE_SERVER";User ID=...	Data Source="LAPTOP-6JVFC03QIVE_SERVER";User ID=...
Data Source="LAPTOP-6JVFC03QIVE_SERVER";User ID=...	Data Source="LAPTOP-6JVFC03QIVE_SERVER";User ID=...	Data Source="LAPTOP-6JVFC03QIVE_SERVER";User ID=...
Data Source="LAPTOP-6JVFC03QIVE_SERVER";User ID=...	Data Source="LAPTOP-6JVFC03QIVE_SERVER";User ID=...	Data Source="LAPTOP-6JVFC03QIVE_SERVER";User ID=...

FIGURE 4.11 – Base de données des cliniques

Étant donné que nous travaillons avec 3 cliniques, il est essentiel d'établir une connexion dynamique entre elles. Pour ce faire, nous avons procédé à la création de variables, comme présenté dans la figure 4.12. Ces variables seront utilisées ultérieurement dans le processus

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

Name	Scope	Data type	Value	Expression
clinique	Package1	String	CliniqueF	
ConnectionSucceeded	Package1	Boolean	False	
Economat	Package1	String	Data Source=LAPTOP-6VFC03Q\VE_SERVER\Initial Catalog=farabigd...	
ErrorMessage	Package1	String		
Gclinique	Package1	String	Data Source=LAPTOP-6VFC03Q\VE_SERVER\Initial Catalog=farabigd...	
GRôle	Package1	String	Data Source=LAPTOP-6VFC03Q\VE_SERVER\Initial Catalog=farabigd...	
sourcelist	Package1	Object	System.Object	

FIGURE 4.12 – Configuration des variables

Par la suite, nous utilisons le composant "tâche d'exécution de requête SQL" pour effectuer une configuration générale de cet élément. Voici les étapes que nous suivons :

- Nous configurons la connexion vers notre entrepôt de données cible
- Nous définissons l'instruction SQL pour effectuer une sélection à partir de la table sourcelist, qui contient les différentes cliniques.
- Nous choisissons l'option "jeu de résultats complet" dans les paramètres du jeu de résultats

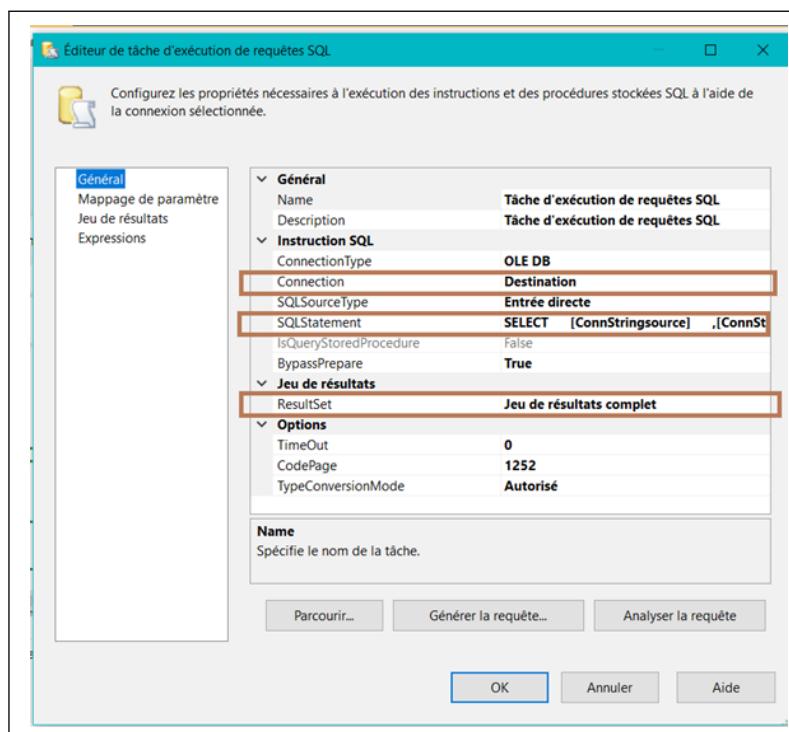


FIGURE 4.13 – Configuration de la partie générale de la tâche d'exécution de requêtes SQL

En ce qui concerne la configuration de l'élément "jeu de résultats" de la tâche d'exécution de requêtes SQL, voici les actions que nous avons effectuées :

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

- Nous avons utilisé la variable nommée "sourceliste", que nous avons créée en tant que variable de type objet.

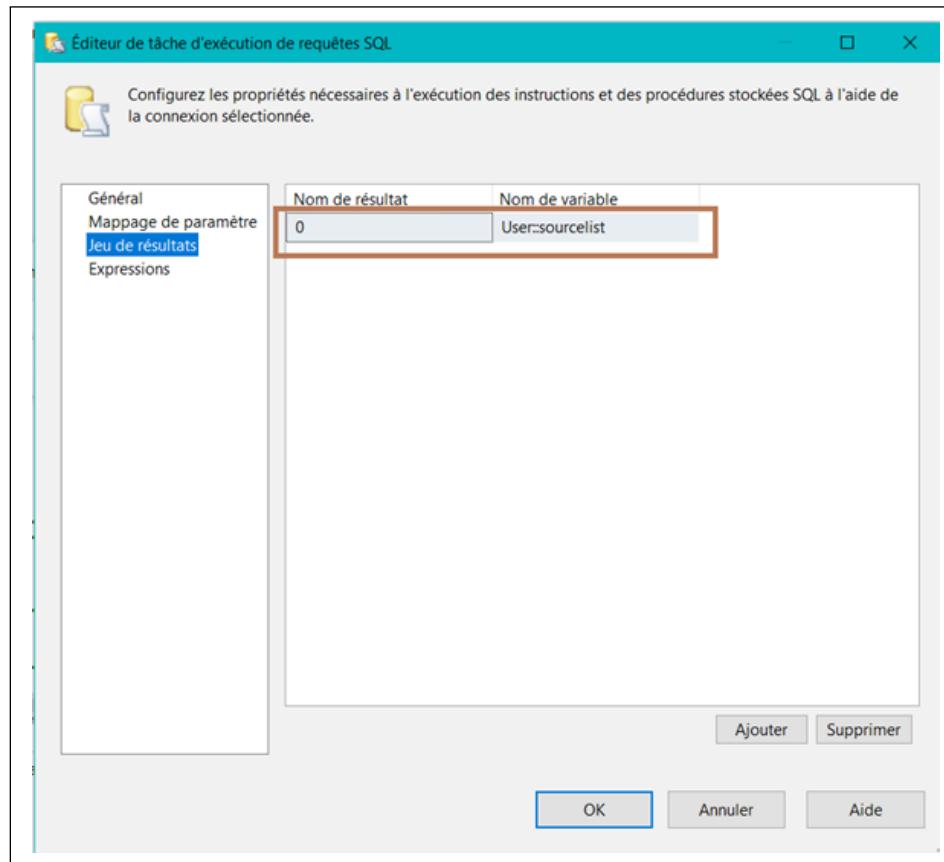


FIGURE 4.14 – Configuration de la partie jeu de résultats de la tâche d'exécution des requêtes SQL

Ensuite, nous avons utilisé le composant "Conteneur des boucles Foreach" et effectué les configurations suivantes pour son élément de collection :

- Nous avons défini la variable source de l'objet ADO comme étant la variable "sourceliste"
- Nous avons choisi le mode d'énumération "Lignes dans la première table".

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

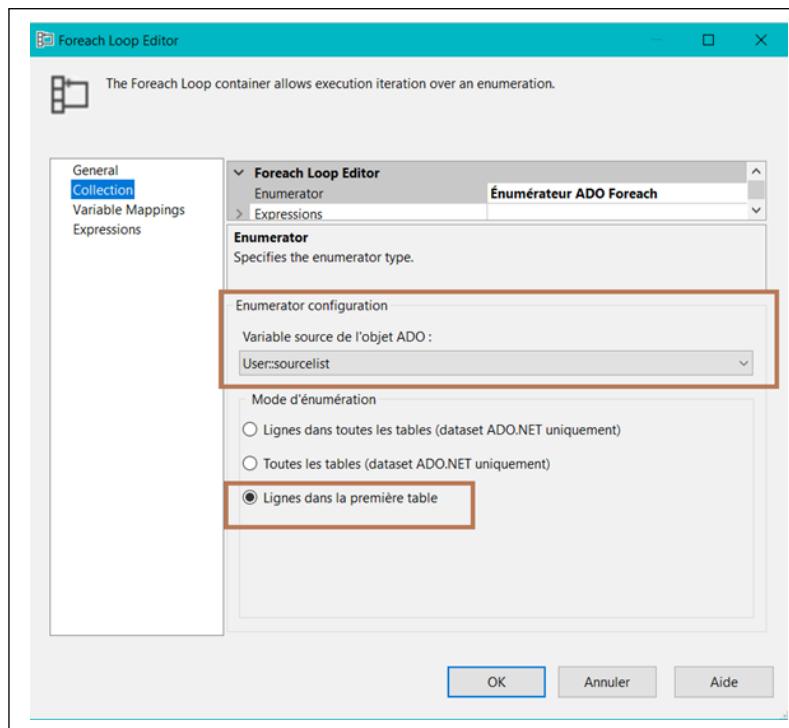


FIGURE 4.15 – Configuration de la collection du conteneur de boucle Foreach

Pour le mappage des variables, nous avons sélectionné 4 variables. Les 3 premières variables correspondent à la connexion dans laquelle nous avons stocké nos informations de connexion. La quatrième variable est le code de la clinique, qui contient les codes des 3 cliniques.

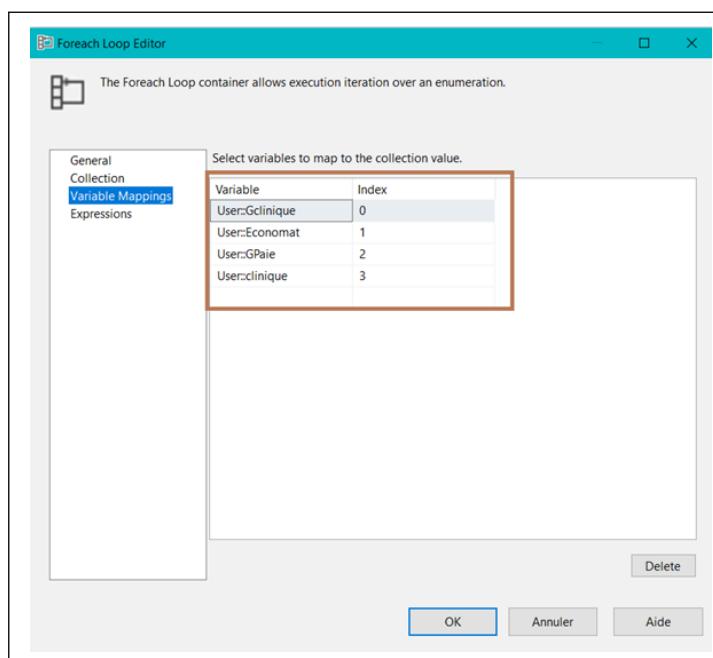


FIGURE 4.16 – Configuration du mappage de variables de conteneur de boucle Foreach

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

Finalement nous utilisons le composant de « test de connexion » dans l'ETL qui est utilisé pour vérifier la connectivité et l'accès à notre source de données. Il permet de s'assurer que l'ETL peut se connecter correctement à la source de données avant de procéder à l'extraction ou à la transformation des données. Ce composant peut effectuer un test de connexion. Il renvoie généralement un résultat indiquant si la connexion a réussi 'ConnectionSucceeded' ou un message d'erreur si elle a échoué, ce qui peut être utilisé pour déclencher d'autres actions dans le flux de l'ETL.

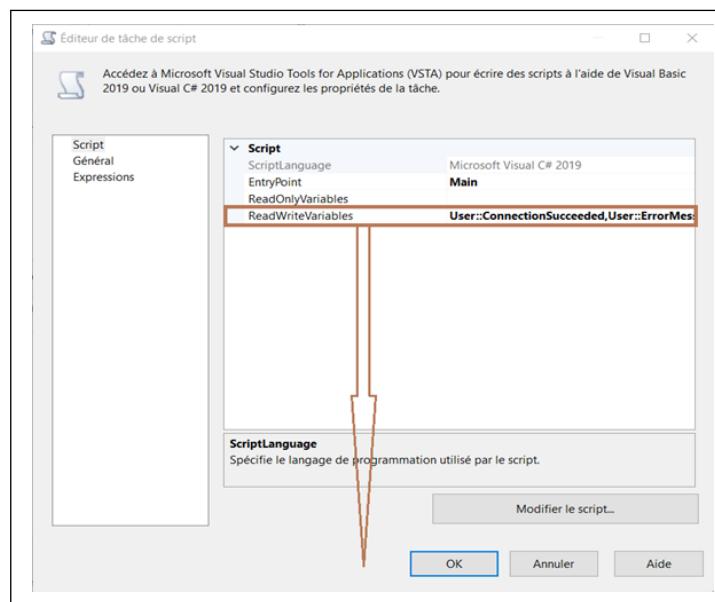


FIGURE 4.17 – Configuration de la test connexion

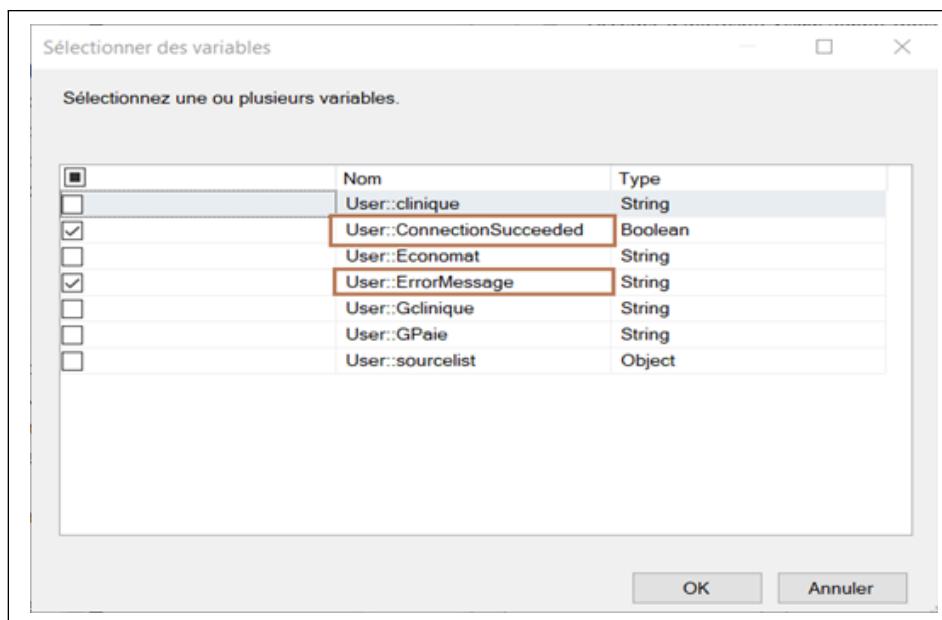


FIGURE 4.18 – Configuration de la test connexion

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

Le diagramme de la figure 4.19 illustre un flux de contrôle comportant un conteneur Foreach, avec dix-sept tâches de flux de données définies au niveau du package. Chaque tâche de flux de données représente une séquence d'opérations visant à extraire les données de la base de données et à les charger dans notre entrepôt de données.

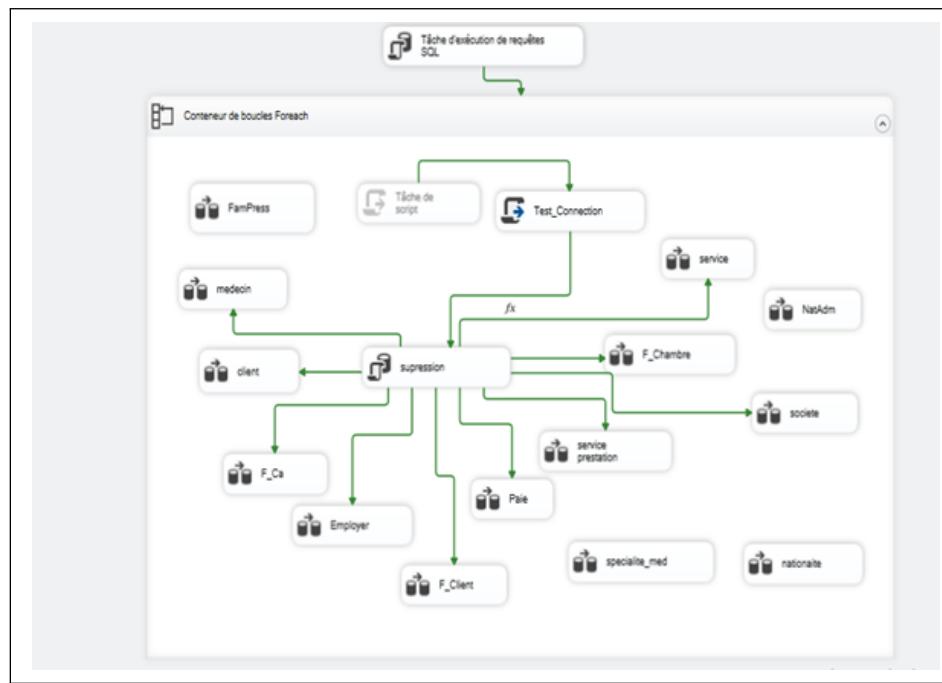


FIGURE 4.19 – Flux de contrôle

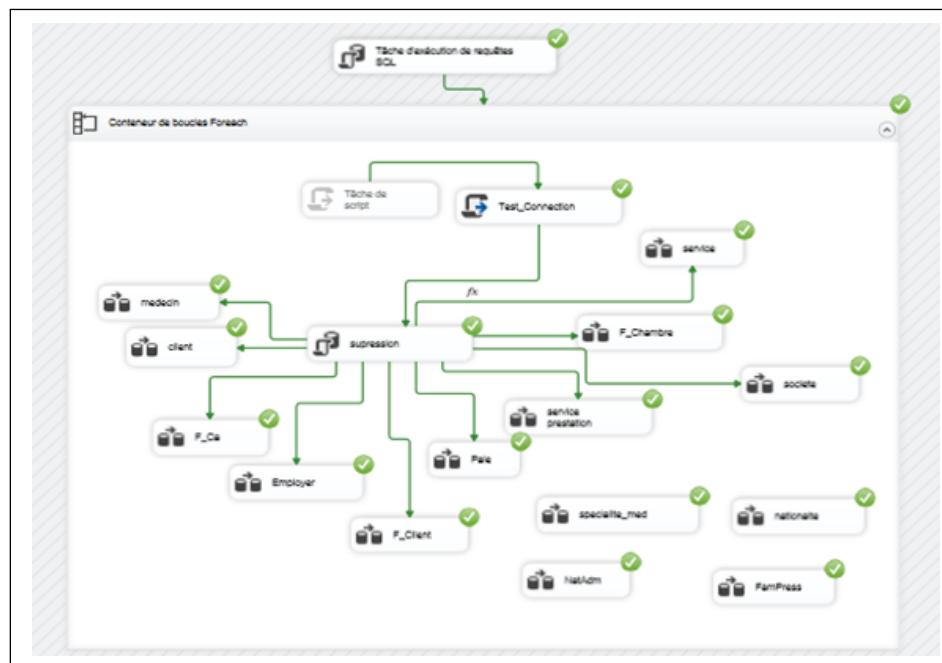


FIGURE 4.20 – Projet SSIS validé

La composition d'une dimension

La composition d'une dimension dans le flux de données d'un ETL (Extract, Transform, Load) fait référence à la création et à la structuration d'une dimension lors du traitement des données. Cela implique l'extraction des données nécessaires à partir des sources, leur transformation selon les besoins de la dimension et leur chargement. Tout cela est réalisé dans le flux de données de l'ETL, où chaque tâche représente une étape spécifique du processus de composition de la dimension. La figure 4.21 présente en détail le processus d'extraction, de transformation et de chargement des nouvelles données d'une tâche de flux de données concernant la dimension "Nationalité".

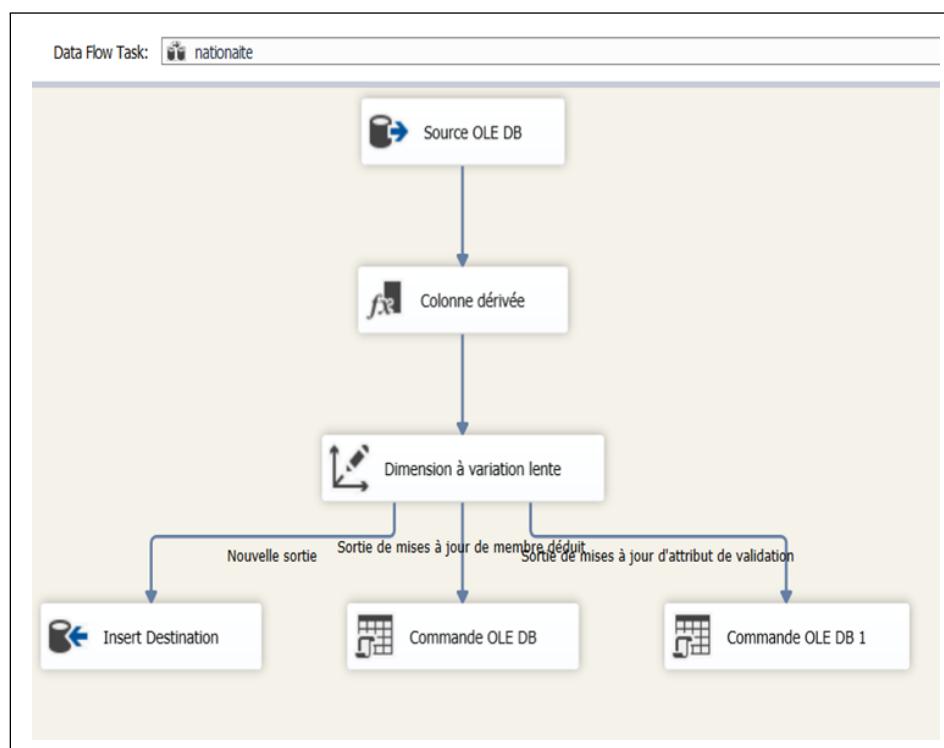


FIGURE 4.21 – Flux de données relative à la dimension "nationalité"

Nous commençons par définir la source de données pour la table "nationalité" dans la base opérationnelle. Cela nous permet de récupérer les données requises qui seront ensuite acheminées vers la dimension "nationalité" en utilisant le composant "Source OLE DB". Ce composant est configuré pour extraire les données de manière efficace et fiable, en garantissant la cohérence et l'intégrité des données lors du chargement dans la dimension.

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

Après cela, nous utilisons le composant "Colonne dérivée" pour ajouter des colonnes à la base de données en transit dans le flux. Nous créons spécifiquement la variable "Clinique" dans ce composant afin de permettre à la dimension de boucler sur les 3 cliniques. Cette approche garantit que les données sont traitées individuellement pour chaque clinique, ce qui facilite l'intégration précise des informations associées à chacune d'entre elles dans la dimension.

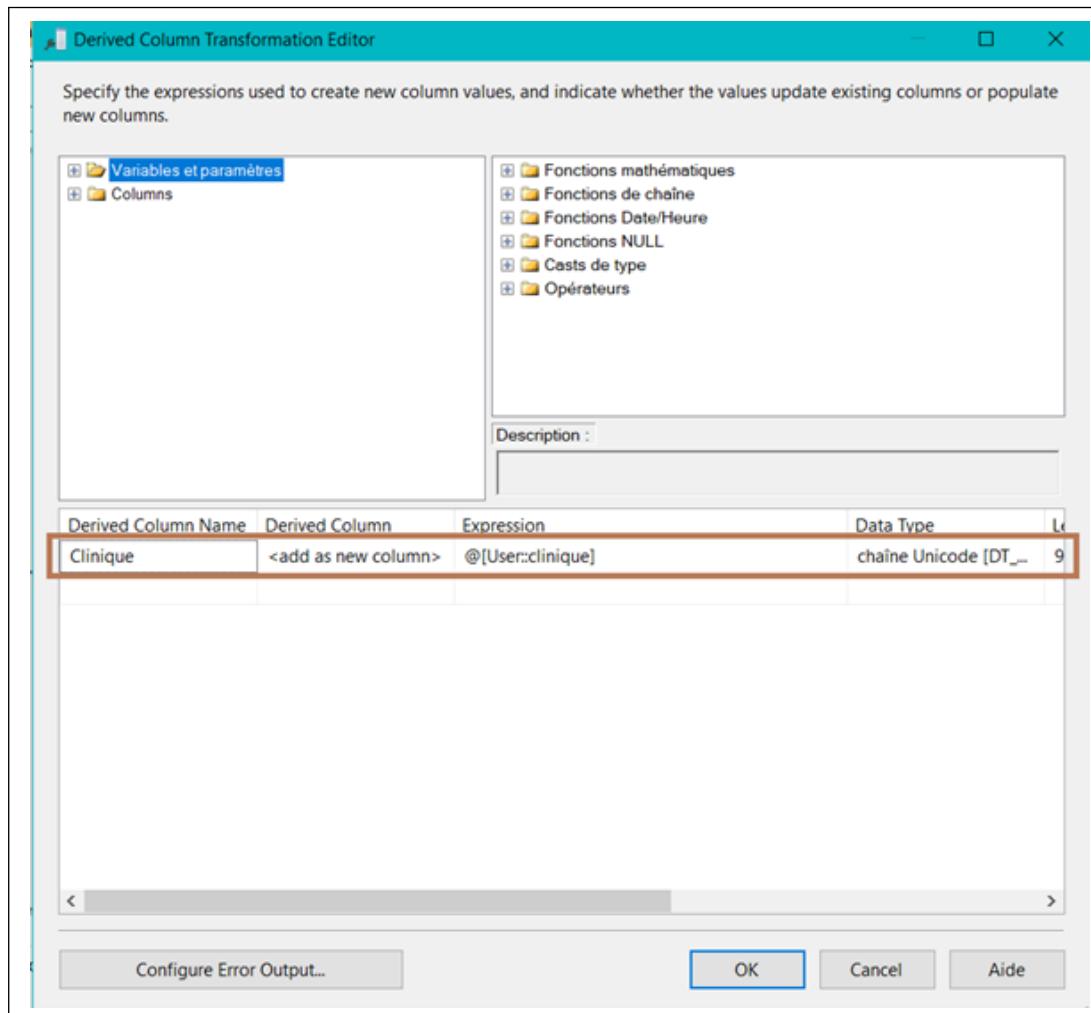


FIGURE 4.22 – Editeur de transformation de colonne dérivé

Dans la phase suivante, nous utilisons le composant "Dimension à variation lente" pour établir la clé primaire de la dimension. Nous sélectionnons spécifiquement l'attribut "modification d'attribut" afin de permettre aux valeurs modifiées de remplacer les valeurs existantes. Cette approche assure une mise à jour précise et cohérente des valeurs modifiées dans la dimension lors du chargement de nouvelles données. En utilisant ce type d'attribut, nous garantissons une gestion efficace des changements et une intégrité optimale de la dimension.

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

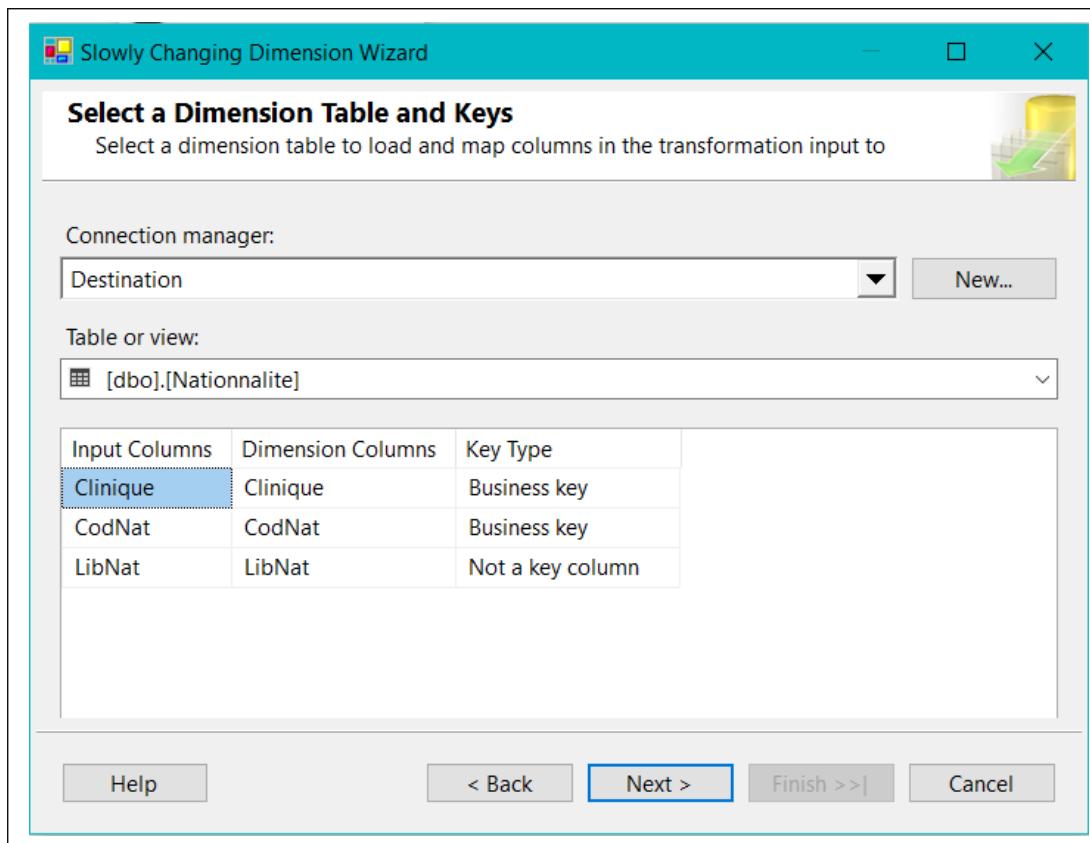


FIGURE 4.23 – Editeur de transformation de la variation lente

Enfin, pour conclure le processus, nous utilisons le composant "Destination OLE DB" pour charger les données dans notre entrepôt de données. Ce composant est responsable de l'acheminement des données transformées vers la destination finale. Grâce à cette étape, les informations pertinentes extraites, transformées et enrichies précédemment sont chargées de manière fiable et sécurisée dans notre entrepôt de données, prêtes à être exploitées pour l'analyse et la prise de décision.

La composition d'un fait Les tables de faits peuvent inclure des colonnes dérivées qui sont des calculs ou des agrégations basées sur les mesures et les attributs existants. Lorsque vous utilisez la tâche de suppression (Delete) dans SSIS en relation avec les tables de faits, vous pouvez spécifier des critères de suppression, y compris ceux basés sur les colonnes dérivées, pour supprimer des enregistrements spécifiques de la table de faits.

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

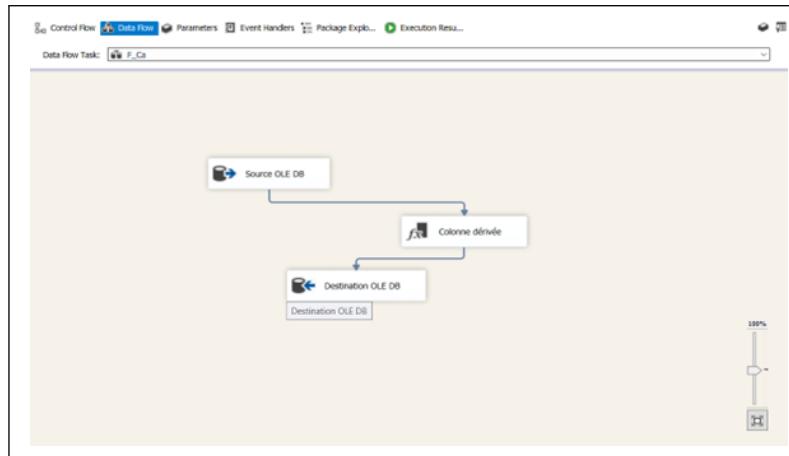


FIGURE 4.24 – Flux de données « F_Ca »

Au sein du flux de contrôle de notre table, nous avons décidé de ne pas inclure la tâche de "recherche", conscient du caractère massif de nos tables de données. Effectuer une recherche ligne par ligne aurait considérablement ralenti l'exécution. Afin de remédier à cette problématique, nous avons préféré adopter une approche différente : la tâche de suppression. Cette solution nous permet de supprimer les données anciennes et de les recharger avec des données à jour. En optant pour cette méthode, nous améliorons les performances globales du processus, en évitant les délais engendrés par une recherche exhaustive

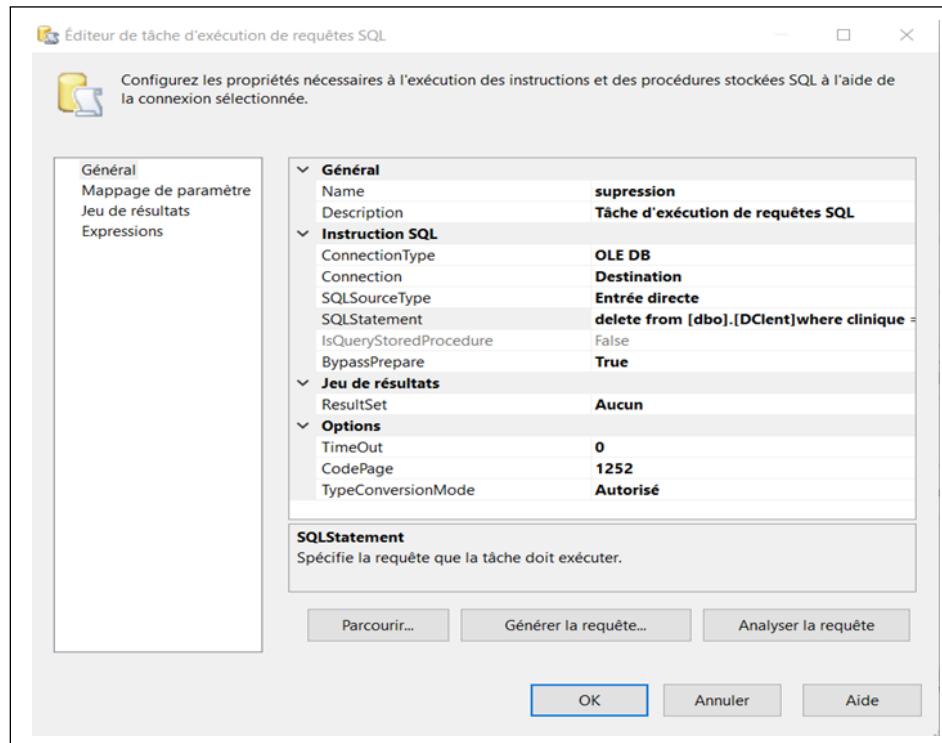


FIGURE 4.25 – Editeur de transformation de la tâche suppression

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

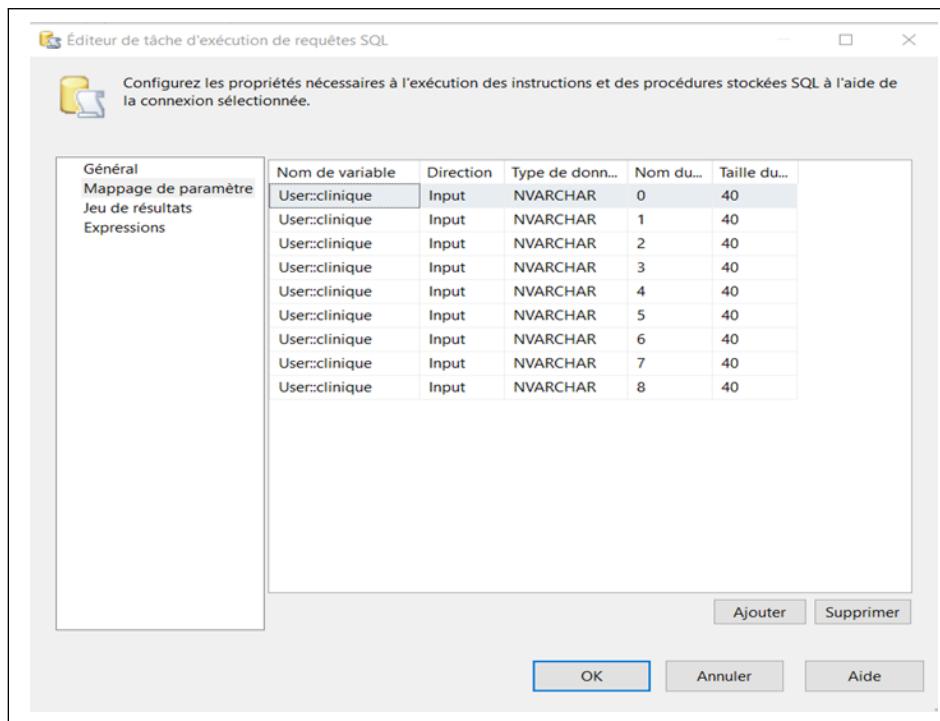


FIGURE 4.26 – Editeur de transformation de la tache suppression

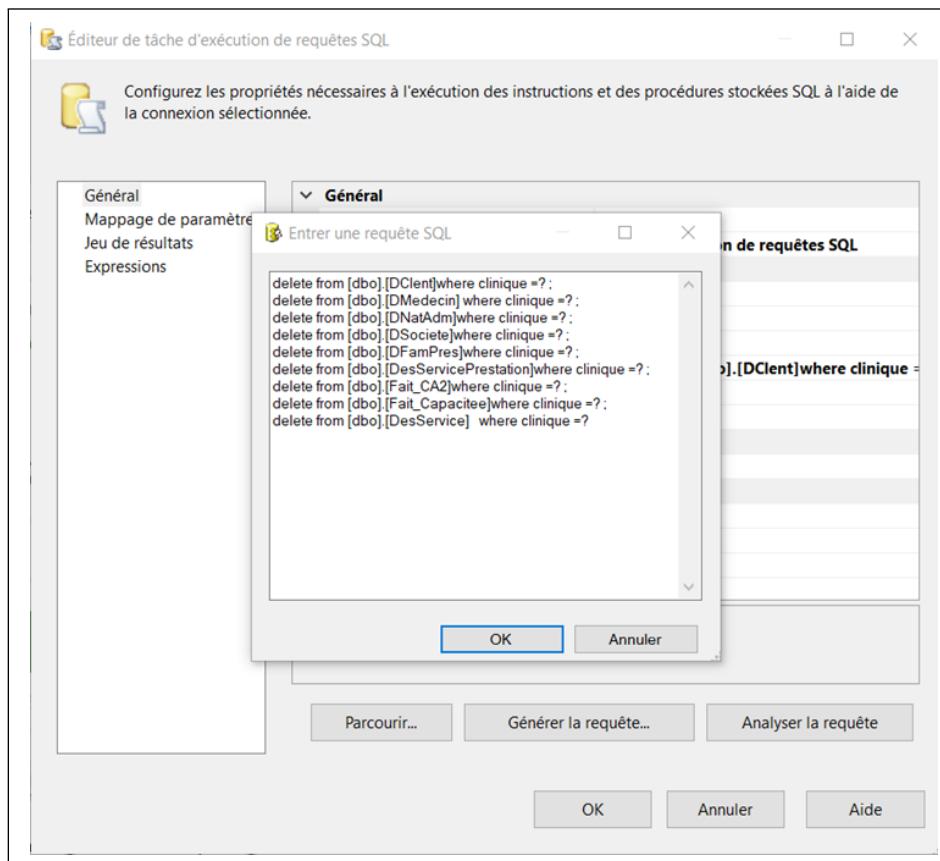


FIGURE 4.27 – Les requêtes SQL de la tache suppression

Déploiement et création d'un job

Une fois que nous avons terminé la création de notre ETL dans un projet SSIS, nous avons procédé au déploiement de celui-ci. Le déploiement consiste à rendre le package accessible à partir du serveur SQL. Cela permet au package d'être exécuté à partir d'un emplacement centralisé, offrant une meilleure gestion et une accessibilité facile pour les autres membres de l'équipe. Le déploiement inclut généralement la configuration des connexions aux sources de données, des paramètres et des variables nécessaires à l'exécution de l'ETL.

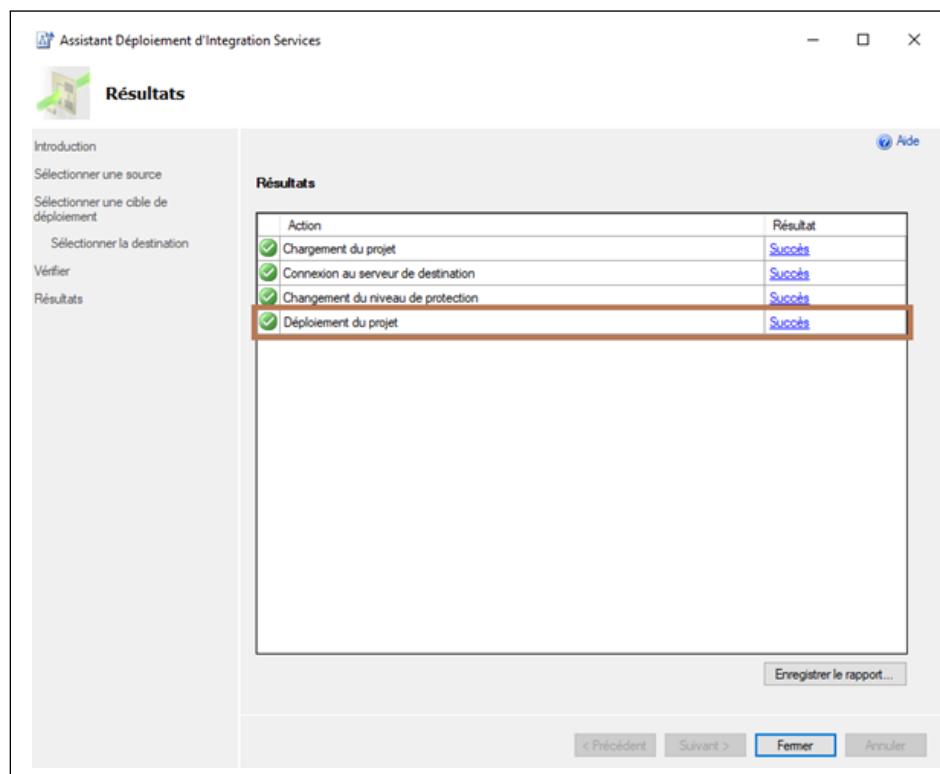


FIGURE 4.28 – Résultats du déploiement validé

Une fois le package déployé, nous avons créé un job dans SQL Server. Nous avons configuré ce job pour qu'il s'exécute tous les jours à minuit. Cette planification régulière garantit que l'ETL est exécuté de manière automatique et prévisible, sans nécessiter d'intervention manuelle. Ainsi, les données sont intégrées quotidiennement dans le système de manière cohérente et opportune. En automatisant l'exécution de l'ETL à travers un job programmé, vous gagnez du temps et vous assurez que le processus d'intégration des données est effectué de manière fiable et régulière. Vous pouvez également surveiller les résultats du job, consulter les journaux d'exécution et prendre des mesures en cas d'erreurs ou d'exceptions.

CONSTRUCTION DE L'ENTREPÔT DE DONNÉES

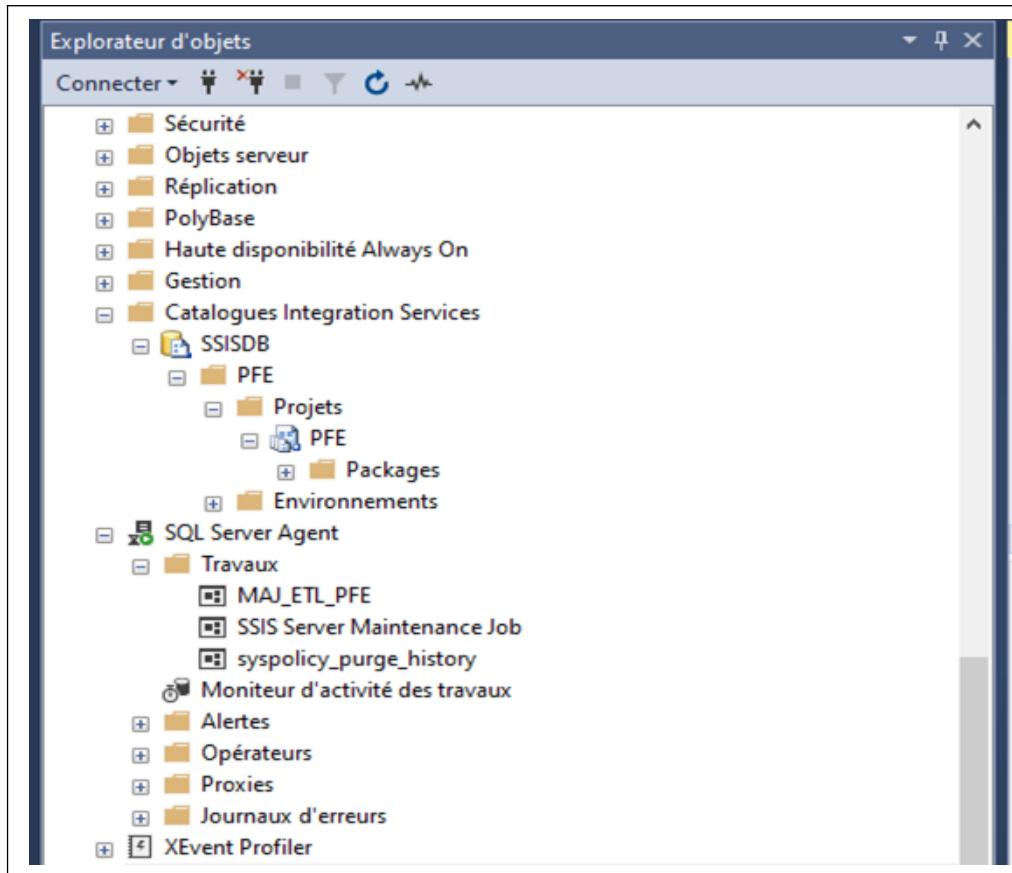


FIGURE 4.29 – Crédit du job dans SQL

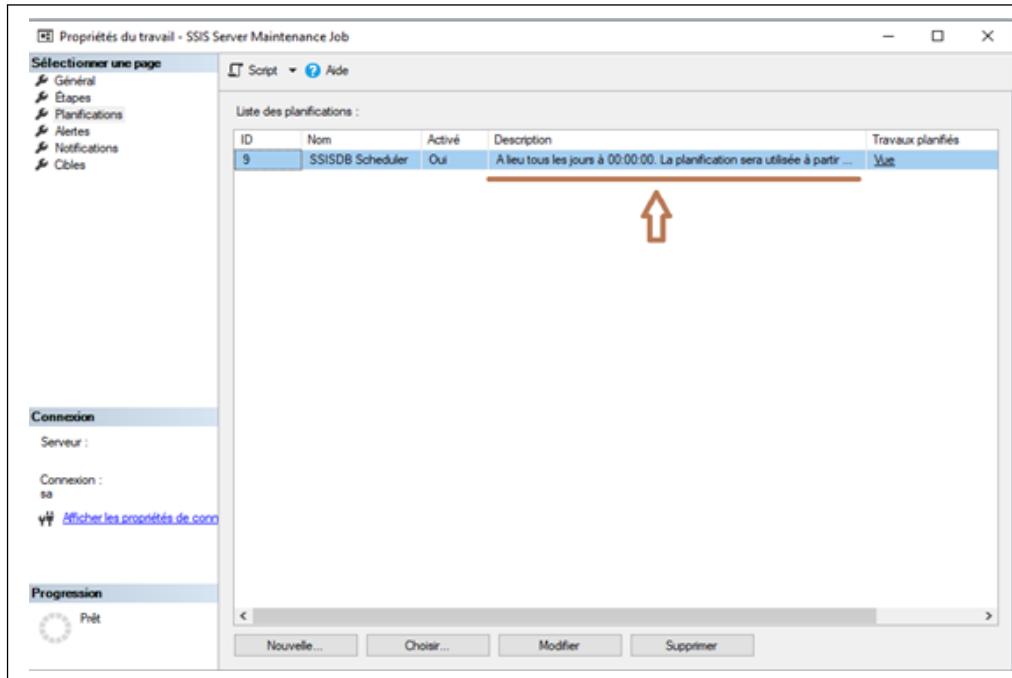


FIGURE 4.30 – La planification du job créé

Conclusion

Dans ce chapitre, nous avons élaboré la modélisation logique pour notre entrepôt de données, en nous basant sur la modélisation physique existante, afin de réaliser la construction concrète de l'entrepôt. Nous avons également décrit en détail la conception et le développement du processus ETL (Extract-Transform-Load). À ce stade, il ne nous reste plus qu'à mettre en place l'application permettant d'interroger efficacement notre entrepôt de données. Cette application fournira les fonctionnalités nécessaires pour explorer, analyser et extraire des informations précieuses à partir des données stockées dans notre entrepôt, afin de prendre des décisions éclairées et d'obtenir des insights pertinents pour notre entreprise.

Restitution des données de l'entrepôt

Sommaire

5.1 Modélisation de l'application BI	68
5.1.1 Types d'analyse	68
5.1.2 Types d'application BI	69
5.1.3 Types d'utilisateurs BI	70
5.1.4 Modèle conceptuel de l'application BI	70
5.2 Développement de l'application BI	74
5.2.1 Cube tabulaire de l'entrepôt de données	74
5.2.2 Projet Power BI	77

Introduction

Ce chapitre se concentre sur la modélisation de l'application Business Intelligence (BI) en examinant ses divers types, ses utilisations et les personnes qui en bénéficient. En outre, nous explorons en détail le processus de conception et de mise en œuvre de cette application.

5.1 Modélisation de l'application BI

La modélisation de l'application Business Intelligence (BI) joue un rôle essentiel dans l'établissement d'un système BI performant. Elle établit une structure cohérente pour organiser et exploiter les données commerciales, ce qui permet aux utilisateurs d'exploiter pleinement les informations disponibles pour éclairer leurs prises de décision de manière optimale.

5.1.1 Types d'analyse

Dans le domaine de la Business Intelligence (BI), on distingue généralement trois types d'analyses : la BI stratégique, la BI tactique et la BI opérationnelle. Voici une définition de chacun de ces types :

1. **Business Intelligence stratégique** : Elle se concentre sur l'analyse des données à long terme et sur la vision globale de l'organisation. L'objectif est d'aider les dirigeants et les décideurs à formuler des objectifs stratégiques, à identifier les opportunités de croissance, à évaluer les performances globales et à prendre des décisions à long terme pour le succès de l'entreprise.
2. **Business Intelligence tactique** : La BI tactique est axée sur l'analyse des données à moyen terme. Elle vise à fournir des informations spécifiques et détaillées pour soutenir les décisions tactiques et les actions opérationnelles.
3. **Business Intelligence opérationnelle** : La BI opérationnelle se concentre sur l'analyse des données en temps réel pour prendre des décisions opérationnelles quotidiennes. Elle fournit du support informationnel aux points d'affaires de l'entreprise. Elle s'appuie sur des indicateurs de performance clés (KPI) pour surveiller les processus métier en

cours, détecter les problèmes, prendre des mesures correctives et améliorer l'efficacité opérationnelle.

5.1.2 Types d'application BI

Il existe différents types d'applications de Business Intelligence (BI) qui sont utilisées pour répondre à des besoins spécifiques d'analyse et de prise de décision. Voici quelques-uns des types d'applications BI couramment utilisés :

- Les tableaux de bord power BI : Un tableau de bord Power BI est une page unique, souvent appelée canevas, qui raconte une histoire au moyen de visualisations. Comme il est limité à une seule page, un tableau de bord bien conçu contient uniquement les éléments clés de cette histoire. Les lecteurs peuvent afficher des rapports connexes pour obtenir plus de détails[7].

- Le reporting : Le reporting en BI concentre les données collectées par une entreprise depuis diverses sources internes ou externes à l'entreprise. Son objectif principal est de rendre ces données plus accessibles et compréhensibles, afin de faciliter la prise de décision éclairée. En regroupant et en analysant les données, le reporting en BI permet de fournir des rapports clairs et concis, présentant les tendances, les performances et les indicateurs clés de manière visuellement attrayante [8].

- Le cube OLAP : C'est une solution technologique qui simplifie la manipulation de vastes volumes de données dans un but décisionnel. Il utilise une structure multidimensionnelle où chaque côté représente une dimension d'analyse, et chaque cellule contient une métrique. Cette représentation permet une visualisation des données et offre la possibilité de les explorer en détail à l'aide d'opérations spécifiques. Le cube OLAP facilite ainsi l'analyse approfondie des données en offrant une vue multidimensionnelle intuitive, permettant aux utilisateurs de découvrir des relations, des tendances et des insights pertinents pour prendre des décisions éclairées.

- Le Data Mining : le terme Data Mining désigne l'analyse de données depuis différentes perspectives et le fait de transformer ces données en informations utiles, en établissant des relations entre les données ou en repérant des patterns. Ces informations peuvent ensuite être

utilisées par les entreprises pour augmenter un chiffre d'affaires ou pour réduire des coûts. Elles peuvent également servir à mieux comprendre une clientèle afin d'établir de meilleures stratégies marketing [9].

5.1.3 Types d'utilisateurs BI

Il existe différents types d'utilisateurs :

- **Analystes** : Ce sont des professionnels spécialisés dans l'analyse des données. Ils utilisent les outils de BI pour extraire, transformer et analyser les données, afin de fournir des informations exploitables aux décideurs de l'entreprise.
- **Développeurs et administrateurs de BI** : Les développeurs et les administrateurs de BI sont responsables de la mise en place et de la gestion des systèmes de BI. Ils développent et maintiennent les applications et les infrastructures de BI, assurent la qualité des données et veillent à la sécurité et à la performance des systèmes.
- **Cadres et décideurs** : Les cadres et les décideurs utilisent la BI pour prendre des décisions stratégiques. Ils se basent sur les rapports, les tableaux de bord et les analyses pour évaluer la performance de l'entreprise, identifier les opportunités de croissance et définir les objectifs stratégiques.

5.1.4 Modèle conceptuel de l'application BI

5.1.4.1 Elaboration du modèle de cas d'utilisation

Les diagrammes de cas d'utilisation décrivent les fonctions générales et la portée d'un système. Ces diagrammes identifient également les interactions entre le système et ses acteurs. Les cas d'utilisation et les acteurs dans les diagrammes de cas d'utilisation décrivent ce que le système fait et comment les acteurs l'utilisent, mais ne montrent pas comment le système fonctionne en interne. Le diagramme de cas d'utilisation de notre application décisionnelle présente les acteurs impliqués, notamment un Admin, développeur BI et un décideur[10].

RESTITUTION DES DONNÉES DE L'ENTREPÔT

La figure 5.1 fournit une illustration visuelle de ces acteurs et de leurs interactions avec notre application décisionnelle.

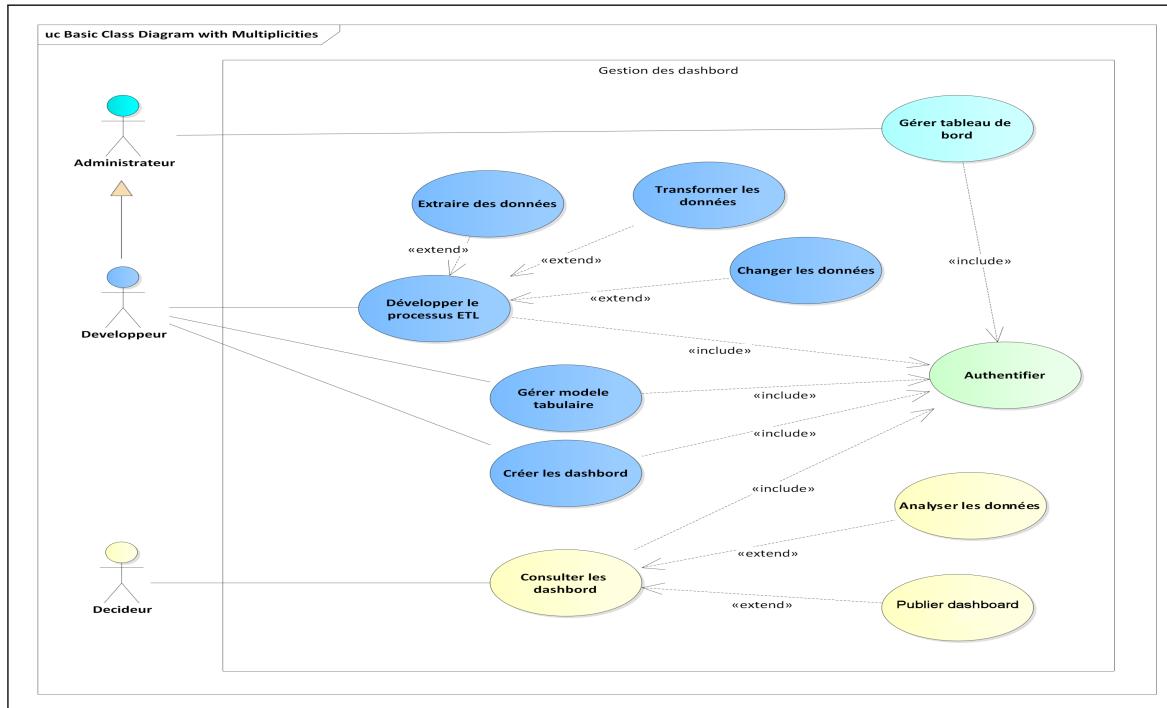


FIGURE 5.1 – Diagramme de cas d'utilisation

5.1.4.2 Description textuelle des cas d'utilisation

Pour chaque cas d'utilisation, nous avons élaboré une description textuelle détaillée. Nous commençons par la description textuelle du cas d'utilisation intitulé "Développer le processus ETL", qui est présentée dans le tableau 5.1

RESTITUTION DES DONNÉES DE L'ENTREPÔT

TABLEAU 5.1 – Description du cas d'utilisation « Développer le processus ETL »

Titre	Développer le processus ETL
Acteur	Développeur BI
Objectif	Permettre aux utilisateurs de manipuler les données
Précondition	L'utilisateur démarre l'application et procède à l'authentification
Postcondition	L'utilisateur traite les données
Scenario nominal	<ol style="list-style-type: none">1. L'utilisateur se connecte à l'application2. Le système présente à l'utilisateur une interface pour manipuler les données3. L'utilisateur sélectionne l'outil à utiliser4. L'utilisateur exécute les différentes opérations sur les données5. Le système affiche les données correspondantes
Scenario alternatif	<ol style="list-style-type: none">1. Problème au niveau de l'authentification4. Problème au niveau des données

RESTITUTION DES DONNÉES DE L'ENTREPÔT

TABLEAU 5.2 – Description du cas d'utilisation « Créer les Dashboard »

Titre	Créer les Dashboard
Acteur	Développeur BI
Objectif	Permettre aux utilisateurs de créer un tableau de bord
Précondition	L'utilisateur crée les tableaux
Postcondition	L'utilisateur traite les données
Scenario nominal	<ol style="list-style-type: none"> 1. L'utilisateur se connecte à l'application 2. Le système présente à l'utilisateur l'interface graphique dédiée à la création d'un nouveau rapport 3. Le système exécute la requête et affiche le résultat
Scenario alternatif	<ol style="list-style-type: none"> 6.1. Problème au niveau de l'authentification 8.1. Problème au niveau de l'affichage 8.2. Requête invalide

TABLEAU 5.3 – Description du cas d'utilisation « Consulter les Dashboard »

Titre	Consulter les Dashboard
Acteur	Décideur
Objectif	Permettre aux utilisateurs de consulter le tableau de bord
Précondition	L'utilisateur démarre l'application et procède à l'authentification
Postcondition	L'utilisateur examine et étudie les tableaux
Scenario nominal	<ol style="list-style-type: none"> 1. L'utilisateur se connecte à l'application 2. Le système présente à l'utilisateur une sélection de tableaux de bord qu'il souhaite consulter 3. Le système affiche le tableau de bord choisi
Scenario alternatif	<ol style="list-style-type: none"> 9.1. Problème au niveau de l'authentification 11. Problème au niveau de l'affichage

5.2 Développement de l'application BI

5.2.1 Cube tabulaire de l'entrepôt de données

Lors de la définition du modèle du cube, nous avons pris soin d'établir des clés étrangères (foreign keys) pour les tables de faits et des clés primaires (primary keys) pour les dimensions. Cette étape est essentielle pour assurer les relations entre les différentes tables du cube tabulaire. Dans le cadre d'un projet SSAS, nous avons utilisé le langage DAX avec la clause "CONCATENATE" pour créer les clés étrangères et les clés primaires. Nous avons concaténé l'identifiant de chaque enregistrement avec l'identifiant de la clinique pour établir ces relations. Cela permet d'assurer l'intégrité des données et d'éviter les redondances entre les différentes cliniques. Dans cette image, nous pouvons observer un exemple concret qui est illustré dans la figure ci-dessous.

	[PK_Med]	f(x)	=CONCATENATE(DMedecin[CodMed],DMedecin[Clinique])							
	CodMed	NomMed	Adresse	TypMed	TelBur	TelAut	MatFisc	NumCpt	Clinique	PK_Med
1	CHIP005	DR ALOUL...		P				45707005	CliniqueM	CHIP005CliniqueM
2	ANES005	DR GARNA...		P				45743005	CliniqueM	ANES005CliniqueM
3	DIEN002	DR DERO...		P				45713002	CliniqueM	DIEN002CliniqueM
4	CARD013	DR HAMMA...		P				45703013	CliniqueM	CARD013CliniqueM
5	DIV0014	HOTEL KH...		P				45714014	CliniqueM	DIV0014CliniqueM
6	INS0004	INS AZZA		P				45720004	CliniqueM	INS0004CliniqueM
7	MED012	DR ARFA F...		P				45724012	CliniqueM	MED012CliniqueM
8	RADIO12	DR KCHOK		P				45738012	CliniqueM	RADIO12CliniqueM
9	MED013	DR AYECH...		P				45724013	CliniqueM	MED013CliniqueM
10	PHEX	PHARMACI...		P				45735001	CliniqueM	PHEXCliniqueM
11	MED0014	DR BEN Y...		P				45724014	CliniqueM	MED0014CliniqueM
12	DERM003	DR JARRA...		P				45712003	CliniqueM	DERM003CliniqueM
13	CARD003	DR BOUZIRI		P				45703003	CliniqueM	CARD003CliniqueM
14	GYNE022	DR GUERM...		P				45717022	CliniqueM	GYNE022CliniqueM
15	ORTHO06	DR BOUZG...		P				45732006	CliniqueM	ORTHO06CliniqueM
16	CHES005	DR GUIGA ...		P				45705005	CliniqueM	CHES005CliniqueM
17	GYNE021	DR GUED...		P				45717021	CliniqueM	GYNE021CliniqueM
18	ORTHO15	DR HEDID...		P				45732015	CliniqueM	ORTHO15CliniqueM
19	DIV019	DIVERS AN...		P				45714019	CliniqueM	DIV019CliniqueM

FIGURE 5.2 – préparation de la clé primaire dans la table " DMedecin"

Maintenant nous avons assuré que les valeurs des clés étrangères dans la table de fait correspondent aux valeurs des clés primaires correspondantes dans la table dimension. La

RESTITUTION DES DONNÉES DE L'ENTREPÔT

dernière étape consiste à créer le modèle tabulaire en reliant les tables entre elles. Cela permet de construire des cubes, des mesures et des rapports pour l'analyse de données.

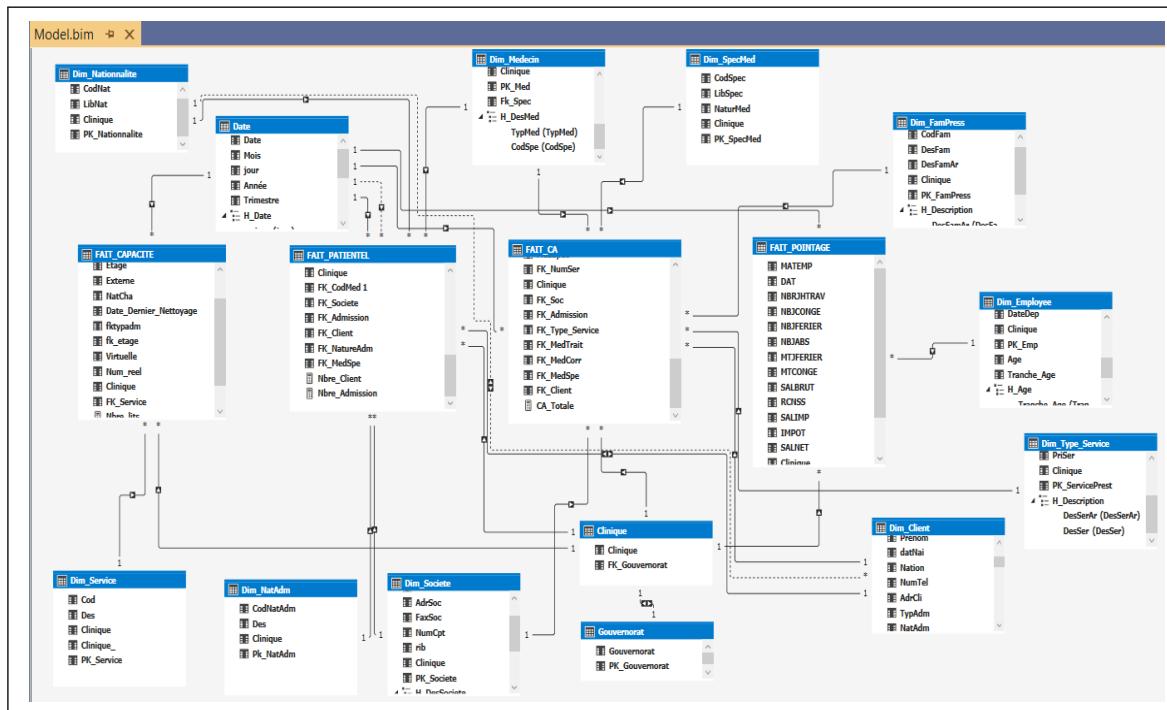


FIGURE 5.3 – Modèle tabulaire

Dans la dimension « DEmployee » nous avons créé deux hiérarchies : hiérarchie « détail employée » et l'autre « détail contrat employée » comme indique la figure 5.4.

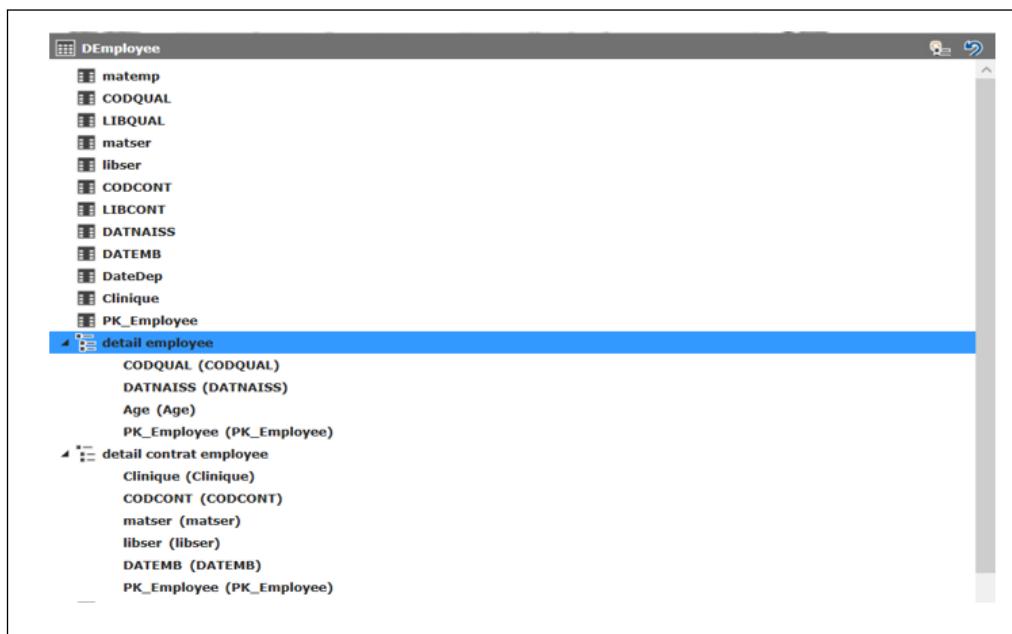


FIGURE 5.4 – Hiérarchie de la table "DEmployee"

RESTITUTION DES DONNÉES DE L'ENTREPÔT

Avec ces deux hiérarchies, nous pouvons analyser les données de la table "DEmployee" en fonction des détails individuels des employés ou des informations spécifiques à leur contrat. Cela facilite l'exploration des données et permet des analyses multidimensionnelles plus approfondies. Ces hiérarchies jouent un rôle crucial dans Power BI, car elles améliorent l'expérience utilisateur et la capacité d'analyse des données. Voici quelques raisons supplémentaires qui illustrent leur importance dans Power BI :

Création de rapports interactifs : Les hiérarchies permettent de créer des rapports interactifs et dynamiques dans Power BI. Les utilisateurs peuvent aisément sélectionner les niveaux de hiérarchie et obtenir des visualisations adaptées en fonction de leurs besoins d'analyse. Cela permet de créer des tableaux de bord intuitifs et conviviaux.

Facilitation des filtres : Les hiérarchies simplifient l'application de filtres dans Power BI. Les utilisateurs peuvent filtrer les données en fonction des niveaux de hiérarchie, ce qui permet de filtrer les données de manière plus précise et contextuelle.

Une fois que nous avons créé notre modèle de cube tabulaire dans SSAS, nous pouvons procéder au déploiement pour rendre le cube accessible dans Power BI.

La figure 5.5 montre de déploiement de notre cube tabulaire

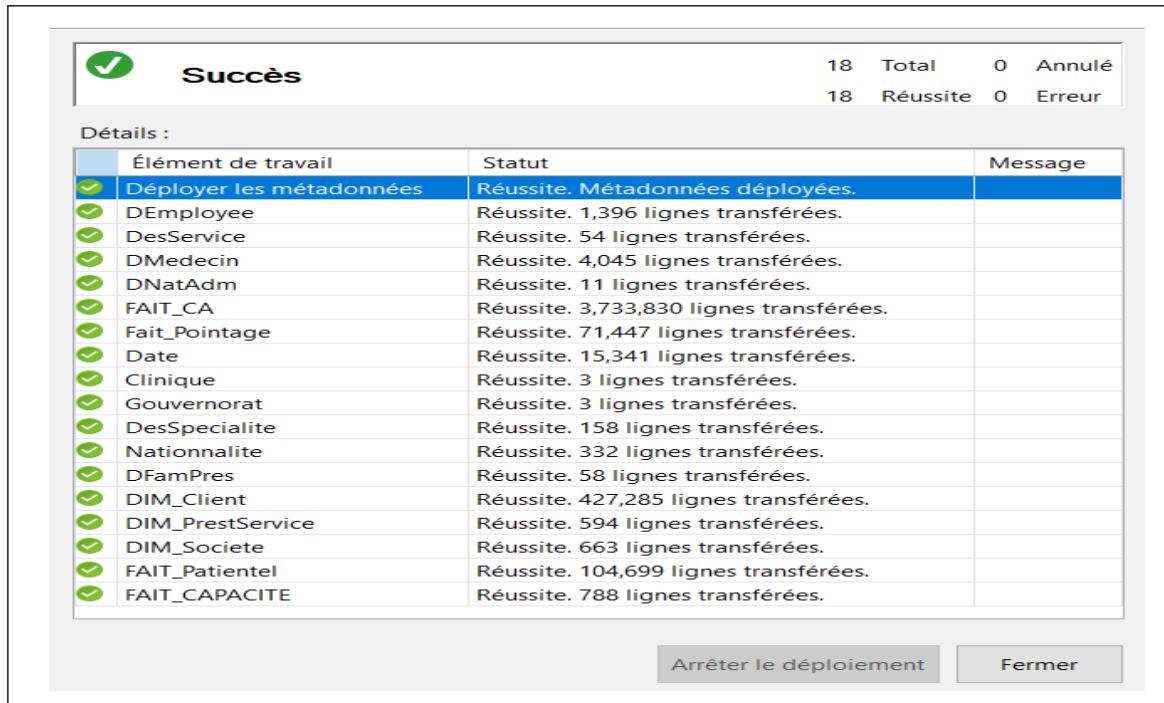


FIGURE 5.5 – Déploiement du cube tabulaire

5.2.2 Projet Power BI

5.2.2.1 Importation de la base de données

Notre première étape a consisté à importer notre base de données dans Power BI comme le montre la figure 5.6.

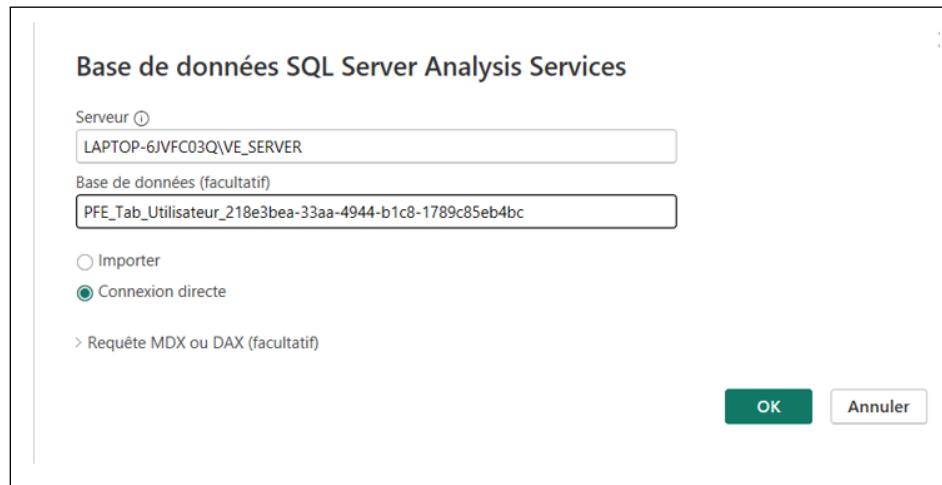


FIGURE 5.6 – Connexion directe de la base de données

Ensuite, nous avons procédé à la sélection des tables de données , puis nous les avons transformées et chargées dans notre projet power BI. Ce processus est illustré dans la figure 5.7 du rapport.

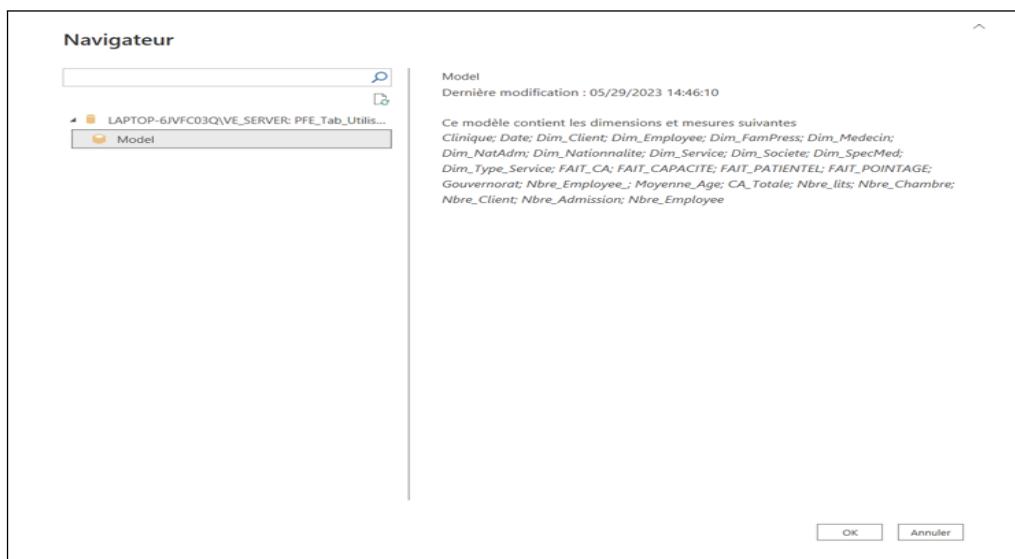


FIGURE 5.7 – Chargement des tables

RESTITUTION DES DONNÉES DE L'ENTREPÔT

La figure 5.8 présente le projet power BI



FIGURE 5.8 – Le projet power BI

5.2.2.2 Présentation du tableau de bord

Les tableaux de bord spécialement conçus pour les cadres supérieurs, offrent une analyse approfondie des principaux indicateurs tels que le chiffre d'affaires, le nombre d'admissions, le nombre d'examens, les prestations et les sous-familles de prestations, ainsi que les sociétés de prise en charge. Grâce à leur interface interactive et rapide , ces tableaux de bord simplifient considérablement la consultation et l'exploitation des données. Ils permettent aux cadres supérieurs d'accéder rapidement aux informations essentielles nécessaires pour prendre des décisions éclairées. Nous avons créé cinq tableaux de bord dans notre projet Power BI, ainsi qu'une page d'accueil qui nous offre une vue globale de notre projet. Les tableaux de bord sont un excellent moyen de visualiser et de présenter vos données de manière claire et concise.

La page d'accueil que nous avons créée peut-être très utile pour donner aux utilisateurs une vue d'ensemble rapide de ce qui est disponible dans notre projet Power BI. Elle peut également servir de point d'entrée central, permettant aux utilisateurs de naviguer facilement vers les tableaux de bord spécifiques qui les intéressent comme indiqué dans la figure 5.9.

RESTITUTION DES DONNÉES DE L'ENTREPÔT



FIGURE 5.9 – Page d'accueil

La figure 5.10 présente un aperçu de tableau de bord concernant le chiffre d'affaires par nationalité des patients, famille de prestation, service, spécialité médecin, type de payeur et par clinique.



FIGURE 5.10 – Tableau de bord « CA »

RESTITUTION DES DONNÉES DE L'ENTREPÔT

L'histogramme est largement utilisé comme outil pour résumer des données discrètes ou continues qui sont regroupées en intervalles de valeurs. C'est pourquoi nous avons choisi d'utiliser l'histogramme ci-dessous pour visualiser le nombre de chiffres d'affaires par spécialité médicale.

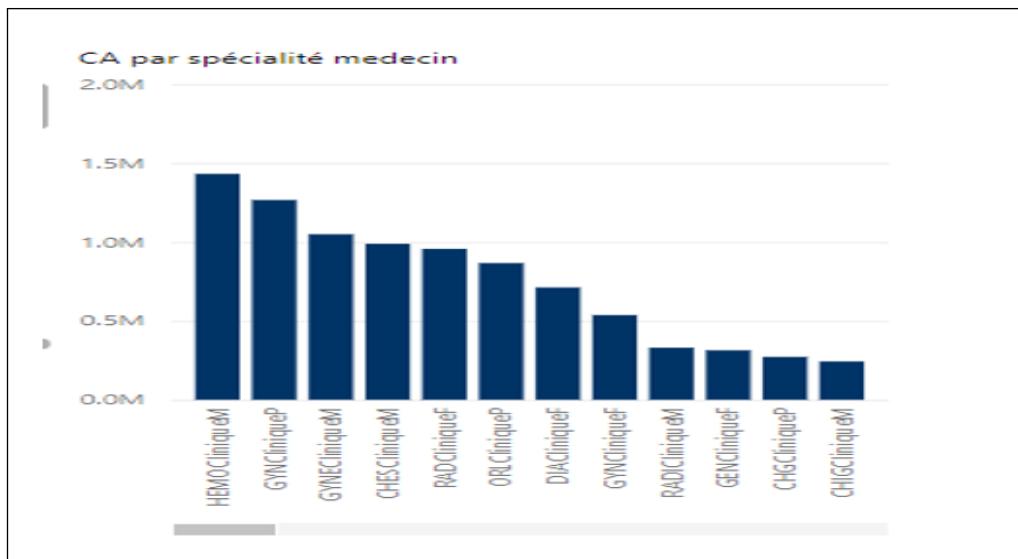


FIGURE 5.11 – Histogramme de chiffre d'affaires par spécialité médecine

Et pour visualiser le chiffre d'affaires par service, nous avons opté pour l'utilisation d'un tableau simple dans le graphique.

CA par service	
CA_ytd	DesSer
225	DINER SUPPLEMENTAIRE
110	1/2 J SEJOUR PATIENT
150	1/2 JOURNEE CHAMBRE VIP
462.068	1/2JOURNEE REA
19011.6355	1/4 J CHAMBRE INDIVIDUELLE
360	1/4 J REA
60	1/4 JOURNEE CHAMBRE INDIVIDUELLE
11054587.488382274	

FIGURE 5.12 – Tableau du chiffre d'affaires par service

RESTITUTION DES DONNÉES DE L'ENTREPÔT

Lors de la création de tableau de bord CA, il est important de déterminer les indicateurs clés de performance (KPI) financiers pertinents que nous souhaitons suivre. Ces KPI peuvent inclure des mesures telles que le chiffre d'affaires comme indiqué dans la figure 5.13.

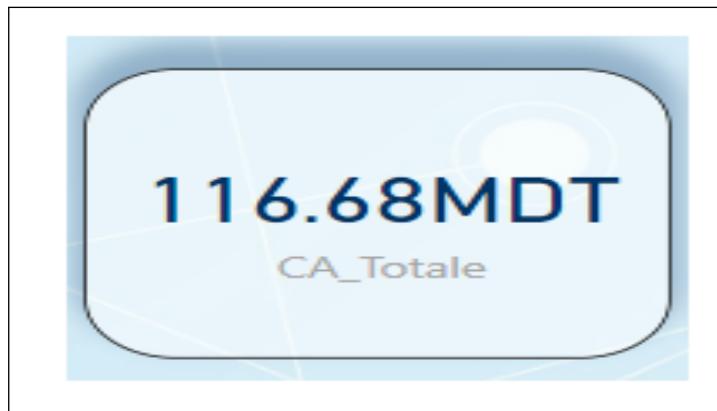


FIGURE 5.13 – Tableau du chiffre d'affaires par service

A l'aide du graphique en courbes représenté par la figure 5.14, nous avons pu suivre l'évolution du chiffre d'affaires par type de paiement des clients soit par société ou paiement directe .

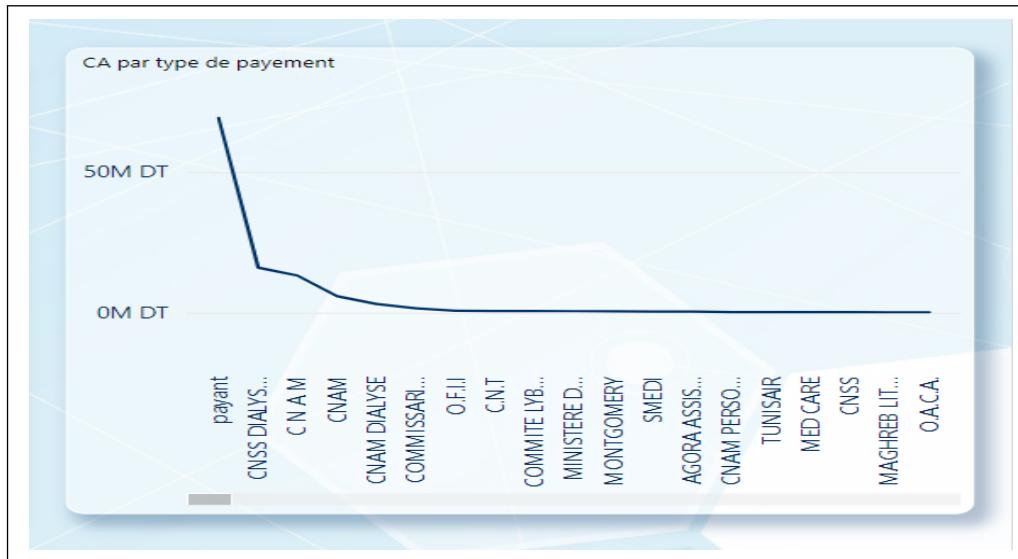


FIGURE 5.14 – Chiffre d'affaires par type de paiement

La figure 5.15 offre un aperçu du tableau de bord « patientèle », résumant le nombre de client et d'admissions par nationalité et par société de prise en charge, pour chaque clinique. Chaque catégorie est accompagnée de son pourcentage d'utilisation, offrant ainsi une vue d'ensemble rapide et informative de la répartition du chiffre d'affaires.

RESTITUTION DES DONNÉES DE L'ENTREPÔT



FIGURE 5.15 – Tableau de bord « patientèle »

La figure 5.16 présente un aperçu général et élégant du tableau de bord, offrant une synthèse visuelle du nombre des employées et de leur répartition par type de contrat et qualification, pour chaque clinique, ainsi que le nombre de jours d'absence par trimestre



FIGURE 5.16 – Tableau de bord « Employée »

RESTITUTION DES DONNÉES DE L'ENTREPÔT

Nous avons utilisé un graphique en anneau dans l'affichage du nombre de jours d'absence par trimestre parce que c'est utile pour illustrer des données de manière claire et concise, en mettant en évidence les parts relatives de chaque catégorie par rapport à l'ensemble. Cela permet de faciliter la compréhension des proportions, des tendances ou des comparaisons.

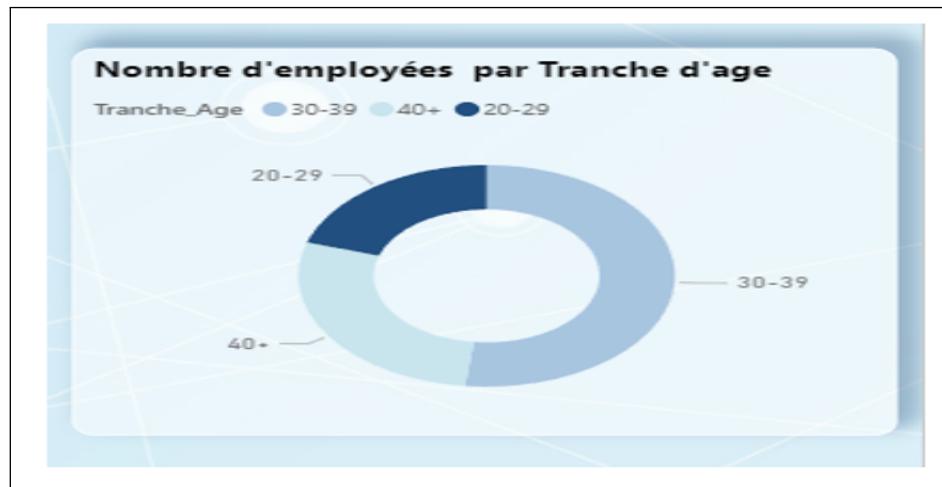


FIGURE 5.17 – Graphique en anneau pour le nombre de jour d'absence par trimestre

La dernière figure présente l'étude générale du tableau de bord « charge personnel », offrant une synthèse visuelle sur le paiement des employées et de leur répartition par type de contrat et qualification, pour chaque clinique, ainsi que le nombre de jours d'absence par trimestre.



FIGURE 5.18 – Tableau de bord " Charge Personnel "

Finalement on a créé des filtres et des options de navigation dans nos tableaux de bord , de sorte que les utilisateurs puissent explorer les données selon leurs besoins spécifiques. Par exemple, ils pourraient vouloir filtrer les données par période (mois, trimestre, année) ou par clinique (CliniqueM, CliniqueP, CliniqueF.). La figure 5.19 montre les filtres de temps et de cliniques.

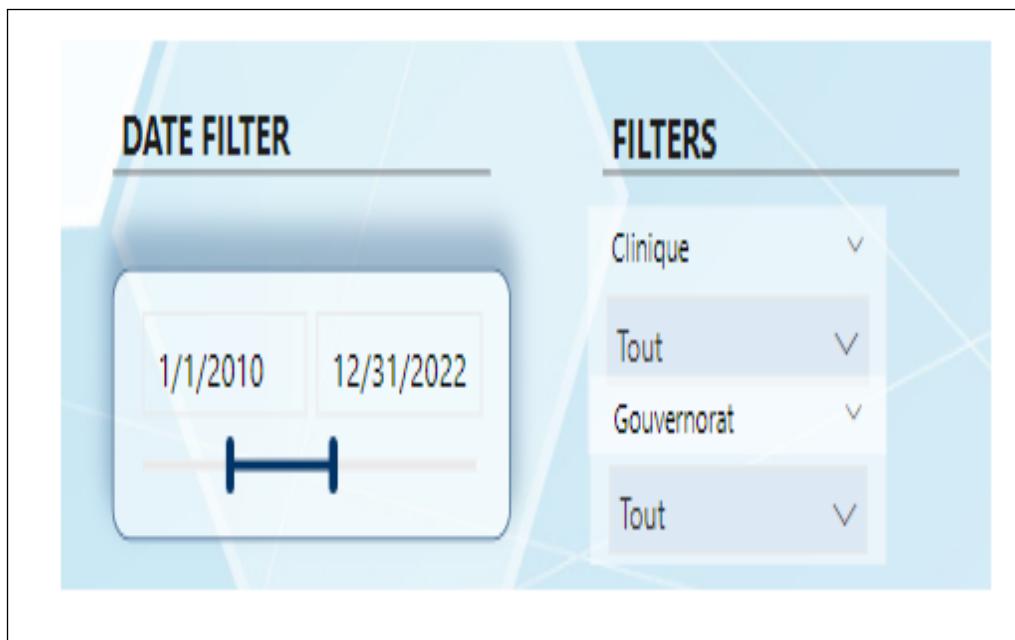


FIGURE 5.19 – Les filtres de temps et de cliniques

Conclusion

En conclusion, ce chapitre de restitution de l'application BI met en évidence l'importance d'une visualisation efficace des données pour une prise de décision éclairée. l'application BI a transformé les données complexes en visualisations claires et concises, offrant ainsi une vue d'ensemble précise des performances des cliniques étudiées. Cette restitution efficace a permis aux utilisateurs de prendre des décisions éclairées, d'identifier les tendances clés et de maximiser l'efficacité opérationnelle. En utilisant les tableaux de bord interactifs, les décideurs ont pu exploiter pleinement le potentiel des données pour optimiser la gestion des cliniques.

CONCLUSION GÉNÉRALE

En conclusion, notre projet de fin d'étude intitulé "Mise en place d'une application décisionnelle pour la gestion des cliniques de CliniSys" a permis la mise en place d'un système décisionnel pour la gestion financière, la gestion de capacité et la gestion des ressources humaines de trois cliniques, en utilisant efficacement l'ETL (Extract, Transform, Load) et Power BI. Au cours de ce projet, nous avons souligné l'importance des tableaux de bord dans la prise de décisions stratégiques informées. Grâce à leur capacité à visualiser et à interpréter les données de manière intuitive, les tableaux de bord facilitent la compréhension et la communication des performances de l'entreprise. Nous avons également mis en évidence le rôle essentiel de l'ETL dans le processus de collecte, de transformation et de chargement des données provenant de sources multiples et hétérogènes. L'ETL permet de créer un entrepôt de données homogène, prêt à être utilisé dans l'analyse et la génération de rapports. Power BI s'est révélé être une plate-forme puissante et conviviale pour la création de tableaux de bord interactifs. Nous avons exploité les fonctionnalités de visualisation et de reporting de Power BI pour fournir aux gestionnaires des cliniques des informations clés sur leurs performances financières, de la capacité d'hébergement et des ressources humaines. En mettant en œuvre notre solution, nous avons adopté une approche mixte de modélisation conceptuelle de l'entrepôt de données, en prenant en compte les besoins spécifiques des cliniques. Nous avons également développé un processus ETL robuste pour extraire, transformer et charger les données de manière efficace. En conclusion, notre projet a abouti à la création d'un tableau de bord efficace et personnalisé pour les cliniques, offrant une visibilité accrue sur leurs opérations financières et administratives. Nous sommes convaincus que notre solution contribuera à améliorer la prise de décisions et à optimiser les performances des cliniques, en exploitant pleinement le potentiel des données disponibles. Nous espérons

CONCLUSION GÉNÉRALE

que notre rapport et notre travail apporteront une valeur ajoutée et significative à l'organisme d'accueil et serviront de base solide pour de futures initiatives d'analyse et de reporting

BIBLIOGRAPHIE

- [1] <https://www.lebigdata.fr/businessintelligencedefinition?fbclid=IwAR1ZfJPyoctpyvvx4tioz7-KEhnI7T9ozZbjeEFz7zoQfqKqaZzN3nXkNtI>
- [2] <https://azure.microsoft.com/fr-fr/resources/cloud-computing-dictionary/what-is-a-data-warehouse/>
- [3] [Ghozzi 2004] Ghozzi F., “Conception et manipulation de bases de données dimensionnelles à contraintes”. Thèse de doctorat, Université Paul Sabatier, Toulouse, France, 2004.
- [4] https://datascientest.com/olap-tout-savoir-sur-cette-puissante-organisation-des-bases-de-donnees?fbclid=IwAR2QF_MIHNTg_ONaJn6ouQBf6Lp7xYK2UajL_FKvsfakQsPvkVzDWTZSvWA
- [5] <https://medium.com/data354/le-langage-dax-1ce3e347659f>
- [6] [Kimball et al., 2002] Ralph Kimball, Margy Ross. The datawarehouse toolkit. Wiley Compute, 2002 modèle conceptuel de l’application BI
- [7] <https://learn.microsoft.com/fr-fr/power-bi/create-reports/service-dashboards>
- [8] <https://www.mediacritik.com/reporting-bi-definition-fonctionnement>
- [9] <https://www.lebigdata.fr/data-mining-definition-exemples>
- [10] <https://www.ibm.com/docs/fr/rational-soft-arch/9.5?topic=diagrams-use-case/>

Mise en place d'une application décisionnelle pour la gestion de cliniques des CliniSys

Sirine BOUZIDI et Taha DAOUD

الخلاصة:

في الختام، أدى مشروع تخرجنا المسمى "تصميم وتطوير أداة تحليل وإعداد التقارير" إلى إنشاء نظام يساعد على أخذ القرارات بكفاءة في ما يخص المالية و الموارد البشرية لثلاثة عيادات، باستخدام إنشاء بيئة مالئمة للتحليل وإنتاج ETL استخراج وتحويل وتحميل التقارير. يوفر هذا النظام رؤية شاملة ودقيقة للبيانات، مما يسمح باتخاذ قرارات مستنيرة .لمسؤولي العيادات

Résumé :

En résumé, notre projet de fin d'étude intitulé "La mise en place d'une application décisionnelle pour la gestion des cliniques de CliniSys" a permis la mise en place d'un système décisionnel pour la gestion financière, gestion de capacité et gestion des ressources humaines de trois cliniques, en utilisant efficacement l'ETL (Extract, Transform, Load) et Power BI.

Abstract:

In conclusion, our end-of-study project entitled "The implementation of a decision-making application for the management of CliniSys clinics" has enabled the implementation of a decision-making system for the financial and human resources management of three clinics, effectively using ETL (Extract, Transform, Load) and Power BI.

الكلمات المفاتيح: BI power , ETL, تطبيق ذكاء الأعمال

Mots clés : Power BI, ETL, Application décisionnels.

Key-words: Power B, ETL, Business intelligence application.