

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université SAAD DAHLAB- BLIDA 1

Faculté des sciences

Département informatique



Mémoire pour l'obtention d'un diplôme de Master en informatique.

Option : Ingénierie du Logiciel

Thème

**La conception et la réalisation d'un système décisionnel
pour une entreprise d'expertise et d'assurance**

Organisme d'accueil :

Société Algérienne d'Expertise et Contrôle Technique Automobile (SAE)

Réalisé par :

- RECHOUM Massinissa Nabil
- BENALI Mohamed Oussama

Présenté le 08 Septembre 2020, devant le jury composé de :

- | | |
|-----------------------|--------------|
| - Mme CHIKHI Imane | Présidente |
| - Mme BEY Fella | Examinatrice |
| - Mme LAHIANI Nesrine | Promotrice |
| - Mr ADGHAR Abdellah | Encadreur |

Promotion 2019-2020

Remerciements

Nos remerciements s'adressent à Dieux tout puissant qui nous a donné le courage, la force et les moyens pour réussir et donner le meilleur de nous-mêmes.

Nos remerciements s'adressent à notre promotrice Madame LAHIANI Nesrine, notre encadreur Monsieur ADGHAR Abdellah et l'équipe informatique de SAE pour avoir accepté de nous diriger et pour leurs compétences qui nous ont été d'une grande aide.

Nous tenons à remercier les membres du jury qui nous font l'honneur d'évaluer ce travail.

Ce travail n'aurait pas été possible sans le soutien de nos familles, c'est pour ça que nous tenons à remercier nos parents ainsi que nos proches pour tous les encouragements.

Nos remerciements les plus chaleureux vont à tous ceux qui nous ont apportés de l'aide de loin ou de prêt.

Et enfin nous tenons à remercier nos amis qui ont contribué à l'accomplissement de ce travail.

Dédicaces

Je dédie tout d'abord ce travail à ma jolie grande famille , mes très chère parents pour leurs patience, encouragement et les valeurs de la vie qui m'ont enseigné pour aller toujours de l'avant, faire de mon mieux et surtout de ne jamais abandonner

A ma fiancée, mes frères et ma sœur qui ont contribué chacun a leurs manières durant ces années d'études à ma réussite

Egalement à l'équipe IMKAN, spécialement à LOUAHEDJ Noureddine et à la deuxième famille qui sont les personnes qui m'entoure et qui m'enrichie chaque jours avec qui je partage mon quotidien difficile ou joyeux soit-il, merci de votre soutiens et de votre présence

Enfin je remercie Oussama un binôme pas comme les autres, courageux, généreux et une force de caractère qui m'a beaucoup motivé et stimulé durant ces derniers mois.

RECHOUM Massinissa Nabil

Dédicaces

Je dédie ce mémoire à ma très chère famille ma mère, mon père mes frères mes tantes ainsi que ma grand-mère qui m'ont beaucoup aidé moralement et financièrement et qui m'ont soutenu pour en arriver là depuis le tout début de mes études.

A mon oncle Karim qui a été décisif dans le stage pratique.

A notre promotrice Madame Lahiani qui nous a tout d'abord acceptés dans des conditions difficiles et nous a assistés jusqu'à la fin du projet avec ses précieux conseils et son travail remarquable.

Et surtout ma fiancée qui a toujours été là dans les bons et mauvais jours à m'encourager et me remonter le moral à chaque fois dans ce parcours.

A tous les employés de la SAE Cheraga pour leur aide précieuse surtout Mr Abdelah, Mr Tarek et Mr Ilias.

Et enfin à mon binôme Rechoum.

Merci à tous....

BENALI MOHAMED OUSSAMA

Table des matières

Résumé	ix
LISTE DES FIGURES	x
LISTE DES TABLEAUX	xii
LISTE DES ABREVIATIONS	xiii
Introduction générale	1
1. Problématique.....	2
2. Objectifs du Projet.....	2
CHAPITRE 1 : GENERALITE SUR L'INFORMATIQUE DECISIONELLE	3
I. Data Warehouse.....	4
I.1. Introduction	4
I.2. Système opérationnel vs système décisionnel	4
I.3. Objectifs d'avoir un système décisionnel.....	5
I.4. Définition.....	5
I.5. Caractéristiques d'un DataWarehouse	6
I.6. Histoire de Data Warehouse	7
I.7. La structure de système décisionnel.....	8
I.8. Architecture de Data Warehouse	12
I.9. Data warehouse management	13
II. Modélisation des données de l'entrepôt.....	15
II.1. La modélisation dimensionnelle.....	15
II.1.1. Principe de fait	16
II.1.2. Dimension	17
II.1.3 Comparaison entre les tables de faits et les tables de dimensions	18
II.2. Différents modèles de la modélisation dimensionnelle.....	18
II.2.1. Le model en étoile (Star schéma)	18
II.2.2. Modèle en flocon	19
II.2.3. Modèle en constellation (fact-flack)	20
II.3. Le concept OLAP	21
II.4. Models de stockage pour OLAP.....	21
II.4.1. MOLAP – Multidimensionnel OLAP	21
II.4.2. ROLAP – Relational OLAP	21
II.4.3. HOLAP – Hybride OLAP.....	22
II.4.4. DOLAP – Desktop OLAP.....	22
II.5. La navigation dans les données	22
II.5.1. Slice & Dice.....	22

II.5.2. Drill-down & Roll-up	24
II.6. Reporting.....	25
II.7. Tableau de bord en BI.....	25
II.8. Data mining.....	26
III. Démarche de Construction d'un Data Warehouse	26
III.1. Définir les besoins commerciaux de l'entreprise	26
III.2. Identifier les sources de données.....	26
III.3. Les approches de modélisation et construction de data warehouse.....	26
III.3.1. Approche ascendante	27
III.3.2. Approche descendante.....	28
III.3.3. Approche Hybride	29
III.4. Identifier l'environnement de développement	30
III.5. Alimentation de l'entrepôt de données	30
III.6. Développement frontal	33
III.7. Tester	34
III.8. Documentations.....	34
III.9. Évolution des exigences.....	34
CHAPITRE 2 : PRESENTATION DE SAE/EXACT ET L'ETUDE D'EXISTENCE.....	35
I. Pr�setation SAE/EXACT	36
II. D�finition de l'expertise.....	36
III. Types d'expertise fournis par SAE/EXACT	36
III.1. Expertise automobile	36
III.2. Expertise B�timent et Risques Divers	37
III.3. Evaluations.....	38
III.4. Expertise M�dicale	38
III.5. Expertise V�t�rinaire	38
III.6 Expertise Agricole.....	38
IV. Organisation de la soci�t�.....	38
V. Les chiffres de la soci�t�.....	39
VI. Etude de l'existence	40
VII. Niveau direction g�n�ral et direction de la finance et comptabilit� (DFC).....	40
IIX. Niveau direction de centre.....	40
IX. Conclusion.....	40
CHAPITRE 3 : ETUDE DES BESOINS ET CONCEPTION DE LA ZONE D'ENTROPOSAGE DES DONNEES	41
I. Etude des besoins.....	42

II. Démarche d'extraction les besoins	42
III. Problèmes et obstacles rencontrés	44
IV. Conception.....	44
V. Les dimensions (axes d'analyse).....	44
V.1. Dimension Temps	44
V.2. Dimension Centre.....	45
V.3. Dimension Client	45
V.4. Dimension Expert.....	46
V.5. Dimension Type d'expertise	47
VI. Modèles en étoile	47
VI.1. Le modèle en Etoile Expertise	48
VI.2. Modèle en étoile des impayés	49
VI.3. Modèle en étoile des règlements	49
VI.4. Modèle en étoile des annulés	50
VII. Conclusion.....	52
CHAPITRE 4 : ALIMENTATION DU DATA WAREHOUSE.....	53
I. Introduction	54
II. Etude pré-alimentaire.....	54
II.1. L'étude de la source des données.....	54
II.2. Détection l'emplacement des informations	55
II.3. Définir la périodicité de chargement des données	55
III. Politique d'alimentation.....	55
IV. Processus de chargement	55
IV.1. Processus de chargement des dimensions	57
IV.2. Processus de chargement des tables de faits	58
IV.3. Processus de chargement particulier	59
V. Conclusion	60
CHAPITRE 5 : REALISATION DU PROJET.....	61
I. Introduction	62
II. Présentation des différents outils	62
II.1. Pentaho Data Integration (PDI)	62
II.2. PostGreSQL	63
II.2.2. Pourquoi choisir postgresql ?.....	63
I.3 Power BI	64
III. Présentation des étapes de réalisation de notre entrepôt de données	67
III.1. Création du DATA WAREHOUSE	67

III.2 Alimentation de l'entrepôt de données	68
III.3 Visualisation des tableaux de bords et des rapports	75
IV. Architecture de système développé.....	80
V. Conclusion	80
Conclusion Générale	81
Bibliographie.....	82

Résumé

L'entreposage de données est une phase du processus décisionnel qui supporte efficacement le processus OLAP, il est né dans les entreprises pour l'aide à la prise de décision. Ainsi, les principaux utilisateurs de ces technologies font partie intégrante de l'entreprise, Nous pouvons citer les secteurs qui peuvent tirer profit des outils d'aide à la décision comme les banques et les assurances, ainsi que ceux de l'automobile et des institutions médicales.

La réalisation de l'entrepôt de données au sein de la SAE, facilitera la prise de décision, et donnera une vision plus claire sur la situation des centres, connaître les compagnies qui augmentent leur chiffre d'affaire, visualisation la liste des experts qui ont un rendement élevé et de savoir la créance des clients.

Abstract

Data warehousing is a phase of the decision-making process that effectively supports the OLAP process, it was born in enterprises to support decision-making. Thus, the main users of these technologies are an integral part of the company. We can cite the sectors that can benefit from decision support tools such as banks and insurance, as well as those of the automobile and institutions medical.

The realization of the data warehouse within the SAE, will facilitate decision-making, and will give a clearer vision on the situation of the centers, know the companies, which increase their turnover, visualize the list of the experts who have high yield and know the debt of customers.

ملخص

تخزين البيانات هو مرحلة من عملية صنع القرار التي تدعم بشكل فعال عملية OLAP ، وقد ولدت في المؤسسات لدعم اتخاذ القرار. وبالتالي يعتبر المستخدمون الرئيسيون لهذه التقنيات جزءاً لا يتجزأ من الشركة، ويمكننا الاستشهاد بالقطاعات التي يمكن أن تستفيد من أدوات دعم القرار مثل البنوك والتأمين، وكذلك قطاعات السيارات والمؤسسات الطبية.

إن تحقيق مستودع البيانات داخل SAE سيسهل اتخاذ القرار وسيعطي رؤية أوضح عن حالة المراكز، ومعرفة الشركات التي تزيد رقم الأعمال، الاطلاع على قائمة بالخبراء الذين لديهم عائد مرتفع ومعرفة ديون العملاء .

LISTE DES FIGURES

Figure 1. 1: Les caractéristiques de Data Warehouse.....	7
Figure 1. 2: La structure de système décisionnel [3].....	8
Figure 1. 3: Exemple de Métadonnées Technique [4].....	10
Figure 1. 4: Les différents types de Data Mart [5].....	11
Figure 1. 5: Architecture de Data Warehouse [6].....	13
Figure 1. 6: Exemple de schéma en étoile [6].....	15
Figure 1. 7: Considération d'un sujet d'analyse comme un cube à plusieurs dimensions [7]..	16
Figure 1. 8: Structure générale d'une table de faits [8].....	17
Figure 1. 9: Structure générale d'une table de dimension [8].....	18
Figure 1. 10: Exemple d'un modèle en étoile [9].....	19
Figure 1. 11: Exemple d'un modèle en flocon de neige [9].....	20
Figure 1. 12: Exemple d'un model en constellation [6].....	20
Figure 1. 13: Principe de l'architecture MOLAP [10].....	21
Figure 1. 14: Principe de l'architecture ROLAP [10].....	22
Figure 1. 15: Exemple sur le Slicing [11].....	23
Figure 1. 16: Exemple du Dicing [11].....	23
Figure 1. 17: Exemple de Roll-Up [11].....	24
Figure 1. 18: Exemple de Drill-Down [11].....	25
Figure 1. 19: Approche de Kimball [6].....	27
Figure 1. 20: Approche de Inmon [6].....	28
Figure 1. 21: Approche Hybrid [13].....	29
Figure 1. 22: Processus de l'ETL [6].....	31
Figure 2. 1: Hiérarchie de l'entreprise.....	39
Figure 2. 2: Graphe d'évolutions de chiffre d'affaire.....	39
Figure 3.1: Dimension Temps.....	44
Figure 3.2: Dimension Centre.....	45
Figure 3.3: Dimension Client.....	46
Figure 3.4: Dimension Expert.....	46
Figure 3. 5: Dimension Type Expertise.....	47
Figure 3. 6: Modèle en étoile Expertise.....	48
Figure 3.7 : Modèle en étoile Impayé.....	49
Figure 3.8: Modèle en étoile Règlement.....	50
Figure 3.9: Modèle en étoile Annulé.....	51
Figure 4.1: Liste des schémas d'une base de données.....	54
Figure 4.2: Diagramme d'activité du processus générale d'alimentation de DW.....	56
Figure 4.3: Diagramme d'activité de chargement de tables de dimensions.....	58
Figure 4.4: Diagramme d'activité de chargement de tables de faits.....	59

Figure 5.1: Pentaho Data Integration (PDI) [16]	63
Figure 5. 2 : Logo PostgreSQL [17].....	63
Figure 5. 3: Processus général de fonctionnement Power BI [18].....	64
Figure 5.4: Composant logiciel Power BI [18].....	65
Figure 5. 5 : Procédure de fonctionnement Power BI [18].....	65
Figure 5.6: Exemple d'une page de rapport Power BI [18].....	66
Figure 5.7: Exemple d'une vignète dans Power BI [17]	67
Figure 5.8 Le schéma de DW.....	68
Figure 5. 9: Palette de création.....	69
Figure 5.10: Composant d'extraction de données	69
Figure 5.11: Paramétrage d'un composant ETL (extract)	70
Figure 5.12: Comparaison entre les données de la dimension et les données sources	70
Figure 5.13: Paramètres de la comparaison	71
Figure 5.14: Paramètres avancé « synchronize »	72
Figure 5.15: Schéma de correspondance	72
Figure 5.16: Paramètre générale.....	73
Figure 5.17: Exemple d'alimentation d'une table de dimension	73
Figure 5.18: Planification l'exécution de la tâche.....	74
Figure 5.19: Exemple d'une tâche.....	74
Figure 5.20: Exemple d'une tâche globale	75
Figure 5. 21: Fenêtre de PowerBI	76
Figure 6.22: Outil de connexion à la source de données	77
Figure 5.23: Liste des capteurs de connexion	77
Figure 5.24: Tableau de bord globale	78
Figure 5.25: Tableau de bord détail fait Expertise	79
Figure 5.26: Tableau de bord détail paiement et annulé.....	79
Figure 5. 27 Architecture de projet	80

LISTE DES TABLEAUX

Tableau 1. 1: Les déférences entre système décisionnel et système opérationnel	5
Tableau 1. 2: Tableau comparatif entre les tables de faits et les tables de dimensions	18
Tableau 1. 3: Avantages et inconvénients d'approche ascendante	27
Tableau 1. 4: Avantages et inconvénients d'approche descendante	29
Tableau 1. 5: Avantages et inconvénients d'approche hybrid.....	30
Tableau 3. 1: Synthétisation des besoins recensés	44
Tableau 3.1: Tableau descriptif de la dimension « Temps »	45
Tableau 3.2: Tableau descriptif de la dimension « Centre »	45
Tableau 3.3: Tableau descriptif de la dimension « Client »	46
Tableau 3.4: Tableau descriptif de la dimension « Expert »	47
Tableau 3.5: Tableau descriptif de la dimension « Décade »	47
Tableau 3.6: Tableau descriptif de fait « Expertise »	48
Tableau 3.7: Tableau descriptif de fait « Impayé »	49
Tableau 3.8: Tableau descriptif de fait « Règlement »	50
Tableau 3.9: Tableau descriptif de fait « Annulé »	51

LISTE DES ABREVIATIONS

BI : Business Intelligence.

DIM : Dimension.

DW : Data Warehouse (Entrepôt de données).

ETL: Extract, Transform and Load (ETC).

FK: Foreign Key.

DG: Direction Générale

DFC: Direction Finance et Comptabilité

HOLAP: Hybrid On Line Analytical Process.

MOLAP: Multidimensional On Line Analytical Process.

OLAP: On Line Analytical Process.

OLTP: On Line Transactional

PK: Primary Key.

ROLAP: Relational On Line Analytical Process.

SI : Systèmes d'Information.

SID : Systèmes d'Information Décisionnels.

SGBD : Système de Gestion de Base de Données.

SQL : Structured Query Language.

Introduction générale

Dans un environnement hautement concurrentiel qu'évolue la majorité des entreprises qui exige à ces entreprises de suivre le marché très étroite afin de répondre aux besoins de leurs clients et partenaires le plus rapide possible.

Pour cela, les dirigeants de l'entreprise devront prendre des décisions les plus opportunes. Ces décisions, qui influenceront grandement sur la stratégie de l'entreprise et donc sur son devenir, ne doivent pas être prises ni à la légère, ni de manière trop hâtive, compte tenu de leurs conséquences sur la survie de l'entreprise. Il s'agit de prendre des décisions basées sur des informations claires, fiables et pertinentes. Le problème est de savoir donc comment identifier et présenter ces informations à qui de droit, sachant par ailleurs que les entreprises croulent d'une part sous une masse considérable de données et que d'autre part les systèmes opérationnels « transactionnels » s'avèrent limités, voire inaptes à fournir de telles informations et constituer par la même un support appréciable à la prise de décision.

C'est dans ce contexte que les « systèmes décisionnels » ont vu le jour. Ils offrent aux décideurs des informations de qualité sur lesquelles ils pourront s'appuyer pour arrêter leurs choix décisionnels. Pour se faire, ces systèmes utilisent un large éventail de technologies et de méthodes, dont les « entrepôts de données » (Data Warehouse) représentent l'élément principal et incontournable pour la mise en place d'un bon système décisionnel.

Ce projet tend à mise en place un système décisionnel pour la Société Algérienne d'Expertise (SAE) en basant sur leur système transactionnel qui comporte une grande masse de données afin d'offrir aux décideurs les données nécessaires et claires pour faciliter la prise des décisions

1. Problématique

SAE est une grande société qui est leader dans le domaine d'expertise, elle a des centres d'expertise un peu partout sur le territoire national. Cette société d'un ERP qui gère leur système d'information où chaque centre a sa propre base de données (32 bases de données).

Dans un pareil contexte, la plus simple des opérations d'analyse devient une tâche ardue. En effet, la société se trouve dans l'incapacité de faire des analyses fiables, efficaces et à des moments opportuns sans engager des moyens considérables sur des périodes plus ou moins longues. Ainsi, les principales difficultés rencontrées peuvent être résumées en :

- L'inexistence d'un système décisionnel
- L'inventaire se fait semi-automatique
- La masse importante de données
- Difficultés dans l'élaboration des rapports d'activité

2. Objectifs du Projet

Afin de pallier aux problèmes précédemment cités, la direction générale d'informatique de la société a initié ce présent projet.

Ce projet a pour but d'introduire, en premier lieu, une informatique décisionnelle au sein d'entreprise tout en conférant aux décideurs un support fiable pour une meilleure prise de décision. Ainsi, les principaux objectifs assignés au projet sont :

- Elaboration des rapports de différents axes d'analyses selon le besoin.
- Donner la main aux décideurs pour générer leurs propres rapports.
- Elaborer des tableaux de bords
- La réduction du nombre d'intervenants lors de la production de rapports.
- Offrir des informations fiables, cohérentes et pertinentes, contenant le langage business souhaité.

CHAPITRE 1 : GENERALITE SUR L'INFORMATIQUE DECISIONELLE

I. Data Warehouse

I.1. Introduction

Toutes les entreprises du monde disposent d'une masse de données plus ou moins considérable. Ces informations proviennent soit de sources internes (générées par leurs systèmes opérationnels au fil des activités journalières), ou bien de sources externes (web, partenaire, ... etc.)

Cette masse de données avec les systèmes opérationnels n'ont pas la capacité d'exploiter des données afin d'analyser ou prendre une décision.

En effet, les décideurs de l'entreprise ont besoin d'avoir une vision sur leur entreprise, l'environnement et ces évolutions, pour cela ils ont investi sur le « Business Intelligence », en français « l'informatique décisionnel ».

Le concept de Data Warehouse, tel que connu aujourd'hui, est apparu pour la première fois en 1980, l'idée consistait alors à réaliser une base de données destinée exclusivement au processus décisionnel. Les nouveaux besoins de l'entreprise, les quantités importantes de données produites par les systèmes opérationnels et l'apparition des technologies aptes à sa mise en œuvre ont contribué à l'apparition du concept « Data Warehouse » comme support aux systèmes décisionnels.

I.2. Système opérationnel vs système décisionnel

Les systèmes « opérationnels » ou « de gestion », également appelés systèmes OLTP (on-line transaction processing), sont dédiés aux métiers de l'entreprise pour les assister dans leurs tâches de gestion quotidiennes et donc directement opérationnels.

Les systèmes « décisionnels », également appelés OLAP (on-line analytical processing), sont dédiés au management de l'entreprise pour l'aider au pilotage de l'activité, et donc indirectement opérationnels. Ils offrent au décideur une vision transversale de l'entreprise. La tendance pour réaliser un système décisionnel est à la mise en place d'un entrepôt de données. [1]

	Système opérationnel (OLTP)	Système décisionnel (OLAP)
Utilisateur	Grand public : ils sont destinés à toute personne participant à la vie quotidienne de l'entreprise.	Petit nombre d'utilisateurs : quelques personnes dans l'entreprise (décideurs).
	Le niveau des besoins analytiques est bas	Le niveau des besoins analytiques est haut
	Une seule vision métier	Plusieurs visions métiers
Données	Données atomiques : on manipule un produit, une ligne de commande, une facture.	Données générales: Les décideurs veulent voir l'ensemble de l'activité.
	BDD normalisée (3 ^{ème} forme normale)	BDD dénormalisée
	Petite volumétrie des données : les systèmes de gestion gèrent des Giga Octets de données.	Gros volumes de données : les SID doivent regrouper toutes les données de l'entreprise.
	Lecture, écriture et modification des données	Données en lecture seule
	Hétérogènes : Les systèmes OLTP sont souvent disparates en termes de technologie utilisée.	Homogènes
	A besoin des informations récentes.	A besoin de garder l'historique des transactions.
Système	Extrêmement rapides	Plus c'est rapide, mieux c'est.
	Fermés : on ne laisse pas la place à l'improvisation dans les OLTP, les choix sont restreints, les utilisateurs sont guidés dans le processus.	Ouverts : les environnements d'un SID doivent permettre d'accéder le plus simplement possible aux données.
	Transactionnels : les OLTP fonctionnent en utilisant le principe de transaction.	Non transactionnels : L'utilisateur doit pouvoir commencer une analyse, revenir en arrière ...
	Fragmentés : ou décentralisés. Sauf dans le cas des ERP.	Centralisés : toutes les données sont regroupées en une même source.

Tableau 1. 1: Les différences entre système décisionnel et système opérationnel

I.3. Objectifs d'avoir un système décisionnel

- Permet le développement d'applications décisionnelles et de pilotage de l'entreprise et de ses processus.
- Joue un rôle de référentiel pour l'entreprise puisqu'il permet de fédérer des données souvent éparpillées dans différentes bases de données.
- Offre une vision globale et orientée métier de toutes les données que manipule l'entreprise.
- Permet de faire face aux changements du marché et de l'entreprise
- Offre une information compréhensible, utile, rapide et à jour.
- Fournit un avantage concurrentiel.
- Génère un retour sur investissement élevé.

I.4. Définition

« Le Data Warehouse est une collection de données orientées sujet, intégrées, non volatiles et évolutives dans le temps, organisées pour le support d'un processus d'aide à la décision. » [2]

I.5. Caractéristiques d'un DataWarehouse

- **Orienté sujet** : le Data Warehouse est organisé autour des sujets majeurs de l'entreprise, contrairement à l'approche transactionnelle utilisée dans les systèmes opérationnels, qui sont conçus autour d'applications et de fonctions telles que : cartes bancaires, solvabilité client..., les Data Warehouse sont organisés autour de sujets majeurs de l'entreprise tels que : clientèle, ventes, produits... Cette organisation affecte forcément la conception et l'implémentation des données contenues dans le Data Warehouse. Le contenu en données et en relations entre elles diffère aussi. Dans un système opérationnel, les données sont essentiellement destinées à satisfaire un processus fonctionnel et obéit à des règles de gestion, alors que celles d'un Data Warehouse sont destinées à un processus analytique [20].
- **Intégrée** : le Data Warehouse va intégrer des données en provenance de différentes sources. Cela nécessite la gestion de toute incohérence.
- **Evolutives dans le temps** : Dans un système décisionnel il est important de conserver les différentes valeurs d'une donnée, cela permet les comparaisons et le suivi de l'évolution des valeurs dans le temps, alors que dans un système opérationnel la valeur d'une donnée est simplement mise à jour. Dans un Data Warehouse chaque valeur est associée à un moment.
- **Non volatiles** : c'est ce qui est, en quelque sorte la conséquence de l'historisation décrite précédemment. Une donnée dans un environnement opérationnel peut être mise à jour ou supprimée, de telles opérations n'existent pas dans un environnement Data Warehouse.

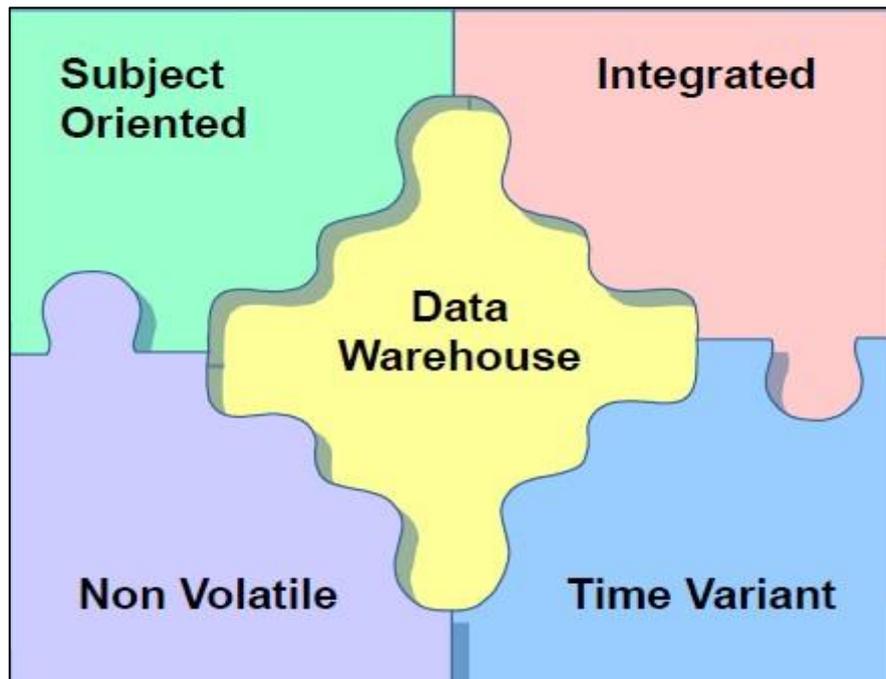


Figure 1. 1: Les caractéristiques de Data Warehouse

I.6. Histoire de Data Warehouse

Au début, le Data Warehouse n'était rien d'autre qu'une copie des données du système opérationnel prise de façon périodique, dédiée à un environnement de support à la prise de décision. Ainsi, les données étaient extraites du système opérationnel, stockées dans une nouvelle base de données « concept d'infocentre », le motif principal étant de répondre aux requêtes des décideurs sans pour autant altérer les performances des systèmes opérationnels. Voici quelques événements d'évolution de Data Warehouse :

- **1960** : Dartmouth et General Mills dans un projet de recherche conjoint, développent les termes dimensions et faits.
- **1970** : A Nielsen et IRI introduisent des magasins de données dimensionnelles pour les ventes au détail.
- **1983** : « Tera Data » Corporation présente un système de gestion de base de données spécialement conçu pour l'aide à la décision.
- **1990** : « Red Brick Systems » crée 'Red Brick Warehouse', un système spécifiquement destiné à la construction de l'entrepôt de données

L'entreposage de données a commencé à la fin des années **1980** lorsque les employés d'IBM, Paul Murphy et Barry Devlin, ont développé le Business Data Warehouse.

Cependant, le vrai concept a été donné par Inmon Bill. Il était considéré comme le père d'un entrepôt de données. Il avait écrit sur une variété de sujets concernant la construction, l'utilisation et l'entretien de Data Warehouse et de l'usine d'information d'entreprise. [3]

I.7. La structure de système décisionnel

L'environnement du Data Warehouse est constitué essentiellement de cinq composantes :

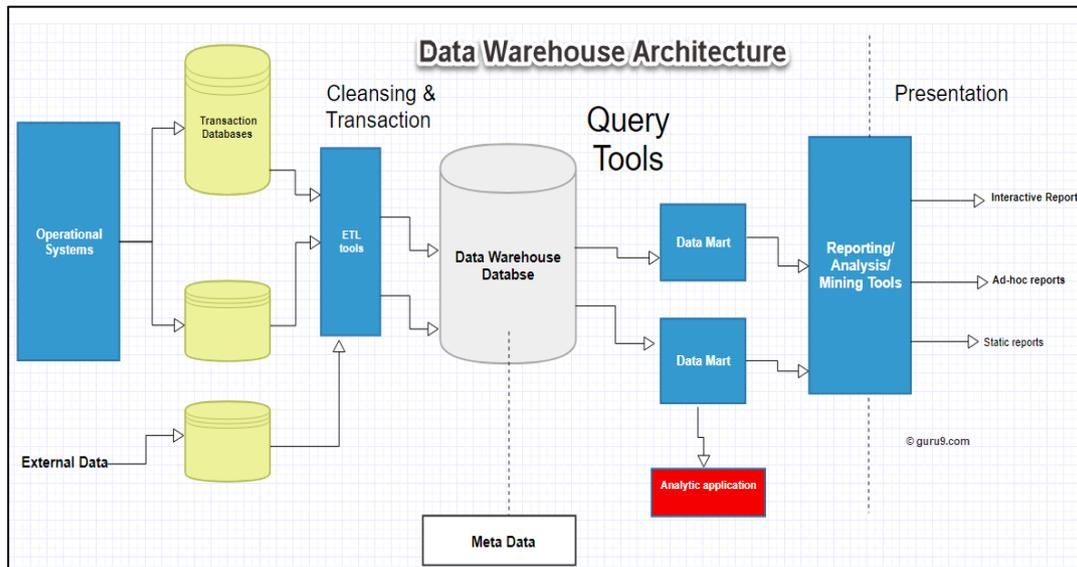


Figure 1. 2: La structure de système décisionnel [3]

- **Source de données :**

Une source commune pour les données dans les entrepôts de données est les bases de données opérationnelles de l'entreprise ou des bases de données externes de l'entreprise, qui peuvent être des bases de données relationnelles ou non relationnelles.

- **Outils d'extraction, de transformation et de chargement (ETL) :**

Les outils d'extraction, de transformation et de migration des données sont utilisés pour effectuer toutes les conversions, les synthèses et toutes les modifications nécessaires pour transformer les données en un format unifié dans le Data Warehouse. Ils sont également appelés outils d'extraction, de transformation et de chargement (ETL).

« Un point très important, dans l'aménagement d'un entrepôt de données, est d'interdire aux utilisateurs l'accès à la zone de préparation des données, qui ne fournit aucun service de requête ou de présentation » [2]

Leurs fonctionnalités incluent :

- Anonymiser les données conformément aux dispositions réglementaires.
- Élimination des données indésirables dans les bases de données opérationnelles du chargement dans l'entrepôt de données
- Recherchez et remplacez les noms et définitions communs pour les données provenant de différentes sources
- Calcul des résumés et des données dérivées

- En cas de données manquantes, remplissez-les avec des valeurs par défaut.
- Données répétées déboulonnées provenant de plusieurs sources de données.

Ces outils d'extraction, de transformation et de chargement peuvent générer des tâches cron, des tâches d'arrière-plan, des programmes Cobol, des scripts shell, etc. qui mettent régulièrement à jour les données dans Data Warehouse. Ces outils sont également utiles pour maintenir les métadonnées.

Ces outils ETL doivent faire face aux défis de l'hétérogénéité des bases de données et des données [21].

- **Métadonnées :**

Les métadonnées sont des données sur les données qui définissent l'entrepôt de données. Il est utilisé pour la construction, la maintenance et la gestion de l'entrepôt de données.

Dans l'architecture de l'entrepôt de données, les métadonnées jouent un rôle important car elles spécifient la source, l'utilisation, les valeurs et les caractéristiques des données de l'entrepôt de données. Il définit également comment les données peuvent être modifiées et traitées. Il est étroitement lié à l'entrepôt de données.

- a. **Métadonnées techniques** : ce type de métadonnées contient des informations sur l'entrepôt qui sont utilisées par les concepteurs et les administrateurs d'entrepôt de données.
- b. **Métadonnées d'entreprise** : ce type de métadonnées contient des détails qui permettent aux utilisateurs finaux de comprendre facilement les informations stockées dans l'entrepôt de données.

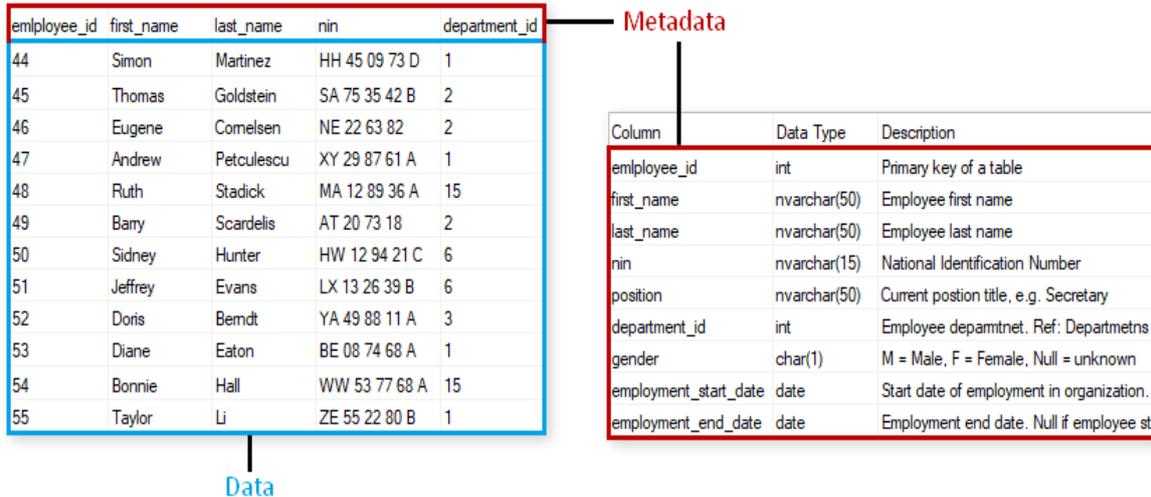


Figure 1. 3: Exemple de Métadonnées Technique [4]

Entrepôt de données d'entreprise/Data Mart :

- I. **Entrepôt de données d'entreprise :** Un entrepôt de données d'entreprise est conçu comme un entrepôt unifié et centralisé contenant toutes les informations transactionnelles de l'organisation, à la fois actuelles et historiques. Un entrepôt de données d'entreprise doit intégrer des données de tous les domaines liés à l'entreprise, tels que le marketing, les ventes, les finances, les ressources humaines.
- II. **Data Mart :** en français « magasin de données » sont des sous-ensembles d'entrepôts de données orientés pour des secteurs d'activité spécifiques, tels que les ventes ou la finance. Un entrepôt de données combine généralement les informations de plusieurs magasins de données dans plusieurs secteurs d'activité, mais un magasin de données contient des données provenant d'un ensemble de systèmes sources liés à un seul secteur d'activité.

Il existe deux types de Data Mart :

- ✓ **Dépendant :** C'est les données sont chargées à partir d'un DW basées sur les besoins départementaux.
- ✓ **Indépendant :** les données à charger dans les data marts proviennent directement des systèmes opérationnels.

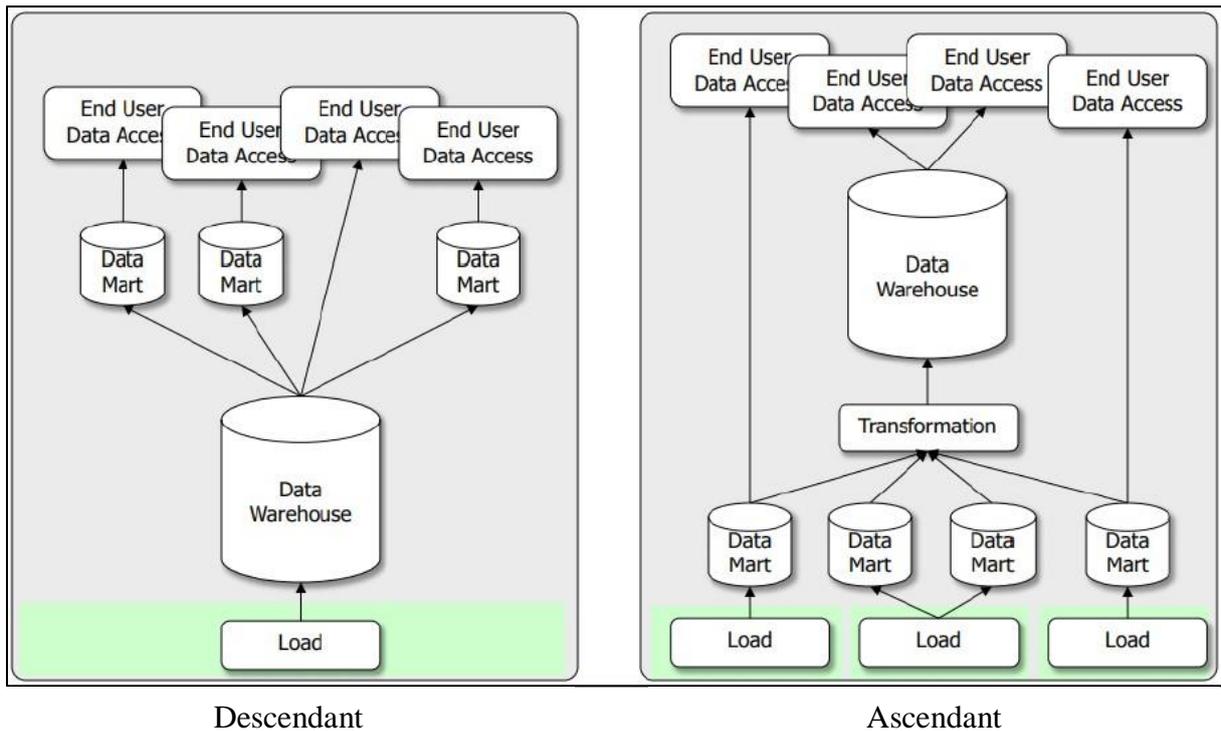


Figure 1. 4: Les différents types de Data Mart [5]

- **Outils d’analyses :**

L'un des principaux objectifs de l'entrepôt de données est de fournir des informations aux entreprises pour prendre des décisions stratégiques. Les outils de requête permettent aux utilisateurs d'interagir avec le système d'entrepôt de données.

Ces outils se répartissent en quatre catégories différentes :

- ❖ **Requêtage ad-hoc :** Le requêtage ad-hoc reste très fréquent dans ce type de projet. En effet, les utilisateurs de l’entrepôt de données, et spécialement les analystes, seront amenés à interagir avec le DW via des requêtes ad-hoc dans le but de faire les analyses requises par leurs métiers et, d’élaborer aussi, des rapports et des tableaux de bords spécifiques.
- ❖ **Executive Information System (EIS) :** est un mode de représentation des données décisionnelles, au sein d'un système d'information. Il se fait par l'intermédiaire de tableaux de bord de haut niveau. Souvent utilisé au plus haut niveau de la pyramide décisionnaire d’une entreprise, l'EIS doit permettre d'avoir une vision générale (Big picture) grâce à un nombre très restreint d'informations.
- ❖ **Data Mining :** Le terme de data mining couvre toutes les méthodes d’analyse non dirigées et partiellement automatisées qui visent à identifier

les modèles, tendances et relations pertinentes dans la base de données. Les outils d'exploration de données sont basés sur des méthodes statistiques et mathématiques ainsi que sur l'intelligence artificielle (AI) et les techniques d'apprentissage de la machine. La quantité de données que les entreprises génèrent, traitent et fusionnent dans un DWH à des fins d'analyse croît de manière exponentielle. Le volume moyen des données dans le monde double tous les deux ans. Dans ce contexte, les méthodes d'exploration des données prennent une importance croissante dans le cadre de l'entreposage de données.

- ❖ **Outils OLAP :** Parmi les outils d'analyse et d'évaluation des données disponibles, les applications OLAP se sont particulièrement imposées comme des interfaces utilisateurs standard pour le data warehousing. Les outils qui sont utilisés dans le contexte du traitement analytique en ligne apportent aux utilisateurs diverses fonctions afin de formuler des requêtes à l'entrepôt de données. Ils sont utilisés pour naviguer dans l'ensemble multidimensionnel de données.

L'affichage via OLAP permet de modéliser des données formatées en fonction d'un nombre quelconque de dimensions prédéfinies. Les analystes peuvent utiliser diverses opérations de base afin d'éditer un cube OLAP.

- ❖ **Tableau de bord :** Le tableau de bord est un outil de pilotage permettant de quantifier l'activité de l'entreprise, d'un service, d'une personne. A partir de données provenant d'une ou plusieurs sources. Le tableau de bord fournit une vision synthétique du passé, du présent et par conséquent du futur de l'entreprise, d'un service, d'une personne.
- ❖ **Reporting :** C'est la présentation périodique de rapports et bilans analytiques sur les activités et résultats d'une organisation, d'une unité de travail ou du responsable d'une fonction, destinée à en informer ceux chargés de les superviser en interne ou en externe, ou tout simplement concernés par ces activités ou résultats.

I.8. Architecture de Data Warehouse

L'architecture de Data Warehouse utilise une structure à trois niveaux composés des niveaux suivants [17] :

- **Niveau inférieur** : ce niveau contient le serveur de base de données utilisé pour extraire les données de nombreuses sources différentes, telles que des bases de données transactionnelles utilisées pour les applications frontales.
- **Niveau intermédiaire** : le niveau intermédiaire abrite un serveur OLAP, qui transforme les données en une structure mieux adaptée à l'analyse et aux requêtes complexes. Le serveur OLAP peut fonctionner de deux manières : soit en tant que système de gestion de base de données relationnelle étendu qui mappe les opérations sur les données multidimensionnelles aux opérations relationnelles standard (OLAP relationnel), soit en utilisant un modèle OLAP multidimensionnel qui implémente directement les données et opérations multidimensionnelles.
- **Niveau supérieur** : le niveau supérieur est la couche client. Ce niveau contient les outils utilisés pour l'analyse de données de haut niveau, les rapports de requête et l'exploration de données.

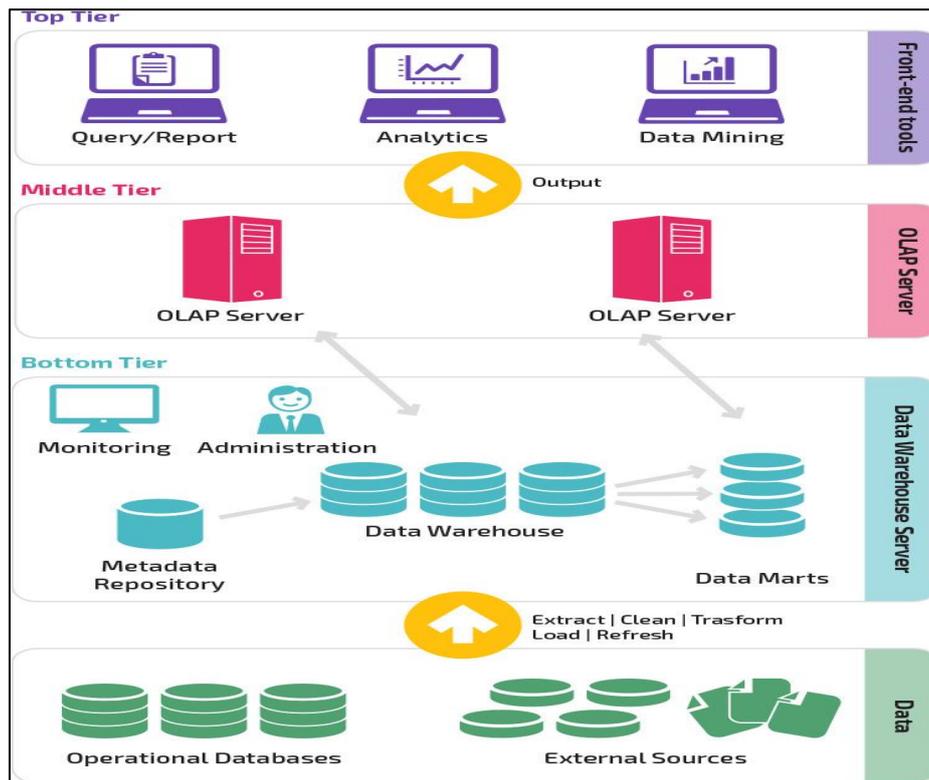


Figure 1. 5: Architecture de Data Warehouse [6]

I.9. Data warehouse management

Des outils spéciaux sont actifs à tous les niveaux du DWH et sont regroupés sous le terme de warehouse management. Le rôle de ces composants consiste à mettre en place, maintenir et exploiter toutes les fonctions d'administration nécessaires à l'entreposage des

données. Les principales tâches du « responsable DWH » sont la planification des processus DWH, la gestion des métadonnées, la gestion de la sécurité ainsi que la gestion du système.

- **Scheduling** : Ceci comprend le contrôle des processus DWH. Les fonctions d'administration dans le contexte de l'ordonnancement peuvent être classées comme suit en fonction des niveaux de l'architecture de l'entrepôt des données :
- **Collecte et entrée des données** : Au niveau de la collecte des données, le manager DWH est responsable de la conception et de la personnalisation des processus ETL. Des fonctions d'administration sont aussi prévues pour contrôler les mises à jour et la gestion de la qualité.
- **Stockage et archivage des données** : Au niveau du stockage des données, le manager DWH surveille l'utilisation de la mémoire, construit des tables d'agrégation et effectue des opérations d'archivage et de sauvegarde.
- **Fourniture et transmission des données** : Les fonctions d'administration à ce niveau comportent la gestion des utilisateurs et le suivi des temps d'exécution des requêtes.
- **Management des méta-données** : Le dépôt de métadonnées (*metadata repository*) est un élément central du manager DWH. Il contient toutes les informations nécessaires à la conception et au fonctionnement des informations relatives à la base de données du DWH. Les métadonnées stockées dans le référentiel comprennent, par exemple, la définition du schéma de base de données sous-jacent, des informations sur les structures de sauvegarde, les chemins d'accès avec la taille des fichiers, les méta-données sur la description des données sources, mais aussi les temps de mise à jour, les règles sur le nettoyage des données ainsi que la transformation des données, les index et les tables de partitions. En outre, le gestionnaire DWH s'occupe de l'échange des métadonnées entre les différentes composantes de l'entrepôt de données et apporte ainsi une base homogène pour les métadonnées.
- **Gestion de la sécurité** : Le management de la sécurité comporte divers services d'authentification, d'autorisation des utilisateurs et de cryptage.
- **Gestion du système** : Dans le cadre de la gestion du système, le manager DW assure diverses fonctions d'administration pour le fonctionnement du DWH. Il s'agit par exemple du monitoring (performance, utilisation etc.), de l'archivage ou de la sauvegarde des données.

II. Modélisation des données de l'entrepôt

II.1. La modélisation dimensionnelle

Technique de conception logique permettant de structurer les données de manière à les rendre intuitives aux utilisateurs d'affaires et offrir une bonne performance aux requêtes qui caractérise de :

- Divise les données en faits et dimensions
- Les faits (mesures) sont généralement des valeurs numériques provenant des processus d'affaires
- Les dimensions fournissent le contexte (qui, quoi, quand, où, pourquoi et comment) des faits

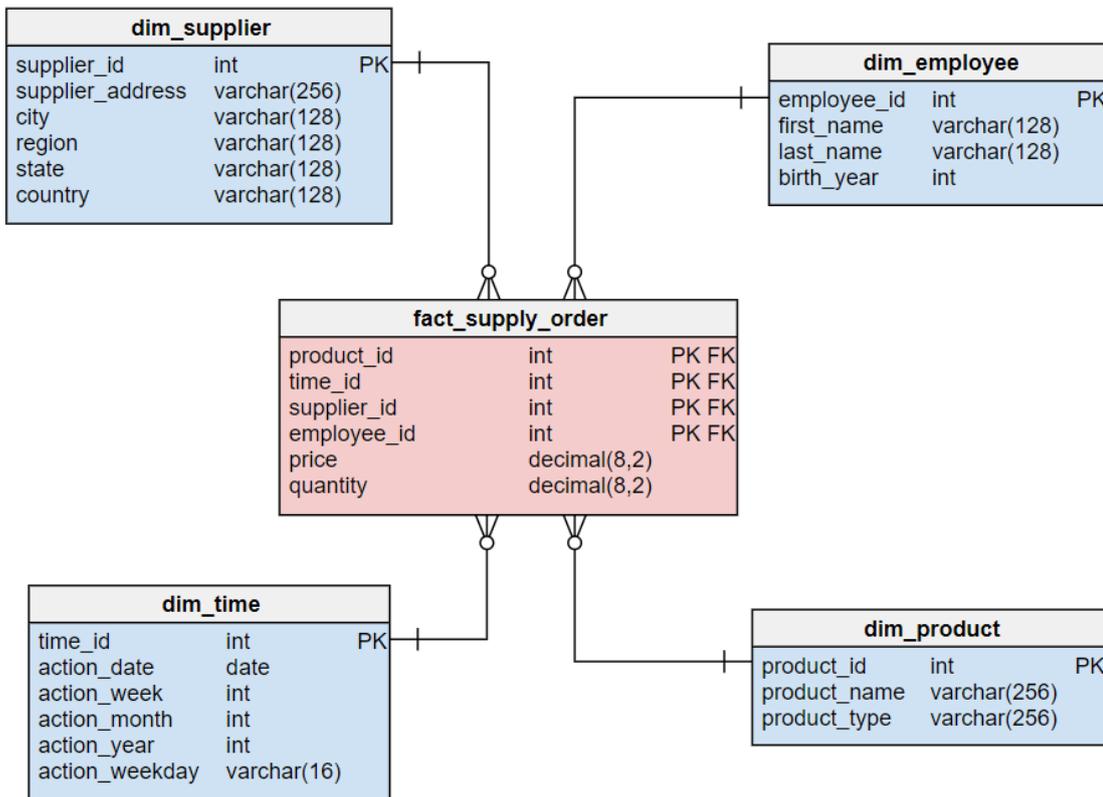


Figure 1. 6: Exemple de schéma en étoile [6]

La modélisation dimensionnelle consiste à considérer un sujet d'analyse comme un cube à plusieurs dimensions, offrant des vues en tranches ou des analyses selon différents axes.

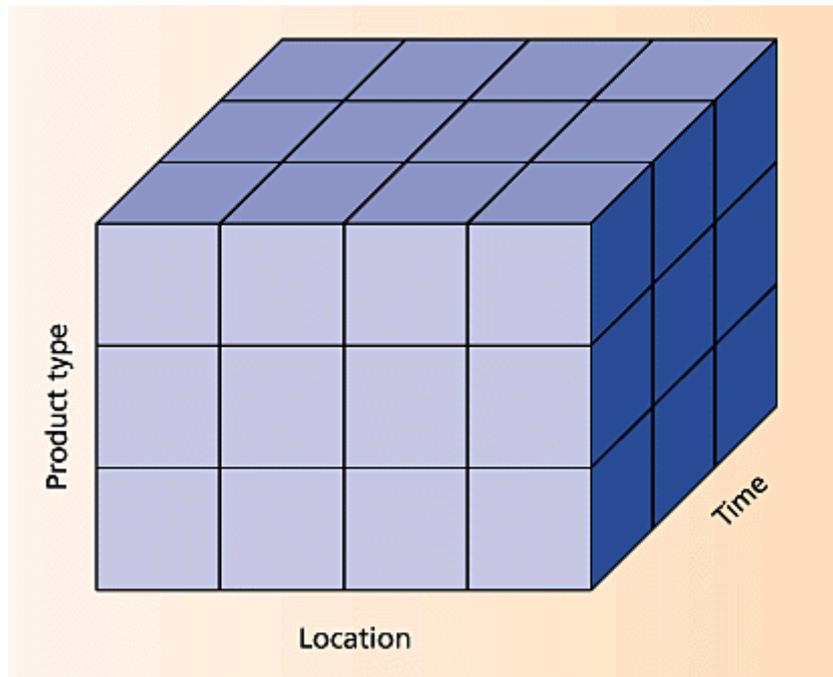


Figure 1. 7: Considération d'un sujet d'analyse comme un cube à plusieurs dimensions [7]

II.1.1. Principe de fait

Une table de faits est la table centrale d'un modèle dimensionnel, où les mesures de performances sont stockées. Une ligne d'une table de faits correspond à une mesure. Ces mesures sont généralement des valeurs numériques (additives), cependant des mesures textuelles peuvent exister mais sont rares.

Le concepteur doit faire son possible pour faire des mesures textuelles des dimensions, car elles peuvent être corrélées efficacement avec les autres attributs textuels de dimensions.

Une table de faits représente une association entre plusieurs tables de dimensions, elle doit donc respecter le principe d'intégrité référentielle. Une table de faits se représente comme suite :

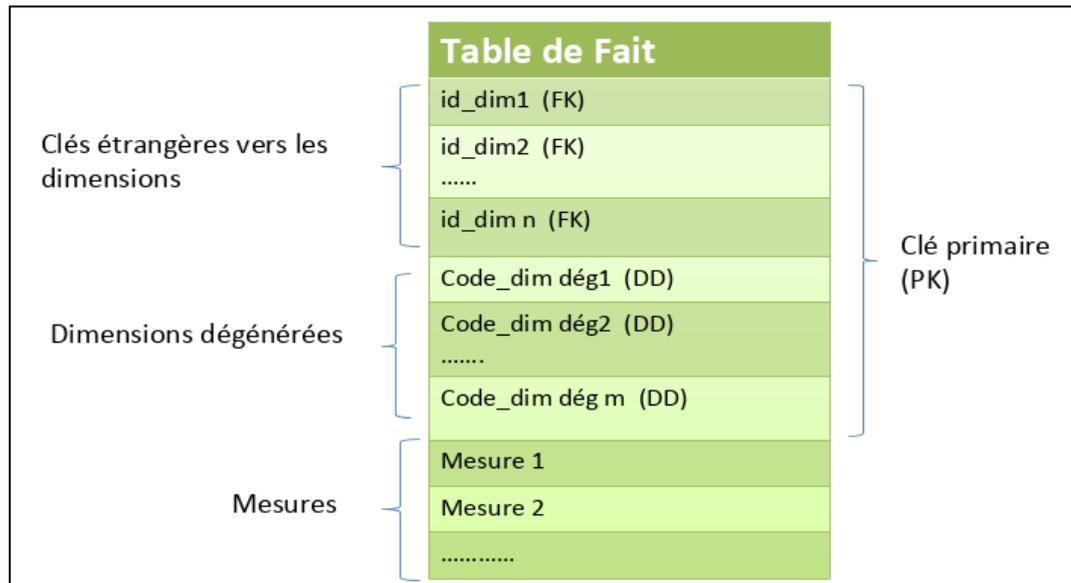


Figure 1. 8: Structure générale d'une table de faits [8]

II.1.2. Dimension

Les tables de dimension sont les tables qui rattachent une table de faits, elles contiennent les descriptions textuelles de l'activité.

Une table de dimension est constituée de nombreuses colonnes qui décrivent une ligne. C'est grâce à cette table que l'entrepôt de données est compréhensible et utilisable, elles permettent des analyses en tranches et en dés.

Une dimension est généralement constituée : d'une clé artificielle, une clé naturelle et des attributs.

« Une table de dimension établit l'interface homme / entrepôt, elle comporte une clé Primaire » [2]

Caractéristiques d'une dimension :

Une table de dimension contient le détail sur les faits.

- Une table de dimension contient les informations descriptives des valeurs numériques de la table de faits.
- Une table de dimension contient en général beaucoup moins d'enregistrements qu'une table de faits.
- Les attributs d'une table de dimension sont souvent utilisés comme « Tête de lignes » et « Tête de colonne » dans un rapport ou résultat de requête.

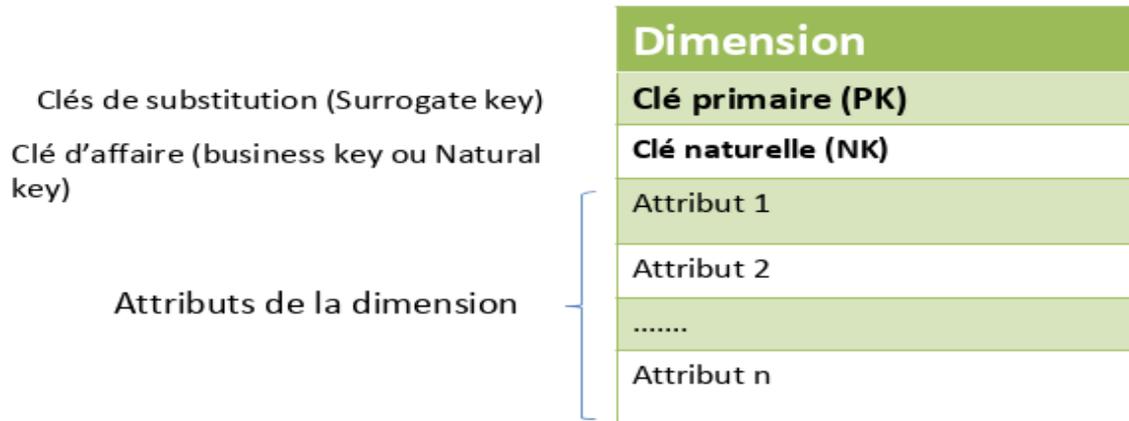


Figure 1. 9: Structure générale d'une table de dimension [8]

II.1.3 Comparaison entre les tables de faits et les tables de dimensions

Le Tableau ci-dessous montre les différences entre la table de fait et la table de dimension :

	Tables de faits	Tables de dimensions
Structure	Peu de colonnes beaucoup de lignes	Peu de lignes beaucoup de colonnes
Données	Mesurable, généralement numérique	Descriptives généralement textuelles
Référentiel	Plusieurs clés étrangères	Une clé primaire
Valeur	Prend de nombreuses valeurs	Plus ou moins constantes
Manipulation	Participe à des calculs	Participe à des contraintes
Signification	Valeurs de mesure	Descriptive
Rôle	Assure les relations entre les dimensions	Assure l'interface homme / entrepôt de données

Tableau 1. 2: Tableau comparatif entre les tables de faits et les tables de dimensions

II.2. Différents modèles de la modélisation dimensionnelle

Il existe 3 formes de modèles multidimensionnels :

1. Le modèle en étoile (Star schema).
2. Le modèle en flocon de neige (Snowflake schema).
3. Le modèle en constellation (Factflake schema).

II.2.1. Le modèle en étoile (Star schéma)

C'est un modèle sous forme d'étoile avec un centre table de fait et des dimensions comme entant des branches, la table de fait identifie les tables de dimension et il y'a absence totale de la notion d'hierarchie.

Avantage :

- Facilité de navigation.
- Performances : nombre de jointures limité.

Inconvénients :

- Dimensions dénormalisées (non en 3 FN).
- Redondances dans les dimensions.

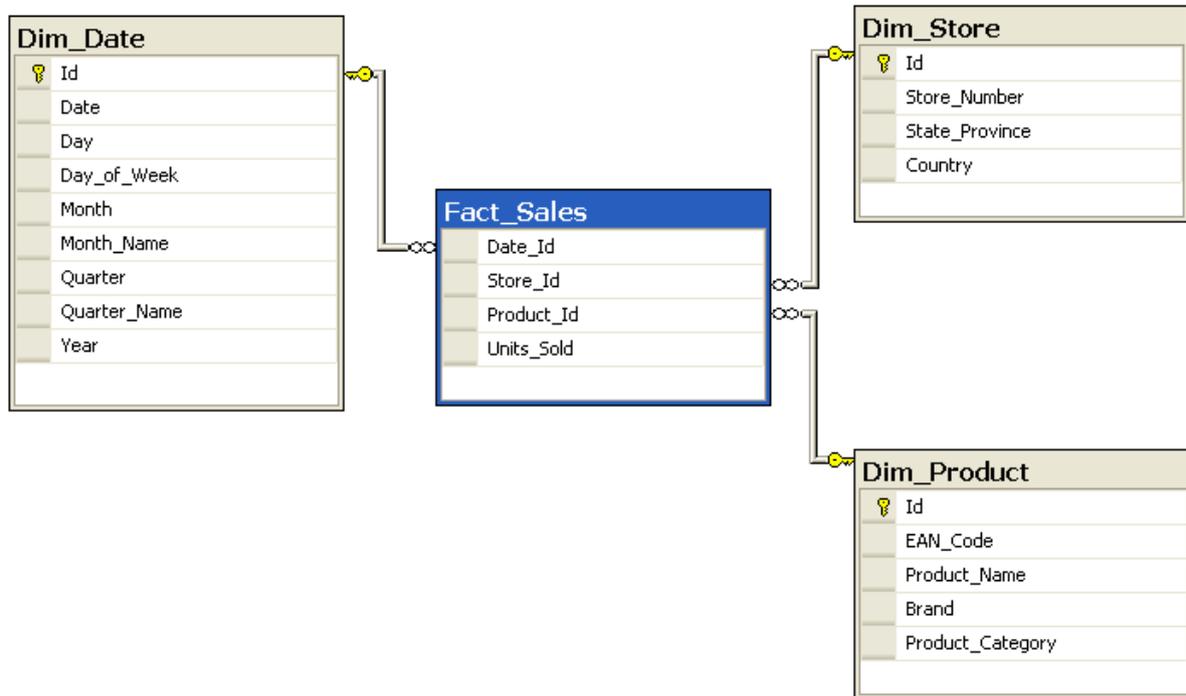


Figure 1. 10: Exemple d'un modèle en étoile [9]

II.2.2. Modèle en flocon

Pour des raisons économie d'espace on opte pour l'éclatement en hiérarchies à partir du model en étoile cela nous amène au model en flocon de neige mais cela risque d'être très couteux en termes de performance mais aussi en termes de compréhension de l'utilisateur finale.

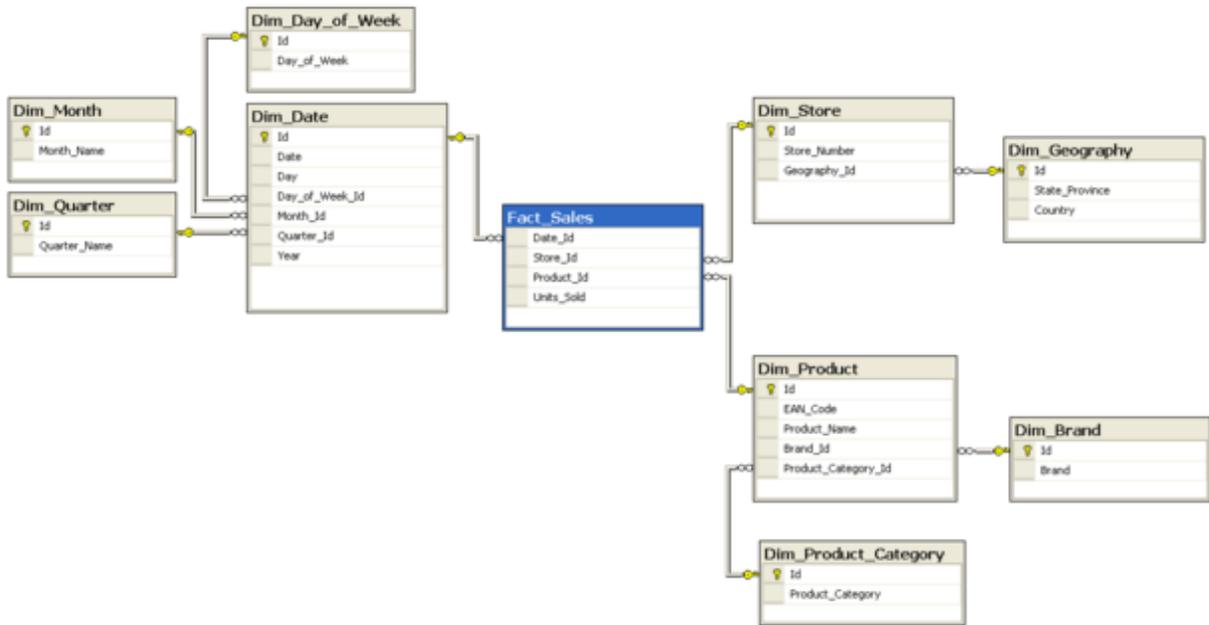


Figure 1. 11: Exemple d'un modèle en flocon de neige [9]

II.2.3. Modèle en constellation (fact-flack)

Il représente un groupement des modèles en étoile ou en flocon avec des tables de dimension communes.

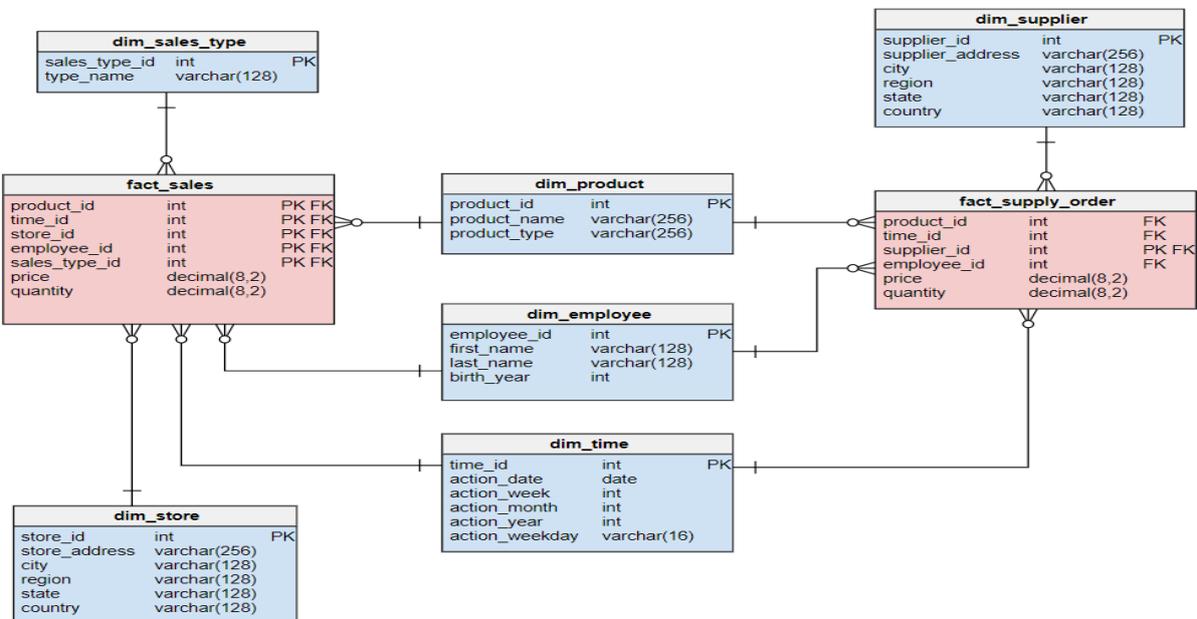


Figure 1. 12: Exemple d'un model en constellation [6]

II.3. Le concept OLAP

Les systèmes OLAP sont des processus analytiques qui permettent d'extraire et de visualiser de manière sélective des jeux de données et de leur appliquer assez facilement différents points de vue.

Pour faciliter l'analyse, les données sont stockées dans une base dite multidimensionnelle, un cube OLAP.

II.4. Modèles de stockage pour OLAP

Il existe plusieurs modèles de stockage de bases de données pour implémenter OLAP on va citer quelques un :

II.4.1. MOLAP – Multidimensionnel OLAP

La base MOLAP (Multidimensional OLAP) est l'application physique du concept OLAP. Il s'agit réellement d'une structure multidimensionnelle. Les bases MOLAP sont rapides et performantes. Elles proposent des fonctionnalités particulièrement évoluées.

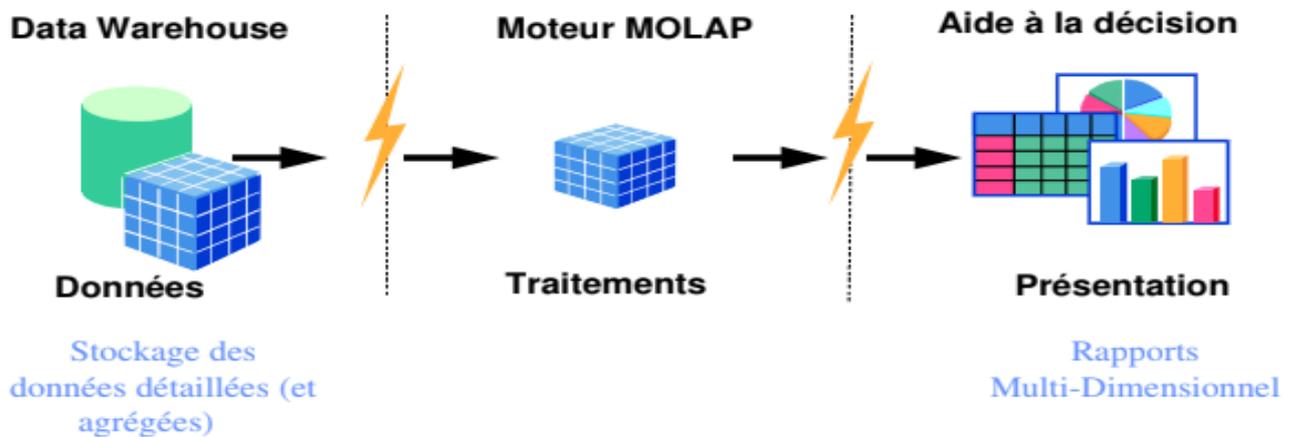


Figure 1. 13: Principe de l'architecture MOLAP [10]

II.4.2. ROLAP – Relational OLAP

La base ROLAP (Relational OLAP) est en fait une base relationnelle organisée pour fonctionner comme une base OLAP. Les bases ROLAP sont bien plus lentes et nettement moins performantes que les bases MOLAP.

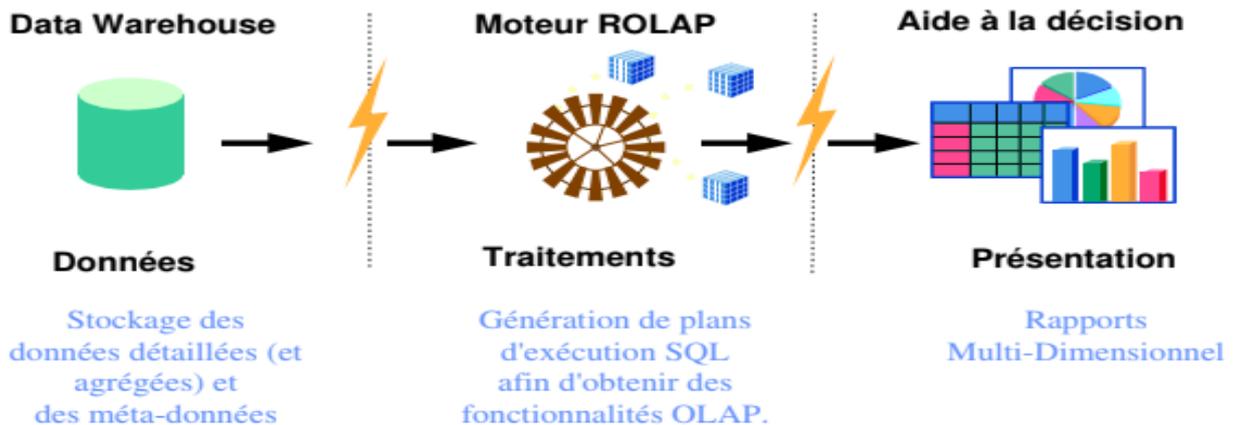


Figure 1. 14: Principe de l'architecture ROLAP [10]

II.4.3. HOLAP – Hybride OLAP

Le modèle HOLAP (Hybride OLAP) est hybride entre MOLAP et ROLAP. Il propose de cumuler les avantages des deux modèles : MOLAP et ROLAP. Les données agrégées sont stockées sous formes multidimensionnelles, alors que les données détaillées sont stockées dans des structures relationnelles.

II.4.4. DOLAP – Desktop OLAP

La base DOLAP (Desktop OLAP) est une base OLAP hébergée sur le poste client et qui est très limitée en taille. Cette utilisation est intéressante pour continuer à faire des analyses sur des données de manière nomade et déconnectée.

II.5. La navigation dans les données

Une fois que le serveur OLAP a construit le cube multidimensionnel « ou simulé ce cube selon l'architecture du serveur », plusieurs opérations sont possibles sur ce dernier offrant ainsi la possibilité de naviguer dans les données qui le constituent. Ces opérations de navigation « Data Surfing » doivent être, d'une part, assez complexes pour adresser l'ensemble des données et, d'autre part, assez simples afin de permettre à l'utilisateur de circuler de manière libre et intuitive dans le modèle dimensionnel. Afin de répondre à ces attentes, un ensemble de mécanismes est exploité, permettant une navigation par rapport à la dimension et par rapport à la granularité d'une dimension.

II.5.1. Slice & Dice

Le « Slicing » et le « Dicing » sont des techniques qui offrent la possibilité de faire des tranches « trancher » dans les données par rapport à des filtres de dimension bien précis, se classant de fait comme des opérations liées à la structure « se font sur les dimensions ». La différence entre eux se manifestent dans le fait que : Le Slicing consiste à faire une sélection de

tranches du cube selon des prédicats et selon une dimension « filtrer une dimension selon une valeur ».

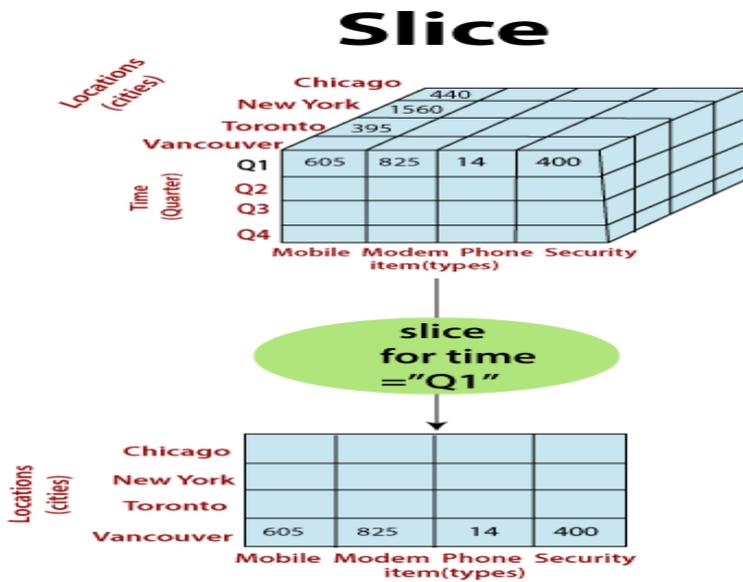


Figure 1. 15: Exemple sur le Slicing [11]

Le Dicing, quant à lui, peut être vu comme étant une extraction d'un sous cube.

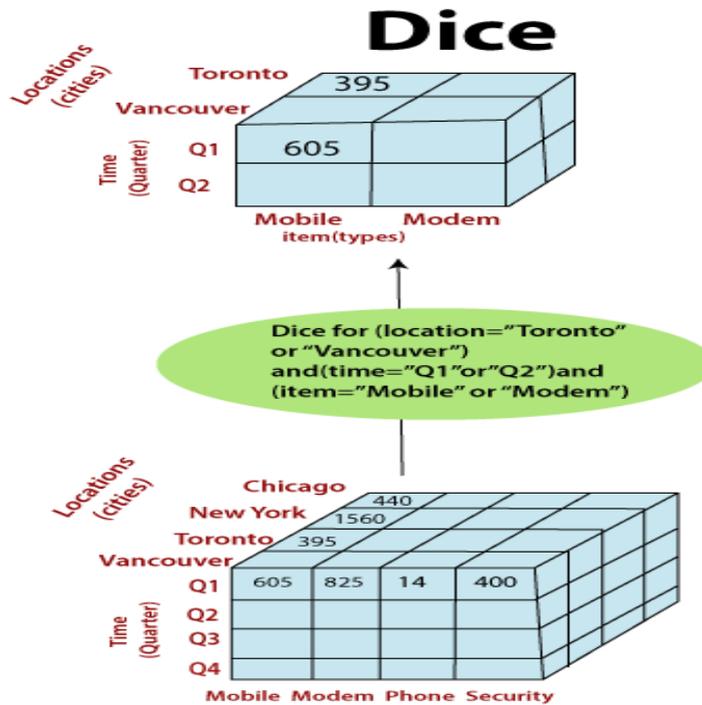


Figure 1. 16: Exemple du Dicing [11]

II.5.2. Drill-down & Roll-up

Ces méthodes, appelées aussi « forage vers le bas/vers le haut », sont les méthodes les plus répandues pour une navigation dans un entrepôt de données. Elles consistent à représenter les données du cube à un niveau de granularité inférieur.

Ces opérations ne sont pas aussi faciles à implémenter car basées sur la notion d'une bonne hiérarchisation des attributs d'une dimension et la différenciation entre tous les niveaux de hiérarchie disponibles dans les différentes dimensions.

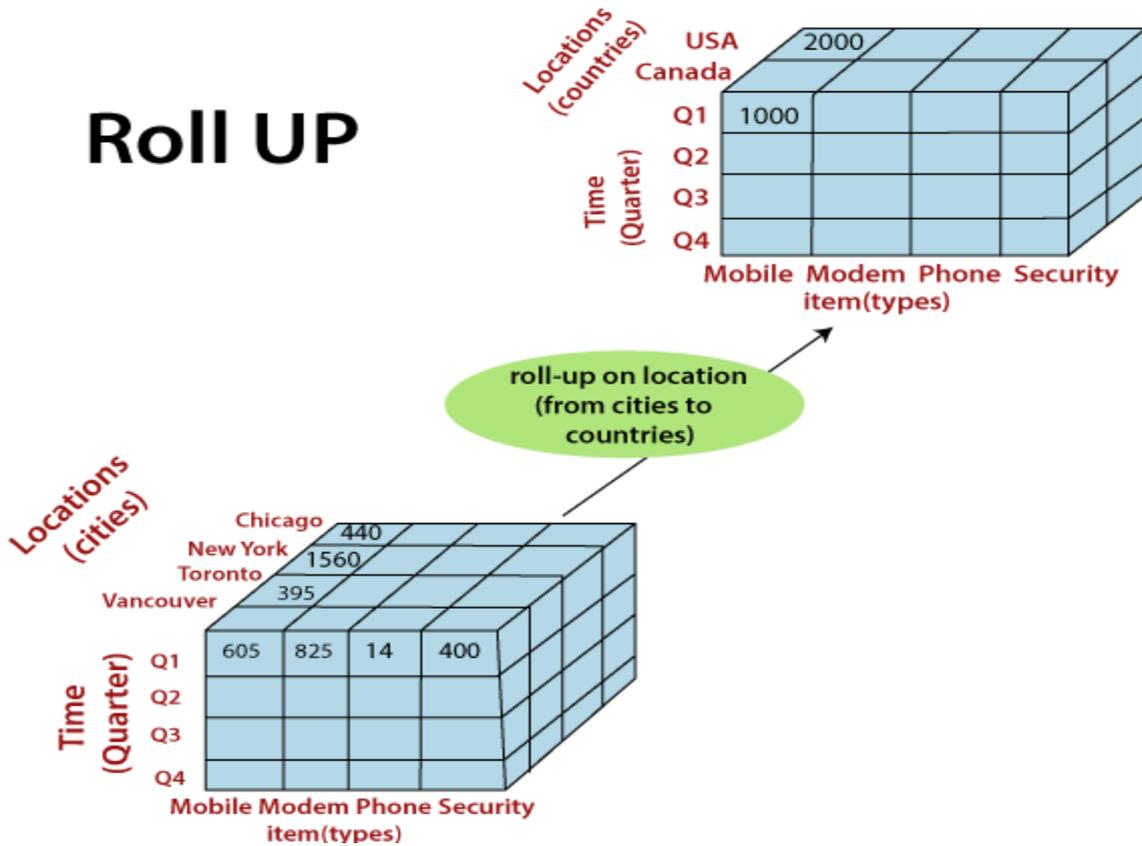


Figure 1. 17: Exemple de Roll-Up [11]

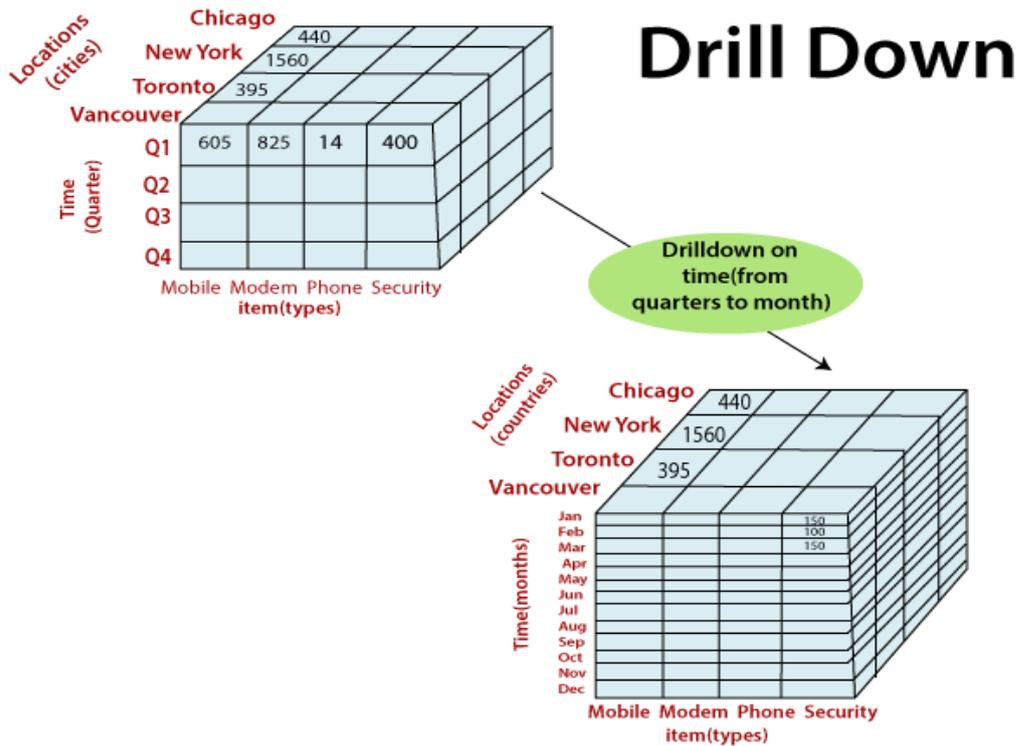


Figure 1. 18: Exemple de Drill-Down [11]

II.6. Reporting

Le terme "Reporting" désigne une famille d'outils de Business intelligence destinés à assurer la réalisation, la publication et la diffusion de rapports d'activité selon un format prédéterminé.

Ils sont essentiellement destinés à faciliter la communication de résultats chiffrés ou d'un suivi d'avancement.

Principe :

L'outil de reporting assure l'interrogation des bases de données selon les requêtes SQL préparées lors de l'élaboration du modèle. Le rapport d'activité peut ensuite être publié sur l'Intranet, périodiquement en automatique ou ponctuellement à la demande.

L'outil offre bien entendu des fonctions spécifiques pour l'élaboration du modèle du rapport, des modules de calcul et de présentation (graphiques) afin de concevoir des comptes rendus particulièrement seyants et pertinents.

II.7. Tableau de bord en BI

Le tableau de bord est au coeur du système de Business Intelligence de l'entreprise. Le soin apporté à son élaboration conditionne la réussite globale du projet décisionnel d'entreprise.

Par définition, un tableau de bord sert à piloter une démarche de progrès. Il s'agit en effet de prendre les bonnes décisions pour atteindre l'horizon fixé. Il semble donc évident que la toute première phase du projet soit nécessairement de fixer et de formaliser ledit horizon.

II.8. Data mining

DM est le processus d'analyse des données provenant de différentes sources et de les résumer en informations pertinentes pouvant être utilisées pour augmenter les revenus et réduire les coûts. Son objectif principal est de trouver des corrélations ou des modèles parmi des dizaines de champs dans de grandes bases de données.

Le logiciel de DM est l'un des nombreux outils analytiques pour lire les données, permettant aux utilisateurs de visualiser les données sous différents angles, de les catégoriser et de résumer les relations identifiées. Le but ultime de l'exploration de données est la prédiction et la découverte. Le processus recherche des modèles cohérents et des relations systématiques entre les variables, puis valide les résultats en appliquant les modèles à de nouveaux sous-ensembles de données.

III. Démarche de Construction d'un Data Warehouse

III.1. Définir les besoins commerciaux de l'entreprise

Dans cette étape on doit communiquer les décideurs et les différents service ou processus métiers de l'entreprise afin de définir des objectifs et des besoins globales et spécifiques de l'entreprise à court, moyen et long terme, ces besoins sont appelés « mesures ». Ces mesures sont les indicateurs de performance clés, une mesure numérique des activités de l'entreprise, telles que les unités vendues, le bénéfice brut, le bénéfice net...etc. [12]

III.2. Identifier les sources de données

Un entrepôt de données extrait généralement des données de diverses sources, nous devons identifier toutes les sources de données nécessaires qui contribueront à fournir les données dont nous avons besoin pour atteindre les objectifs et en choisir les points / éléments de données nécessaires. Les sources de données peuvent être de tout type - autres bases de données (SQL / NoSQL), applications, médias sociaux, enquêtes, capteurs / IoT, fichiers Excel / CSV, formulaires opérationnels, etc. Si les données sont nécessaires, elles doivent être introduites dans l'entrepôt. [12]

III.3. Les approches de modélisation et construction de data warehouse

Pour la construction de Data Warehouse il existe des approches qui ont été proposé par les pères de Data Warehouse qui sont :

III.3.1. Approche ascendante

D'après Kimball [6] cette approche consiste à créer les data marts pour fournir des rapports et des capacités analytiques pour spécifiques départements ou les processus métiers de l'entreprise après avec ces data marts peuvent ensuite les intégrer grâce à des dimensions communes entre des faits de deux data marts différents pour créer un entrepôt de données complet.

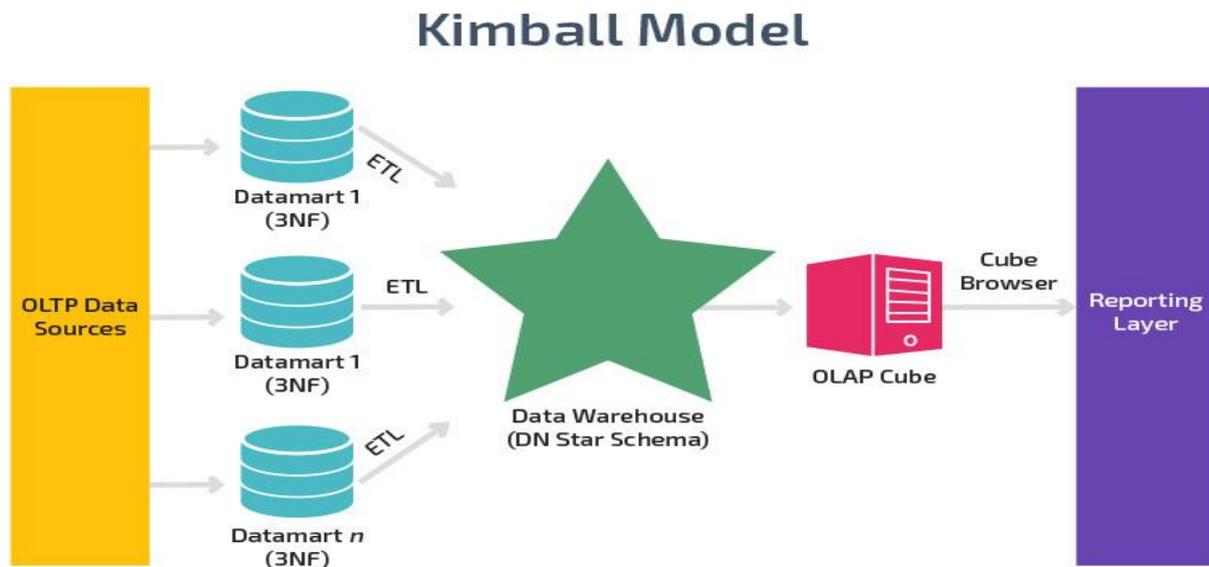


Figure 1. 19: Approche de Kimball [6]

Avantages	Inconvénients
<ul style="list-style-type: none"> ➤ Ce modèle contient des data marts cohérents et ces data marts peuvent être livrés rapidement. ➤ Les magasins de données sont créés en premier pour fournir une capacité de rapport. ➤ Il est plus facile d'étendre l'entrepôt de données car il peut facilement accueillir de nouvelles unités commerciales. Il s'agit simplement de créer de nouveaux data marts, puis de les intégrer à d'autres data marts. ➤ Cette approche prend moins de temps. La configuration initiale est très rapide. 	<ul style="list-style-type: none"> ➤ Le modèle étant orienté processus métier, au lieu de se concentrer sur l'entreprise dans son ensemble, cette approche ne peut pas gérer toutes les exigences de reporting BI. ➤ Le processus d'intégration de grandes quantités de données héritées dans l'entrepôt de données est complexe. ➤ La redondance est difficile à gérer.

Tableau 1. 3: Avantages et inconvénients d'approche ascendante

III.3.2. Approche descendante

D'après Inmon [6] cette approche consiste à savoir au début les besoins des décideurs en suite créer un data warehouse complet après des magasins de données sont créés contenant les données nécessaires à des processus métier spécifiques ou à des services spécifiques.

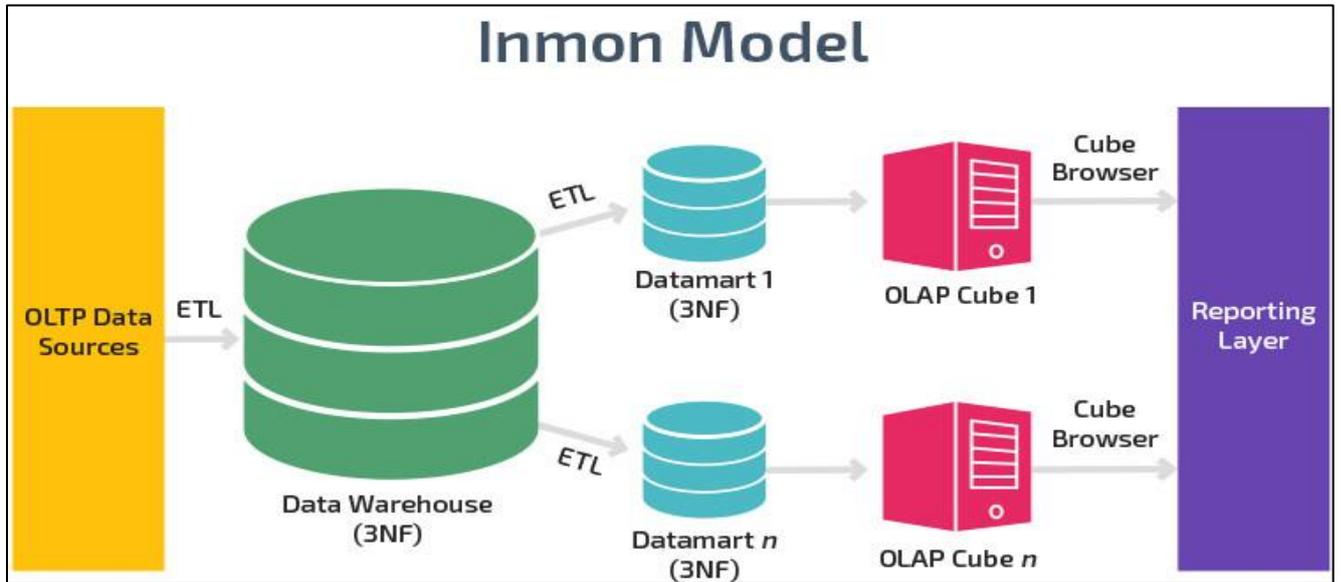


Figure 1. 20: Approche de Inmon [6]

Avantages	Inconvénients
<ul style="list-style-type: none"> ➤ L'entrepôt de données constitue une source de vérité unifiée pour l'ensemble de l'entreprise, où toutes les données sont intégrées. ➤ Cette approche a une très faible redondance des données. Il y a donc moins de risques. ➤ D'irrégularités dans la mise à jour des données, ce qui rend le processus ETL plus simple et moins susceptible d'échouer. ➤ Cela simplifie les processus métier, car le modèle logique représente des objets métier détaillés. ➤ Cette approche offre une plus grande flexibilité. Car il est plus facile de mettre à jour l'entrepôt de données en cas de changement des exigences commerciales ou des données source. 	<ul style="list-style-type: none"> ➤ Il s'agit d'un très grand projet et le coût de mise en œuvre du projet est important. ➤ Cela prend du temps et nécessite plus de temps pour la configuration initiale. ➤ Personnel hautement qualifié requis pour l'installation. ➤ Cette technique est inflexible à l'évolution des besoins ministériels. ➤ Le coût, le temps de conception et sa maintenance sont très élevés.

➤ Il peut gérer diverses exigences de reporting à l'échelle de l'entreprise.

Tableau 1. 4: Avantages et inconvénients d'approche descendante

III.3.3. Approche Hybride

Cette approche est conseillée par les professionnels du BI. Cette approche consiste à construire des schémas dimensionnels à partir des structures des données du système opérationnel, et les valider par rapport aux besoins analytiques. Cette approche cumule les avantages et quelques inconvénients des deux approches déjà citées, telles que la complexité des sources de données et la difficulté quant à la détermination des besoins analytiques.

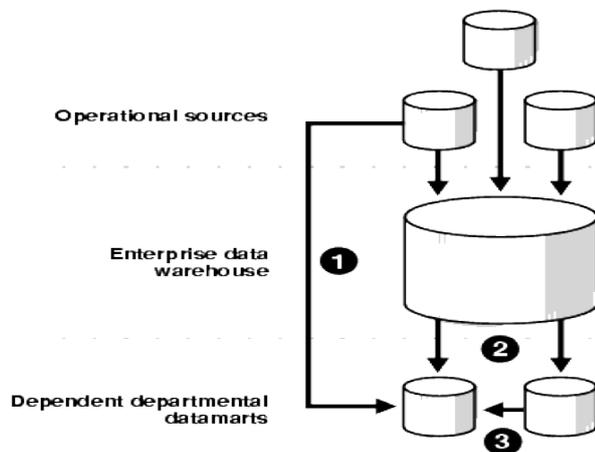


Figure 1. 21: Approche Hybrid [13]

Avantages	Inconvénients
<ul style="list-style-type: none"> ➤ Capacité à ingérer, traiter et analyser des données en streaming ➤ Permettre aux utilisateurs professionnels d'explorer, de découvrir et de libre-service ➤ Améliorer considérablement les performances des données intégrées ➤ Disponibilité plus rapide des données actuellement inaccessibles données ➤ Capacité de stocker de grands ensembles de données et des données semi-structurées 	<ul style="list-style-type: none"> ➤ Exige que les organisations appliquent l'utilisation standard des entités et des règles. ➤ Le remplissage d'un DW est perturbateur, nécessitant l'engagement de l'entreprise, le financement et la réécriture des applications. ➤ Peu d'outils de requête peuvent interroger dynamiquement des données atomiques et récapitulatives dans DW et DM.

- | | |
|---|--|
| ➤ Fournir une passerelle de source unique pour l'accès à toutes les données | |
|---|--|

Tableau 1. 5: Avantages et inconvénients d'approche hybrid

Les méthodes Inmon et Kimball peuvent être utilisées pour concevoir avec succès des entrepôts de données. En fait, plusieurs entreprises utilisent un mélange de ces deux approches (appelé le modèle hybride).

Dans le modèle hybride, la méthode Inmon est utilisée pour former un entrepôt de données intégré. Tandis que l'approche de Kimball est suivie pour développer des dépôts de données en utilisant le schéma en étoile.

Il n'est pas possible de déterminer quelle approche est la meilleure car les deux méthodes présentent des avantages et des inconvénients et fonctionnent bien dans des situations différentes. Le concepteur d'entrepôt de données doit choisir une méthode en fonction des divers facteurs abordés dans cet article.

Enfin, pour que toute méthode soit efficace, elle doit être bien pensée, explorée en profondeur et développée de manière à satisfaire les besoins de votre entreprise en matière de reporting en matière de renseignements commerciaux.

A la fin de cette étape nous devons avoir un modèle de modélisation des données et le schéma de la structure du système décisionnel dépendamment de l'approche adapté.

III.4. Identifier l'environnement de développement

Cette identification consiste à déterminer l'environnement physique de l'entrepôt de données. Au minimum, il doit y avoir des serveurs physiques d'application et de base de données distincts ainsi que des processus ETL / ELT, OLAP, cube et de reporting distincts configurés pour le développement, les tests et la production. La création d'environnements physiques distincts garantit que tous les changements peuvent être testés avant de les déplacer vers la production, le développement et les tests peuvent se produire sans arrêter l'environnement de production, et si l'intégrité des données devient suspecte, le personnel informatique peut enquêter sur le problème sans impact négatif sur l'environnement de production. [14]

III.5. Alimentation de l'entrepôt de données

A ce stade nous avons identifié les sources de données et les éléments de données d'une part et la base de données de l'entrepôt de l'autre. Comment pouvons-nous obtenir les données dans la base de données pour analyse. Le processus de cette opération est appelé Extract-Transform-Load (ETL) [18].

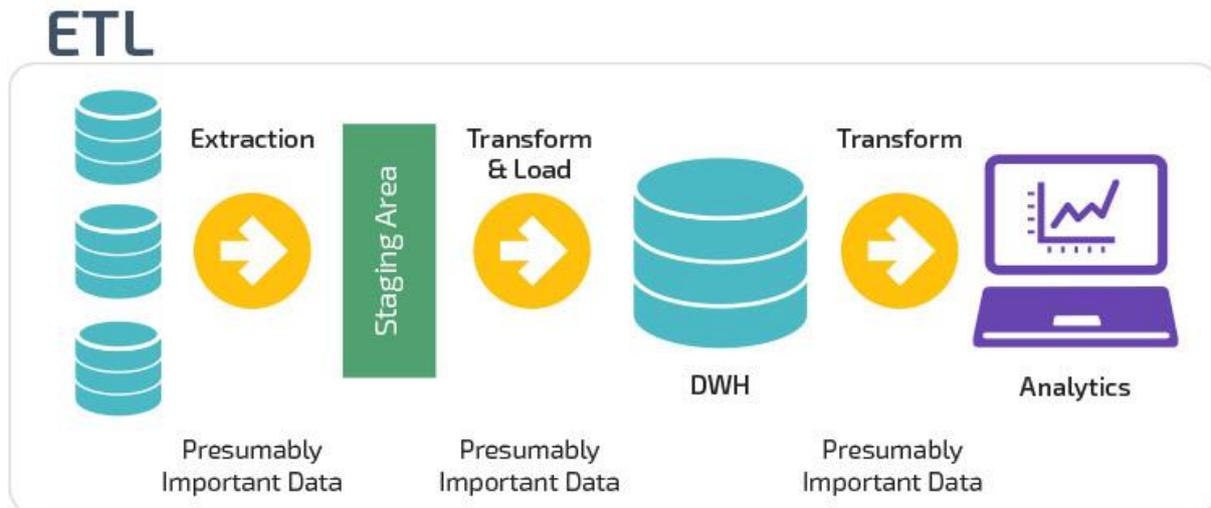


Figure 1. 22: Processus de l'ETL [6]

Le processus de l'alimentation peut se faire de différentes manières. Le choix de la politique de chargement dépend des sources : disponibilité et accessibilité.

Ces politiques sont :

- **Push** : Dans cette méthode, la logique de chargement est dans le système de production. Il " pousse " les données vers la zone de préparation quand il en a l'occasion. L'inconvénient est que si le système est occupé, il ne poussera jamais les données.
- **Pull** : Contrairement de la méthode précédente, le Pull " tire " les données de la source vers la zone de préparation. L'inconvénient de cette méthode est qu'elle peut surcharger le système s'il est en cours d'utilisation.
- **Push-pull** : C'est la combinaison des deux méthodes. La source prépare les données à envoyer et indique à la zone de préparation qu'elle est prête. La zone de préparation va alors récupérer les données.
- ❖ **Extraction** : il s'agit d'extraire les données des données de différents systèmes sources de l'entreprise ou récupérer des données à partir de sources externes par des moyens tels que l'araignée Web ou le grattage d'écran, chaque source peut également utiliser une organisation et un format de données différents cela représente l'aspect le plus important d'ETL, car l'extraction des données ouvre correctement la voie au succès des processus ultérieurs. Ces données extraites sont chargées dans une base de données temporaire (Staging Area).

Lors du chargement des données, il faut extraire les nouvelles données ainsi que les changements intervenus sur ces données. Pour cela, il existe trois stratégies de capture changement :

- **Trigger ou déclencheur** : si les systèmes opérationnels d'une entreprise prennent en charge les déclencheurs de base de données, ils peuvent alors être utilisés pour automatiser l'extraction des données. Les déclencheurs vous permettent de définir des opérations qui sont exécutées automatiquement lorsque certains événements se produisent. En règle générale, les événements déclencheurs se rapportent à des changements dans la base de données du système source, qui conduisent à une extraction des données modifiées dans le DWH.
- **Fichiers journaux** : si le système opérationnel ne prend pas en charge le trigger ou déclencheur, le niveau d'acquisition des données d'un DWH peut contenir des programmes capables d'évaluer les fichiers journaux (ou fichiers log) des systèmes sources et d'extraire les opérations enregistrées dans ces derniers.
- **Programme de monitoring** : si pour l'extraction il n'est pas possible de recourir au déclencheur ou aux fichiers journaux, les programmes de monitoring sont alors généralement utilisés. Ces derniers extraient les changements dans l'ensemble de données d'un système opérationnel à l'aide d'algorithmes qui créent des instantanés des données à surveiller (snapshots) à intervalles réguliers et les synchronisent avec les précédents.

Si aucune des techniques décrites ci-dessus n'est prise en charge, car l'accès aux données dans le système d'exploitation n'est pas possible, ces modifications doivent être enregistrées de manière indépendante et les modifications correspondantes communiquées à l'entrepôt de données.

- ❖ **Transformation** : les données extraites des sources sont brutes et non utilisables dans leur forme d'origine, Il doit donc les nettoyer, les filtrer et les transformer en un format (schéma de donnée) d'entrepôt de données. En fait, c'est l'étape clé où le processus ETL ajoute de la valeur et modifie les données de sorte d'avoir des données d'analyses, pertinentes et cohérentes pour les charger dans le DW. En effet afin avoir ces données il faut appliquer des procédures qui sont :
 - Filtrage - Sélectionnez uniquement certaines colonnes à charger

- Utilisation de règles et de tables de recherche pour la normalisation des données.
 - Conversion de jeux de caractères et gestion de l'encodage
 - Conversion d'unités de mesure comme la conversion date-heure, les conversions de devises, les conversions numériques, etc.
 - Vérification de la validation du seuil de données. Par exemple, l'âge ne peut pas dépasser deux chiffres.
 - Validation du flux de données de la zone de transfert vers les tables intermédiaires.
 - Les champs obligatoires ne doivent pas être laissés en blanc.
 - Nettoyage (par exemple, mappage de NULL à 0 ou du sexe masculin à "M" et de la femme à "F", etc.)
 - Fractionnez une colonne en plusieurs et fusionnez plusieurs colonnes en une seule colonne.
 - Utilisation d'une validation de données complexe (par exemple, si les deux premières colonnes d'une ligne sont vides, elle rejette automatiquement la ligne du traitement).
- ❖ **Loading (Chargement)** : Le chargement des données dans la base de données cible qui est le datawarehouse ou data marts est la dernière étape du processus ETL. Dans un entrepôt de données typique, un énorme volume de données doit être chargé dans une période relativement courte (nuits). Par conséquent, le processus de chargement doit être optimisé pour les performances.

Il existe 3 types de chargements :

- **Charge initiale** : remplir toutes les tables de l'entrepôt de données.
- **Charge incrémentielle** : appliquer périodiquement les modifications en cours, au besoin.
- **Actualisation complète** : effacement du contenu d'une ou plusieurs tables et rechargement avec de nouvelles données.

A la fin, on doit automatiser ces processus c-à-dire il faut avoir un calendrier de lancement d'ETL pour avoir des données récente et pertinentes.

III.6. Développement frontal

À ce stade, les exigences commerciales ont été saisies, l'environnement physique est terminé, le modèle de données a été décidé et le processus ETL a été documenté. L'étape

suyvante consiste à déterminer comment les utilisateurs accéderont à l'entrepôt de données. Le développement frontal est la manière dont les utilisateurs accéderont aux données pour l'analyse. Il existe de nombreuses options disponibles, y compris la construction de votre front-end. Ces options-là ont été déjà motionné précédemment (outils d'analyse). [12]

III.7. Tester

Une fois que le système d'entrepôt de données a été développé en fonction des besoins de l'entreprise, l'étape suivante consiste à le tester. Le test, ou assurance qualité, est une étape à ne pas ignorer car elle permettra à l'équipe de l'entrepôt de données d'exposer et de résoudre les problèmes avant le déploiement initial. Le fait de ne pas terminer la phase de test pourrait entraîner des retards de mise en œuvre ou la fin du projet d'entrepôt de données. [14]

III.8. Documentations

C'est plus opérationnel que technique. Un entrepôt de données performant doit être un élément fonctionnel des opérations d'une organisation et évoluer au fur et à mesure que l'entreprise et les sources de données évoluent. Par conséquent, une bonne documentation sur la façon dont les choses ont été mises en place, les politiques et les conventions pour un développement ultérieur est essentielle pour assurer la continuité et la maintenance facile. [14]

III.9. Évolution des exigences

Du point de vue du processus de livraison, les exigences sont toujours modifiables. Ils ne sont pas statiques. Le processus de livraison doit prendre en charge cela et permettre à ces changements de se refléter dans le système.

Ce problème est résolu en concevant l'entrepôt de données autour de l'utilisation des données dans les processus métier, par opposition aux exigences de données des requêtes existantes.

L'architecture est conçue pour changer et croître pour répondre aux besoins de l'entreprise, le processus fonctionne comme un processus de développement de pseudo-applications, où les nouvelles exigences sont continuellement introduites dans les activités de développement et les livrables partiels sont produits. Ces livrables partiels sont renvoyés aux utilisateurs, puis retravaillés en veillant à ce que le système global soit continuellement mis à jour pour répondre aux besoins de l'entreprise.

CHAPITRE 2 : PRESENTATION DE SAE/EXACT ET L'ETUDE D'EXISTENCE

I. Pr  s  tation SAE/EXACT

La Soci  t   Alg  rienne d'Expertise et de Contr  le Technique Automobiles par abr  viation SAE/EXACT a   t   cr  e le 1er F  vrier 1998 sous la forme juridique de EURL, dot  e d'un capital social de 85,5 MDA divis   en 1710 actions d  tenues par le propri  taire unique. Elle a 28 centres d'expertises et 4 antennes au niveau national.

II. D  finition de l'expertise

L'expertise est une science, un art, et une technique consistant en l'ex  cution d'une mission confi  e dans le but de fournir des renseignements ou informations qu'il est impossible pour l'assureur d'avoir directement et qui font d  faut pour r  soudre un pr  judice mat  riel caus      la suite d'un sinistre, l'expertise est le seul moyen    m  me de permettre    l'assurance de proc  der    l'indemnisation exacte des assur  s. L'expert joue un r  le d  terminant dans la prestation de service et se r  percute sur la relation entre l'assureur et l'assur  , l'expert intervient en amont lors de visite technique du risque avant la souscription des garanties dommages, comme il intervient en aval pour expertiser Les sinistres, l'expert intervient   galement durant la validit   du contrat pour la v  rification de la tenue du risque assur  .

III. Types d'expertise fournis par SAE/EXACT

III.1. Expertise automobile

L'expertise automobile consiste    faire analyser le dommage de v  hicule apr  s une identification technique de v  hicule qui se base sur ses caract  ristiques (marque, mod  le, type, ...), en suite une identification administrative afin de v  rifier les documents de v  hicule. Il existe 4 sortes d'expertise automobile :

➤ **L'expertise simple :**

La proc  dure est r  alis  e g  n  ralement au niveau des centres d'expertises ou au niveau des antennes rattach  es administrativement aux centres d'expertises ou hors de centres o   l'expert sera amen        tablir un proc  s-verbal.

➤ **L'expertise contradictoire :**

Dans ce cas l'expert doit dresser son proc  s-verbal et adresser ensuite une invitation    son confr  re repr  sentant la compagnie adverse dans laquelle il doit indiquer les renseignements du v  hicule et le lieu exact o   celui-ci sera visit   conform  ment    la convention interentreprises.

➤ **La contre-expertise :**

En cas l'assuré n'est pas satisfait au remboursement il fait un recours pour réétudier le dossier par un autre expert.

➤ **L'expertise judiciaire :**

Est une mesure d'information technique confiée à un homme de l'art par une juridiction, il doit donner un avis technique sur les points décrits dans la mission, il n'a pas à décrire les circonstances de l'accident Alors que pour l'estimation des dommages matériels engendrés par les accidents de la circulation

➤ **La tierce expertise ou l'arbitrage :**

En cas de désaccord par l'une ou l'autre partie, un expert doit être désigné pour arbitrer.

III.2. Expertise Bâtiment et Risques Divers

➤ **Visite de risques :**

Dans le cadre des visites de risques, la mission confiée à l'expert consiste à informer son mandant, l'assureur sur l'ensemble des risques que peut encourir son potentiel client afin de lui permettre de prendre en charge ce risque ou de le récuser.

➤ **Vérification de risques :**

La vérification de risques intervient durant la validité du contrat d'assurance et consiste à contrôler l'état général du risque (tenue du risque), à vérifier si le risque a subi des modifications par rapport à la visite initiale, prise en compte dans l'élaboration du contrat d'assurance (aggravations, changements dans sa composition etc..), formuler des recommandations pour une meilleure gestion du risque assuré.

➤ **Expertise de sinistres :**

Dans ce cadre, il s'agit à la fois de vérifier l'exactitude des déclarations et allégations de l'assuré et la véracité des faits, de déterminer les causes et circonstances du sinistre connues ou présumées de rechercher l'origine exacte des sinistres, d'estimer les dommages provoqués, et d'identifier les cas de négligence carences ou cas de fraudes ou fausses déclarations.

➤ **Evaluation de patrimoine :**

Cette mission consiste à procéder à l'évaluation du patrimoine du client d'un assureur. L'expert procédera sur la base des documents graphiques (plans), administratifs (titres de propriété) et autres factures d'acquisition des mobiliers et équipements de toutes natures à l'évaluation patrimoine dans ses différents composants (immobilier, mobilier, équipements, engins et autres). L'expert doit tenir compte des valeurs de remplacement ou de reconstitution.

➤ **Actualisation des capitaux :**

Il s'agit, dans ce cadre, de mettre à jour les valeurs assurées. L'expert, pour la réalisation de cette mission, doit tenir compte d'un ensemble de paramètres.

Entre autres, nous citons, à titre d'illustration, le facteur vétusté, l'entretien, les opérations de rénovation réalisées (pour les équipements et engins) ainsi que l'ensemble des éléments constitutif des prix unitaires à appliquer.

III.3. Evaluations

➤ **Expertise immobilière :**

L'expert tiens compte de l'existant physique et des valeurs moyennes usitées localement, dans son évaluation ainsi que des textes législatifs régissant les différents types d'évaluation.

➤ **Suivi financement :**

Cette mission est exclusivement commandée par des organismes financiers et consiste à l'expert de suivre l'évolution des travaux financés et d'en informer régulièrement le mandant.

III.4. Expertise Médicale

L'expertise médicale constitue un acte médical particulier par lequel le rôle du médecin n'est ni de soigner ni de tenter de guérir une maladie, mais seulement de faire des constatations médicales dans le but d'établir une preuve du dommage corporel, d'évaluer les séquelles fonctionnelles des victimes d'accident.

III.5. Expertise Vétérinaire

Le domaine d'expertise vétérinaire est un domaine spécialisé nécessitant la mobilisation rapide du vétérinaire c'est dans ce sens que l'expert intervient dans les domaines suivants : Avicole, Apicole, Bovin, Ovin et caprin, Chevalin, Aviaire.

III.6 Expertise Agricole

L'expertise agricole concerne :

- **Les productions animales :** évaluation de cheptels, élevage spécialisé, pisciculture, etc
- **Productions végétales :** évaluation de récoltes, viticulture, arboriculture, horticulture, culture sous serres, dommages aux cultures et aux récoltes, accidents culturels, avarie de denrées agricoles, produits phytosanitaires, etc.

IV. Organisation de la société

La société est organisée en 5 directions qui gèrent un ensemble de centres et d'antennes d'expertise sur le territoire nationale

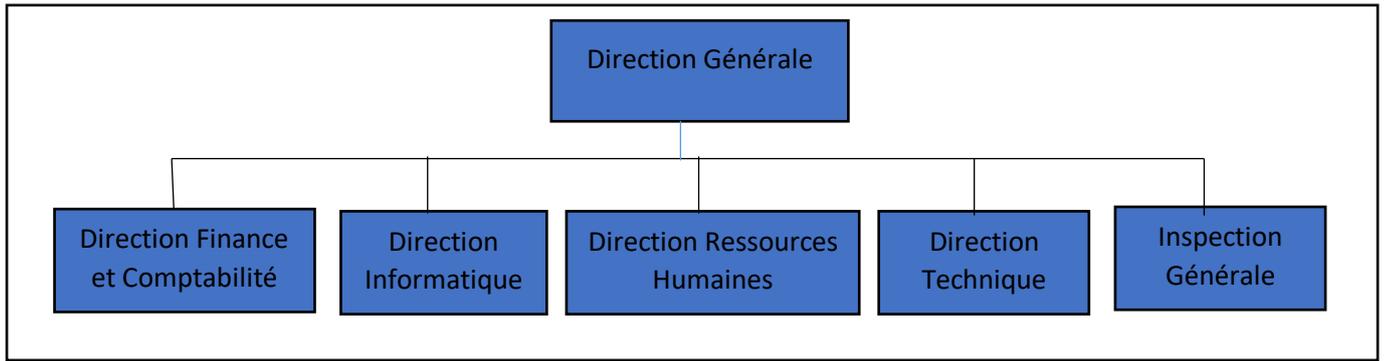


Figure 2. 1: Hiérarchie de l'entreprise

- **Direction générale** : gère et pilote la société, prise des décisions.
- **Direction Finance est comptabilité** : prise en charge la gestion d'argent, des dépenses
- **Direction Informatique** : prise en charge le développement informatique, la gestion de système d'information.
- **Direction Ressources Humaines** : prise en charge d'aborder la relation de l'organisation avec son collaborateur depuis son recrutement jusqu'à son départ de l'entreprise, organiser des formations, gestion de la paie, ...
- **Direction Technique** : prise en charge la gestion d'expertise de différent types.

V. Les chiffres de la société

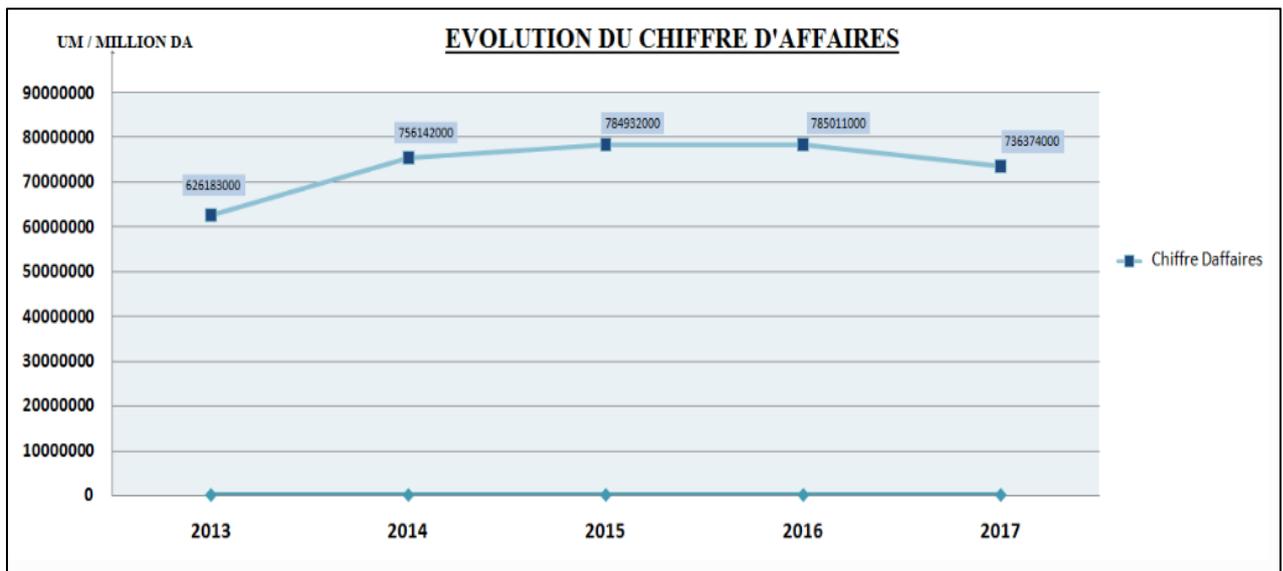


Figure 2. 2: Graphe d'évolutions de chiffre d'affaire

VI. Etude de l'existence

Ce projet vise à remplir un manque décisionnel au sein de l'entreprise car il existe soit des outils d'aide à la décision semi-automatique dans une partie du progiciel métier qui génère des tableaux de bord et qui manque d'axes d'analyse, mais aussi il y'a un manque d'outils de reporting efficace en mesure de fournir des informations adéquates en temps voulu.

De cela nous allons essayer à travers ce chapitre de donner une analyse de l'existant décisionnel au sein de l'entreprise dans le cadre directionnel (DG, DFC) mais aussi au niveau métier (direction de centre).

VII. Niveau direction général et direction de la finance et comptabilité (DFC)

L'entreprise ne possède pas d'outil unique qui a la fonctionnalité d'avoir une information qui représente tous les centres au niveau national dans une même mesure, donc le système décisionnel au sein de la DG est semi-automatique car ils ont l'accès à chaque session progicielle de chaque centre.

Ainsi la consolidation des données se fait en calculant toujours de façon manuelle une mesure qui représente tout l'entreprise en additionnant sa valeur de chaque centre.

IIX. Niveau direction de centre

Le directeur de centre et le responsable de la prise de décision au sein d'un centre, après avoir effectué une analyse de l'existant décisionnel au sein du centre via une interview avec le directeur de centre nous avons pu constater que l'outil d'aide à la décision à ce niveau est inclus dans le progiciel métier, de plus il n'est pas assez riche en termes de mesure à analyser car seul le chiffre d'affaire et le nombre d'actes sont prise en charge par cette option décisionnelle du progiciel.

La richesse d'axes d'analyse est aussi mise en question à ce niveau car seule la notion de temps est prise en charge, c'est-à-dire que le chiffre d'affaire ainsi que le nombre d'acte est analysé soit par année, par mois ou par décade (10 jours) seulement.

IX. Conclusion

Dans cette étude nous avons pu détecter les mesures et les axes d'analyse existante, ainsi la manier de générer les rapports et la consolidation des données au sein de l'entreprise. Cette étape nous a permis de connaître les insuffisances du système actuel en soulignant ses points faibles et de nous donner la procédure à suivre dans l'étape d'étude de besoin, elle est considérée comme le point de départ du projet.

CHAPITRE 3 : ETUDE DES BESOINS ET CONCEPTION DE LA ZONE D'ENTROPOSAGE DES DONNEES

I. Etude des besoins

Le data warehouse est une base de données immense qui répond aux besoins des utilisateurs, pour le créer il faut y avoir un schéma de données, ce dernier doit être obtenu par une étude de besoins qui se base sur les interviews des utilisateurs et la source de données pour alimenter le DW, dans ce chapitre nous allons expliquer comment extraire et présenter les besoins d'analyses.

II. Démarche d'extraction les besoins

Afin d'avoir une étude de qualité nous avons adopté une approche hybride qui combine entre l'approche ascendante et l'approche descendante pour la création de schéma de DW, pour cela nous avons suivi les étapes suivantes :

a. Étude préliminaire des systèmes sources :

Nous avons fait un tour sur l'entreprise et le centre d'expertise pour comprendre le fonctionnement de l'entreprise, le métier d'expertise et détecter les postes qui doivent utiliser ce projet, à la fin de la tour un des membres de la DI nous a présenté et expliqué les systèmes opérationnels utilisés au sein de l'entreprise.

b. Détection des postes susceptibles d'être porteurs d'informations utiles :

Après avoir une vue sur l'entreprise, nous avons ciblé des postes qui peuvent nous aider à avoir des informations utiles pour notre étude de besoins qui sont :

- Direction générale
- Direction de finance et comptable
- Direction informatique
- Direction technique
- Directeur de centre

Vu que la dispersion géographique des centres, il a été obligé d'avoir interviewé qu'un seul directeur de centre et prendre ses besoins comme un modèle pour les autres centres.

c. Planification, préparation et conduite des interviews :

Avant de détailler cette étape, il est nécessaire de justifier notre choix de l'entretien (interviews) comme méthode de collecte des besoins. Bien qu'il existe d'autres méthodes d'identification des besoins, les entretiens s'imposent comme étant la valeur sûre dans un tel projet. En effet, cette méthode a l'avantage d'être, plus ou moins, facilement planifiable

et d'ouvrir le dialogue avec les interviewés, qui sont pour la plupart des décideurs et des analystes.

Dans cette étape nous avons fait plan pour visiter les postes ciblés précédemment pour extraire leurs besoins d'analyses, avant les questionnés nous avons présenté notre projet et son but pour qu'ils puissent nous bien comprendre et répondre à nôtres questions clairement afin obtenir des informations pertinentes sur leurs besoins.

Cette étape nous a pris une semaine car les postes ont en charge de travail.

d. Utilisation d'autres moyens pour la détection des besoins :

Les entretiens représentent une source importante d'informations et aident grandement à l'identification des besoins des utilisateurs, leur utilisation exclusive n'est pas conseillée dans la construction d'un entrepôt de données. Cela tient principalement au fait que les utilisateurs ne peuvent, même avec la meilleure volonté du monde, exprimer tous leurs besoins.

De cela, il est fait appel à l'étude des rapports déjà demandés et des données disponibles a même de fournir des informations exploitables.

L'étude des rapports offre un certain apport à notre démarche d'études des besoins, dans la mesure où les utilisateurs peuvent ne pas mentionner des besoins qui leur paraissent évidents ou qui ne leur viennent pas à l'esprit durant nos interviews, ces derniers peuvent être, en effet, influencés par nos questions.

e. Rédaction et validation de besoins :

Cette dernière étape est une conclusion des étapes précédentes qui décrit les besoins analytiques et les utilisateurs concernés.

Besoins	Utilisateurs
Le chiffre d'affaire et le nombre d'acte selon l'expert, centre et le temps (décade, mois, année)	Direction générale Direction de finance et comptable Directeur de centre Direction technique
Le montant et le nombre de PV réglés selon le client, le centre et le temps (décade, mois, année)	Direction générale Direction de finance et comptable Directeur de centre
Le montant et le nombre de PV impayés selon le client, le centre et le temps (durée, date d'acte)	Direction générale Direction de finance et comptable Directeur de centre

Le nombre de PV annulés selon le client, le centre et le temps (décade, mois, année)	Direction générale Directeur de centre
--	---

Tableau 3. 1: Synthétisation des besoins recensés

III. Problèmes et obstacles rencontrés

Comme chaque travail lors de notre étude de besoins nous avons rencontré des problèmes et des obstacles qui nous ont retardés un peu le travail qui est :

- ✓ L'indisponibilité de personnes concernées par les entretiens.
- ✓ La rétention d'informations sous couvert de confidentialité.
- ✓ La compréhension de projet par les interviewés.
- ✓ La charge de travail qui nous incombe.
- ✓ L'emploi du temps chargé des interviewés.

IV. Conception

Dans le développement logiciel l'étape conception se déclenche après une étude des besoins qui schématise ou modélise les besoins détectés, pour cela nous avons choisi la modélisation dimensionnelle en étoile pour modéliser le Data Warehouse de cette entreprise vu les avantages qu'elle apporte.

V. Les dimensions (axes d'analyse)

Les dimensions sont des axes d'analyses où chaque clé primaire d'une dimension sera une clé étrangère dans la table de fait où la dimension intervient, ils ont pour objectifs de décrire les faits.

V.1. Dimension Temps

La dimension temps est entrepôt de données, car en pratique tout entrepôt de données est une série temporelle. Le temps est le plus souvent la première dimension dans le classement sous-jacent de la base de données » [Kimball, 2001].

La dimension temps se présente comme suit :

Dim_Temps		
<u>date_exp</u>	<u>DATE</u>	<u><pk></u>
mois	VARCHAR(6)	
annee	VARCHAR(4)	
decade	VARCHAR(8)	

Figure 3.1: Dimension Temps

Le tableau suivant décrit les attributs de la dimension :

Désignation	Détail
Date	Clé primaire format date
Mois	Représente le mois et l'année
Année	L'année
Décade	Concaténation du l'année et le mois et le numéro de la décade

Tableau 3.1: Tableau descriptif de la dimension « Temps »

V.2. Dimension Centre

Cette dimension représente un centre d'expertise parmi tous les centres du territoire national, la dimension Centre se présente comme suit :

Dim_Centre		
<u>id dim centre</u>	<u>VARCHAR(15)</u>	<u><pk></u>
wilaya	VARCHAR(25)	
nom_centre	VARCHAR(20)	
type_centre	VARCHAR(10)	

Figure 3.2: Dimension Centre

Le tableau suivant décrit les attributs de la dimension :

Désignation	Détail
Id_dim_centre	Clé artificielle de centre
Wilaya	La wilaya du centre
Nom_centre	Le nom complet attribué au centre
Type_centre	Décrit le type de centre (centre ou antenne)

Tableau 3.2: Tableau descriptif de la dimension « Centre »

Il est tout de même important d'avoir un axe d'analyse centre afin d'avoir une vue globale sur le rendement de tous les centre et arriver à faire une comparaison et même un classement des centres.

V.3. Dimension Client

Un client chez la SAE est une agence d'assurance qui appartient à une compagnie d'assurance (ex : saa1403 qui appartient à la saa), la dimension Client se présente comme suit :

Dim_Client		
<u>id_dim_client</u>	<u>VARCHAR(25)</u>	<u><pk></u>
nom_agence	VARCHAR(25)	
nom_compagnie	VARCHAR(20)	
wilaya	VARCHAR(20)	

Figure 3.3: Dimension Client

Le tableau suivant décrit les attributs de la dimension :

Désignation	Détail
Id_dim_client	Clé primaire de client
Nom_agence	Commence par le nom de la compagnie suivie d'un numéro de séquence
Nom_compagnie	Nom de la compagnie d'assurance
Wilaya	Nom de la wilaya de l'agence

Tableau 3.3: Tableau descriptif de la dimension « Client »

Le client s'impose comme un élément important dans l'analyse, car il intéresse fortement les analystes et les décideurs de l'entreprise, l'analyse du comportement du client permet à l'entreprise de mieux le satisfaire.

V.4. Dimension Expert

Cette dimension décrit l'information qui concerne un expert, la dimension Expert se présente comme suit :

Dim_Expert		
<u>id_dim_exp</u>	<u>VARCHAR(100)</u>	<u><pk></u>
mat_expert	VARCHAR(25)	
nom_expert	VARCHAR(50)	

Figure 3.4: Dimension Expert

Le tableau suivant décrit les attributs de la dimension :

Désignation	Détail
Id_dim_exp	Clé primaire d'expert
Nom_expert	Nom complet de l'expert
Mat_expert	Matricule de l'expert

Tableau 3.4: Tableau descriptif de la dimension « Expert »

L'expert est un employé décisif au sein de l'entreprise car il est impliqué directement dans l'activité principale (l'expertise), de cela l'analyse de l'efficacité de ce dernier est très importante pour les décideurs.

V.5. Dimension Type d'expertise

La SAE a plusieurs type d'expertise et elle base sur 3 types d'expertise essentielle : automobile, risque divers et évaluation car ces types existent dans tous les centre cela nécessite une dimension type-d 'expertise qui se présente comme suit :

Dim_Type_exp		
<u>id dim type exp</u>	<u>VARCHAR(10)</u>	<u><pk></u>
type_expertise	VARCHAR(25)	

Figure 3. 5: Dimension Type Expertise

Le tableau suivant décrit les attributs de la dimension :

Désignation	Détail
Id_dim_type_exp	Clé primaire de type d'expertise
Type_expertise	D'écrit le type d'expertise exercicer

Tableau 3.5: Tableau descriptif de la dimension « Décade »

Il est important d'analyser le type d 'expertise afin de savoir quelle branche est la plus efficace en terme de nombre d'acte ou de chiffre d'affaire et le décalage entre ces deux dernier et d'avoir des statistique exacte sur leur évolution.

VI. Modèles en étoile

Nous avons opté pour la modélisation dimensionnelle en étoile pour son principal avantage est qu'il permet d'optimiser la performance et la simplicité des requêtes effectuées sur de grands volumes de données.

VI.1. Le modèle en Etoile Expertise

La table de fait que nous allons présenter contient comme mesure le chiffre d'affaire ainsi que le nombre d'acte par temps, centre et expert, de cela les analystes et décideurs peuvent savoir en temps réel quelle centre d'expertise effectue le plus d'expertise ainsi de savoir celui qui a le meilleur chiffre d'affaire par exemple.

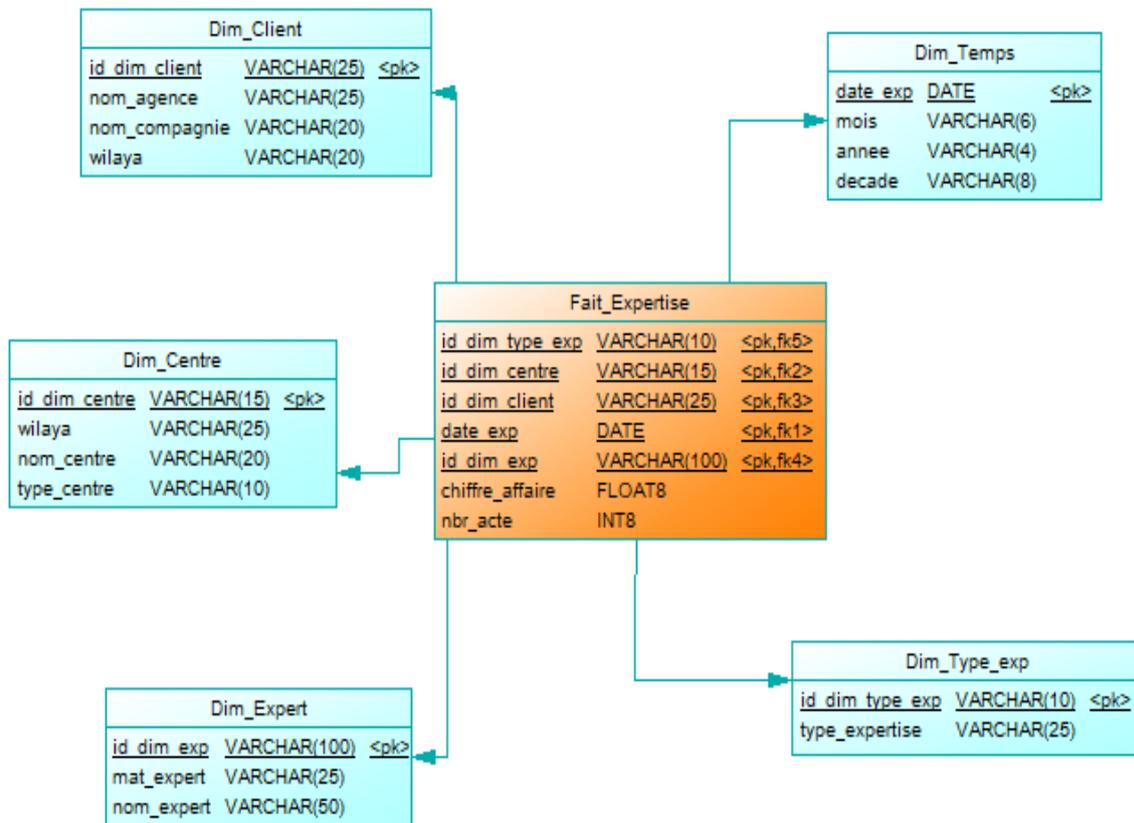


Figure 3. 6: Modèle en étoile Expertise

Désignation	Détail
Id_dim_type_exp	Clé étrangère de la dimension « type_expertise » et clé primaire de Fait_Expertise
Id_dim_centre	Clé étrangère la dimension « centre » et clé primaire de Fait_Expertise
Id_dim_client	Clé étrangère la dimension « client » et clé primaire de Fait_Expertise
Id_dim_exp	Clé étrangère la dimension « expert » et clé primaire de Fait_Expertise
Date_exp	Clé étrangère la dimension « temps » et clé primaire de Fait_Expertise
Chiffre_affaire	Somme des montants hors taxe des PV
Nbr_act	Nombre des PV fait

Tableau 3.6: Tableau descriptif de fait « Expertise »

VI.2. Modèle en étoile des impayés

Un PV peut être en trois statuts possibles (réglé, annuler ou impayé) où il est important d'analyser le montant et le nombre des PV impayés par rapport au temps, client et centre afin de savoir par exemple les clients qui ne payent pas, les clients qui payent en retard, pour cela nous représentons le modèle comme suit :

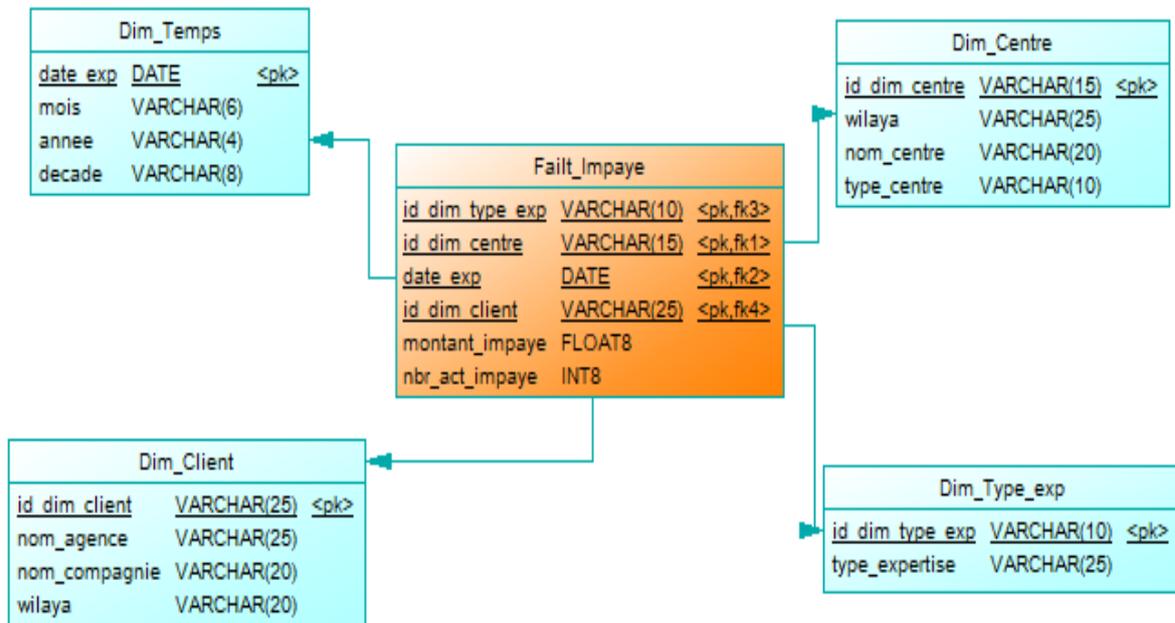


Figure 3.7 : Modèle en étoile Impayé

Désignation	Détail
Id_dim_type_exp	Clé étrangère de la dimension « type_expertise » et clé primaire de Fait_Expertise
Id_dim_centre	Clé étrangère la dimension « centre » et clé primaire de Fait_Expertise
Id_dim_client	Clé étrangère la dimension « client » et clé primaire de Fait_Expertise
Date_exp	Clé étrangère la dimension « temps » et clé primaire de Fait_Expertise
Montant_impayé	Somme des montants des PV non payés
Nbr_act_impayé	Nombre des PV fait et non payés

Tableau 3.7: Tableau descriptif de fait « Impayé »

VI.3. Modèle en étoile des règlements

Lors le client paye sont PV l'état de ce dernier devient « réglé », alors nous voulons analyser le nombre et le montant réglé des PV par rapport au temps, client et centre pour savoir

par exemple les clients fidèles, pour cela nous représentons le modèle comme suit :

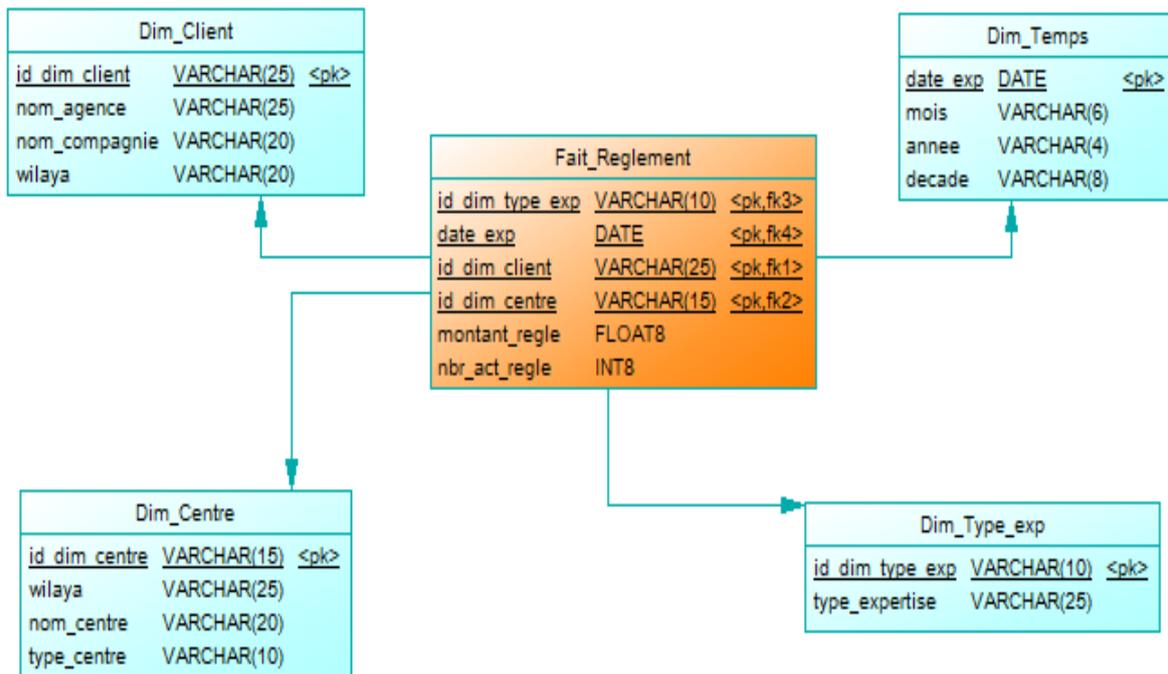


Figure 3.8: Modèle en étoile Règlement

Désignation	Détail
Id_dim_type_exp	Clé étrangère de la dimension « type_expertise » et clé primaire de Fait_Expertise
Id_dim_centre	Clé étrangère la dimension « centre » et clé primaire de Fait_Expertise
Id_dim_client	Clé étrangère la dimension « client » et clé primaire de Fait_Expertise
Date_exp	Clé étrangère la dimension « temps » et clé primaire de Fait_Expertise
Montant_regle	Somme des montants des PV payés
Nbr_act_regle	Nombre des PV fait payés

Tableau 3.8: Tableau descriptif de fait « Règlement »

VI.4. Modèle en étoile des annulés

Parfois, les experts annulent des PV donc l'état de PV devient « annulé », nous allons analyser le nombre des PV annulés par rapport au temps, client, centre et expert pour avoir par exemple une vision sur la croissance des annulations, les clients qui annulent beaucoup de PV, les causes d'annulations pour cela nous représentons le modèle comme suit :

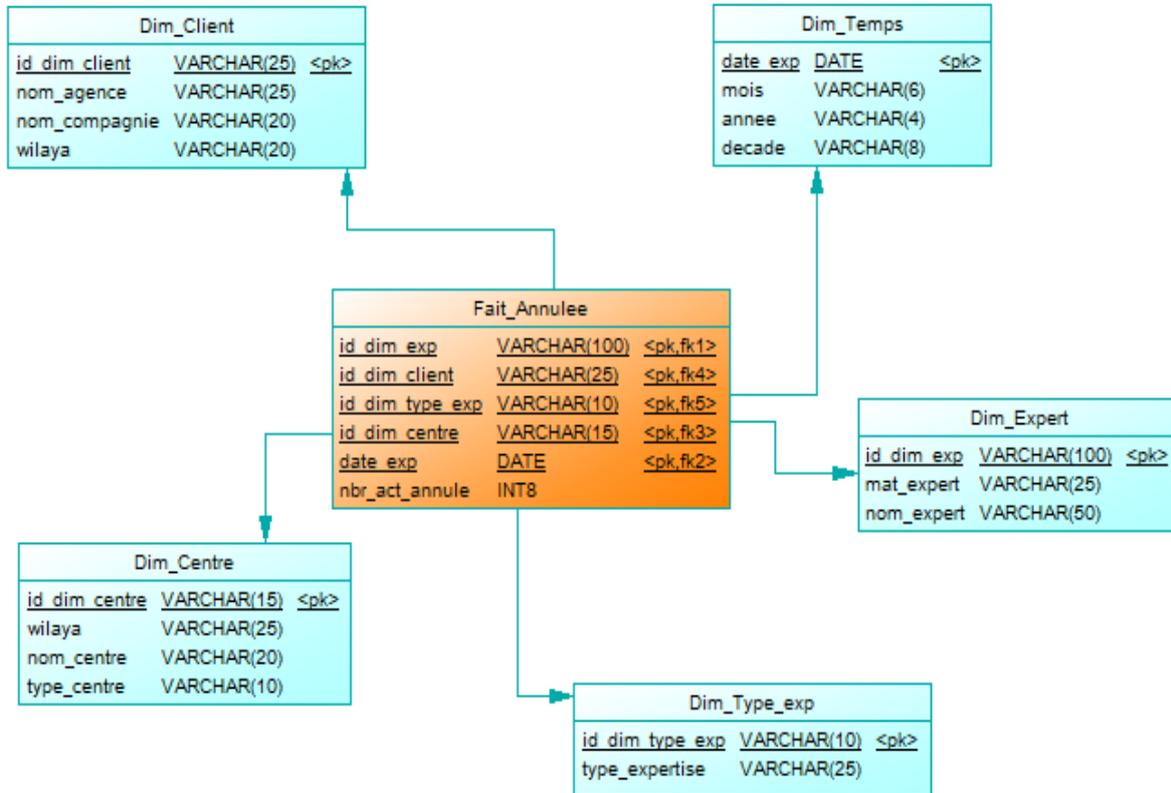


Figure 3.9: Modèle en étoile Annulé

Désignation	Détail
Id_dim_type_exp	Clé étrangère de la dimension « type_expertise » et clé primaire de Fait_Expertise
Id_dim_centre	Clé étrangère la dimension « centre » et clé primaire de Fait_Expertise
Id_dim_client	Clé étrangère la dimension « client » et clé primaire de Fait_Expertise
Id_dim_exp	Clé étrangère la dimension « expert » et clé primaire de Fait_Expertise
Date_exp	Clé étrangère la dimension « temps » et clé primaire de Fait_Expertise
Nbr_act_anuule	Nombre des PV annulé

Tableau 3.9: Tableau descriptif de fait « Annulé »

VII. Conclusion

L'étude des besoins est une étape nécessaire et très importante pour la construction de l'entrepôt de données car elle définit la vision de schéma et l'architecture de DW, alors avec cette étude on peut passer à la conception et l'implémentation de DW dans les chapitres suivants.

La conception de l'entrepôt de données est l'étape suivante de l'étude des besoins schématise les besoins d'analyse de l'entreprise et décrit les dimensions et les faits en utilisant le modèle en Etoile en expliquant les différentes tables et leurs attributs.

Cette étape de conception nous aide à la création de l'entrepôt de données savoir le son schéma.

CHAPITRE 4 : ALIMENTATION DU DATA WAREHOUSE

I. Introduction

L'alimentation d'un entrepôt de données est étape très importante dans un projet d'un système décisionnel, elle représente 80% de la charge de travail. Cette étape se déroule essentiellement par la migration des données de la base de production vers l'entrepôt de données qui se fait à l'aide des outils ETL en passant par les différentes phases de nettoyage et de transformations nécessaires.

II. Etude pré-alimentaire

Cette phase représente une phase préliminaire à l'ensemble du processus. Elle consiste de :

II.1. L'étude de la source des données

Les sources de données de notre entrepôt sont les 32 bases de données PostgresSql éparpillées sur l'ensemble du territoire national ont les mêmes schémas qui sont illustré sur cette figure :

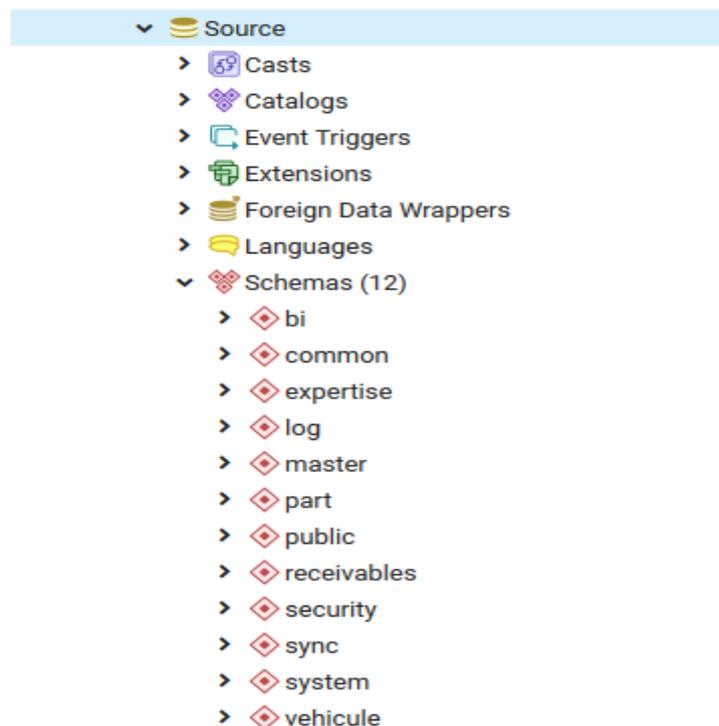


Figure 4.1: Liste des schémas d'une base de données

Chaque schéma contient des tables qui sauvegardent un sujet de données par exemple le schéma « expertise » contient 88 tables qui portent des données sur le métier expertise quel que soit le type d'expertise.

II.2. Détection l'emplacement des informations

Afin d'avoir des informations nécessaires et fiables nous avons analysé avec l'équipe informatique les sources qui contiennent un nombre important de tables, la redondance de données et immenses d'informations, ce travail nous a aidé à :

- Lister les données nécessaires pour les chargés
- Lister les emplacements des données qui situent sur 3 schémas sont « expertise » contient des données sur le métier expertise (numéro de pv, montant, ...), « master » contient des données communs entre tous les types d'expertise (données sur l'expert, le centre, agence,...) et « recevables » contient des données sur les paiements, les versements
- Choisir la source fiable des données pour le chargement

II.3. Définir la périodicité de chargement des données

Pour décider la périodicité il faut prendre en considération les critères suivants :

- La quantité des données à charger
- Le temps de non activité de système
- Chaque décade le système ouvert une nouvel session et ferme la précédente

Après une étude de ses critères nous avons choisi de configurer l'ETL avec une périodicité journalière pendant les heures non activité sur le système (de dix-huit heures à sept heures du matin).

III. Politique d'alimentation

Il est nécessaire de mettre en place une politique d'alimentation consistante à prendre en charge toutes les contraintes auxquelles on doit faire face, pour cela il existe différent politique, dans notre cas nous choisissons la méthode push-pull (pousse-tire) qui permet de gagner de temps, la vitesse de chargement de données et garante un chargement fiable.

IV. Processus de chargement

Pour décrire le processus général de chargement de Data Warehouse nous avons créé un diagramme d'activité comme suite :

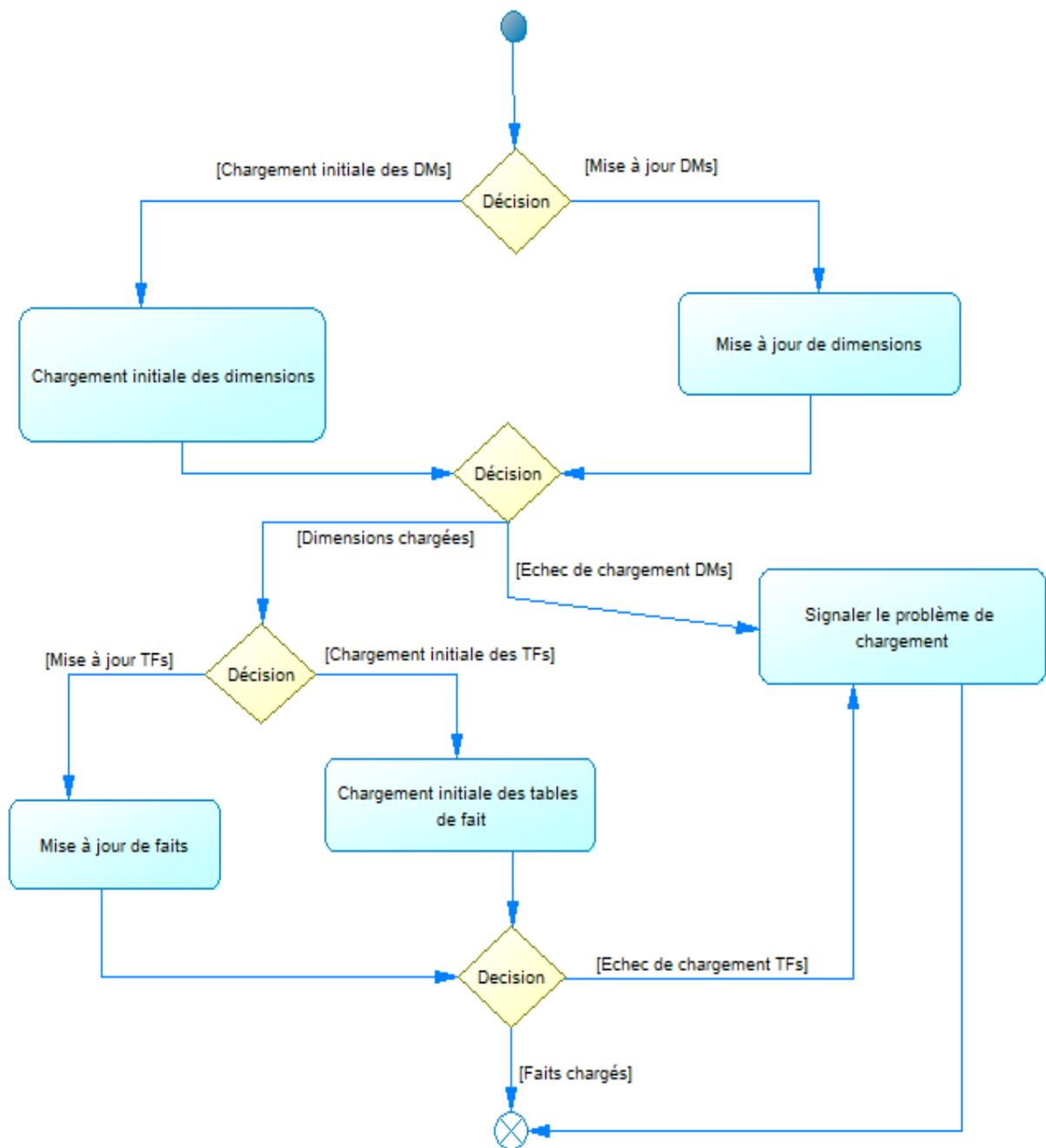


Figure 4.2: Diagramme d'activité du processus générale d'alimentation de DW

Pour alimenter un entrepôt de donnée il y a deux types de tables à charger (faits et dimensions) où chaque type a ses propres informations, pour cela nous distinguons deux sous processus qui exécute un avant l'autre car les tables de faits dépendent de tables de dimension.

Ce processus parfois trouve des problèmes de chargement ou des empannes pour cela il va signaler le problème en envoyant un email à l'administrateur de système.

IV.1. Processus de chargement des dimensions

Les tables de dimension constituent le contexte de la table de faits. Elles représentent le point d'entrée au Data Warehouse. Une dimension est généralement constituée : d'une clé artificielle, d'une clé naturelle et des attributs.

Le processus de chargement de dimension doit, outre le chargement des données, assurer :

- **La gestion des clés artificielles** : affectation des clés et mise en correspondance avec les clés naturelles.
- **La gestion de l'évolution de dimension** : gérer les changements que subissent les dimensions. Il existe deux types de traitement par rapport à l'évolution d'une dimension :
 - **Mise à jour** : consiste à mettre à jour l'enregistrement ou un attribut subissant un changement.
 - **Création d'un nouvel enregistrement** : consiste à créer un nouvel enregistrement afin consiste à mettre à jour l'attribut subissant un changement. de sauvegarder tout le cycle d'évolution de la dimension.

La stratégie adoptée pour la détection des changements se fait grâce à la comparaison entre le dernier chargement et l'extraction actuelle qui la précède un tri de données pour faciliter cette comparaison. Cette méthode est la plus fiable pour la capture des changements.

Le diagramme d'activité suivant décrit le processus de chargement des tables de dimensions.

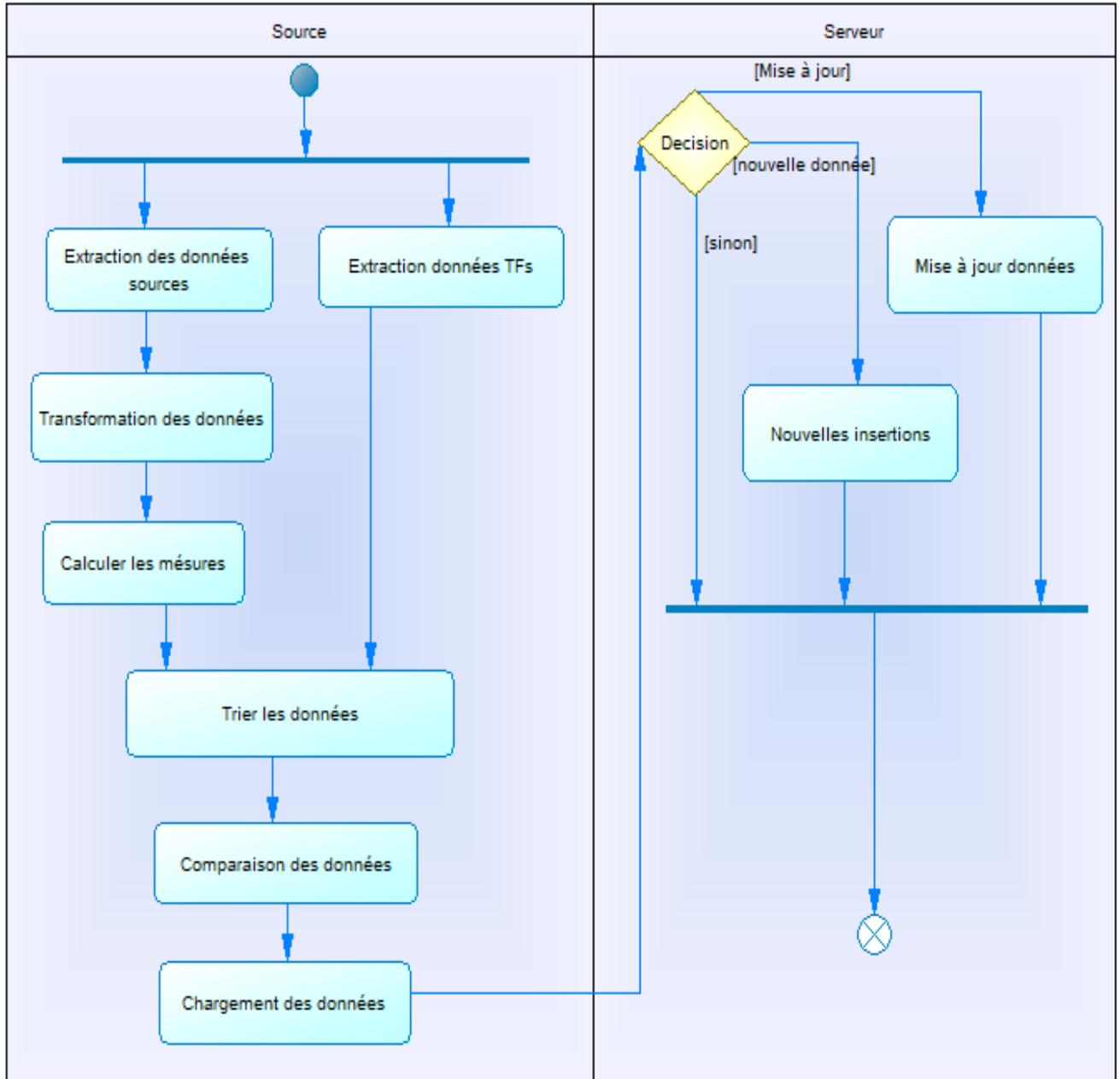


Figure 4.3: Diagramme d'activité de chargement de tables de dimensions

IV.2. Processus de chargement des tables de faits

Les tables de faits contiennent des mesures d'analyses et un ensemble des clés étrangères qui référencent aux dimensions qui sont des axes d'analyses où cette l'ensemble est une clé primaire de la table de fait, alors le processus de chargement des tables de faits doit garantir l'intégrité vis à-vis des tables de dimensions.

Le diagramme d'activité suivant décrit le processus de chargement des tables de faits :

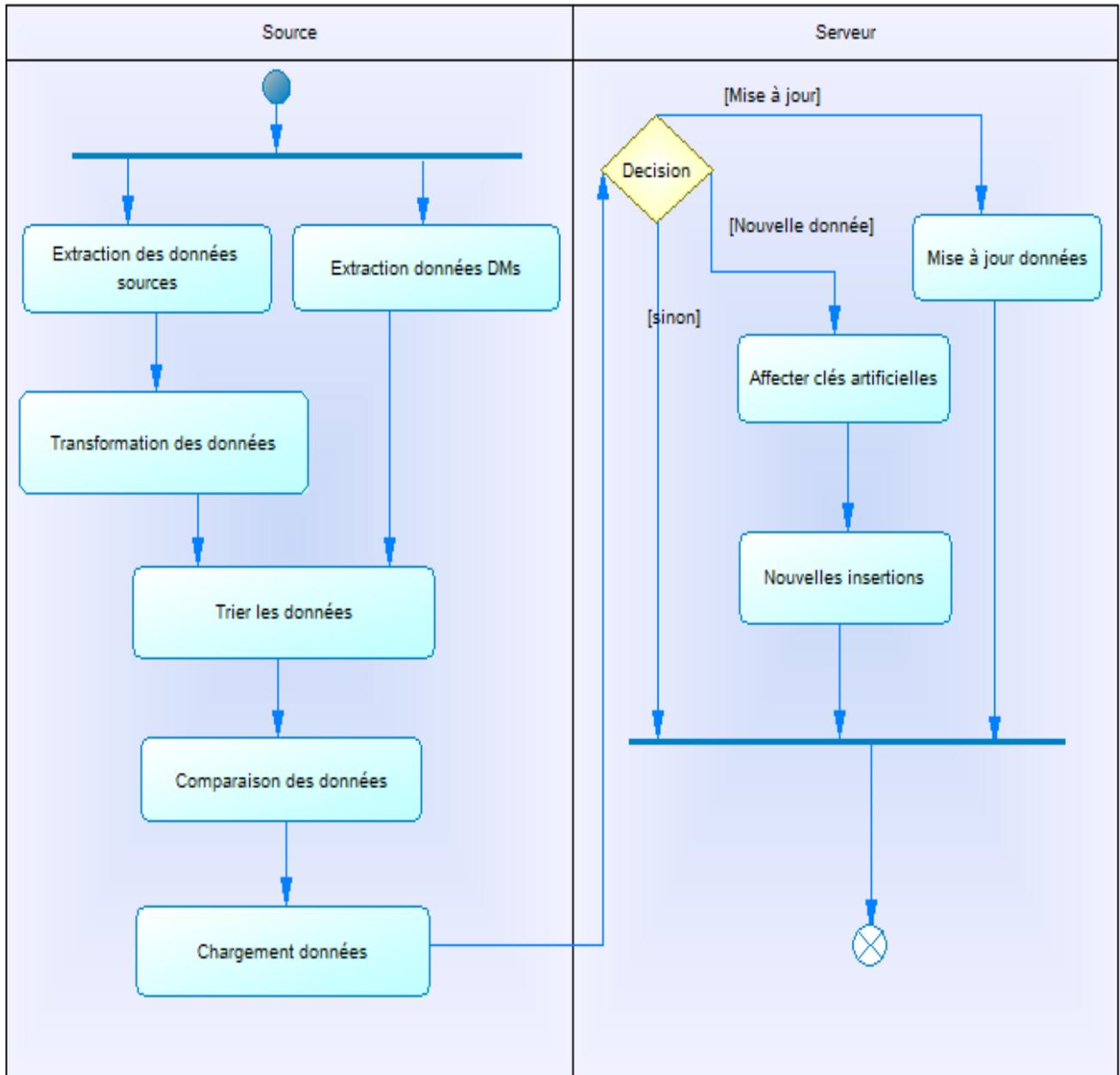


Figure 4.4: Diagramme d'activité de chargement de tables de faits

IV.3. Processus de chargement particulier

Dans un entrepôt de données il y a des tables particulières, soit : la table de la dimension temps

Contrairement aux autres dimensions, la dimension temps contient uniquement des dates qui ne sont pas forcément extraites à partir des systèmes sources. Cette dimension doit contenir, en effet, toutes les dates qui coïncident, ou coïncideront, avec un fait donné. Elle

participe à toutes les étoiles et assure l'historisation. Dès lors, il est préférable de construire un calendrier

« La dimension date est plus souvent construite comme étant un calendrier avec une granularité journalière ». [15]

V. Conclusion

Un processus E.T.L a pour mission d'assurer la livraison de données conformes, cohérentes et correctes tout en assurant des performances acceptables. Lors de la conception du processus de l'ETL il a été envisagé d'atteindre les objectifs suivants :

- Assurer le chargement des données et leur qualité.
- Assurer la transparence des processus de chargement et de consolidation qui assure la qualité des données.
- Ne pas nuire aux performances des systèmes sources.
- Permettre un suivi de l'avancement des chargements, et corrections en cas d'erreur.
- Offrir une documentation complète du processus, afin de faciliter la maintenance ainsi que l'amélioration.
- Offrir une performance optimale grâce aux chargements parallèles et séparation des données selon l'opération de chargement (mise à jour ou nouvelle insertion).

CHAPITRE 5 : REALISATION DU PROJET

I. Introduction

Ce dernier chapitre est consacré à la présentation des différents outils qui ont permis la réalisation de notre projet et l'ensemble des fonctionnalités qu'offre ce dernier sous forme de prises d'écran avec des descriptions pour bien comprendre chaque étape de la réalisation et de la mise en place de notre entrepôt de donnée.

II. Présentation des différents outils

II.1. Pentaho Data Integration (PDI)

Pentaho Data Integration (PDI) permet aux entreprises d'extraire les données, à partir des sources complexes et hétérogènes et de créer des informations cohérentes et de bonne qualité pour les applications critiques. Grâce à une interface utilisateur graphique complète et un moteur de traitement parallèle, PDI offre une solution ETL (extraction, transformation, chargement) de haute performance qui couvre tous les besoins en intégration de données. [16]

Les cas d'utilisation de PDI comprennent :

- La Business Intelligence et l'entreposage de données ;
- La migration des données ;
- La consolidation des applications.

La synchronisation des données PDI se compose de 3 applications :

- **Spoon** : est l'outil qui permet grâce à son interface graphique de créer des transformations, les exécuter et les sauvegarder. Les composants permettant la manipulation des données sont nommés « étapes » (steps en anglais). Spoon comprend un grand nombre d'étapes, par exemple il existe une étape permettant d'extraire des données de diverses bases de données, un autre aidant à l'extraction depuis des fichiers. [16]
- **Pan** : application en ligne de commande permettant de lancer l'exécution d'une transformation donnée. [16]
- **Kitchen** : application en ligne de commande permettant de lancer l'exécution d'une tâche donnée. Dans notre cas nous avons utilisés l'application Spoon. [16]



Figure 5.1: Pentaho Data Integration (PDI) [16]

II.2. PostGreSQL

PostgreSQL est un Système de Gestion de Base de Données (SBGD) libre disponible sous licence BSD. Ce système multiplateforme est largement connu et réputé à travers le monde, notamment pour son comportement stable et pour être très respectueux des normes ANSI SQL. Ce projet libre n'est pas géré par une entreprise mais par une communauté de développeurs. [17]



Figure 5.2 : Logo PostgreSQL [17]

II.2.2. Pourquoi choisir postgresql ?

- **Déploiement illimité**

Vous pouvez déployer PostgreSQL sur autant de serveurs avec autant de CPU que vous le souhaitez. Non seulement le coût d'investissement est nul, mais il n'y a pas de maintenance annuelle à payer sur le long terme, l'économie est très importante.

- a. **Excellent support**

Le support assuré par la communauté PostgreSQL est excellent et gratuit. De plus, de nombreuses SLL peuvent vous offrir un contrat de support formel sur mesure.

- b. **Economies significatives sur les coûts de personnel**

PostgreSQL nécessite beaucoup moins de maintenance et de paramétrage que les grandes bases de données commerciales, tout en proposant la plupart de leurs fonctionnalités, et surtout la fiabilité et les performances que l'on attend d'un tel produit.

Contrairement à beaucoup de bases de données commerciales, PostgreSQL ne nécessite pas de suivre plusieurs semaines de formation, ni d'avoir un administrateur de bases de données à plein temps. C'est un avantage majeur pour beaucoup de PME.

c. Fiabilité et stabilité

Il est très courant que des sociétés rapportent que PostgreSQL n'a jamais crashé, même pendant des années. Pas une seule fois. Tous les SGBDR ne peuvent pas en dire autant. En particulier, PostgreSQL ne craint pas les coupures électriques.

d. Très bonne compatibilité SQL

Comme beaucoup d'outils Open Source, PostgreSQL met un point d'honneur à suivre les normes, et en particulier les normes SQL 92 et 99. C'est un gage de pérennité et de portabilité.

I.3 Power BI

Power BI est une suite d'outils permettant d'analyser des données et de partager des informations. Quelle que soit la nature de vos données, Power BI vous permet de vous connecter facilement à vos sources de données, de visualiser les informations importantes et de les partager. [18]



Figure 5. 3: Processus général de fonctionnement Power BI [18]

Power BI se compose d'une application appelée Power BI Desktop, d'un service SaaS (Software as a Service) en ligne appelé service Power BI, ainsi que des applications Power BI mobiles.

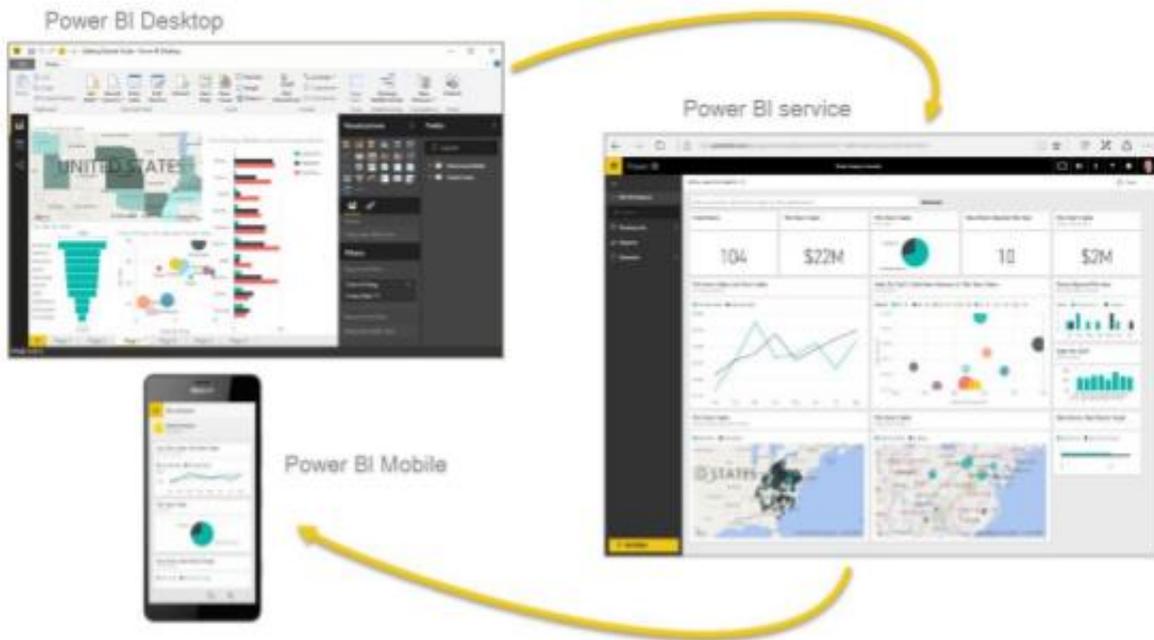


Figure 5.4: Composant logiciel Power BI [18]

Power bi fonctionne selon les étapes suivantes :

- Extraction des données depuis la source depuis power bi desktop
- Faire la mise à jour à partir de power bi getways
- La publication à partir de power bi service
- Enfin l'accès aux résultats comme le démontre la figure suivantes

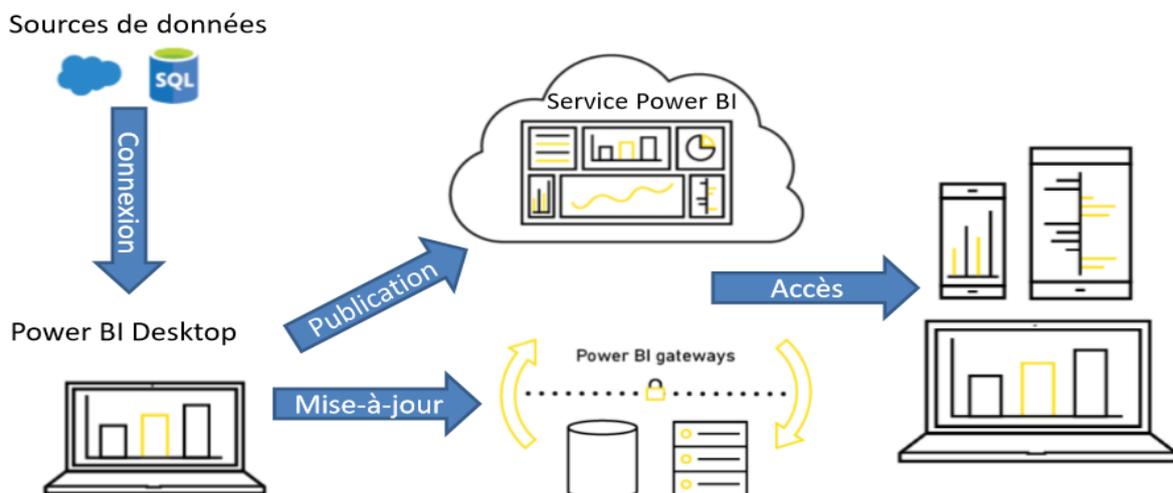


Figure 5. 5 : Procédure de fonctionnement Power BI [18]

Power BI permet nous de faire des tableaux de bords qu'on peut exercer du ROLAP sur les visualisations et aussi créer des rapports sur 2 formats Excel ou PDF.

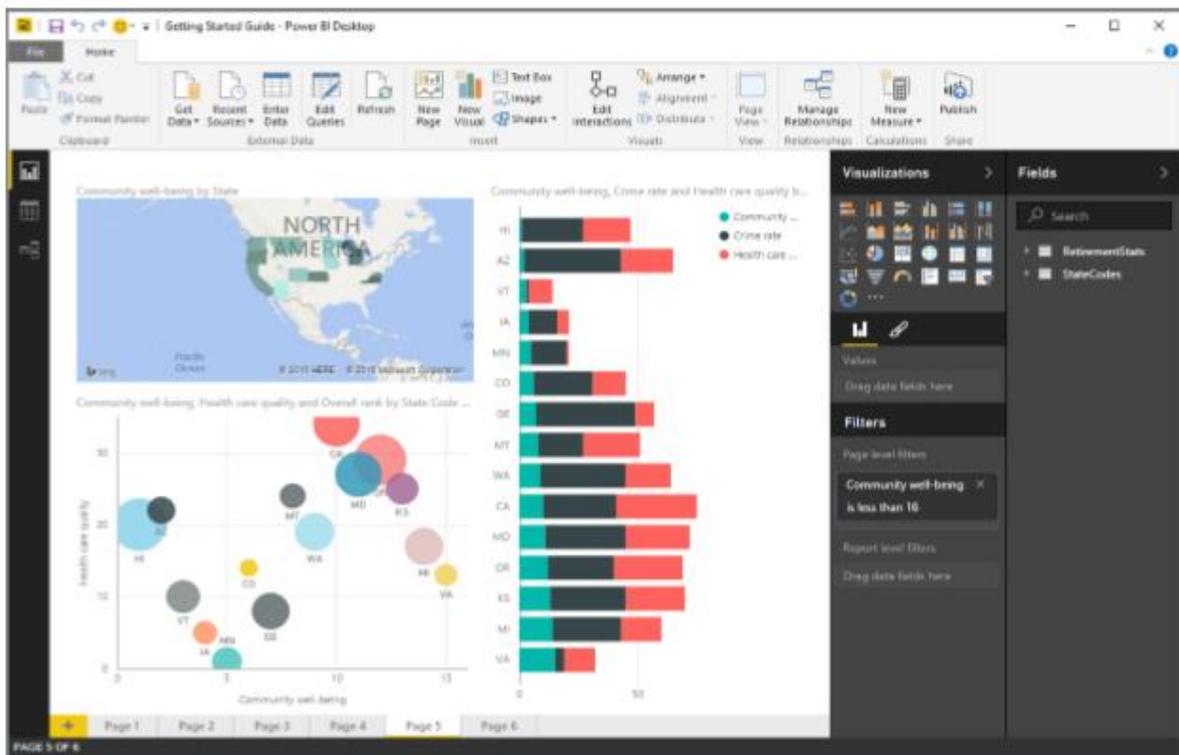


Figure 5.6: Exemple d'une page de rapport Power BI [18]

Comme le tableau de bord d'une voiture, un tableau de bord Power BI est un ensemble d'éléments visuels contenus dans une page que vous pouvez partager avec d'autres utilisateurs.

Un tableau de bord doit tenir sur une seule page, souvent appelée canevas. Vous pouvez partager les tableaux de bord avec d'autres utilisateurs ou groupes, qui peuvent alors interagir dans le service Power BI ou sur leur appareil mobile.

Dans Power BI, une vignette est une visualisation unique qui se trouve dans un rapport ou sur un tableau de bord.

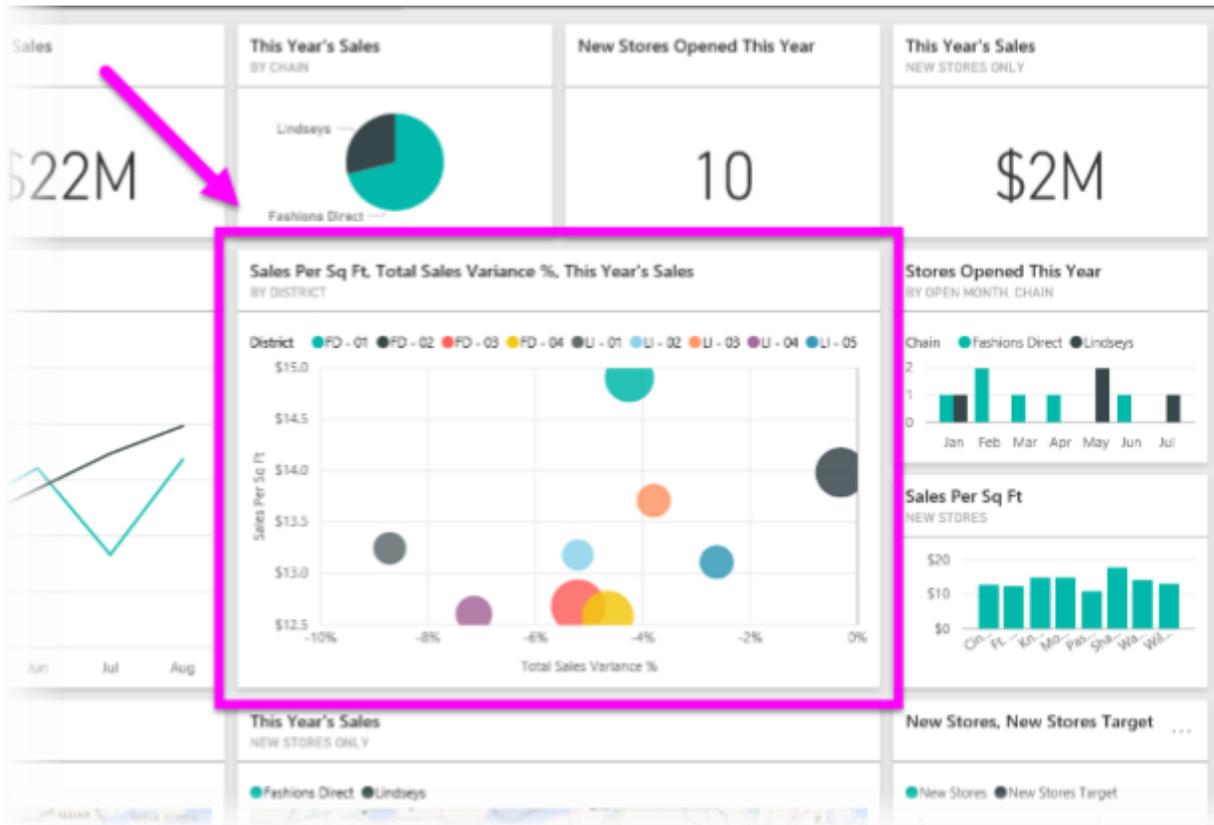


Figure 5.7: Exemple d'une vignette dans Power BI [17]

III. Présentation des étapes de réalisation de notre entrepôt de données

La réalisation de notre entrepôt de données s'effectue en 3 étapes essentielles comme suit :

III.1. Création du DATA WAREHOUSE

La création du data-warehouse ou entrepôt de donnée se fait selon sa conception, en utilisant query tool de postgresql on exécute une requête de création de la base de données intitulée (DATA-WAREHOUSE) avec ses différentes tables de dimensions et tables de faits

On aura qu'a créé un seul schéma qui regroupe toutes les tables

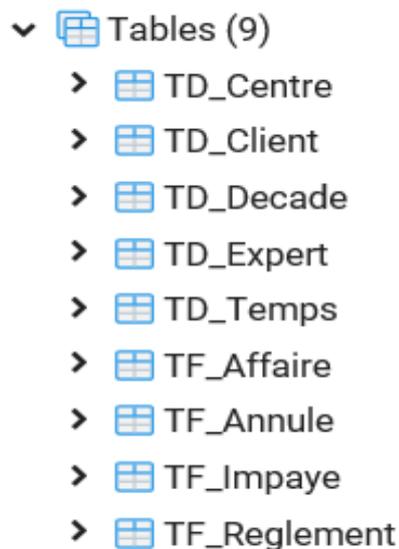


Figure 5.8 Le schéma de DW

III.2 Alimentation de l'entrepôt de données

Intégration (spoon) est l'étape la plus critique et la plus importante pour l'alimentation de notre entrepôt.

L'extraction des données, leurs transformations ainsi que leurs chargements sont réalisés de la manière suivante :

Lors du lancement de ETL, on a le panneau « navigateur » qui nous permet de créer une transformation dans laquelle une « palette de création » apparaît qui donne la possibilité de faire une extraction de n'importe quel type de fichier, dans notre cas c'est une base de données Postgresql.

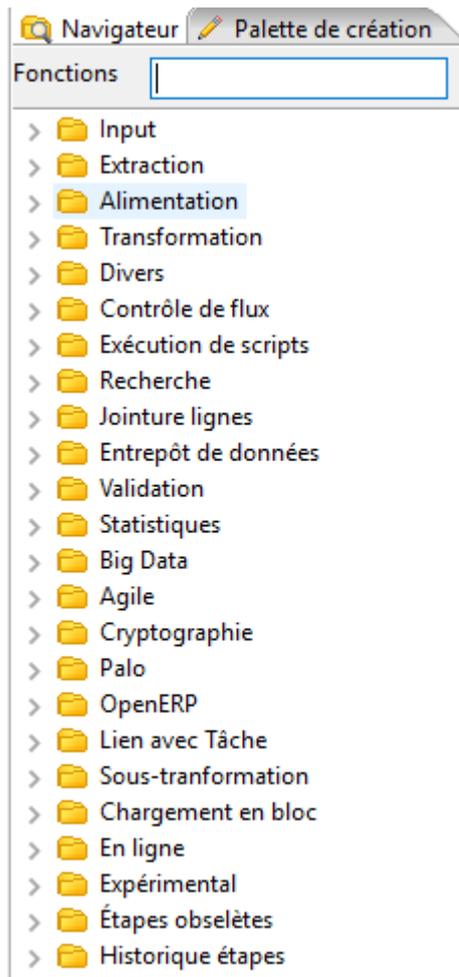


Figure 5. 9: Palette de création

La première étape de notre processus ETL est l'extraction des données en utilisant le composant d'extraction d'après une base de données.



Figure 5.10: Composant d'extraction de données

Pour le bon fonctionnement d'un composant « extract » nous le paramétrons, tout d'abord il faut donner un nom à l'étape ensuite définir la connexion à la source de données tout en assurant le type de connexion et enfin définir la requête SQL adapté à notre SGBD qui sélectionnera nos données.

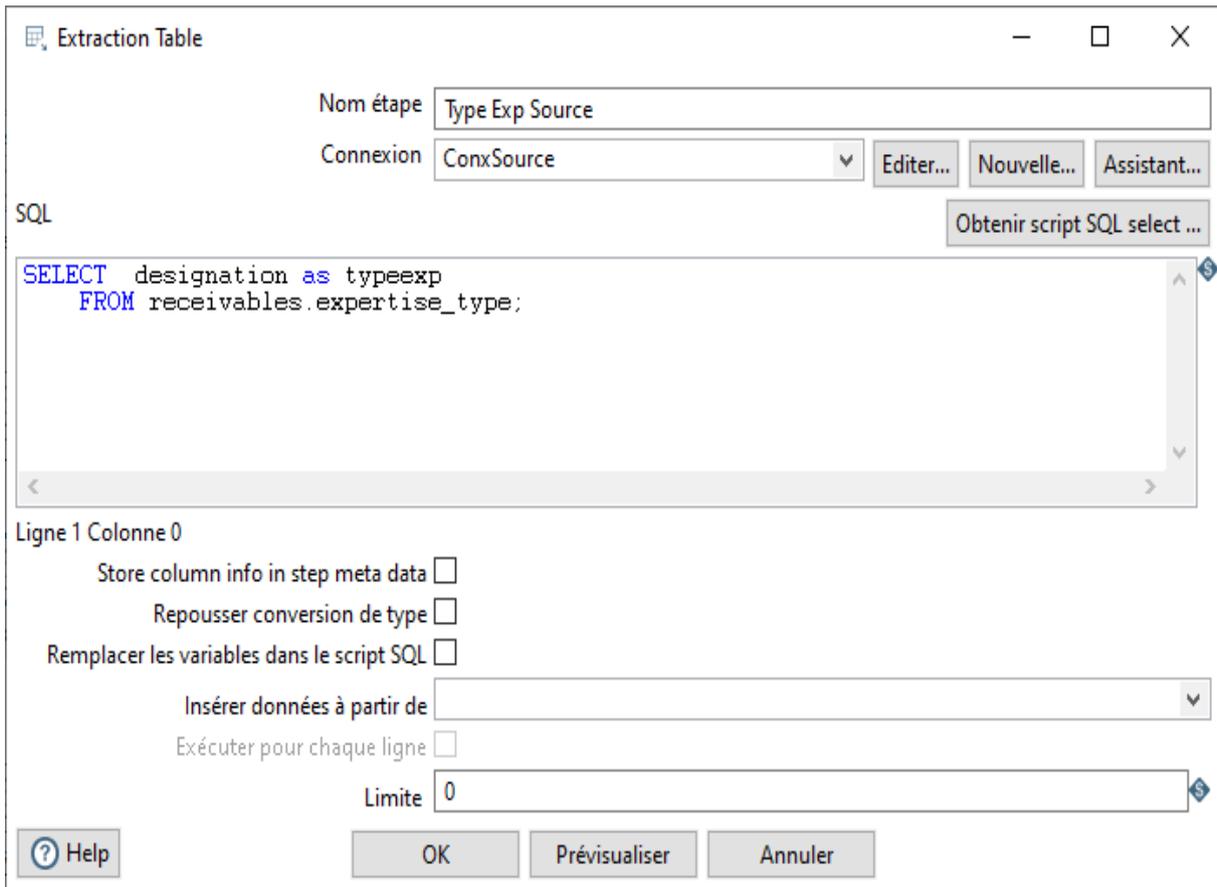


Figure 5.11: Paramétrage d'un composant ETL (extract)

La deuxième étape est la comparaison on utilise une technique bien précise qui consiste à extraire les données de la source en plus d'extraire les données de notre table de dimension ou de fait, trier les deux résultats et leur faire une étape de comparaison.

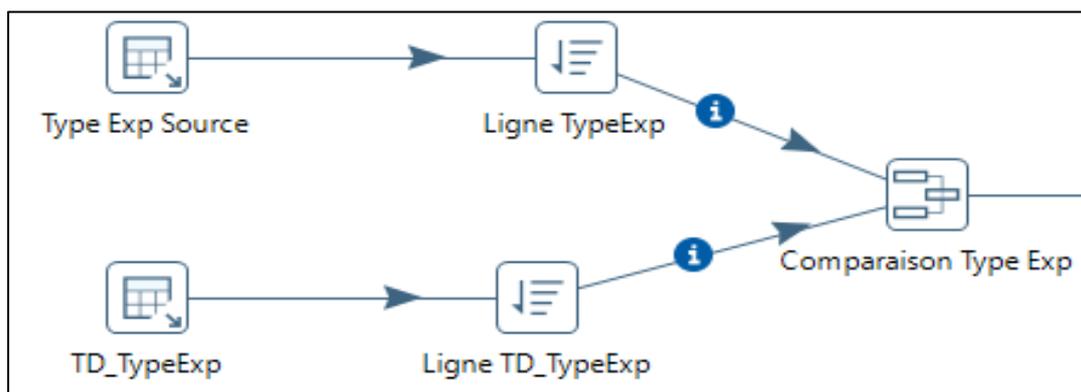


Figure 5.12: Comparaison entre les données de la dimension et les données sources

Le composant « compare ligne » crée un nouvel attribut qui peut prendre les valeurs suivantes :

- **Identique** : signifie que l'enregistrement dans les deux extractions est le même.

- **Supprimer** : signifie que l'enregistrement dans la destination a été supprimé dans la source.
- **Modifier** : qui signifie que l'enregistrement a été mise à jour dans la source.

Comparaison lignes

Nom étape: Comparaison Type Exp

Origine lignes référence: Ligne TD_TypeExp

Origine ligne comparaison: Ligne TypeExp

Nom champ résultat: flagfield

#	Champ clé
1	typeexp

#	Valeur champ
1	typeexp

Récupérer champs clés Récupérer valeurs champs

Help OK Annuler

Figure 5.13: Paramètres de la comparaison

Une fois la comparaison faite on utilise un nouveau composant « synchronize after merge », ce dernier sera le responsable des différentes actions à faire tout en prenant compte le résultat de l'attribut de comparaison.

Comme le montre la figure suivante cette partie va nous permettre de choisir l'action à activer à chaque résultat de notre champ opération, cela nous permet de contrôler l'évolution des données et à avoir un entrepôt mise à jour par rapport aux bases de données source.

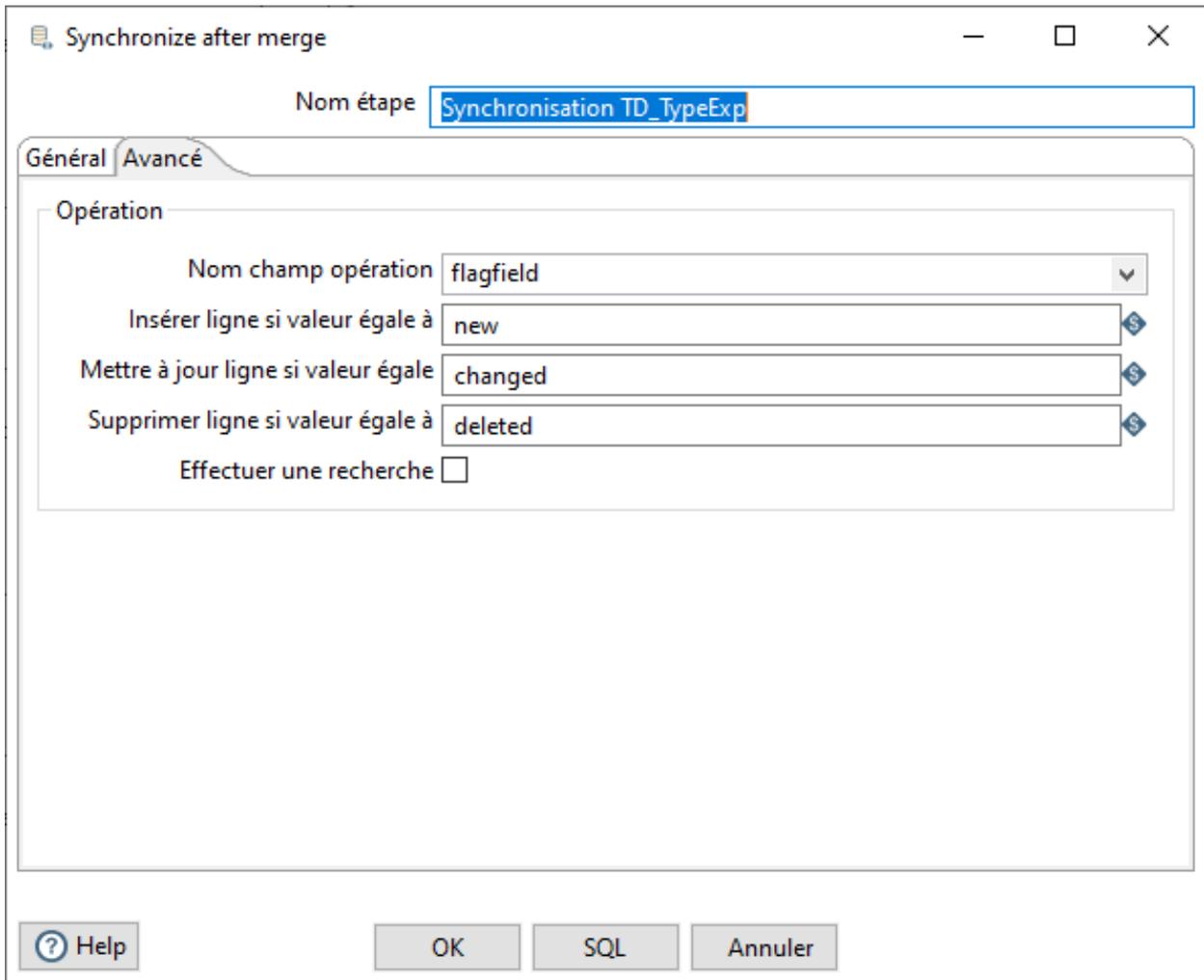


Figure 5.14: Paramètres avancé « synchronize »

Ce n'est pas les seuls paramètres à contrôler dans ce composant, la figure suivante montre le paramétrage essentiel à effectuer où il faut faire une correspondance entre les champs de la source et les champs de la destination (dimension ou fait).

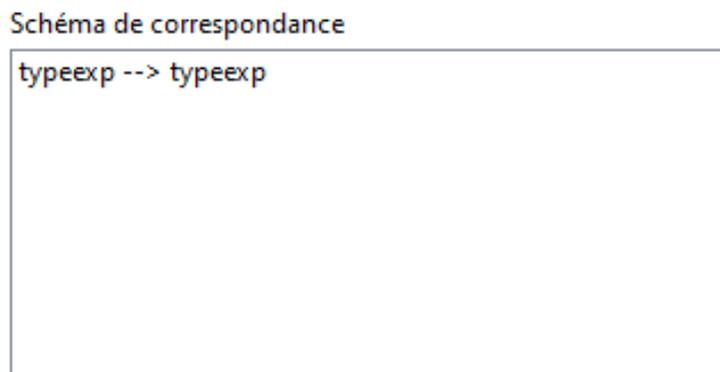


Figure 5.15: Schéma de correspondance

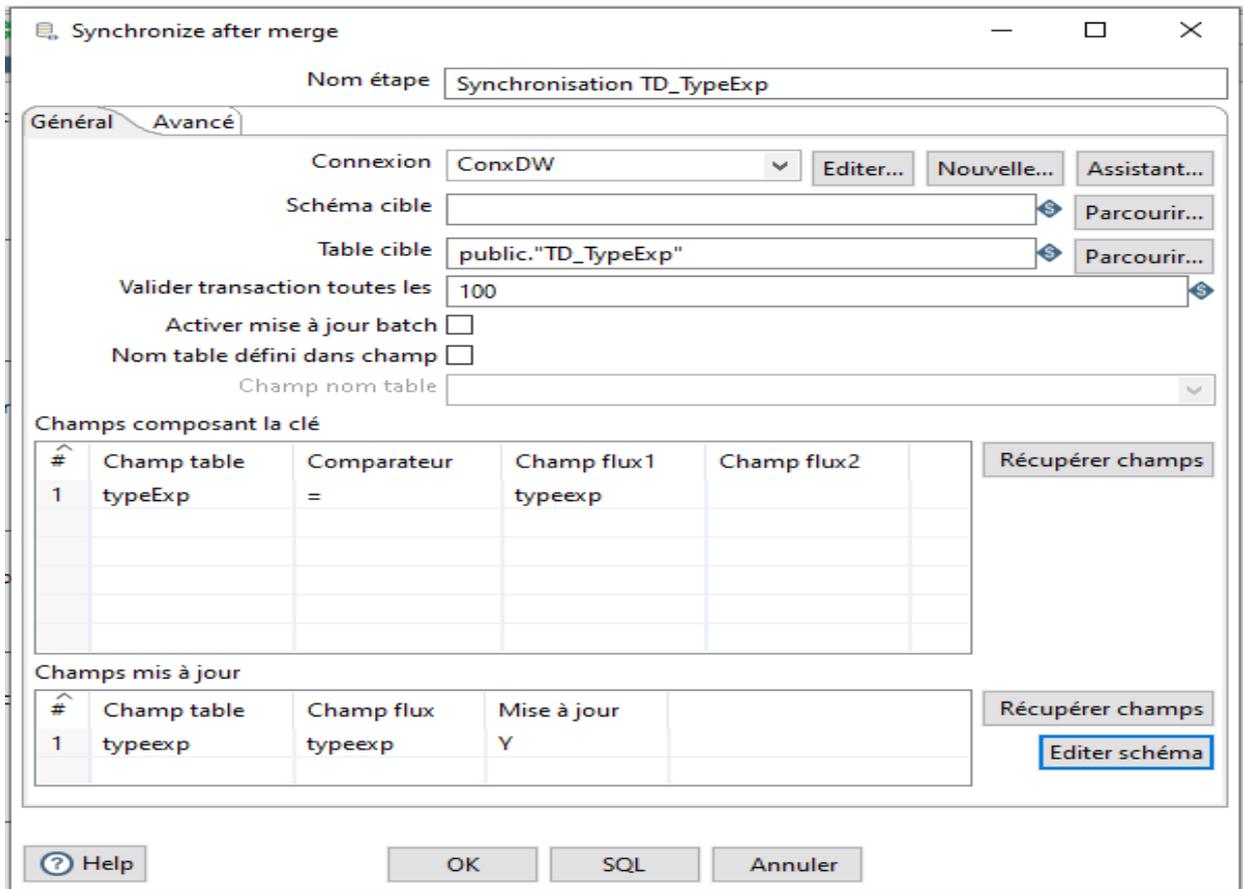


Figure 5.16: Paramètre générale

Exemple de processus d'alimentation d'une table de dimension ou d'une table de fait.

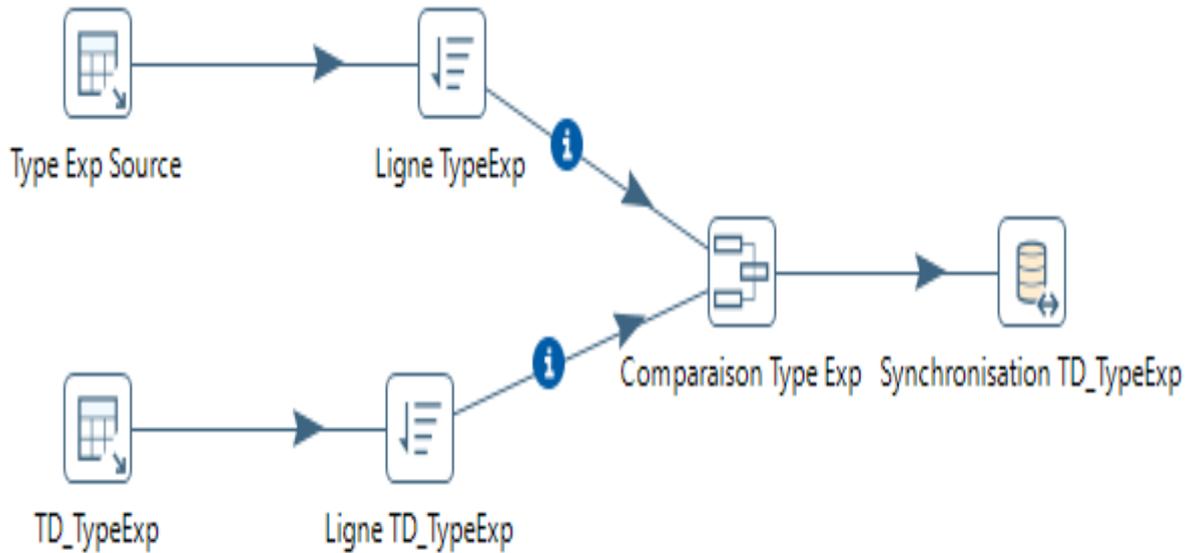


Figure 5.17: Exemple d'alimentation d'une table de dimension

Alors comme nous avons expliqué précédemment l'alimentation d'une dimension on fait le même traitement pour les autres dimensions et même les tables de faits, l'ensemble de ses traitements sont groupé dans une transformation, en effet nous obtenons deux type de transformation une pour l'alimentation des dimensions et l'autres l'alimentation des faits, alors nous créons les mêmes transformations pour les 32 bases de données.

Dans le chapitre alimentation du Data Warehouse nous avons expliqué qu'il faut alimenter les dimensions puis les faits et pentaho nous permette faire se traitement à l'aide de la création d'une tâche (Job).

Une tâche a un composant Start unique où elle commence l'exécution, ce composant permet nous de configuré la périodicité comme suite :

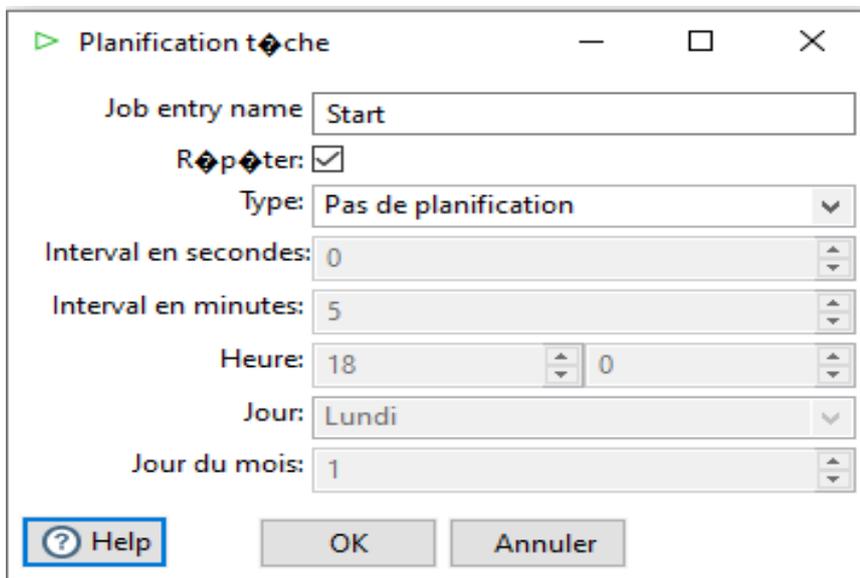


Figure 5.18: Planification l'exécution de la tâche

Nous configurons la périodicité pour toutes tâches d'exécutions des transformations à « pas de planification » car il existe une tâche qui englobe toutes ses tâches et nous ordonnons les transformations à exécuter, la figure suivante illustre la tâche Cheraga :



Figure 5.19: Exemple d'une tâche

A la fin de la création de toutes les tâches nous les groupons dans une tâche globale pour la gestion de la périodicité et les problèmes de chargement en envoyant des courriels d'email vers les admins de ce système, la figure suivante représente ce travail :

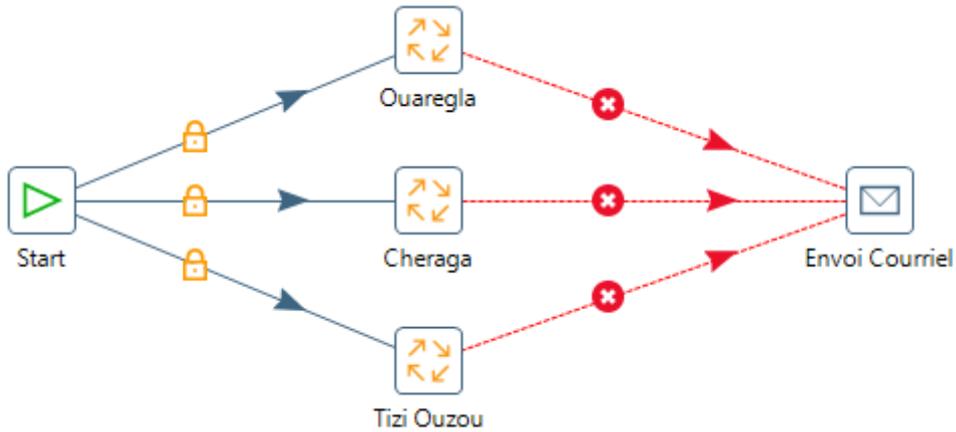


Figure 5.20: Exemple d'une tâche globale

III.3 Visualisation des tableaux de bords et des rapports

Maintenant l'entrepôt de données est bien chargé il faut visualiser ces données aux décideurs afin de bien comprendre, pour cela nous devons créer de tableaux de bords et des rapports avec PowerBI que nous allons expliquer prochainement.

Tous d'abord nous décrivons l'interface de PowerBI et expliquons comment connecter avec l'entrepôt de données.

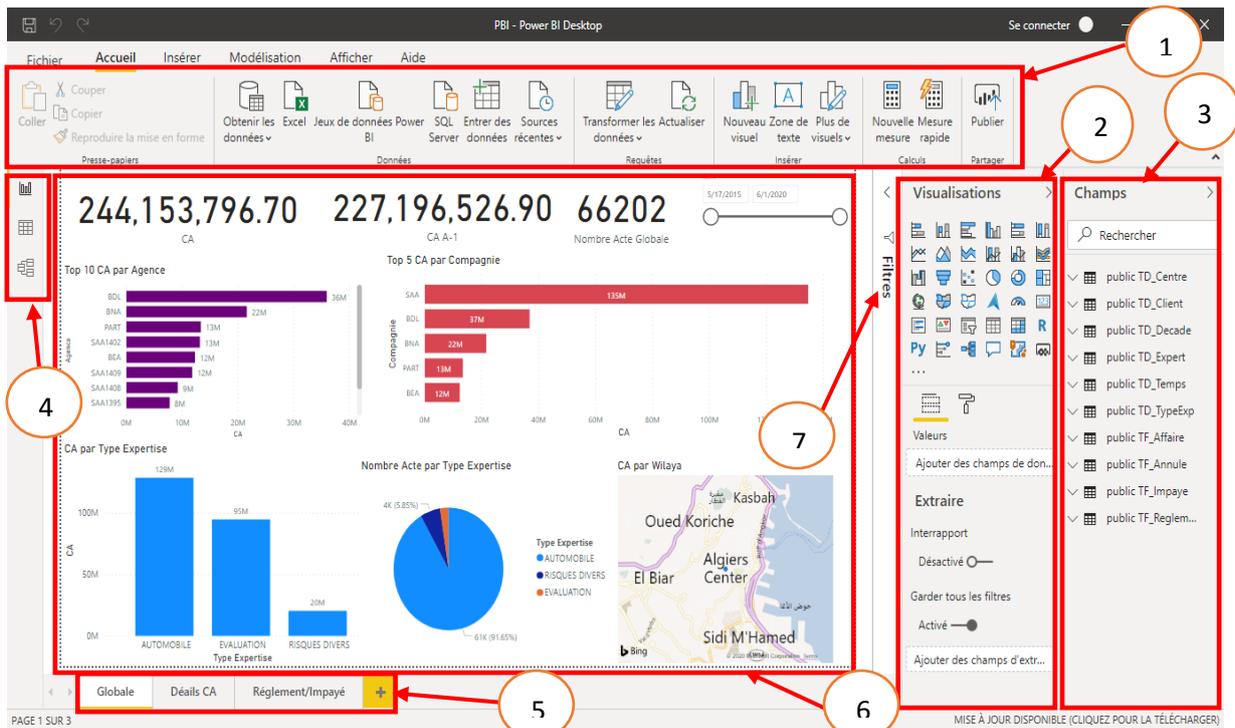


Figure 5. 21: Fenêtre de PowerBI

- 1- Barre d’outils :** Contienne des outils nous aide à naviguer sur power bi comme connecter aux sources de données, actualiser, publier les rapports et accéder au power bi query pour faire des transformations sur données récupérer
- 2- Visualisations :** Contient la liste des étiquettes de visualisations comme histogramme, carte géographique, courbe et leurs styles
- 3- Chapms :** Contient les tables récupérer d’après la source pour choisir les champs que nous voulons visualiser
- 4- Vue :** Contienne 3 vues : Rapport, Données et Relations
- 5- Pages :** Contient les pages (tableaux bords)
- 6- Cadre des étiquettes :** Contient les étiquettes qu’on veut les affichés
- 7- Filtres :** Panneau qu’on peut paramétrer les axes d’analyses

Pour connecter à une source de données il existe deux méthodes :

Méthode 1 :

- a. Nous utilisons un outil qui situe dans la barre d’outils « Obtenir des données »
- b. Choisir un capteur dans liste des capteurs de connexion
- c. Configurer la connexion

d. Charger les données

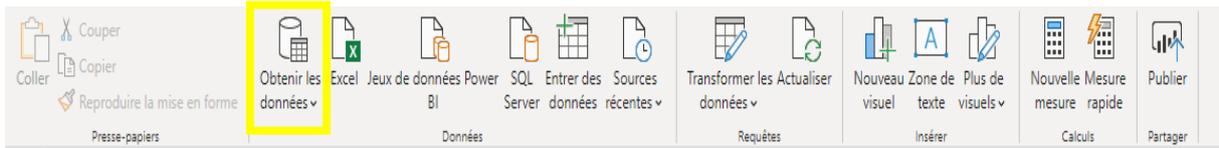


Figure 6.22: Outil de connexion à la source de données

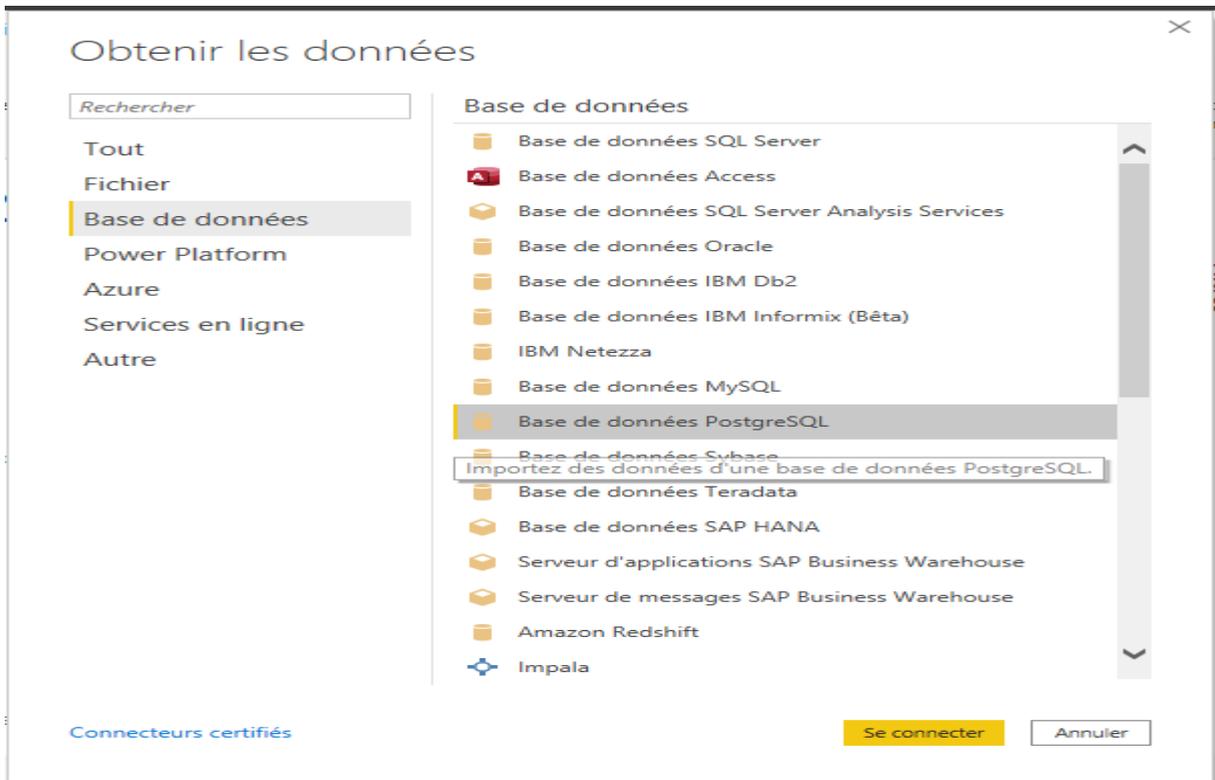


Figure 5.23: Liste des capteurs de connexion

Méthode 2 :

- a. Nous accédons au power bi query
- b. Dans la barre d'outils choisir « Nouvelle Source » ou « Source Récente »
- c. Choisir un capteur dans liste des capteurs de connexion
- d. Configurer la connexion
- e. Appliquer cette connexion et charger les données sur la fenêtre des rapports

Une fois que nous avons bien chargé les données donc nous pouvons créer des tableaux de bords, des rapports selon les besoin de client et ils sont comme suite :

La première page est une page globale porte des étiquettes sur les chiffres générales de l'entreprise comme chiffre d'affaire et le nombre globale par tous les axes d'analyse, la figure suivante illustre la page :

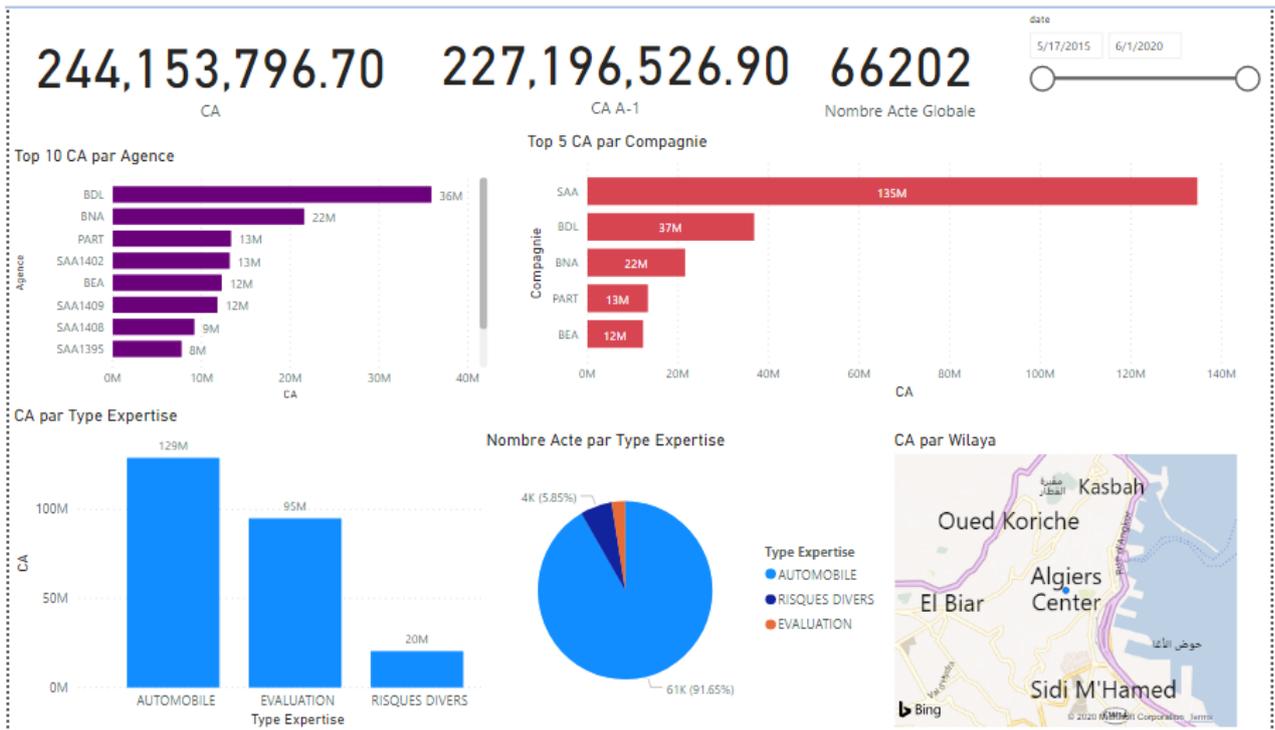


Figure 5.24: Tableau de bord globale

La deuxième page visualise le détail de fait Expertise qui comporte deux tableaux, le premier contient la liste 20% des agences qui réalise 80% de chiffre d'affaire et leur taux d'augmentation par rapport à l'année précédente, le deuxième tableau sur les experts les plus actifs en basant sur les axes d'analyses Centre, Temps et Type d'expertise, la figure suivante illustre la page :

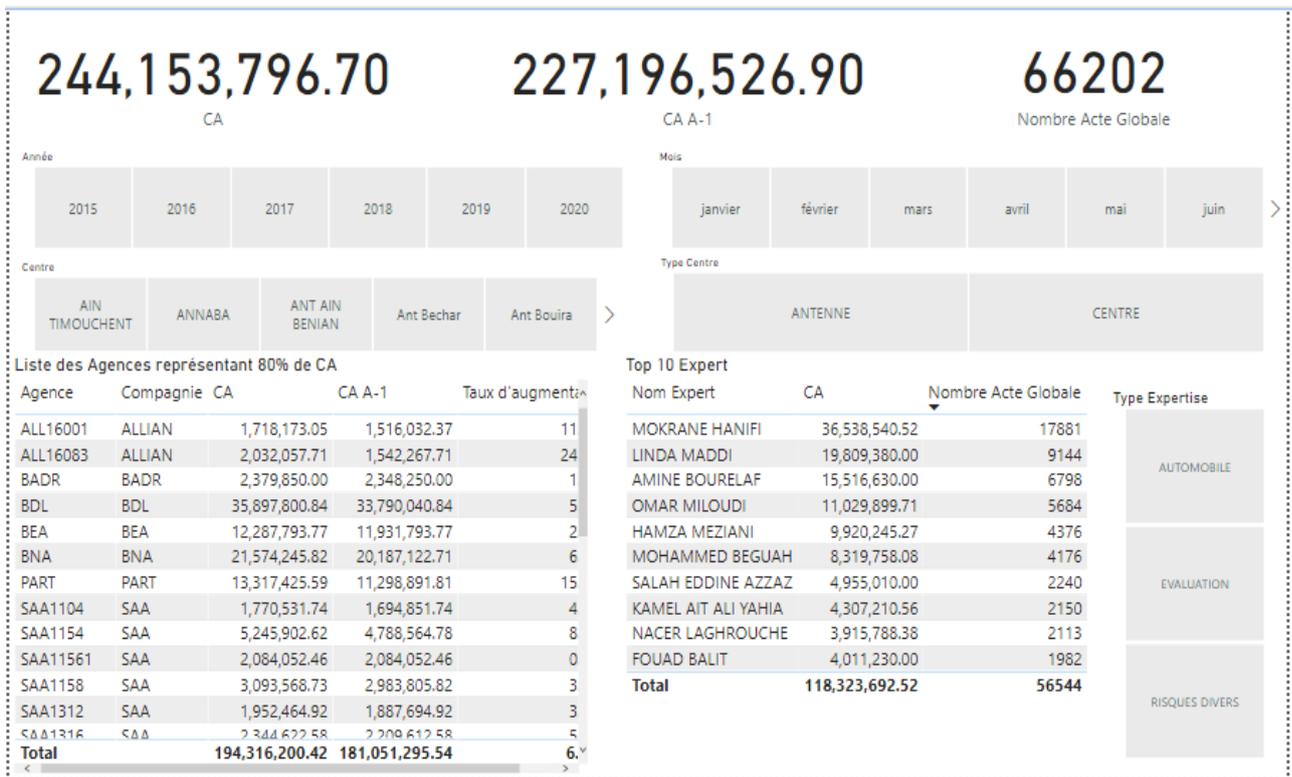


Figure 5.25: Tableau de bord détaillé fait Expertise

La troisième page visualise le détail des faits Impayé, Règlement et Annulé qui comporte deux tableaux, le premier contient la liste des agences et leurs montants et le nombre de PV réglés, montants et le nombre de PV impayés, le deuxième tableau sur les experts et le nombre de PV qui les annulent en basant sur les axes d'analyses Centre, Temps et Type d'expertise, la figure suivante illustre la page :

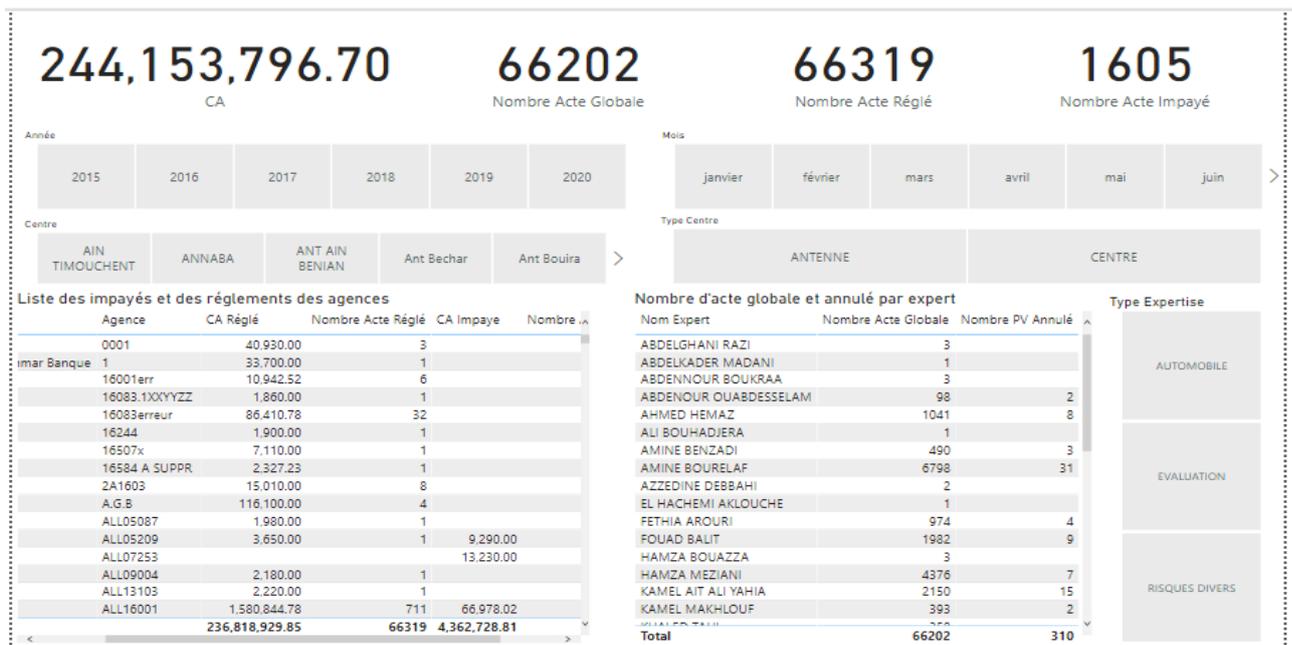


Figure 5.26: Tableau de bord détaillé paiement et annulé

IV. Architecture de système développé

L'architecture de système est composée de trois sites qui sont :

- Site 1 : Les sources des données qui sont des bases de données PostgreSQL (nous avons utilisé que 3 bases de données comme exemple)
- Site 2 : Ici l'ETL Pentaho s'exécute pour extraire, transformer et charger les données dans le DW qui est dans le même site
- Site 3 : C'est le client terminal où le PowerBI visualise des tableaux de bords et des rapports d'après les données de DW

La figure suivante illustre cette architecture :

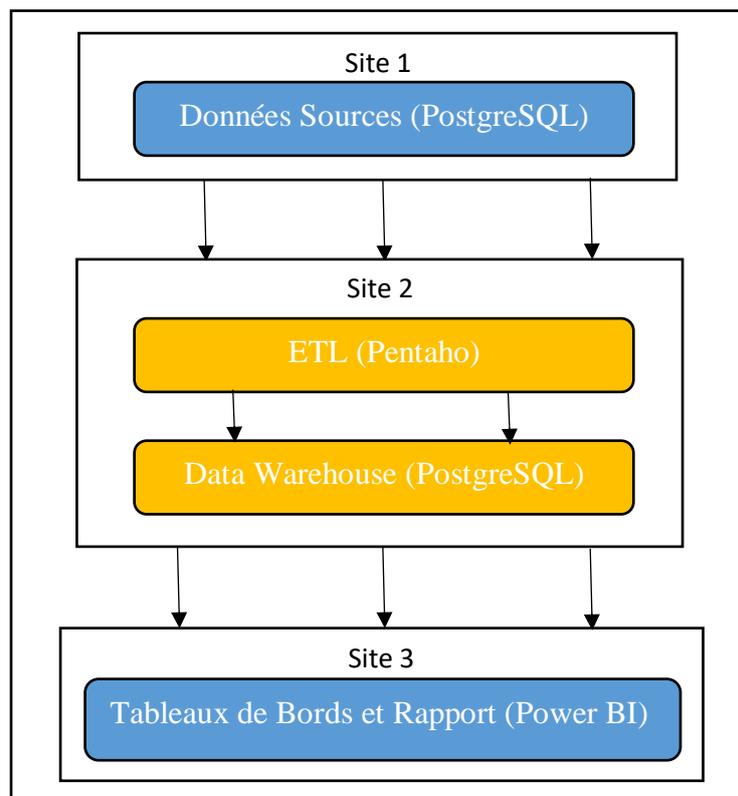


Figure 5. 27 Architecture de projet

V. Conclusion

Dans ce chapitre nous avons détaillé toutes les étapes avec des illustrations les technologies et les logiciels que nous avons utilisé pour réaliser ce projet où nous avons choisi les programmes qui sont open source, gratuit et professionnel afin d'avoir la main pour développer librement et aussi d'avoir une bonne présentation à notre client terminal.

Conclusion Générale

A travers ce mémoire nous avons expliqué la conception et la mise en place d'une DATAWAREHOUSE afin de permettre aux décideurs et aux analystes de la SAE de filtrer et d'analyser les informations nécessaires à la prise de décision et remplir le vide de l'existant décisionnelle au sein de l'entreprise.

Nous avons utilisé des moyens accessible pour réaliser ce projet notamment les logiciel open source comme PDI, POWER BI ... etc.

Dans les perspective à venir, nous projetons de :

- ✓ Créé un portail web (power bi service) pour une meilleure exploitation des rapports.
- ✓ Exploitation des rapports à partir d'application mobile.
- ✓ Implémenter le projet au sein d'un serveur linux haute gamme.
- ✓ Connecter la DATAWAREHOUSE aux 32 bases de données source.
- ✓ Et enfin faire des mise à jours continue au système relativement au changement de l'entreprise.

Bibliographie

- [1] N. B. Darmawan, «nicobudidarmawan,» 03 Janvier 2014. [En ligne]. Available: <http://www.nicobudidarmawan.com/2014/01/business-intelligence-oltp-vs-olap.html>. [Accès le Février 2020].
- [2] W. H. Inmon, Building the Data Warehouse Third Edition, Wiley Computer Publishing, 2002.
- [3] R. Krishna, «Data Warehouse Tutorial for Beginners: Learn in 7 Days,» Mai 2018. [En ligne]. Available: <https://www.guru99.com/data-warehousing-tutorial.html>. [Accès le Février 2020].
- [4] T. C. Programmer, «The Crazy Programmer,» Décembre 2019. [En ligne]. Available: <https://www.thecrazyprogrammer.com/2019/12/metadata-in-dbms.html>. [Accès le Février 2020].
- [5] F. Almeida, «Concepts and Fundaments of Data Warehousing and OLAP,» 2017.
- [6] Cartelis, «Architecture Data Warehouse – Approches Traditionnelle vs Cloud,» Avril 2018. [En ligne]. Available: <https://www.cartelis.com/blog/architecture-data-warehouse/>. [Accès le Février 2020].
- [7] H. Hasan et P. Hyland, «Using OLAP and Multidimensional Data for Decision Making,» *ITProfessional*, vol. 3, pp. 44-50, Septembre 2001.
- [8] M. Bala, *Chapitre 2 : Modélisation dimensionnelle*, 2020.
- [9] Diffen, «Snowflake Schema vs. Star Schema,» [En ligne]. Available: https://www.diffen.com/difference/Snowflake_Schema_vs_Star_Schema. [Accès le Février 2020].
- [10] F. ABDERRAHMANE et K. SOFIANE, Conception et réalisation d'un Data Warehouse pour la mise en place d'un système décisionnel, Alger, 2010.
- [11] JavaTPoint, «OLAP Operations in the Multidimensional Data Model,» [En ligne]. Available: <https://www.javatpoint.com/olap-operations>. [Accès le Février 2020].
- [12] M. Smallcombe, «The Ultimate Guide to Data Warehouse Design,» 02 Août 2019. [En ligne]. Available: <https://www.xplenty.com/blog/the-ultimate-guide-to-data-warehouse-design/>. [Accès le Février 2020].
- [13] P. Lane, Data Warehousing Guide, Oracle, 1999.
- [14] A. Bui, «Data Warehouse Design: The Good, The Bad, The Ugly,» 28 Novembre 2017. [En ligne]. Available: <https://blog.panoply.io/data-warehouse-design-the-good-the-bad-the-ugly>. [Accès le Février 2020].
- [15] R. K. e. J. Caserta, The Data warehouse ETL Toolkit, Wiley Publissing, 2004.
- [16] H. Vantara, «Pentaho documentation,» [En ligne]. Available: <https://help.pentaho.com/Documentation>. [Accès le Juillet 2020].

- [17] PostgreSQL, «Documentation,» [En ligne]. Available: <https://www.postgresql.org/docs/>. [Accès le Juillet 2020].
- [18] Microsoft, «Power BI Documentation,» [En ligne]. Available: <https://docs.microsoft.com/fr-fr/power-bi/>. [Accès le Juillet 2020].
- [19] W. H. Inmon, *Building the data warehouse*, fourth edition, Wiley, 2005.
- [20] R. Kimball et J. Caserta, *The data warehouse ETL toolkit*, Wiley, 2004.
- [21] P. Ponniah, *Data warehousing fundamentals*, Wiley, 2001.
- [22] A. Vaisman et E. Zimányi, *Data warehouse systems design and implementation*, Springer, 2014.
- [23] B. Yellin, «Data Warehouse Fundamentals for Storage Professionals-What You Need To Know,» EMC², 2011.
- [24] S. Crozat, «Data warehouse et outils décisionnels,» utc Formation, 2016.
- [25] C. Imhoff, N. Glemmo et J. G. Geiger, «Mastering data warehouse design relational and dimensional techniques,» Wiley, 2003.
- [26] R. Kimball et M. Ross, *The data warehouse toolkit second edition*, Wiley, 2002.
- [27] R. Kimball et M. Ross, *The data warehouse toolkit the definitive guide to dimensional modeling 3th edition*, Wiley, 2013.
- [28] D. Jones, *Guide concis de mise en oeuvre de la Business Intelligence au sein des PME*, Realtime Publishers.
- [29] L. Matthieu, *Entrepôt de données : De l'alimentation des données au décisionnel de l'établissement*, HAL, 2012.