

# Modeling and Implementation of a Decision-Making System For Personnel Management Within an Appraisal and Insurance Company: Case of The National Security Company.

<sup>1</sup> Dr YENDE RAPHAEL Grevisse, <sup>2</sup>KABEYA ILUNGA Paulin, <sup>3</sup>NTUMBA KALUMBA Joseph, <sup>4</sup>KASEREKA KIHOKOLO Ephrem, <sup>5</sup>BASUA BASUA Sebastien.

<sup>1</sup>Département d'Informatique de l'Université de Bas-Uélé (UNIBAS), Buta, (RDC). Email : [grevisse29@gmail.com](mailto:grevisse29@gmail.com)

<sup>2</sup>Département d'Informatique de l'Université Saint Laurent de Kananga (USLKA), Kananga, Kasai Central (RDC).

<sup>3</sup>Département d'Informatique de l'Institut Supérieur Technique de Kananga (ISTKA), Kasai Central (RDC).

<sup>4</sup>Département d'Informatique des Facultés Africaines BAKHITA de Butembo, Nord-Kivu (RDC).

<sup>5</sup>Département d'Informatique de Gestion de l'Université Presbytérienne Shepard et Lapsley du Congo (UPRECO), Kananga, Kasai Central (RDC).

**Abstract:** *This study aims to model and implement a decision-making system within the National Insurance Company (SONAS) in order to facilitate the management of its personnel and optimal decision-making. The main objective of this study is to assist the SONAS authorities in making ideal decisions and to enable them to have a clear vision of the evolution of the company and its influence on its staff in terms of performance ... To achieve this objective, we used a Data Warehouse based on the PERT method specialised in task scheduling in order to filter, sort and analyse the data related to the operation of SONAS in the DRC. On the other hand, we have used the Unified Modeling Language (UML) to model our system, while the implementation of the system has been carried out using the open source software Power BI and PDI (Pentaho Data Integration); and the data source has been designed using PostgreSQL software. Ultimately, this study envisages a continuous updating of the system in relation to the change and evolution of the company's activities.*

**Keywords:** *Modelling, Implementation, System, Decision Support, Management, Personnel, Enterprise, Expertise, Insurance, Company, National, Security, DRC.*

## Modélisation et implémentation d'un système décisionnel pour la gestion du personnel au sein d'une entreprise d'expertise et d'assurance : Cas de la Société Nationale de Sécurité.

### Résumé

La présente étude vise à modéliser et à mettre en place un système décisionnel au sein de la Société Nationale d'Assurance (SONAS, en sigle) afin de faciliter la gestion de son personnel ainsi qu'une prise de décision optimale. L'objectif principal de cette étude est d'assister les autorités de la SONAS à la prise de décision idéale ainsi que de permettre à ces dernières (Autorités) d'avoir une vision claire et nette de l'évolution de l'entreprise et de son influence exercée sur son personnel en termes des rendements ... Pour parvenir à l'objectif poursuivi, nous nous sommes servis d'une mise en place d'un Data Warehouse qui sera basé sur la méthode PERT spécialisée dans l'ordonnancement des tâches afin de permettre de filtrer, trier et analyser les données en rapport avec le fonctionnement de la SONAS en RDC. D'un autre côté, nous avons utilisé le langage UML (*Unified Modeling Language*) pour modéliser notre système, tandis que l'implémentation du système a été réalisée grâce au logiciel Open Source Power BI et PDI (*Pentaho Data Integration*); et la source de données a été conçue grâce au logiciel PostgreSQL. En définitive, la présente étude envisage une mise à jour continue du système relativement au changement et à l'évolution des activités de l'entreprise.

**Mots-clés :** *Modélisation, Implémentation, Système, Décisionnel, Gestion, Personnel, Entreprise, Expertise, Assurance, Société, Nationale, Sécurité, RDC.*

<sup>1</sup> Computer Science Department of the Bas-Uélé University (UNIBAS), Buta, Bas-Uélé (DRC).

<sup>2</sup> Computer Science Department of Saint Laurent University of Kananga (USLKA), Kananga, Kasai Central (DRC).

<sup>3</sup> Computer Science Department of the Higher Technical Institute of Kananga (ISTKA), Kasai Central (DRC).

<sup>4</sup> Computer Science Department of the BAKHITA African Schools of Butembo, North Kivu (DRC).

<sup>5</sup> Computer Sciences Department of the Shepard and Lapsley Presbyterian University of Congo (UPRECO), Kananga, (DRC).

## I. INTRODUCTION

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 au déploiement d'un système décisionnel pour la Société Nationale d'Assurance Congolaise (SONAS) 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.

En fait, l'informatique qui s'est déjà glissée de façon imperceptible dans la quasi-totalité de nos activités quotidiennes n'a pas encore son mot d'ordre dans les structures d'assurance de la RDC et pourtant, les sociétés nationales congolaises d'assurance ne se soucient pas vraiment des outils de NTIC en adoptant le système décisionnel alors qu'au sein de ces sociétés, c'est l'information qui est la pièce maîtresse. Soucieux de cela, il y a lieu de se poser les questions telles : Quel type d'outil informatique peut-on développer afin de gérer cette société d'expertise et d'assurance pour la prise de décision ? ; Quel peut être l'avantage d'un tel outil dans cette gestion ? Pour répondre anticipativement à ce questionnement, nous avons pensé que : **Déploiement d'un système décisionnel** serait le choix idéal pour cette gestion de l'entreprise d'expertise et d'assurance vu que le système décisionnel semble être un outil très puissant dans la prise de décision. Aussi est-il que le web offrirait des avantages incommensurables, vu qu'il fonctionne en le processus OLAP; une possibilité qui assure l'accès facile aux données dans la prise de décision au sein de cette société. Dans la même optique, rappelons que la présente recherche s'est assigné les objectifs suivants :

- 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 ;
- 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 logique business souhaité.

## II. DEMARCHE METHODOLOGIQUE

Cette partie nous serons obligés de vous présenter les méthodes, les approches et les techniques que nous avons utilisées pour la réalisation de ce modeste travail. Il s'agit en fait, de faire l'optimisation du délai de réalisation du projet et d'évaluer son coût financier par la même occasion.

### 1. Le langage UML

Comme nous le savons UML est un langage utilisé dans la conception de système d'information, ce pourquoi nous l'avons choisi comme méthode conceptuelle à travers ses multiples modèles pour la conception de notre système d'information de ce travail. Le langage UML nous a permis d'analyser et de concevoir le système d'information. Il s'est agi essentiellement d'identifier les différents besoins (besoins fonctionnels et besoins non fonctionnels), de recueillir les besoins utilisateurs via le diagramme de cas d'utilisation. En fait, le recours à la modélisation est depuis longtemps une pratique indispensable au développement, car un modèle est prévu pour anticiper les résultats du développement. Ce modèle est en effet une abstraction du résultat, et va nous permettre de documenter, de prévoir, d'étudier, de collecter ou d'estimer les informations. Avec UML, nous allons arriver à modéliser par graphique et texte destiné, à comprendre et décrire des besoins, à spécifier et à documenter des systèmes, à esquisser des architectures logicielles, à concevoir des solutions et communiquer des points de vue.

### 2. La méthode PERT

La méthode PERT, acronyme de « Program Evaluation and Review Technique », est une technique d'ordonnement des tâches utilisées pour gérer les projets. La méthode PERT permet d'obtenir un ordonnancement optimum des tâches les unes par rapport

aux autres pour minimiser la durée totale du projet. Signalons qu'il s'appuie sur la notion de graphe des antécédents et permet d'inclure dans la planification le risque et l'incertitude attachée à chaque tâche ... c'est aussi une méthode de gestion de projet visant à prévoir les propriétés d'un projet en termes de temps et de délais. La méthode PERT nous a permis de décomposer un projet en un ensemble d'actions appelés tâches et de les représenter sous forme graphique selon un graphe de dépendance. En effet, grâce à la chronologie et à l'interdépendance de chacune des tâches, on va pouvoir structurer l'ensemble du projet et ainsi, planifier la réalisation de chacune des tâches les unes par rapport aux autres, afin de minimiser les délais ainsi que réduire l'impact des retards lors de l'exécution des différentes tâches.

### 3. Le prototypage

Un prototype est un petit programme représentant ou réalisant les principaux concepts, caractéristiques et réalisation d'un modèle. Le prototype valide et termine éventuellement l'étape de la conception générique. Cette phase va consister alors à donner un prototype (une maquette) de notre modèle à partir duquel nous allons démontrer que le modèle pourra être implémenté et qu'il sera correct. Voilà pourquoi, il nous faut avoir un style qui traitera de la manière dont les solutions aux problèmes ou les algorithmes seront formulés. C'est ainsi que nous nous sommes fixé la programmation Orientée Objet, toujours couplée avec le langage UML, comme paradigme de programmation c'est-à-dire une vue de l'exécution du programme que nous aurons à exploiter tout au long de l'activité de programmation.

### 4. Planning prévisionnel de notre projet

Un projet est un ensemble d'action à réaliser pour pouvoir atteindre un objectif défini dans le cadre d'une mission précise et pour la réalisation desquels on a identifié un début et une fin. Il est nécessaire de mettre en place un planning du travail à réaliser afin d'optimiser la durée du projet. Signalons que cela se passe avant tout par la décomposition du projet en tâche. C'est ainsi que dans la rubrique ci-dessous, nous allons décomposer notre projet en différentes tâches.

#### a) Détermination des tâches

La détermination des tâches est la première phase pour établir un réseau PERT, elle consiste à identifier les tâches nécessaires à la réalisation d'un projet. C'est en quelque sorte un listing des tâches. Pour chaque tâche, on associe une durée estimée dans une unité de temps. Ainsi, le tableau suivant illustre les tâches de notre projet.

Tâches	Indices	Durée en jours
Etude de faisabilité	A	30
Analyse des besoins	B	5
Conception du système	C	20
Implémentation	D	160
Test du site	E	20
Intégration	F	40
<b>Total des jours du projet</b>		<b>275</b>

#### b) Détermination des antériorités

Dans cette phase nous nous posons la question suivante : Quelles sont les tâches devant être terminées pour pouvoir commencer la nouvelle tâche ? Elle consiste, en fait, à déterminer les antériorités des tâches précédemment établies. Les tâches A et B ne nécessitent aucune condition préalable pour être effectuées. Elles ne possèdent pas d'antériorité. La tâche C nécessite que la tâche A soit réalisée. La tâche C possède la tâche B comme antériorité. La tâche D nécessite le finissage des tâches C et B. La tâche D possède les tâches B et C comme antériorités. La tâche E nécessite le finissage de la tâche D. La tâche E possède D comme antériorité. La tâche F nécessite le finissage de la tâche E.

Indices	Durée en jours	Antériorités
A	30	Aucune
B	5	Aucune
C	20	A,B
D	160	B,C
E	20	D

F

40

E

## c) Détermination des postériorités

Dans le tableau ci-dessous, nous présentons les tâches postérieures de nos tâches :

Indices	Durée en jours	Antériorités	Postériorités
A	30	Aucune	C
B	5	Aucune	C,D
C	20	A,B	D
D	160	B,C	E
E	20	D	F
F	40	E	Aucune

## d) Calcul des niveaux

La détermination des niveaux permet facilement la mise en évidence des liaisons qui existent entre différentes tâches ainsi que les réalisations de tâches antérieures.

Indices	Durée en jours	Antériorités	Postériorités	Niveau
A	30	Aucune	C	1
B	5	Aucune	C,D	1
C	20	A,B	D	2
D	160	B,C	E	3
E	20	D	F	4
F	40	E	Aucune	5

## e) Construction du diagramme PERT

Le PERT permet de représenter l'ensemble des tâches sur un graphe orienté, à partir duquel il sera possible d'identifier leurs dates au plus tôt et au plus tard et de calculer leurs marges. Un graphe orienté est un réseau composé d'une entrée et d'une sortie, ainsi que de points (appelés « sommets ») reliés entre eux par des flèches (appelées « arcs »).

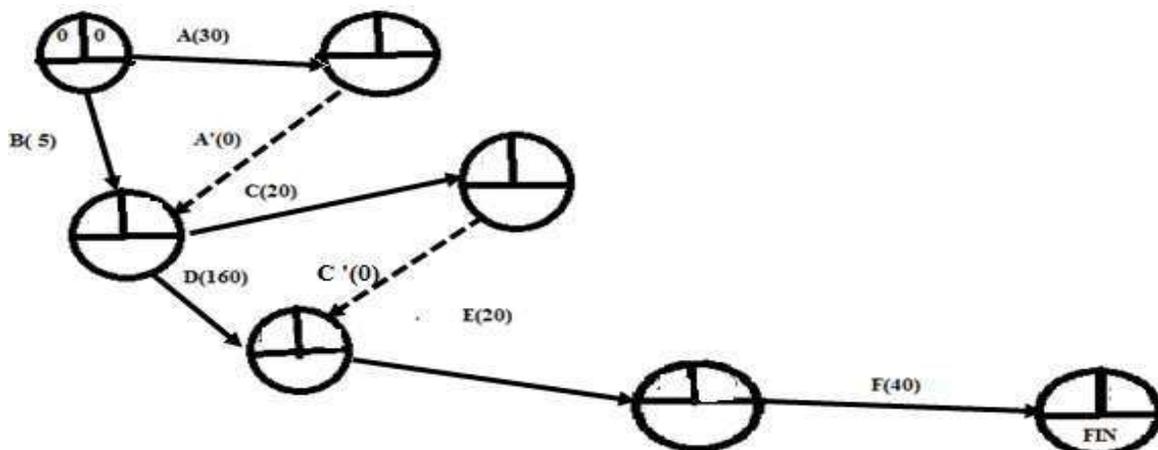
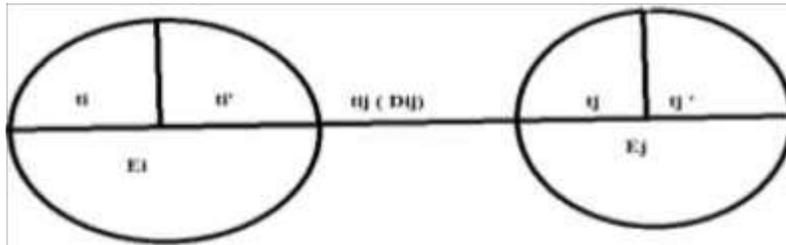


Figure 1. Représentation du réseau PERT.

## f) Calcul des dates

La date au plus tôt d'un réseau PERT correspond à la date à laquelle une étape peut être atteinte au plus tôt. Elle s'obtient en ajoutant à la date au plus tôt de l'étape précédente, la durée de la tâche qui les sépare. Tandis que la date au plus tard d'un réseau PERT correspond à la date à laquelle une étape doit être atteinte au plus tard pour que la durée globale du projet reste minimum. Elle s'obtient

en retirant de la date au plus tard de l'étape qui lui succède la durée de la tâche qui les relie. Et, soit le graphe ci-dessous, la date au plus tôt est calculée avec la formule :  $t_j = \{\max (t_i + d_{ij})\}$  et la date au plus tard avec la formule :  $t_i = \{\min (t_i - d_{ij})\}$ .



Ce qui nous donne les résultats suivants :

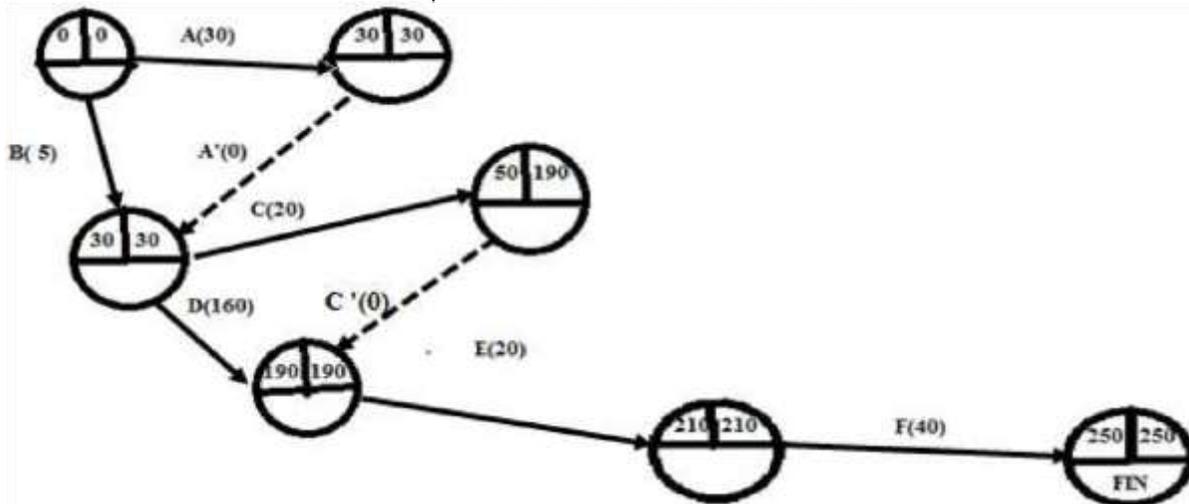


Figure 2. Représentation des dates

**g) Calcul des marges**

Une marge est la possibilité qu'a une tâche d'être retardée sans impacter le projet. Nous distinguons deux types de marge : la marge libre et la marge totale. La marge totale est le retard maximal de la tâche ne perturbant pas la fin du projet. Tandis que la marge libre est le retard maximal ne perturbant pas la date au plus tôt de la tâche suivante.

**[1]. Les marges Totales**

Pour calculer les marges totales, on utilise la formule suivante :  $MT (t_{ij}) = t'_j - t'_i - d_{ij}$ , Ce qui donne :

- $MT(A) = 30 - 0 - 30 = 0$
- $MT(B) = 30 - 0 - 30 = 0$
- $MT(A') = 30 - 30 - 0 = 0$
- $MT(C) = 190 - 30 - 20 = 140$
- $MT(D) = 190 - 30 - 160 = 0$
- $MT(C') = 190 - 190 - 0 = 0$
- $MT(E) = 210 - 190 - 20 = 0$
- $MT(F) = 250 - 210 - 40 = 0$

Les marges libres sont calculées par la formule :  $ML (t_{ij}) = t'_j - t_i - d_{ij}$ , Ainsi :

- $ML(B) = 30 - 0 - 5 = 25$
- $ML(A') = 30 - 30 - 0 = 0$
- $ML(C) = 190 - 30 - 20 = 140$
- $ML(D) = 190 - 30 - 160 = 0$
- $ML(C') = 190 - 50 - 0 = 140$
- $ML(E) = 210 - 190 - 20 = 0$
- $ML(F) = 250 - 210 - 40 = 0$

$ML(A) = 30 - 0 - 30 = 0$

**h) Détermination du chemin critique**

Un chemin critique est l'ensemble des tâches critiques. C'est aussi l'ensemble des tâches dont la date au plus tôt est égale à la date au plus tard, c'est le chemin le plus long du projet. Ainsi, le graphe suivant illustre les tâches qui constituent notre chemin critique. Le chemin critique est donc, pour notre projet : A, A', D, E, F.

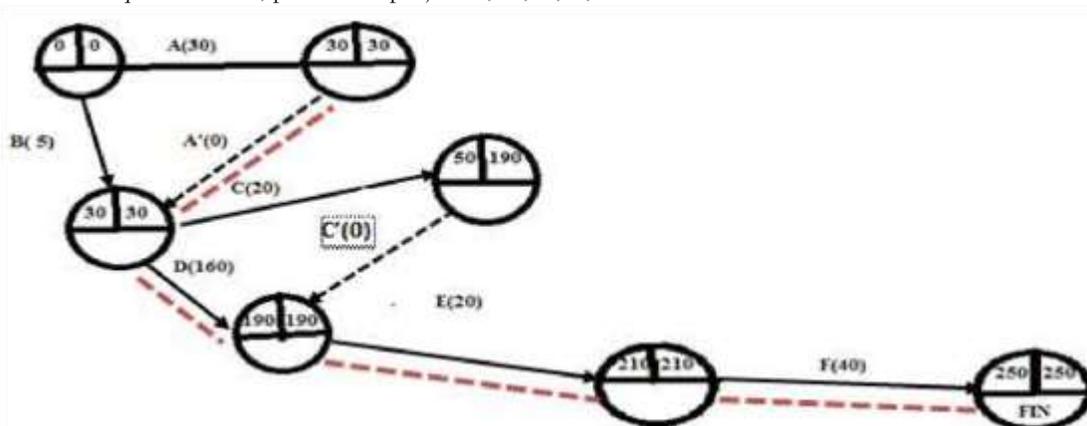


Figure 3.

Représentation du chemin critique

### 5. Evaluation du coût du projet

La mise en œuvre de tout projet exige qu'on évalue les besoins en termes de cout et de matériels pour assurer un bon déroulement et une bonne fin. Cette gestion des coûts correspond à la définition, au contrôle et à l'ajustement éventuel du budget du projet.

#### a) Evaluation du coût de développement

Tout en prenant en compte l'existence des plusieurs méthodes d'estimation du cout d'un projet, nous avons jugé bon utiliser la méthode COCOMO. En effet, COCOMO permet d'obtenir la charge de la réalisation en mois-personne ainsi que le délai normal recommandé. C'est le modèle de construction des coûts. Cela nous indique la taille moyenne de l'équipe qui est égale à charge/délai. Estimant les lignes de ce projet à environ quatre milles, cela nous pousse à dire que c'est un projet simple. Ce qui exige d'utiliser les formules ci-dessous : la charge =  $2,4 \times (KISL) \times 1,05$  et le délai =  $2,5 \times (charge) \times 0,38$ .

Le KISL correspond au nombre de lignes des codes du logiciel/1000. La taille moyenne de l'équipe se calcule par la formule :  $N = Charge / Délai$ . Ainsi, le coût total de développement est égal à la charge en mois-personne multiplié par la valeur mois-personne, c'est-à-dire coût = Charge \* Valeur MP. Valeur MP représente le salaire moyen d'un informaticien en RDC. Nous avons opté pour le salaire de 500USD.

$$Kisl = 4000/1000 = 4. \quad Charge = 2,4 (4) 1,05 = 10 \text{ mois-personnes} \quad Délai = 2,5 (10)0,38 = 5,99 \text{ soit } 6 \text{ mois} \quad N = charge / délai = 10/6 = 2 \text{ personnes} \quad Coût = 10*500 = 5000USD.$$

#### b) Evaluation du coût des matériels à acquérir

Matériels	Nombres	PU	PT
Logiciels	3	25\$	75\$
Ordinateur (HDD : 500Go, RAM : 4Go, Processeur : 4GHZ)	1	650\$	650\$
Fournisseur internet	12 mois	100\$	1200\$

Tableau 5. Tableau d'évaluation des coûts des matériels.

#### c) Coût total du projet

Désignation	Coût en USD
Coût de développement	5000\$
Coût des matériels	1705\$
Coût total	6705\$

Tableau 6. Tableau du coût total du projet.

### III. APPROCHE CONCEPTUELLE

#### 1. Définition des concepts clés

##### a) Information

Une information est une donnée ou un ensemble de données qui a ou ont été interprétée(s). Elle est davantage une interprétation de données dans un contexte particulier. Sa définition fait référence à la signification d'une donnée, mais plus généralement d'un ensemble de données c'est-à-dire à la sémantique, comme le montre la définition précédente. L'information amène à la décision. Nous dirons que l'information induit assez souvent une action, soit une décision. Par analogie, nous pouvons dire que les données constituent la matière brute à partir de laquelle l'information est produite.

##### b) Donnée

Une donnée est une représentation d'un élément d'information, tel qu'un chiffre ou un fait, codé dans un format permettant son stockage et son traitement par ordinateur. Bref, une donnée est la description d'un élément ponctuel de la réalité, comme une mesure ou une observation.

##### c) Système d'information

Un système d'information de l'organisation comporte, outre le système informatique, un ensemble organisé d'autres ressources : humaines, organisationnelles et immatérielles, comme des méthodes, des règles, des procédures ... Le système d'information est destiné à faciliter le fonctionnement de l'organisation en lui fournissant les informations utiles pour atteindre ses objectifs.

##### d) Système Informatique

Un système informatique est un ensemble de moyens informatiques, matériels et logiciels, ayant pour finalité de collecter, de traiter, de stocker, d'acheminer et de présenter des données. Les concepts de système informatique et d'ordinateur ne doivent pas être confondus. L'ordinateur n'est que l'un des composants, certes central, des systèmes informatiques. Il existe beaucoup d'autres, parmi lesquels on peut citer les matériels réseaux, les capteurs et actionnaires, les machines spécialisées, comme les guichets automatiques bancaires ou boîtiers de stockage en réseau, les robots, les Smartphones, les cartes à puce (smartcards), etc.

##### e) 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.

##### f) Base de données (BD)

Une BD est une collection de fichiers reliés par les pointeurs multiples, aussi cohérents entre eux que possibles, organisés de manière à répondre efficacement à une grande variété de questions. Elle est aussi une collection d'informations modélisant le monde réel [4]. C'est ensemble structuré d'éléments d'information, souvent agencés sous forme de tables, dans lesquelles les données sont organisées selon certains critères en vue de permettre leur exploitation pour répondre aux besoins d'information d'une organisation.

##### g) 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.

##### h) 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.

- **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.
- **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.

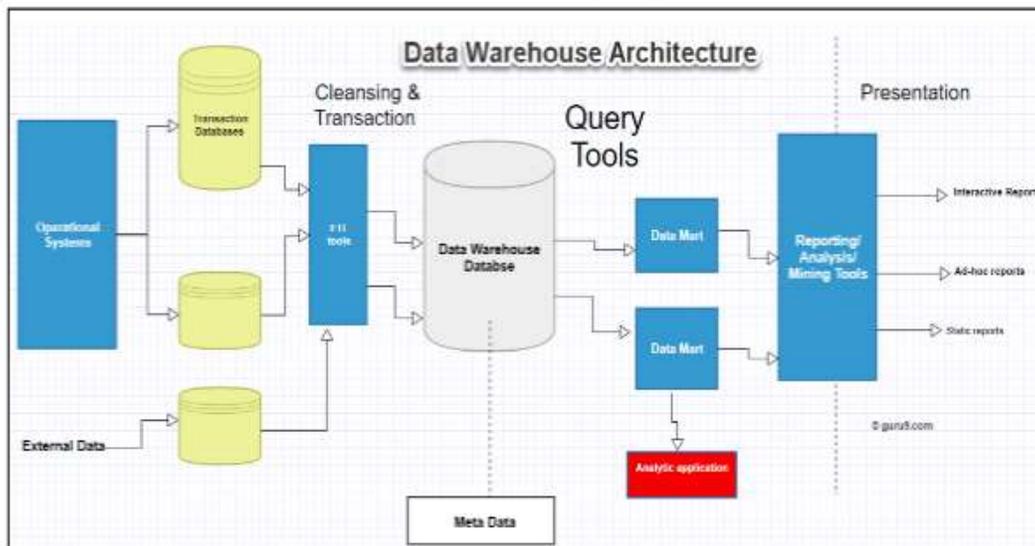
### i) La gestion

La gestion allie formalisations théoriques et pratiques opérationnelles pour piloter une entreprise ; c'est donc simultanément une science de choix et une science de l'action utilisant de nombreuses techniques pour aider à ces décisions. Ainsi, il existe plusieurs types de gestion tels que gestion de l'information du projet, gestion de la chaîne logistique, gestion de la configuration, etc. Pour un logiciel informatique, la gestion est l'organisation et l'actualisation (la mise à jour) de ce logiciel.

## 2. Paradigmes de la création d'un système décisionnel

Un système décisionnel est donc avant tout un moyen qui a pour but de faciliter la définition et la mise en œuvre de stratégies gagnantes. Mais il ne s'agit pas de définir une stratégie une fois pour toute, mais d'être à même de continuellement s'adapter à son environnement, et de le faire plus vite que ses concurrents. Pour cela, il convient de bien comprendre son environnement, d'ajuster ses interactions avec lui en faisant les meilleurs choix de cibles et d'actions. Concrètement le chemin à suivre peut être caractérisé par les quatre objectifs suivants : comprendre son environnement, se focaliser sur des cibles, aligner son organisation et mettre en œuvre les plans d'actions nécessaires.

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



Chaque environnement du Data Warehouse utilise une source de données qui est 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. En dehors de cette source de données il utilise encore les outils d'extraction, de transformation et de chargement (ETL), ces outils effectuent 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 [30].

### Les fonctionnalités de ces outils 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.

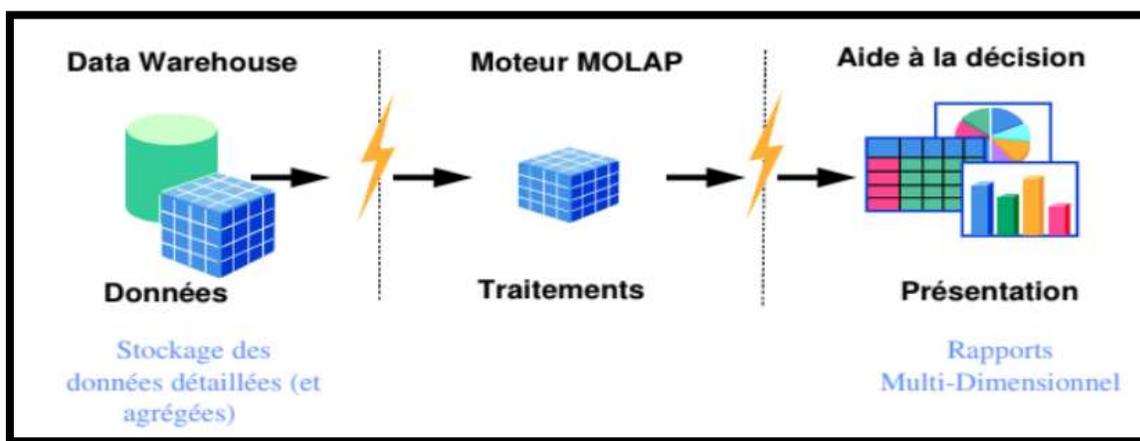
## Le concept OLAP

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.

### ➤ Modèles de stockage pour OLAP

Il existe plusieurs modèles de stockage de bases de données pour implémenter OLAP, on peut citer par exemple : **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.



En guise de conclusion nous allons utiliser les principaux outils d'une architecture décisionnelle suivante :

- ETL : Ils permettent de concevoir et d'organiser les processus de migration du système transactionnel vers le système décisionnel.
- Outils de reporting Ils permettent : la création graphique de rapport et l'accès aux sources de données via des API dédiées.
- Outils d'exploration : Ils permettent de manipuler interactivement des cubes multidimensionnels (choix des dimensions à croiser et des types d'agrégations à effectuer) ;
- Outils d'analyse : Ils permettent l'analyse statistique de données.

En dehors des outils énumérés ci-haut pour le stockage de données nous allons utiliser le SGBD orientés décisionnel il est possible d'utiliser une base relationnelle classique pour implémenter un entrepôt de données modélisé en étoile (c'est même aujourd'hui encore la forme la plus largement mobilisée). Il existe également des technologies dédiées (qui peuvent s'appuyer sur des bases relationnelles ou sur des structures de données dédiées). Le mouvement NoSQL réintègre progressivement des problématiques décisionnelles, reconfigurant petit à petit les approches technologiques liées à ce domaine.

## IV. RESULTATS ET DISCUSSION

### IV.1. Etude des besoins et conception de la zone d'entreposage des données

#### 1. 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 [30].

#### 2. Démarche d'extraction des 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.
- 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.
- 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

### 3. 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.

### 4. 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.

- a. 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>	DATE	<pk>
mois	VARCHAR(6)	
annee	VARCHAR(4)	
decade	VARCHAR(8)	

Figure 3.1: Dimension Temps

- b. 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>	VARCHAR(15)	<pk>
wilaya	VARCHAR(25)	
nom_centre	VARCHAR(20)	
type_centre	VARCHAR(10)	

Figure 3.2: 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.

- c. 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>	VARCHAR(25)	<pk>
nom_agence	VARCHAR(25)	
nom_compagnie	VARCHAR(20)	
wilaya	VARCHAR(20)	

Figure 3.3: 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.

- d. 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>	VARCHAR(100)	<pk>
mat_expert	VARCHAR(25)	
nom_expert	VARCHAR(50)	

Figure 3.4: 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.

- e. 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 :

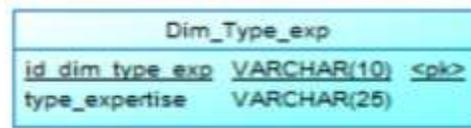


Figure 3. 5: Dimension Type Expertise

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.

### 5. 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.

- a. 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.

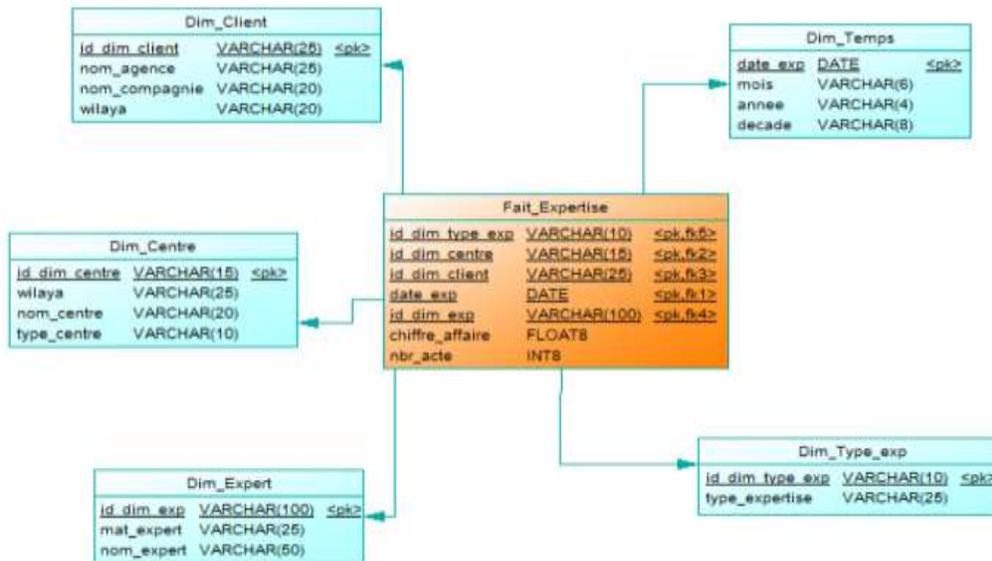


Figure 3. 6: Modèle en étoile Expertise

- b. Modèle en étoile des impayés Un PV peut être en trois statuts possibles (régulé, 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 :

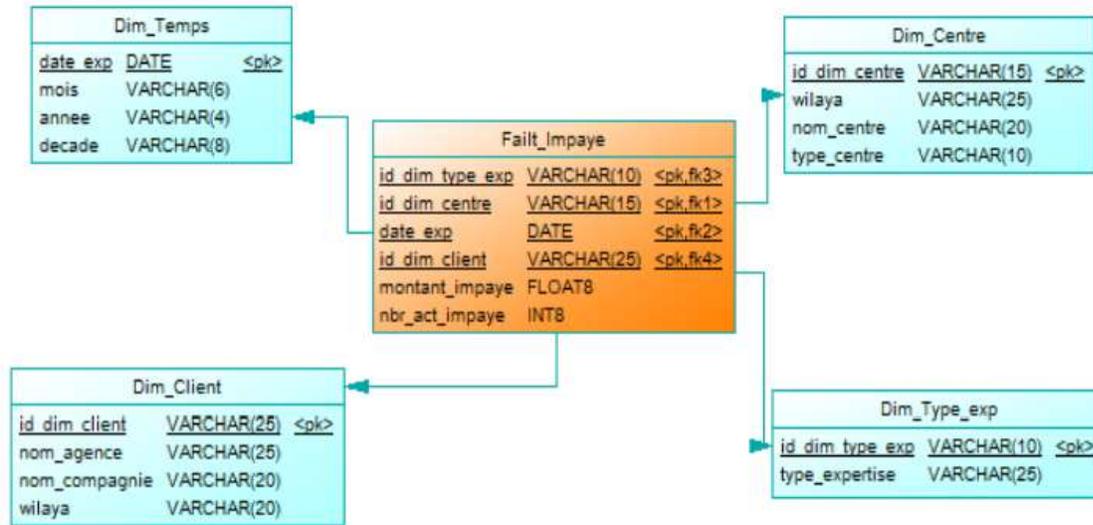


Figure 3.7 : Modèle en étoile Impayé

- c. 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 :

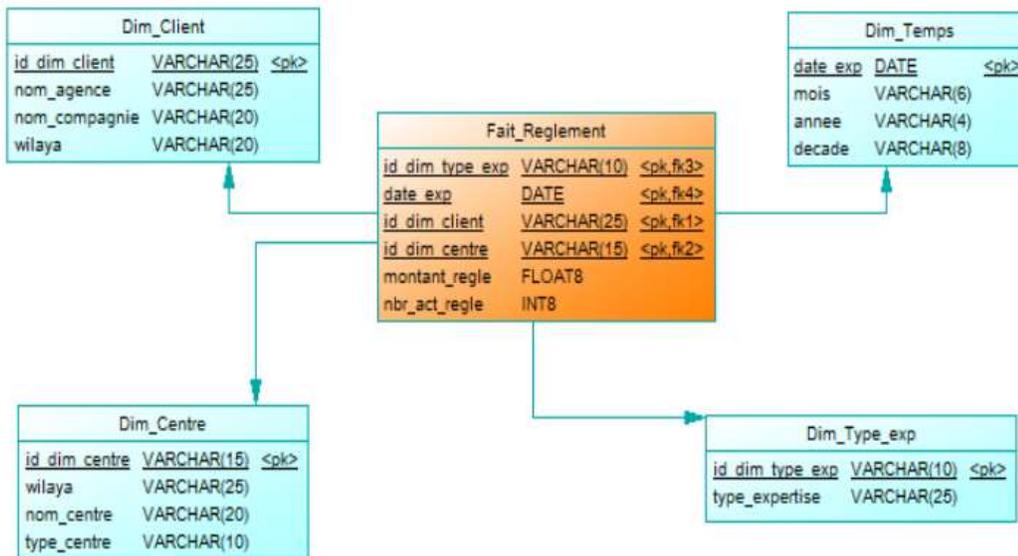


Figure 3.8: Modèle en étoile Règlement

- d. 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 :

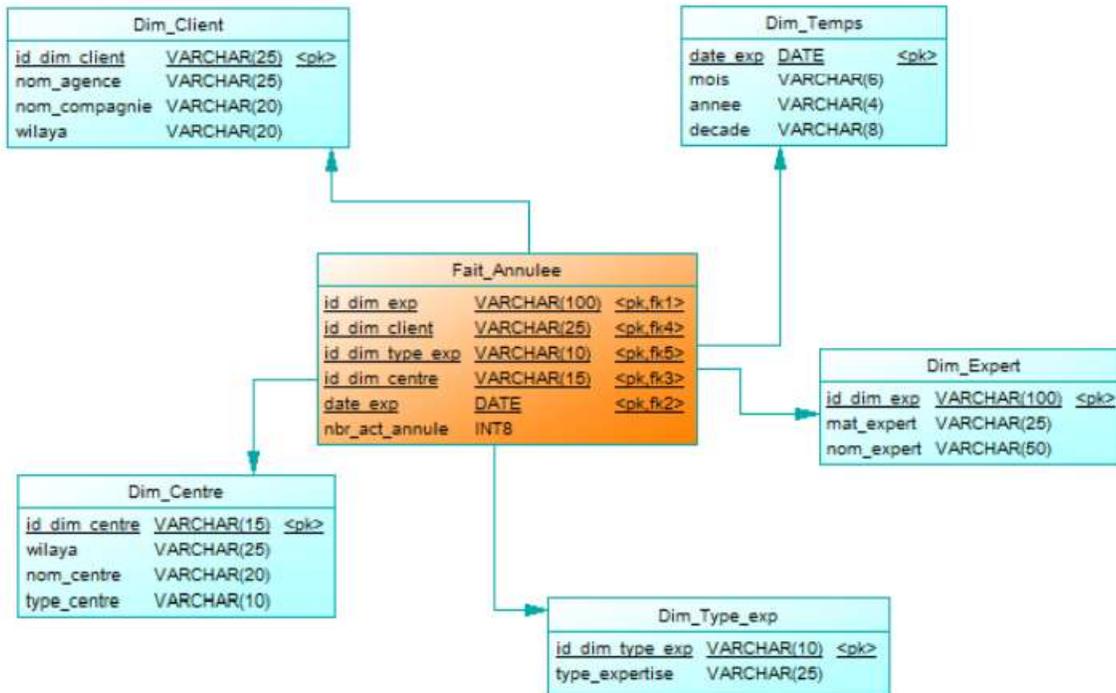


Figure 3.9: Modèle en étoile Annulé

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 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 donnée savoir le son schéma.

#### IV.2. Alimentation du data Warehouse

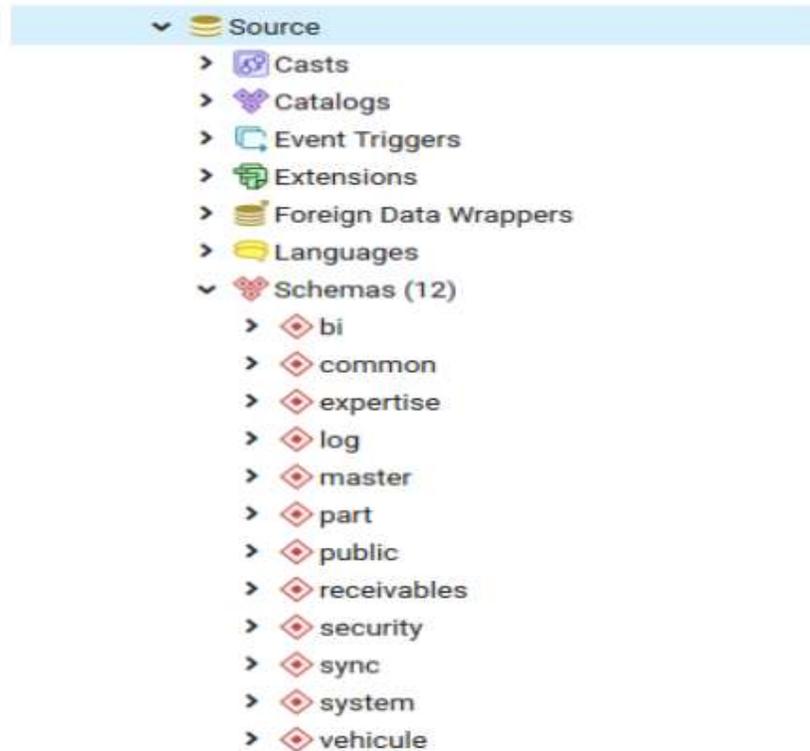
L'alimentation d'un entrepôt de données est étape très importante dans un projet d'un système décisionnel, elle représente 80% de la charge de travail. Cette étape se déroule essentiellement par la migration des données de la base de production vers l'entrepôt de données qui se fait à l'aide des outils ETL en passant par les différentes phases de nettoyage et de transformations nécessaires[30].

##### 1. Etude pré-alimentaire

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

###### a. L'étude de la source des données

Les sources de données de notre entrepôt sont les 32 bases de données PostgreSQL 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.

#### **b. 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 « receivables » contient des données sur les paiements, les versements ;
- Choisir la source fiable des données pour le chargement.

#### **c. Définir la périodicité de chargement des données**

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

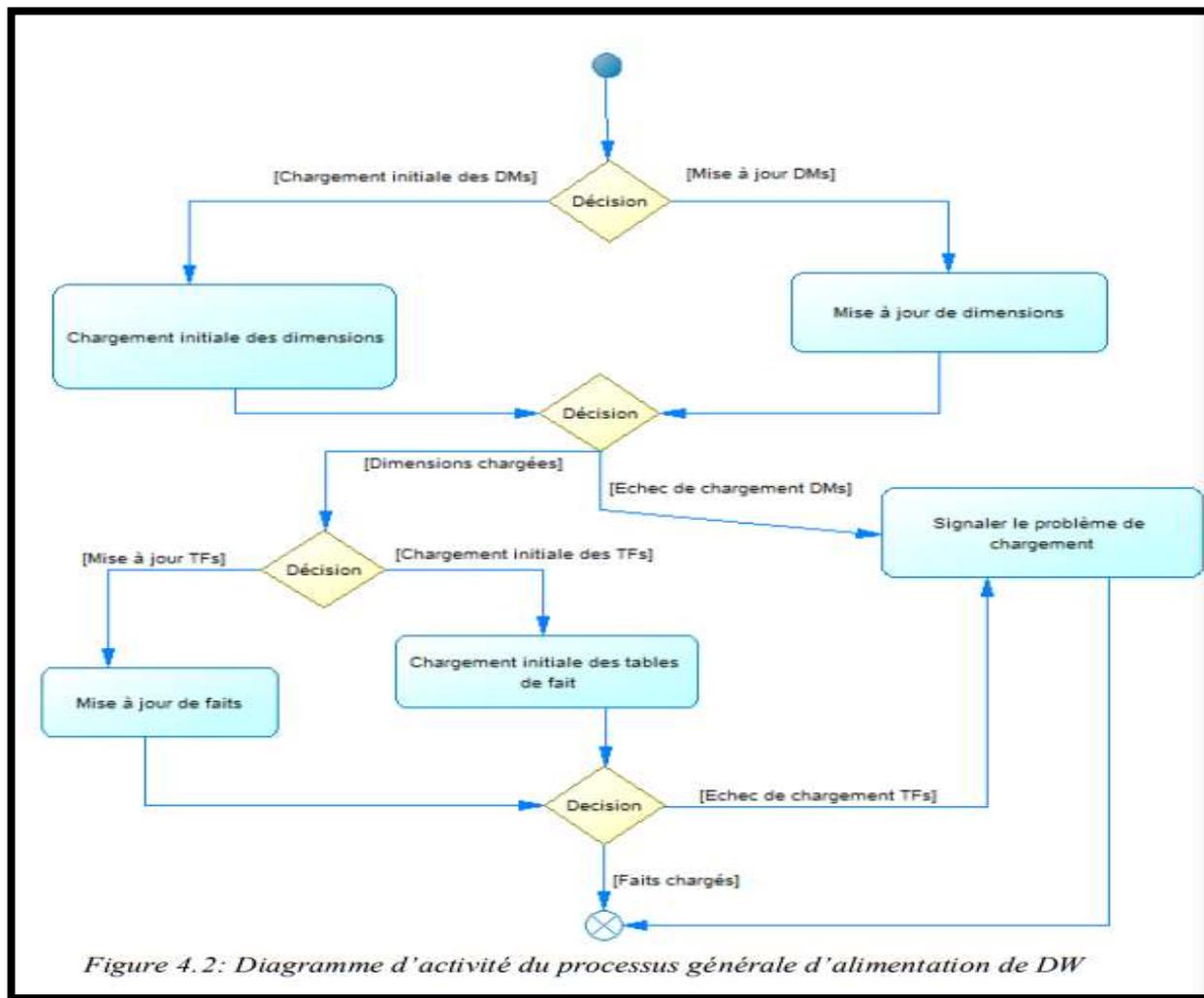
- La quantité des données à charger ;
- Le temps de non activité de système ;
- Chaque décade le système ouvert une nouvel session et ferme la précédente Après une étude de ses critères nous avons choisi de configurer l'ETL avec une périodicité journalière pendant les heures non activité sur le système (de dix-huit heures à sept heures du matin).

## **2. 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.

### 3. Processus de chargement

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



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.

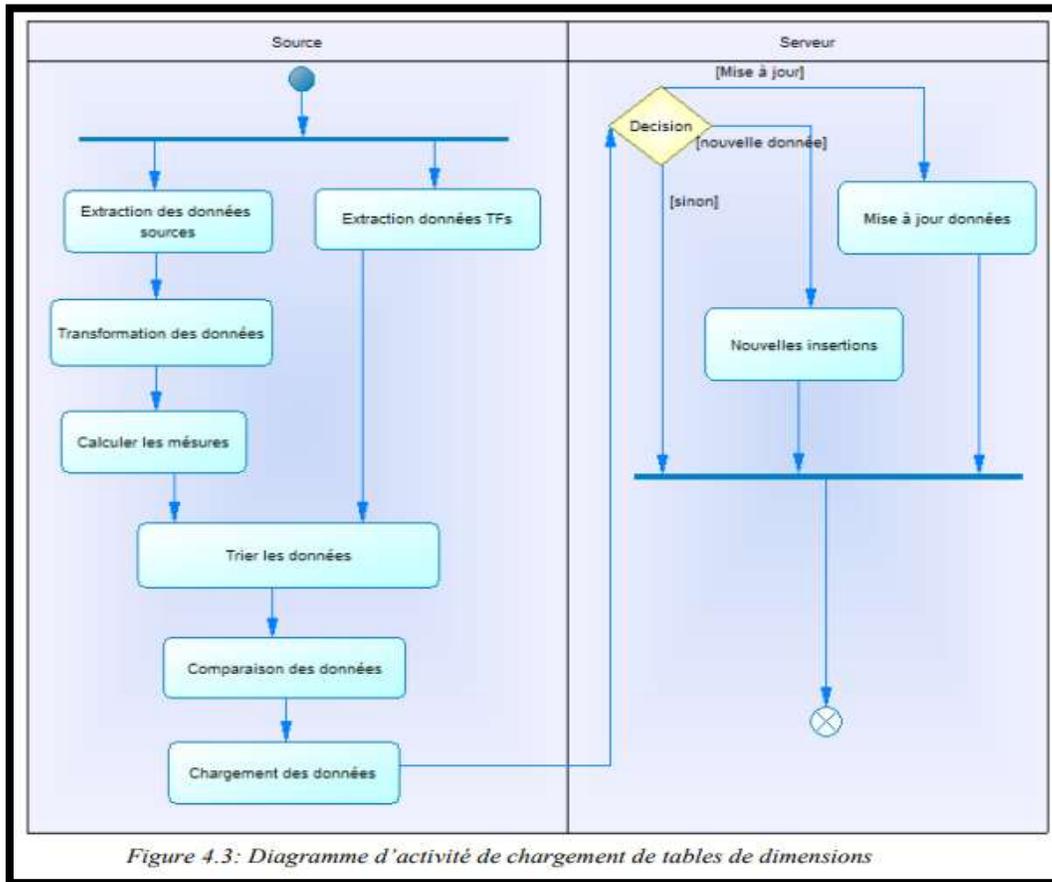
#### a. Processus de chargement des dimensions

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.



### b. 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 :

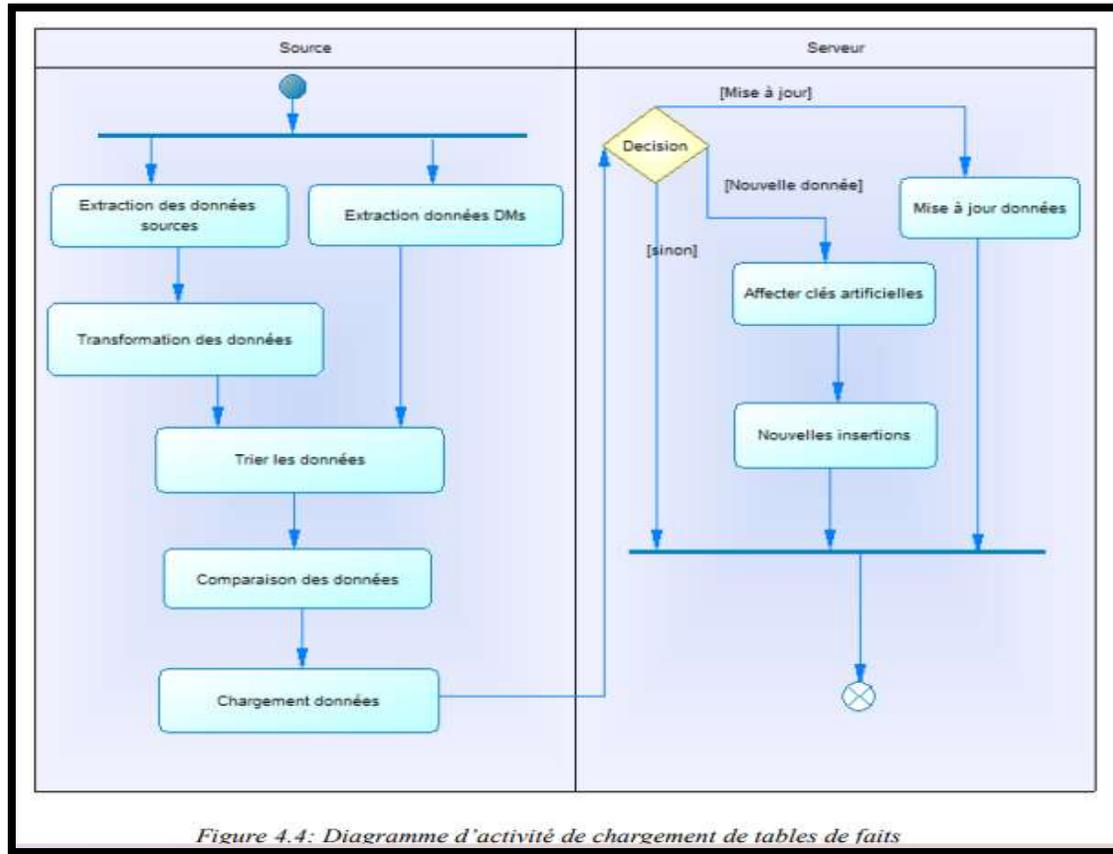


Figure 4.4: Diagramme d'activité de chargement de tables de faits

### c. 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]

## VI.3. Réalisation du projet

La réalisation de cette étude est consacrée à la présentation des distincts outils qui ont permis l'exécution de ce 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.

### a. Présentation Pentaho Data Integration (PDI)

Le 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.

### b. 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 :

- **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

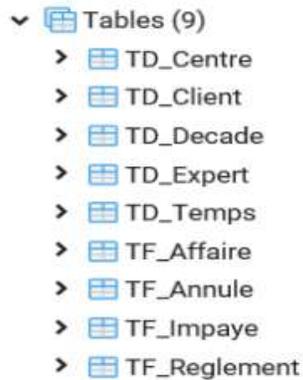


Figure 5.8 Le schéma de DW

- **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 qu'elle 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.

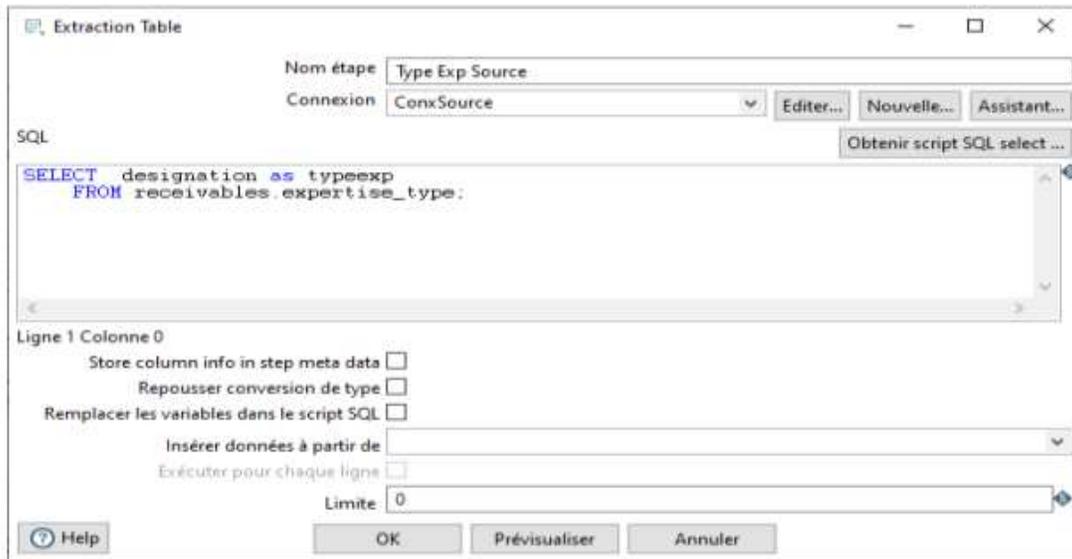


Figure 5.11: Paramétrage d'un composant ETL (extract)

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.

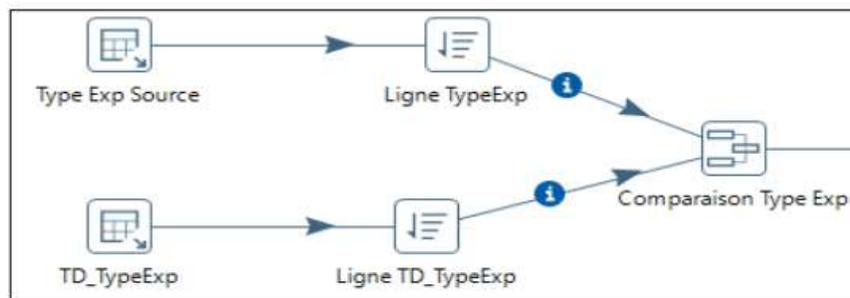


Figure 5.12: Comparaison entre les données de la dimension et les données sources

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.

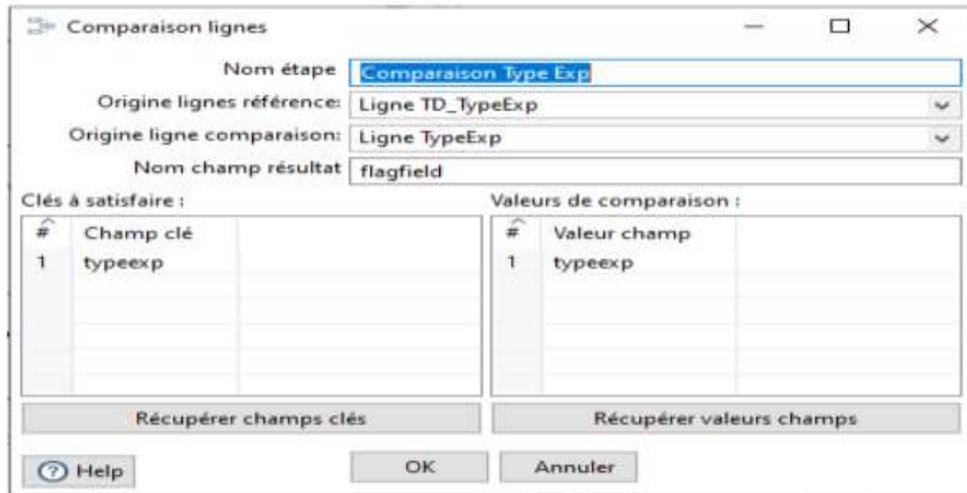


Figure 5.13: Paramètres de la comparaison

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.

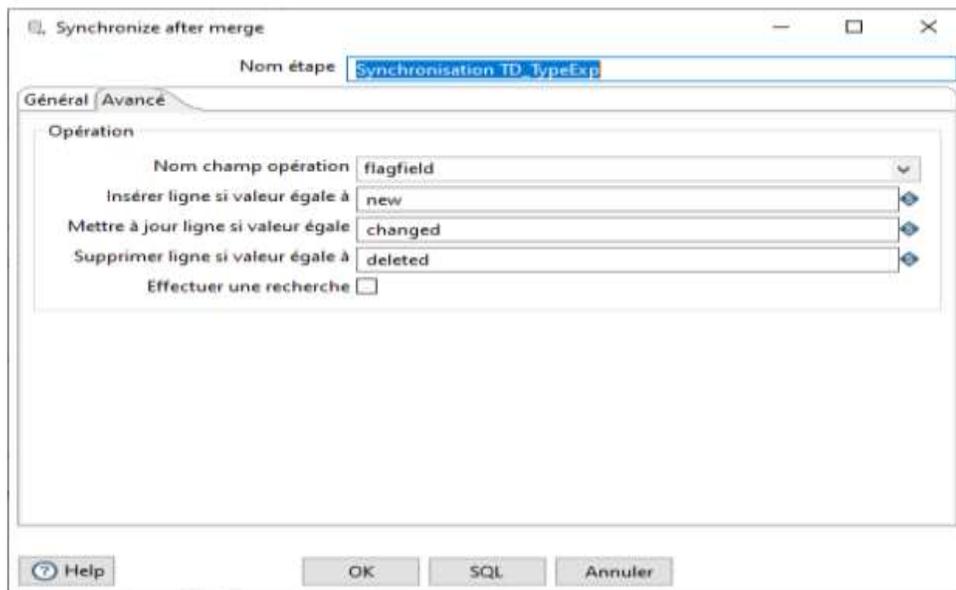


Figure 5.14: Paramètres avancé « synchronize »

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

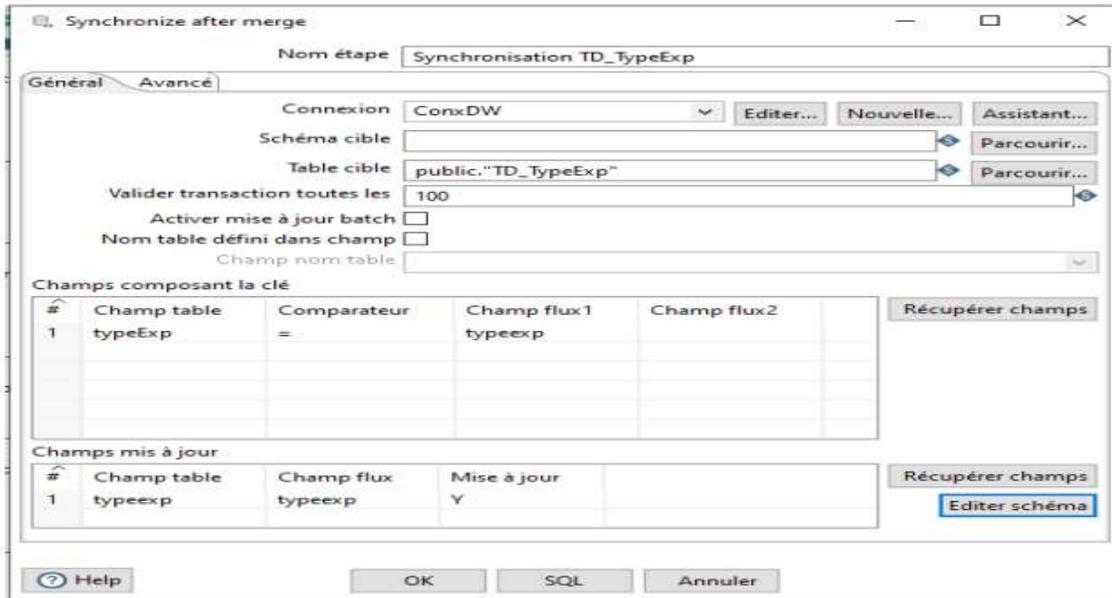


Figure 5.16: Paramètre générale

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

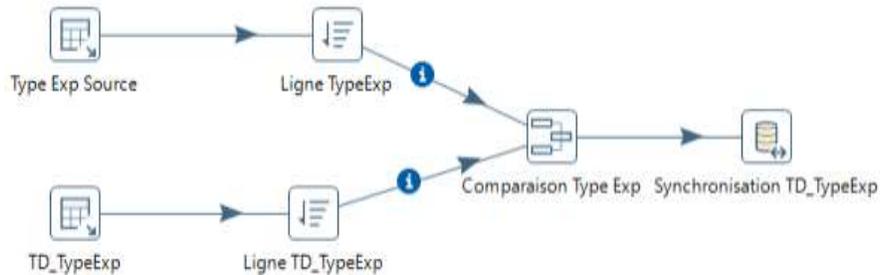


Figure 5.17: Exemple d'alimentation d'une table de dimension

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 :

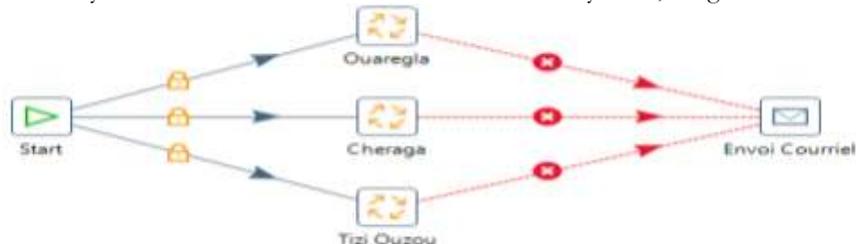


Figure 5.20: Exemple d'une tâche globale

➤ **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.

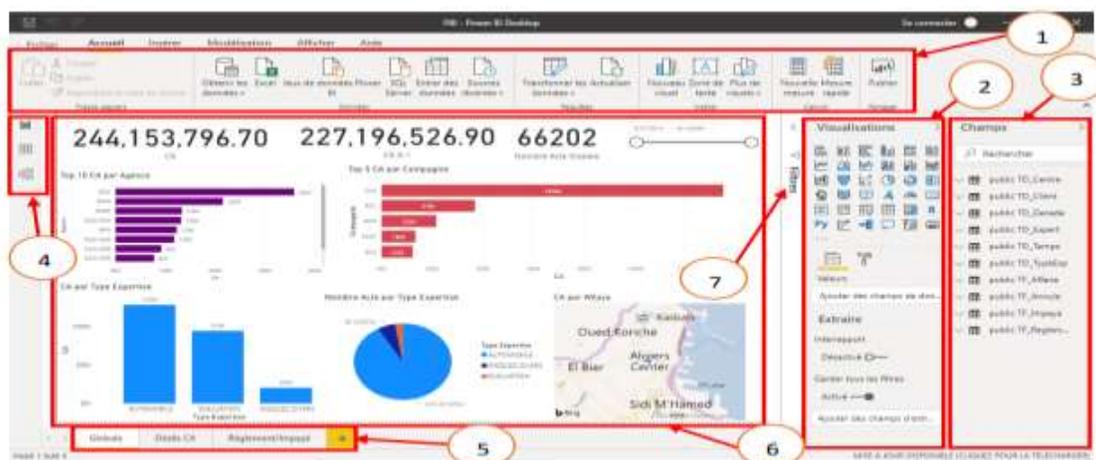


Figure 5.21: Fenêtre de PowerBI

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 qui porte sur des étiquettes sur les chiffres généraux de l'entreprise comme chiffre d'affaire et le nombre globale par tous les axes d'analyse, la figure ci-dessous illustre cela.

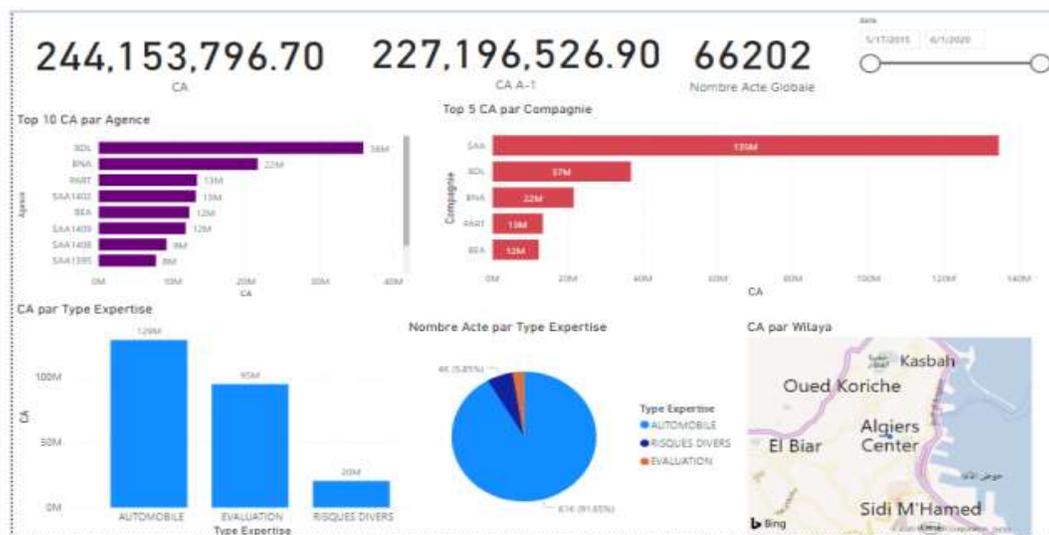


Figure 5.24: Tableau de bord globale

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 :

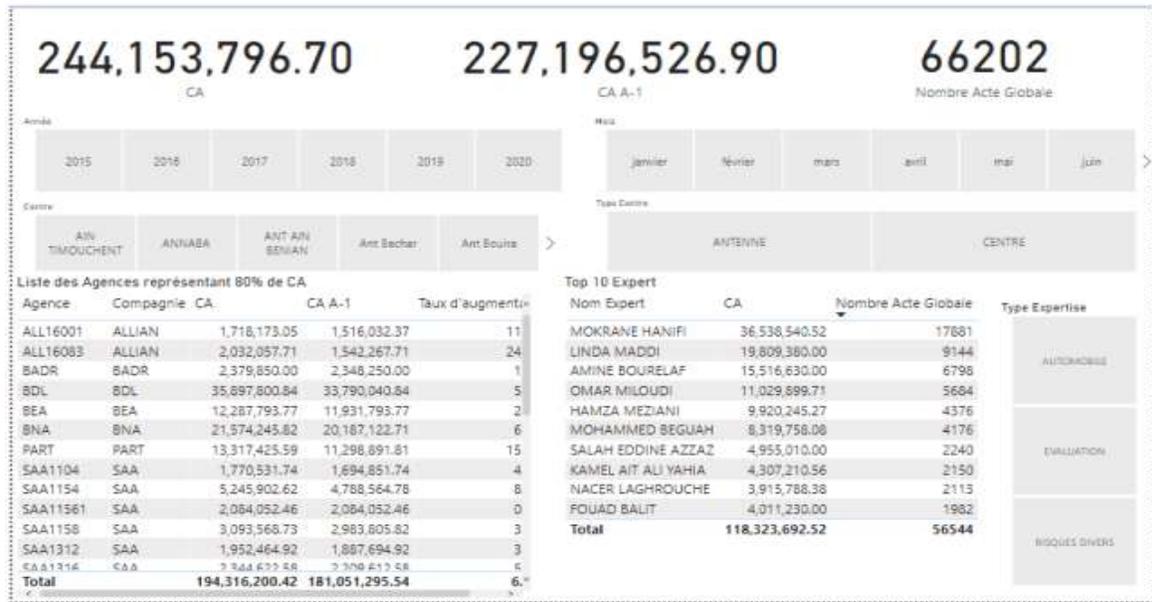


Figure 5.25: Tableau de bord détail fait Expertise

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 détail paiement et annulé

- **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 :

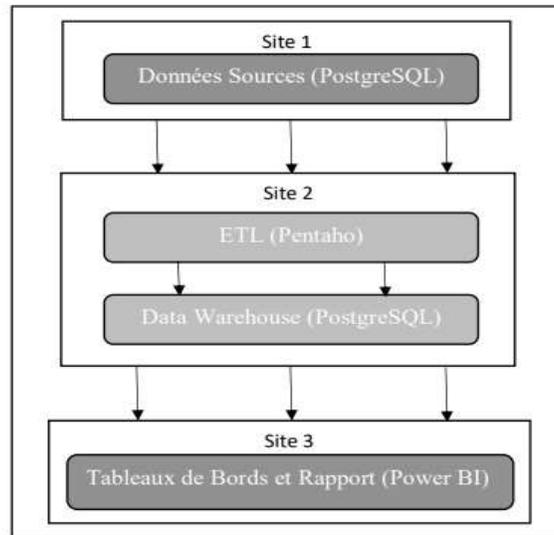


Figure 5. 27 Architecture de projet

### Conclusion

A travers cette étude, nous avons tenté d'expliquer les principes fondamentaux de la conception et la mise en place d'une Data Warehouse afin de permettre aux décideurs de la SONAS de filtrer et d'analyser les informations nécessaires à la prise de décision et remplir le vide de l'existant décisionnelle au sein de cette entreprise. Nous avons utilisé des moyens accessibles pour réaliser ce projet notamment les logiciel open source comme PDI, POWER BI, etc. Au regard à venir, nous projetons de créer un portail web (Power BI service) pour une meilleure exploitation des rapports à partir d'une application mobile ; Implémenter le projet au sein d'un serveur linux haute gamme ; Connecter la Data Warehouse aux 32 bases de données source de la SONAS en RDC ; et enfin, faire des mise à jours continue au système relativement au changement de cette entreprise.

### Références bibliographiques

- [1] 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>. Février 2020.
- [2] A. Vaisman et E. Zim'anyi, « *Data warehouse systems design and implementation* », Srpinger, 2014.
- [3] B. Yellin, « *Data Warehouse Fundamentals for Storage Professionals-What You Need To Know* », EMC<sup>2</sup>, 2011.
- [4] C. Imhoff, N. Galleo and J. G. Geiger, « *Mastering data warehouse design relational and dimentional techniques* », Wiley, 2003.
- [5] Cartelis, « *Architecture Data Warehouse – Approches Traditionnelle vs Clouds* », Avril 2018. [En ligne]. Available: <https://www.cartelis.com/blog/architecture-data-warehouse/>. Février 2020.
- [6] D. Jones, « *Guide de mise en œuvre de la Business Intelligence au sein des PME* », Realtime Publishers, 2016
- [7] Diffen, « *Snowflake Schema vs. Star Schema* », [En ligne]. Available : [https://www.diffen.com/difference/Snowflake\\_Schema\\_vs\\_Star\\_Schema](https://www.diffen.com/difference/Snowflake_Schema_vs_Star_Schema). [Accès le Février 2020].
- [8] F. ABDERRAHMANE et K. SOFIANE, « *Conception et réalisation d'un Data Warehouse pour la mise en place d'un système décisionnel* », Alger, 2010.
- [9] F. Almeida, « *Concepts and Fundaments of Data Warehousing and OLAP* », 2017.
- [10] H. Hasan et P. Hyland, « *Using OLAP and Multidimensional Data for Decision Making* », ITProfessional, vol. 3, pp. 44-50, Septembre 2001.
- [11] H. Vantara, « *Pentaho documentation* », [En ligne]. Available: <https://help.pentaho.com/Documentation>. [Accès le Juillet 2020].
- [12] JavaTPoint, « *OLAP Operations in the Multidimensional Data Model* », [En ligne]. Available: <https://www.javatpoint.com/olap-operations>. [Accès le Février 2020].
- [13] L. Matthieu, « *Entrepôt de données : De l'alimentation des données au décisionnel de l'établissement* », HAL, 2012.

- [14] M. Bala, « *Chapitre 2 : Modélisation dimensionnelle* », 2020.
- [15] 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/>. Février 2020.
- [16] MicroSoft, « *Power BI Documentation* », [En ligne]. Available: <https://docs.microsoft.com/frfr/power-bi/>. [Accès le Juillet 2020].
- [17] N. B. Darmawan, « *nicobudidarmawan* », 03 Janvier 2014. [En ligne]. Available: <http://www.nicobudidarmawan.com/2014/01/business-intelligence-oltp-vs-olap.html>. Février 2020.
- [18] P. Lane, “*Data Warehousing Guide*”, Oracle, 1999.
- [19] P. Ponniah, “*Data warehousing fundamentals*”, Wiley, 2001.
- [20] PostgreSQL, « *Documentation* », . Available: <https://www.postgresql.org/docs/>. [Accès le Juillet 2020].
- [21] R. K. e. J. Caserta, “*The Data warehouse ETL Toolkit*”, Wiley Publisshing, 2004.
- [22] R. Kimbal et J. Caserta, “*The data warehouse ETL toolkit*”, Wiley, 2004.
- [23] R. Kimball and M. Ross, “*The data warehouse toolkit second edition*”, Wiley, 2002.
- [24] R. Kimball and M. Ross, “*The data warehouse toolkit the definitive guide to dimensional modeling*”, 3th edition, Wiley, 2013.
- [25] 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].
- [26] S. Crozat, « *Data warehouse et outils décisionnels* », UTC Formation, 2016.
- [27] 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].
- [28] W. H. Inmon, “*Building the Data Warehouse Third Edition*”, Wiley Computer Publishing, 2002.
- [29] W. H. Inmon, “*Building the data warehouse*”, fourth edition, Wiley, 2005.
- [30] YENDE R. Grevisse et KASIKA K. Viviane, « *Divergence possible des processus de Data mining et Knowledge Discovery in Databases* », European Journal of Natural and Social Sciences, NOVUS, 01(10), 2023
- [31] YENDE RAPHAEL Grevisse, KABEYA ILUNGA Paulin et al, “*Deployment of a Web Application For Managing the Movements of Secondary School Students in the City of Butembo(DRC)*”. International Journal of Academic Multidisciplinary Research (IJAMR), Vol.6, No.5, pp.208-232, 2022