

MA-004-516-1

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



Université Saad Dahleb de Blida (U.S.D.B)

Faculté des Sciences

Département d'Informatique



Mémoire présenté par : TADBIRT BENNACER

En vue d'obtenir le diplôme de Master

Domaine : Mathématique et informatique. MI

Filière : Informatique.

Option : Ingénierie des logiciels.

Thème :

Conception et réalisation d'un DATA Warehouse pour la gestion des processus de (Facturation, Relève, Raccordement) à base d'un système décisionnel

Promoteur : Mr. BALA MAHFOUD

Encadreur : Mme. REDOUANE DOUNIA

Soutenu le 01/07/2017 devant le jury constitué de :

Mme Madani, Présidente
Mme Hadj Henni, Examineur

MA-004-516-1

Organisation d'accueil : ELIT EL Djazair Information Technology une filiale du Groupe NELGAZ.

Promotion 2016-2017



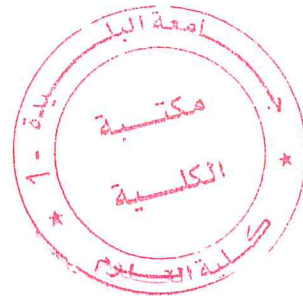
Remerciement

Premièrement je tiens à remercier DIEU de m'avoir donné le Courage et la patience sans lesquels je n'aurais jamais réussi à Élaborer ce mémoire.

Je tiens à exprimer mes vifs remerciements à mon promoteur Monsieur Bala Mahfoud professeur à l'université de Blida, pour sa compréhension, ses précieux conseils, ainsi sa disponibilité et sa confiance en moi.

Je tiens à remercier très vivement mon encadreur Mme Redouane Dounia , ingénieur et développeur a ELIT pour son entière disponibilité, son amabilité et simplicité, ses conseils et encouragements. D'avoir tout mis en œuvre pour que je puisse donner le meilleur de moi-même et pour tous les moyens qu'il a mis à ma disposition durant ma période de stage, je le remercie sincèrement. Qu'il trouve ici l'expression de ma profonde gratitude d'avoir acceptés de diriger et de corriger ce travail.

En fin, nous remercions l'ensemble des enseignants et tout le personnel, du département d'informatique de l'université de Blida.



Dédicace

Sont nombreux ceux à qui je dédie ce mémoire :

A mes chers parents, Faiza et mon père Ali qui m'ont toujours encouragé tout au long de ma vie qu'Allah les protégé.

A mes frères Mhamed Zakaria Et Nouh

A mon promoteur Mahfoud Bala et mon encadreur Mme Redouane Dounia

A tous mes camarades et amis (es) mehdi,talha,ihcen,sofiane

À avec qui j'ai partagé de longues années pleins de souvenirs,

A tous ceux qui, par un mot, m'ont donné la force de continuer

Résumé :

Le groupe Sonelgaz, premier opérateur énergétique en Algérie, est composé d'un ensemble de sociétés, qui assurent différentes missions dans le domaine de l'énergie, ces derniers, allant de la gestion du réseau électrique et gazier à la distribution et commercialisation de l'électricité et du gaz.

Les sociétés de distribution du groupe Sonelgaz, rencontrent quelques problèmes dans leur politique de Reporting Clientèle. Ces difficultés sont dus à l'absence d'un moyen de création et d'édition des rapports en toute autonomie, à l'absence d'un moyen de navigation permettant aux décideurs de visualiser les analyses et les résultats selon le niveau de détail voulu, les difficultés sont liées notamment au niveau élevé de détail des données qui rend l'analyse et la prise de décisions plus difficiles. Dans ce contexte, et afin de répondre à ces attentes grandissantes, le groupe Sonelgaz a sollicité sa filiale spécialisée dans les systèmes d'information et les nouvelles technologies « Elit ».

Ce document, présente une technique efficace de consolidation et de présentation des informations sous forme de rapports, en se basant sur la réalisation d'un Data Warehouse pour le système gestion de la clientèle « SGC », avec un outil du générateur d'état automatique, selon les besoins des décideurs, cet outils offre à la fois l'aspect instrument de contrôle et permet d'améliorer la qualité, l'intégrité et la transparence des informations utilisées pour éclairer d'une manière efficace et sans délai le processus décisionnel du SGC, par catégories organisationnelle et à différents niveaux de privilèges

Le périmètre fonctionnel de la solution proposée dans ce mémoire, prend en charge les processus facturation, relève et raccordement

Notre système est développé en JAVA IEE.

Mots-clés : Data Warehouse, générateur, état, automatique, reporting, rapports, analyses, décisionnel, décideurs, navigation, organisationnelle.

Abstract:

The Sonelgaz group is the first energy provider in Algeria. It is composed of several companies which carry out various tasks in the field of energy, ranging from the management of the electricity and gas network to the distribution and marketing of Electricity and gas.

The distribution companies of the Sonelgaz group, encounter some problems in their Customer Reporting policy. These difficulties are due to the lack of a means of creating and editing reports independently, also to the absence of a means of navigation allowing decision-makers to visualize the analyzes and the results according to the level of detail required, The difficulties relate in particular to the high level of detail of the data which makes analysis and decision-making more difficult.

In this context, and in order to meet these growing expectations, the Sonelgaz Group has asked its subsidiary specialized in information systems and new technologies "Elit".

This document presents an efficient technique for consolidating and presenting information in the form of reports, based on the realization of a Data Warehouse for customer management systems "SGC", with an automatic state generator tool, According to the needs of decision-makers, this tool offers both an instrument of control and improves the quality, integrity and transparency of the information used to inform the decision-making process of SGC in a timely and efficient manner, by organizational categories and at different levels of privilege

The functional perimeter of the solution proposed in this brief, supports the billing process, succession process and connection process.

Our system is developed in JAVA IEE.

Keywords: Data Warehouse, generator, state, automatic, reporting, reports, analytics, decision-makers, decision makers, navigation, organizational.

Lexique :

BI : Business Intelligence.

BT : Basse Tension.

BP : Basse Pression.

DD : Direction de Distribution.

DED : Département Etudes et Développement.

DGDS : Direction Générale du Développement et de la Stratégie.

DR : direction régionale(DD).

DRD : Direction de Distribution Régionale.

DW : Data Warehouse (Entrepôt de données).

EGA : Electricité et Gaz d'Algérie.

ELIT : EL-djazaïr Information Technology.

ETL : Extract, Transform and Load (ETC).

FK: Foreign Key.

FTP : File Transfer Protocol.

HOLAP: Hybrid On Line Analytical Process.

HP : Haute Pression.

HT : Haute Tension.

MOLAP: Multidimensional On Line Analytical Process.

MP: Moyenne Pression.

HTA : Haute Tension.

OLAP : On Line Analytical Process.

OLTP: On Line Transactional Process.

PK : Primary Key.

ROLAP : Relational On Line Analytical Process.

SD : Société de Distribution.

SGC : Système de Gestion de la Clientèle.

SI : Systèmes d'Information.

SID: Systèmes d'Information Décisionnels.

SID : Systèmes d'information de la distribution.

SIAD : Systèmes d'Information d'Aide à la Décision

SGBD : Système de Gestion de Base de Données.

SMTP : Server Mail Transfer Protocol.

SONELGAZ : Société Nationale de l'Electricité et du GAZ.

SPA : Société Par Action.

SQL : Structured Query Language.

Liste des figures

Figure 1 : Architecture d'un entrepot de donnée[3].....	9
Figure 2 : Schéma en étoile.....	12
Figure 3 : Schéma en flocon de neige.....	12
Figure 4 : Schéma en constellation.....	13
Figure 5 : Exemple d'un tableau de bord[4].....	15
Figure 6 : Les trois phases ETL[5].....	16
Figure 7 : Extraction temps-réel[Chafki ,2011].....	18
Figure 8 : Extraction différée[Chafki ,2011].....	19
Figure 9 : Un cube OLAP et le principe d'ogégation[9].....	26
Figure 10 : Architecture d'un produit MOLAP.....	28
Figure 11 : Architecture d'un produit ROLAP[Nakache,1998].....	29
Figure 12 : Architecture d'un produit H-OLAP[11].....	30
Figure 13 : Architecture d'un produit D-OLAP[11].....	31
Figure 14 : Structure du groupe SONELGAZ.....	33
Figure 15 : Structure d'ETL.....	36
Figure 16 : Structure de la DSID.....	37
Figure 17 : Organisations des SD.....	39
Figure 18 : Hiérarchie de la SD.....	40
Figure 19 : Facture cyclique.....	43
Figure 20 : Processus de reporting à travers ELIT.....	48
Figure 21 : Reporting en utilisant un entrepot de données.....	60
Figure 22 : Reporting en utilisant un entrepot de données avec un générateur d'état.....	62
Figure 23 : L'architecture de la solution proposée.....	62
Figure 24 : Fait « FACTURATION_BTBP ».....	64
Figure 25 : Fait « FACTURATION ».....	65

Figure 26 : DATA-MART « FACTURATION »	66
Figure 27 : Fait « RELEVÉ_BTBP ».....	66
Figure 28 : Fait « RELEVÉ ».....	67
Figure 29 : Fait « RELEVÉ_ABOMALIE_BTBP ».....	68
Figure 30 : Fait « RELEVÉ_ANOMALIE »	69
Figure 31 : DATA-MART « RELEVÉ »	70
Figure 32 : Fait « RCN_BRANCHEMENT »	71
Figure 33 : Fait « RCN_EXTENSION »	72
Figure 34 : Fait « RCN_DELAIS_MOYEN »	73
Figure 35 : DATA-MART « RACCORCHEMENT ».....	74
Figure 36 : Schéma en constellation final de l'entrepot de données DW_Client.....	75
Figure 37 : Diagramme de cas d'utilisation général	82
Figure 38 : Diagramme de cas d'utilisation gestion administration.....	82
Figure 39 : Diagramme de cas d'utilisation détaillé.....	83
Figure 40 : Diagramme de cas d'utilisation MAJ d'un rapport	83
Figure 41 : Diagramme de cas d'utilisation gestion des utilisateurs.....	84
Figure 42 : Diagramme de cas d'utilisation consultation des rapports décisionnels.....	84
Figure 43 : Diagramme de séquence créer un rapport.....	85
Figure 44 : Diagramme de classe	86
Figure 45 : Architecture de générateur de rapport décisionnels.....	89
Figure 46 : Le modèle MVC du Framework JSF	95
Figure 47 : L'architecture technique de la solution proposée	96
Figure 48 : PostgreSQL.	97
Figure 49 : Job de chargement processus de la RELEVÉ.....	98
Figure 50 : L'interface qui permet à l'utilisateur de s'authentifier	99
Figure 51 : L'interface permettent la création des rapports	100
Figure 52 : L'interface permettant la génération des rapports	101

Figure 53 : Deux fichiers seront créer l'un sous format jasper	101
Figure 54 : L'interface permettant la création des utilisateurs.....	102
Figure 55 : L'interface de la liste des utilisateurs	102
Figure 56 : L'interface de l'authentification de décideur SDA.....	103
Figure 57 : L'interface des statistiques de facturation, relève et raccordement.....	103
Figure 58 : L'interface permettant la consultation de rapport.....	104
Figure 59 : L'interface de décideur sélectionne l'année et la saison désirées.....	104
Figure 60 : L'interface de rapport en format PDF	105

Liste des tableaux

Tableau 1 : Les principales différences entre les SIO et les SID [Ferragu, 2013]	5
Tableau 2 : Molèle conceptuel d'une table de faits	10
Tableau 3 : Modèle conceptuel d'une table de dimension.	11
Tableau 4 : Un tableau récapitulatif des différences entre OLTP et OLAP[13]	27
Tableau 5 : Catégories d'utilisateurs du système de reporting à travers ELIT	53
Tableau 6 : Mesures du fait « FACTURATION_BTBP »	64
Tableau 7 : Mesures du fait « FACTURATION »	65
Tableau 8 : Mesures du fait « RELEVE_BTBP »	67
Tableau 9 : Mesures du fait « RELEVE »	67
Tableau 10 : Mesures du fait « RELEVE_ANOMALIE_BTBP »	68
Tableau 11 : Mesures du fait « RELEVE_ANOMALIE »	69
Tableau 12 : Mesures du fait « RCN_BRANCHEMENT »	71
Tableau 13 : Mesures du fait « RCN_EXTENSION »	72
Tableau 14 : Mesures du cube « RCN_DELAITS_MOYEN »	73
Tableau 15 : Définition des niveaux et hiérarchies des dimensions « FACTURATION »	77
Tableau 16 : Définition des niveaux et hiérarchies des dimensions « RELEVE »	78
Tableau 17 : Définition des niveaux et hiérarchies des dimensions « RACCORCHEMENT.	79
Tableau 18 : Description des classes de génération de rapports décisionnels.....	89

Sommaire

▪ INTRODUCTION GENERALE

1-	Contexte	1
2-	Problématique	1
3-	Objectifs	2
4-	Organisation du mémoire	3

Partie 1 : Les fondamentaux sur les systèmes Décisionnels.

▪ CHAPITRE I : LES FONDAMENTAUX SUR LES SYSTEMES DECISIONNELS

1.	Introduction	4
2.	Systèmes SIO versus systèmes SID	4
3.	Entrepôts de données(DATA Warehouse)	6
3.1	Définition	6
3.2	Les classes de données	8
4.	Architecture décisionnelle	8
5.	Modélisation d'un entrepot de données(ED)	9
5.1	La modélisation dimensionnelle	9
5.2	La modélisation dimensionnelle logique	13
6.	Exploitation des données	14
6.1	Le reporting	14
6.2	Le DATA Mining	14
6.3	Outils OLAP	14
6.4	Tableau de bord	15
7.	Conclusion	15

▪ CHAPITRE II : PROCESSUS D'INTEGRATION DES DONNEES (ETL)

1.	Introduction	16
2.	Définition du processus ETL	16
2.1	Extraction	17
2.2	Transformation	20
2.3	Chargement.....	20
3.	Les structures de données ETL.....	21
3.1	Sources de données.....	21

3.2	Base de données relationnelles	21
4.	Zone de préparation des données « DSA ».....	22
5.	Conclusion	23
▪ CHAPITRE III : ANALYSE DE DONNEES EN LIGNE(OLAP)		
1.	Introduction	24
2.	Définion	24
3.	Comparaison entre OLAP & OLTP	27
3.1	Définition de OLTP	27
3.2	Tableau comparatif	27
4.	Les différents outils OLAP	27
4.1	M-OLAP	28
4.2	R-OLAP.....	29
4.3	H-OLAP.....	30
4.4	S-OLAP	31
4.5	D-OLAP.....	31
5.	Conclusion	31

Partie 2 : Conception de l'entrepôt de données DW_CLIENT.

▪ CHAPITRE IV PRESENTATION DE L'ORGANISME D'ACCUEIL		
1.	Introduction	32
1.1	Le groupe SONELGAZ	32
1.2	El Djazair Information Technology (ELIT)	34
1.3	Les sociétés de distribution SONELGAZ	38
2.	Conclusion	41
▪ CHAPITRE V :ETUDE DE L'Existant		
1.	Introduction	42
2.	Existant opérationnel	42
2.1	Système de Gestion de la Relation Client (CRM).....	42
2.2	Présentation du système opérationnel actuel (SGC).....	43
3.	Existant décisionnel.....	47
3.1	Processus de reporting	47
3.2	Procédure de reporting au niveau du group	47
3.3	Reporting à traves ELIT.....	48
3.4	Bilan.....	50

4.	Conclusion.....	50
▪	CHAPITRE VI :BESOINS D'ANALYSE DES DONNEES ET AIDE à LA DECISION	
1.	Introduction.....	51
2.	Identification des besoins.....	51
2.1	Démarche de collecte des besoins.....	51
2.2	Identification des utilisateurs finaux du système décisionnel.....	52
2.3	Récapitulatif des besoins.....	53
2.4	Difficultés rencontrées.....	59
2.5	Proposition de la solution.....	59
2.6	Les solutions proposées.....	59
2.7	La solution sélectionnée.....	62
3.	Conclusion.....	63
▪	CHAPITRE VII : Modélisation dimensionnelle de l'entrepot DW_CLIENT	
1.	Introduction.....	64
2.	La modéliation dimensionnelle en étoile.....	64
2.1	Data-mart « Facturation »	64
2.2	Data-mart « Relève »	66
2.3	DATA-MART « Raccordement »	70
2.4	Schéma en constellation final de l'entrepôt de donnée « DW_CLIENT »	74
3.	Définition des niveaux et hiérarchies des dimensions	76
3.1	Data-Mart « FACTURATION »	76
3.2	Data-Mart « RELEVE »	77
3.3	Data-Mart « RACCORDEMENT »	78
4.	Conclusion	79
	Partie 3 : Conception et mise en œuvre du système d'analyse	
▪	CHAPITRE VIII CONCEPTION DU GÉNÉRATEUR DE RAPPORTS DECISIONNELS	
1.	Introduction.....	81
2.	Démarche suivie pour la conception.....	81

2.1	Présentation des acteurs et des cas d'utilisation.....	81
2.2	Diagrammes de séquence.....	85
2.3	Diagramme de classe	86
3.	Mécanisme du générateur de rapports décisionnels	89
3.1	Architecture du générateur de rapports	89
4.	Conclusion.....	91
▪ CHAPITRE IX : Description des applications et mise en œuvre		
1.	Introduction.....	92
2.	Environnement du travail.....	92
2.1	Environnement matériel.....	92
2.2	Outil de modélisation UML.....	92
2.3	ETL	93
2.4	Environnement de développement intégré	94
2.5	Serveur d'application	94
2.6	Outil d'administration de la base de données	94
2.7	Outil de reporting	94
2.8	Plate-forme de développement « JEE »	95
3.	Architecture technique de la solution	96
4.	Réalisation du projet	96
4.1	Construction de l'entrepôt de données	96
4.2	Extraction, transformation et chargement	97
5.	Implémentation des interfaces du système	99
5.1	Authentification	99
5.2	Page d'accueil	99
6.	Conclusion	105

CONCLUSION GENERALE

- BIBLIOGRAPHIE
- WEB

1. Contexte :

A nos jours, et dans un environnement hautement concurrentiel, la simple logique de production ne suffit plus à pérenniser l'activité des entreprises. Pour cela, les différentes entreprises font tout leur possible pour améliorer la rentabilité de leurs investissements qui influence positivement sur leur compétitivité. Les décideurs, de leur part, souhaitent tirer avantage des données de l'entreprise pour prendre des décisions efficaces sur la meilleure façon de gérer leur organisation. Mais aussi, sur le bon déroulement de leurs processus métiers (ressources humaines, relation client, finance, facturation...).

Pour se faire, les entreprises utilisent actuellement des systèmes décisionnels en vue de soutenir le processus de prise de décision. Les systèmes décisionnels reposent sur un espace de stockage centralisé, appelé entrepôt de données, qui permet de collecter des informations provenant de sources différentes afin de les exploiter aux travers d'interfaces.

Le système décisionnel vient pour compléter le système opérationnel par une meilleure exploitation de ses données.

Le Groupe Sonelgaz est composé d'un ensemble de sociétés de production du transport et de la distribution de l'électricité ainsi que la distribution publique de gaz, en Algérie.

La restructuration de Sonelgaz, a permis depuis quelques années la création d'une Société holding, dénommée Sonelgaz.Spa et plusieurs sociétés, toutes juridiquement autonomes les unes des autres. Les Sociétés de Distribution du Groupe Sonelgaz, à savoir : SDA, SDC, SDO et SDE disposent d'un Système d'Information pour la Gestion de la Clientèle. Ce système permet à travers ses modules de prendre en charge la gestion de la clientèle depuis la demande de raccordement jusqu'à la résiliation du contrat, automatisant ainsi plusieurs processus de gestion. En 2016, ELIT la filiale spécialisée dans les systèmes d'information et les nouvelles technologies de Sonelgaz, a procédé à la modification du système de gestion de la clientèle SGC, tout en changeant les systèmes décisionnels de tous les processus (achat, évènement, facturation, relève...). Pour se faire, ELIT a décidé de deviser le projet de système décisionnel de SGC en une série de sous-projets, chacun traite trois processus. L'ensemble des sous projets est classé par priorité dont la priorité 1 est affectée aux « facturation, relève, raccordement ». Ce dernier est le thème de notre projet de fin d'étude.

2. Problématique

Les Sociétés de Distribution et en particuliers leurs Direction Commerciales sont principalement préoccupées par la réalisation d'objectifs fixés par la direction générale. Pour réaliser ces objectifs, elles ont besoin de connaître à tout moment leur situation actuelle (et prévisionnelle) par rapport à cet objectif, afin d'apporter éventuellement les modifications nécessaires à leur accomplissement en temps utile : réorganisation des secteurs, embauches, campagnes marketing, séminaires, etc. Or il est toujours difficile de récupérer les chiffres des acteurs de terrain. Si cette difficulté est surmontable dans le cadre d'une structure locale par

des moyens traditionnels (réunions, ...), elle est beaucoup plus complexe à traiter dans le cadre d'équipes délocalisées (agence, division, direction régionales, sociétés, partenaires).

Ces difficultés sont dues à :

- L'absence d'un moyen permettant aux décideurs d'éditer leurs rapports en toute autonomie.
- L'indisponibilité des rapports en temps voulu.
- L'absence d'un moyen de navigation permettant aux décideurs de visualiser les analyses et les résultats selon le niveau de détail voulu.
- Le niveau élevé de détail des données qui rend l'analyse et la prise de décisions plus difficile.

Les directions commerciales doivent donc disposer d'un système décisionnel efficace, leur permettant de prendre des décisions informées et non empiriques.

Le périmètre fonctionnel de la nouvelle solution doit prendre en charge :

- **Processus Facturation, Processus Relève et Processus Raccordement.**

3. Objectifs : classés par ordre de priorité :

1. Garantir une analyse multidimensionnelle des données et élargir son périmètre pour englober toutes les directions commerciales .
2. Faciliter l'édition des rapports d'analyse et permettre une certaine flexibilité dans la construction de ces derniers.
3. Offrir une vision synthétisée sur les performances de l'entreprise à travers un ensemble d'indicateurs.
4. Intégrer la solution au niveau du processus décisionnel de la gestion de la clientèle de Sonelgaz.

Résultats attendus :

- Concevoir et réalisation un data warehouse pour les données système de gestion clientèle (périmètre fonctionnel ci-dessus).
- Offrir une version synthétisée sur les performances de l'entreprise à travers des indicateurs.
- Concevoir une application web qui Permettra aux utilisateurs d'accéder à un ensemble de fonctionnalités offertes : rapports, exportations et publications vers les outils bureautiques..etc. Et ce, à partir d'une interface unique, adaptée à chaque profil utilisateur, et un outil permettra de générer n'importe quel état selon les besoins des utilisateurs.
- Publier des documents statiques (manuel d'utilisation).
- Exporter des documents analytiques vers d'autres formats (PDF, WORD, XLS...).

4. Organisation du mémoire

Afin de présenter toutes les phases parcourues en vue de mettre en place le système décisionnel du SGC, nous avons organisé le présent mémoire de la manière suivante.

Première partie : Les fondamentaux sur les systèmes Décisionnels

Cette partie présente les généralités sur le domaine décisionnel et, plus particulièrement, sur le processus d'intégration de donnée ETL . Elle est composée de trois chapitres :

Chapitre I : Généralités sur les Systèmes d'Information Décisionnel, Ou nous avons détaillé les systèmes d'information décisionnels (SID) ainsi que la notion d'entreposage de données, le concept OLAP et la phase restitution.

Chapitre II : Processus d'intégration des données (ETL), Ce chapitre présent des définitions sur les approches d'intégration des données et processus ETL.

Chapitre III : Analyse de données en ligne (OLAP) : Dans ce chapitre, nous avons présenté les technologies OLAP qui sont des outils de Business Intelligence.

Deuxième partie : Conception de l'entrepôt de données DW_CLIENT

Chapitre IV : Présentation de l'organisme d'accueil : Le chapitre un, présente l'organisme d'accueil, sa structure organisationnelle, son activité et la culture de l'entreprise en matière d'utilisation des technologies de l'information.

Chapitre V : ÉTUDE DE L'EXISTANT, Il expose l'étude des feuilles de calcul Excel de direction du foncier et des affaires juridiques et l'identification de leurs besoins d'analyses .

Chapitre VI : Besoins d'analyse des données et aide à la décision : présente une synthèse de la collecte des besoins des utilisateurs, ainsi que son déroulement.

Chapitre VII : Modélisation dimensionnelle de L'entrepôt DW_CLIENT : contient la conception de la partie d'entreposage de notre solution. Il présente entre autre les modèles dimensionnels des processus recensées.

Troisième Partie : Conception et mise en œuvre du système d'analyse

Chapitre VIII : Conception du générateur de rapports décisionnels : Ce chapitre constitue la finalité de notre travail dont nous citons la conception du générateur de rapports décisionnels ainsi que des métadonnées

Chapitre IX : Description des applications et mise en œuvre : l'objectif de ce chapitre est de décrire l'environnement de développement de l'application. Nous allons aussi présenter l'implémentation de la solution proposée pour la réalisation de l'application en citant ses différentes interfaces et fonctionnalités.

Partie 1

Les fondamentaux sur les systèmes Décisionnels

Chapitre I

Généralités sur les Systèmes d'Information Décisionnel

1. Introduction

Avec la généralisation de l'informatique dans tous les secteurs d'activité, les entreprises produisent et manipulent de très importants volumes de données électroniques. Ces données sont stockées dans les Systèmes d'Information Opérationnels (SIO) de l'entreprise au sein de bases de données, de fichiers... L'exploitation de ces données dans un but d'analyse et de support à la prise de décision s'avère difficile et fastidieuse, elle est réalisée le plus souvent de manière imparfaite par les décideurs grâce à des moyens classiques (requêtes SQL, vues, outils graphiques d'interrogation...).

Ces systèmes sont utilisés pour faciliter l'accès, l'interrogation et l'analyse de l'information d'une organisation pour ses décideurs. La dernière évolution notable des Systèmes d'Information Décisionnels (SID) repose sur les concepts d'Entrepôt de Données (DW : Data Warehouse) et d'analyse des données en ligne (OLAP : On Line Analytical Processing) [1]. L'entrepôt de données est le cœur du SID : il intègre et stocke d'importants volumes de données issues des différents domaines fonctionnels d'une organisation pour les rendre facilement accessibles aux processus d'interrogations et d'analyses décisionnelles. L'entrepôt de données est défini comme « une collection de données intégrées, orientées sujets, non volatiles, historiées, résumées et disponibles pour l'interrogation et l'analyse »

[2]. L'approche OLAP consiste ainsi à permettre la navigation au sein d'espaces de données entreposées afin de mesurer, à différents niveaux de granularités, les phénomènes remarquables.

2. Systèmes SIO versus systèmes SID

Chaque SIO possède sa propre base de données, d'où une très grande redondance de données au sein de l'ensemble du système d'information de l'organisation. De plus, les données redondantes (c'est-à-dire présentes dans plusieurs SIO) sont souvent incohérentes car elles ne peuvent pas être mises à jour en même temps dans chaque SIO dans lesquels elles sont présentes, elles ne sont synchronisées que périodiquement.

- **Conséquence** : les informations nécessaires à la prise de décision sont le plus souvent éparpillées dans de multiples SIO, il est donc nécessaire de les rassembler dans un endroit unique et de les mettre en cohérence pour pouvoir les exploiter de manière optimale.

Le constat de l'inaptitude des SIO à restituer les données qu'ils stockent a amené les organisations à construire des systèmes à part, dédiés à la restitution d'informations, que l'on appelle systèmes d'information décisionnels (SID).

Un SID est donc un système d'information dédié aux décideurs d'une organisation et permettant, au moyen d'une base de données et d'une interface d'accès aux données, d'obtenir des informations utiles à la prise de décision. [3]

Le tableau 1 illustre les principales différences entre les SIO et les SID :

Critère	SIO	SID
Objectif	Exécution de processus métier	Évaluation de la performance des processus métier
Mode d'interaction entre les utilisateurs et la base de données	Insertion, mise à jour, suppression et sélection de données	Sélection de données
Périmètre d'interaction entre les utilisateurs et la base de données	Transactions unitaires	Sélection de données en masse
Type d'utilisation	Prédéfinie, prévisible	Non prédéfinie, imprévisible
Complexité des requêtes des utilisateurs	Faible	Elevée
Optimisé pour	La performance des transactions unitaires	La performance des requêtes de sélection des données en masse
Fréquence de mise à jour de la base de données	Mises à jour en temps réel, au fur et à mesure de l'exécution des processus métier	Mises à jour périodiques en mode « batch » ³
Historique des données utilisées	Données courantes	Données courantes mais aussi et surtout historiques
Degré de normalisation des données	Hautement normalisé (3 ^e forme normale)	Dénormalisé

Tableau 1 : Les principales différences entre les SIO et les SID. [3]

Le système opérationnel représente les tâches quotidiennes, répétitives et atomiques (insertion, modification, suppression) qui sont effectuées par les employés de l'entreprise pour le support et la gestion efficace de son activité.

Le SIO qui est basé, en ce qui concerne la gestion des données, sur un processus transactionnel (OLTP : *On Line Transaction Processing*) est un système de gestion de production dédié pour assister les opérations d'une entreprise.

Avec l'accroissement du volume des données, les systèmes opérationnels se trouvent limités face aux besoins des dirigeants qui, eux, veulent plutôt des informations synthétisées pour l'analyse et la prise de décisions, d'où la naissance des environnements d'analyse en ligne (OLAP : *On Line Analytical Processing*).

Les activités transactionnelles (OLTP) et d'analyse (OLAP) ne peuvent coexister sur des données dans le même système d'information vu que leurs objectifs de performance sont exactement opposés :

- Les requêtes OLAP, considérées très complexes et lourdes, dégradent les performances des systèmes transactionnels.
- Les données temporelles réparties entre données actuelles et données archivées, rendent la vue historique des données très difficile, voire impossible.

Le support efficace d'une activité OLAP nécessite la constitution d'un système d'information dédié appelé Système d'Information Décisionnel (SID) qui repose sur une technologie de stockage appelée Entrepôt de données. [20]

3. Entrepôt de données (*Data Warehouse*)

3.1 Définition

Un entrepôt de données (Data Warehouse) est une collection de données thématiques, intégrées, non volatiles et historiées pour la prise de décisions [2]., C'est une vision centralisée et universelle de toutes les informations de l'entreprise. En termes plus simples, il s'agit d'une base de données multidimensionnelle qui a pour but, contrairement aux bases de données classiques, de regrouper les données de l'entreprise pour des fins analytiques et d'aide à la décision stratégique. En effet, cette dernière est une action entreprise par les décideurs de l'entreprise visant à améliorer, quantitativement ou qualitativement, la performance de l'entreprise.

Les entrepôts sont physiquement séparés des systèmes de production et ce pour des raisons de :

- **Performance** : les données des systèmes de production ne sont pas organisées pour pouvoir répondre efficacement aux requêtes des systèmes d'aide à la décision. Même les requêtes simples peuvent dégrader sérieusement les performances.
- **Accès aux données** : un entrepôt doit pouvoir accéder aux données uniformément,

quelle que soit la provenance des données.

- **Formats des données** : les données des entrepôts sont transformées, et doivent être disponibles sous un format simple et unique.
- **Qualité des données** : les données d'un entrepôt sont propres et validées. [21].

Les points clefs garantissant le succès d'un entrepôt de données sont les suivants :

- Les informations d'un entrepôt de données doivent être accessibles et fiables (de qualité).
- La conception d'un entrepôt de données doit répondre à un besoin de ROI levé.
- La réponse aux demandes très diverses des utilisateurs.
- L'entrepôt de données doit évoluer avec les besoins des utilisateurs et du système d'information. [22]

3.1.1 Données orientées sujet

L'entrepôt de données est organisé autour des sujets majeurs et des métiers de l'entreprise. Les données sont organisées par thème, contrairement aux données des systèmes de production, organisées par processus fonctionnels. L'avantage de cette représentation demeure dans le fait qu'il devient possible de réaliser des analyses sur des sujets transversaux aux structures fonctionnelles et organisationnelles de l'entreprise. Et ainsi, de pouvoir analyser un processus dans le temps à différentes étapes de sa conception au sein du SI. Cette orientation permet également de faire des analyses par itération, sujet après sujet. L'intégration dans une structure unique est indispensable pour éviter aux données concernées par plusieurs sujets d'être dupliquées. Dans la pratique il existe également des Datamart pouvant supporter l'orientation sujet. [22]

3.1.2 Données intégrées

Un Entrepôt de données est un projet d'entreprise et concerne les différents services et métiers de l'entreprise. L'intégration de données, au sein d'un entrepôt de données, est donc un processus déterminant sur la qualité et la quantité d'informations disponibles aux utilisateurs pour le processus de décision. Cette phase, que nous verrons plus en détail avec les outils ETL, implique que les données doivent être mises en forme et unifiées afin d'avoir un état cohérent. Pour parfaire cette cohérence, l'intégration nécessite une forte normalisation de données. Mais aussi la maîtrise de la sémantique, la prise en compte des contraintes référentielles et des règles de gestion. Ces notions sont énoncées, détaillées et administrées au sein des métadonnées de

l'entrepôt de données. C'est ainsi que l'on pourra donner une bonne vision de l'entreprise via l'utilisation d'indicateurs. [22]

3.1.3 Données historisées

L'historisation est nécessaire pour suivre dans le temps l'évolution des différentes valeurs des indicateurs à analyser. Ainsi, un référentiel temps doit être associé aux données afin de permettre l'identification dans la durée de valeurs précises. [22]

3.1.4 Données non volatiles

Afin de conserver la traçabilité des informations et des décisions prises, les informations stockées au sein de l'entrepôt de données ne peuvent être supprimées. [22]

3.2 Les classes de données

Un entrepôt de données peut se structurer en quatre classes de données organisées selon un axe historique et un axe de synthèse. [22]

3.2.1 Les données agrégées

Les données agrégées correspondent à des éléments d'analyse représentant les besoins des utilisateurs. Elles constituent déjà un résultat d'analyse et une synthèse de l'information contenue dans le système décisionnel, et doivent être facilement accessibles et compréhensibles. [22]

3.2.2 Les données détaillées

Les données détaillées reflètent les événements les plus récents. Les intégrations régulières des données issues des systèmes de production vont habituellement être réalisées à ce niveau. [22]

3.2.3 Les métadonnées

Les métadonnées constituent l'ensemble des données qui décrivent des règles ou processus attachés à d'autres données. Ces dernières constituent la finalité du système d'information. [22]

4. Architecture décisionnelle

Différentes architectures existent pour les systèmes d'entreposage de données. Certaines sont basées sur un entrepôt opérationnel (ODS : Opérationnel Data Store), alors que d'autres peuvent avoir de multiples magasins de données (Data marts). Par ailleurs, certains SID peuvent exploiter un petit nombre de sources de données, alors que d'autres sont alimentés à partir de

dizaines de sources de données. Compte tenu de cela, il est beaucoup plus raisonnable de présenter les différentes couches d'une architecture générale d'entrepôt de données plutôt que de discuter des détails de ces différentes architectures. La figure 1 montre les relations entre les différentes composantes de la suite décisionnelle.

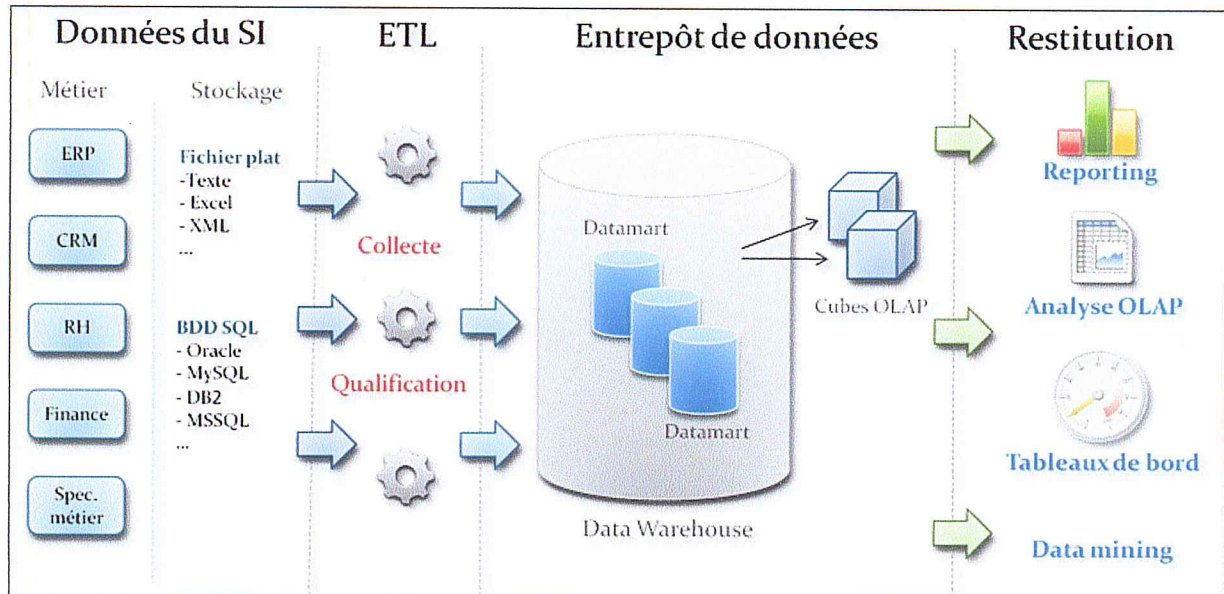


Figure 1 : Architecture d'un entrepôt de données.[23]

5. Modélisation d'un entrepôt de données (ED)

Les données dans l'entrepôt doivent être organisées d'une manière à faciliter leur exploitation ainsi que leur analyse par les décideurs. Ces analyses nécessitent l'exécution de requêtes particulières produisant comme résultat sous la forme d'indicateurs ((analyse des tendances, préférences d'achat, etc.)) selon plusieurs axes d'analyses. Les données manipulées dans ce contexte sont représentées sous forme multidimensionnelle qui est mieux adaptée pour le support des processus d'analyse et d'aide à la décision. [4]

La modélisation dimensionnelle se présente sous deux niveaux : dimensionnelle et logique

5.1 La modélisation dimensionnelle

La modélisation dimensionnelle (ou multidimensionnelle) souvent appelée modélisation OLAP [5] .se présente comme une alternative au modèle relationnel. Elle correspond mieux aux besoins du décideur tout en intégrant la modélisation par sujet. C'est une méthode de modélisation conceptuelle qui vise à présenter les données sous une forme

standardisée intuitive permettant des accès hautement performants. Elle aboutit à présenter les données non plus sous forme de tables mais de cube centré sur une activité. Un cube de dimension n ($n > 3$) est aussi dit hypercube. La modélisation dimensionnelle repose sur deux concepts essentiels « Faits » et « dimensions ».

5.1.1 La table de faits

La table de faits est la clef de voûte du modèle dimensionnel où sont stockés les indicateurs de performances. Le concepteur s'efforce de considérer comme indicateurs les informations d'un processus d'entreprise dans un système d'information. Les indicateurs étant les données les plus volumineuses d'un système d'information, on ne peut se permettre de les dupliquer dans d'autres tables mais de les rationaliser au sein de la table de faits.

Table de faits des ventes journalières
Clé date (CE)
Clé produit (CE)
Clé magasin (CE)
Quantité vendue
Montant des ventes (€)

Tableau 2 : Modèle conceptuel d'une table de faits.

Le terme de fait est utilisé pour représenter une mesure économique (ou indicateur). Pour exemple, lors de la vente de produits sur un marché, on comptabilise les types de produits vendus, leur quantité et le montant de chaque vente au jour le jour et ce, pour chaque produit et pour chaque magasin. La mesure des quantités et des prix est réalisée à l'intersection de toutes les dimensions (produit, magasin, temps). Le nombre des dimensions détermine la finesse, la granularité de la table et indique la portée de la mesure.

Les indicateurs les plus utiles d'une table de faits sont numériques et additifs. L'additivité des attributs d'une table de faits est cruciale pour les outils décisionnels. Les utilisateurs demandent rarement l'analyse d'une seule ligne. Dans notre exemple, constater les ventes de produits sur une année pour les magasins d'une région demande l'analyse de plusieurs milliers de lignes à la fois. Pour autant, tous les attributs utiles ne sont pas additifs. Certains sont semi additifs et ne peuvent être additionnés que pour certaines dimensions. D'autres sont non additifs et ne peuvent pas être additionnés par dimensions. Pour cette dernière catégorie, on utilise des fonctions d'agrégations tel que, le calcul de moyenne, le ratio ou le comptage de lignes.

5.1.2 Les dimensions

Les tables de dimensions sont les entités complémentaires à la conception de la table de faits. Elles contiennent, autant que possible, des attributs sous forme de descriptions textuelles permettant de qualifier ou d'expliquer l'activité. Des attributs de dimensions, nombreux, permettent de varier les possibilités d'analyse (par tranches ou en dés). Ces attributs rendent utilisables et intelligible les données de l'entrepôt de données. Ils établissent, en quelque sorte, une interface homme/entrepôt de données. En général, les tables de dimensions tendent à être peu profondes mais elles sont larges (l'inverse de la table de faits), en d'autres termes elles ont peu de lignes mais beaucoup de colonnes. [22]

Tables de dimension "Produit"
Clé produit (CP)
Numéro US (clé naturelle)
Description de la marque
Description de la catégorie
Description du produit
Description du rayon
... et bien d'autre attributs

Tableau 3 : Modèle conceptuel d'une table de dimension

5.1.3 Les trois schémas de la modélisation dimensionnelle

Au sein de l'entrepôt, les données sont redondantes et dé normalisées, nous sommes loin de la modélisation en troisième forme normale (3NF) et pour cause, cela permet de faciliter l'utilisation et d'améliorer les performances lors de l'analyse des données. Trois types de schémas sont fréquemment rencontrés, le schéma en étoile, le schéma en flocon et le schéma en constellation de faits. [22]

5.1.3.1 Le schéma en étoile

Dans ce modèle, chaque groupe de dimensions est placé dans une table de dimension ; les faits sont placés dans une table des faits. Comme le montre la figure 2, le résultat de cette classification est un schéma en étoile où la table des faits se trouve au milieu de l'étoile et les tables de dimension dans les côtés [4].

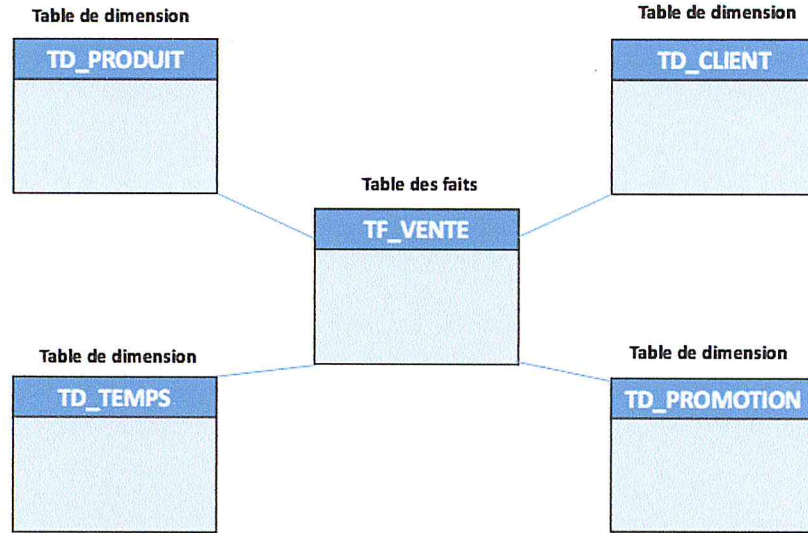


Figure 2 : Schéma en étoile.

5.1.3.2 Le schéma en flocon

La dé-normalisation des tables de dimension dans un schéma en étoile ne présente pas de façon explicite les hiérarchies associées à chaque dimension. Pour mettre en évidence cette hiérarchie, le modèle en flocon de neige a été proposé. Chaque table de dimension comme le montre la figure 3, est éclatée en un ensemble de hiérarchies. Ce schéma normalise les dimensions, réduit la taille de chacune des relations et permet ainsi de formaliser la notion de hiérarchie au sein d'une dimension [6].

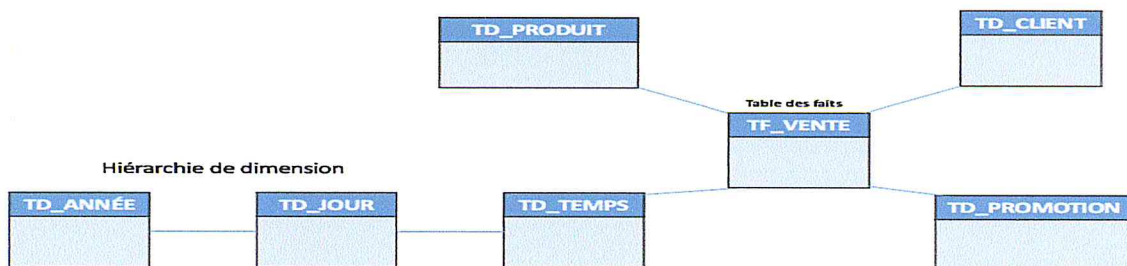


Figure 3 : Schéma en flocon.

5.1.3.3 Schéma en constellation

Il est possible d'avoir plusieurs relations de faits pour représenter les situations dans lesquelles les faits (mesures) ne sont pas déterminés par exactement le même ensemble de dimensions. Dans ce cas, les relations de faits forment une famille qui partage plusieurs

relations de dimension mais où chaque membre possède ses propres dimensions [4]. Le schéma résultant (voir figure 4) s'appelle constellation de faits.

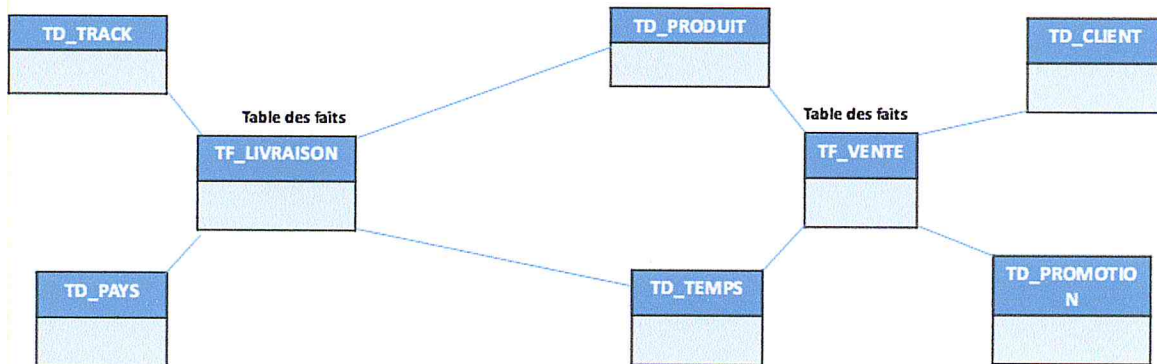


Figure 4 : Schéma en constellation.

5.2 La modélisation dimensionnelle Logique

C'est la description de la base multidimensionnelle suivant la technique utilisée. Il existe principalement trois techniques :

R-OLAP : Les systèmes ROLAP (*Relational On Line Analytical Processing*) stockent les données en utilisant un SGBD relationnel. Chaque dimension est représentée par une table de dimension et chaque fait est représenté par une table de fait. Les mesures sont stockées dans les tables des faits. [8]

M-OLAP : Les systèmes MOLAP (*Multidimensional On Line Analytical Processing*) implémentent le cube (qui est une intersection de données dans un espace à plusieurs dimensions) sous forme d'un tableau multidimensionnel. Chaque dimension du tableau représente une dimension du cube. Les données représentent chaque cellule du cube. C'est une solution de stockage qui garantit un temps de réponse rapide. [8]

H-OLAP : Les systèmes HOLAP (*Hybrid On-Line Analytical Processing*) sont des systèmes où les données fréquemment utilisées (données agrégées) sont maintenues par un SGBD multidimensionnel, et les données non fréquemment utilisées dans un SGBD relationnel. [8]

6. Exploitation des données

C'est la dernière étape d'un projet d'entrepôt de données, soit son exploitation. L'exploitation de l'entrepôt de données se fait par le biais d'un ensemble d'outils analytiques développés autour du Data Warehouse. Donc cette étape nécessite l'achèvement du

développement, ou de la mise en place, de ces outils qui peuvent accomplir les fonctions suivantes :

6.1 Le reporting

Les outils de reporting et de requêtes permettent la mise à disposition de rapports périodiques, pré-formatés et paramétrables par les opérationnels. Ils offrent une couche d'abstraction orientée métier pour faciliter la création de rapports par les utilisateurs eux-mêmes en interrogeant l'entrepôt de données grâce à des analyses croisées. Ils permettent également la production de tableaux de bord avec des indicateurs de haut niveau pour les managers, synthétisant différents critères de performance. [24]

6.2 Le Data Mining

Appelé aussi fouille de données ou bien extraction de connaissance à partir des données, est un domaine de recherche qui a pris son essor à partir des années 90. Le Data Mining est défini comme le traitement d'une grande quantité d'information afin d'y extraire des connaissances non triviales, implicites (non extraite manuellement), potentiellement utiles. Le Data mining permet 'extraction de corrélations, tendances et modèles par un passage au crible des bases de données volumineuses.

6.3 Outils OLAP

L'analyse multidimensionnelle est sans doute celle qui exploite et fait ressortir au mieux les capacités de l'entrepôt de données. Le but de l'analyse multidimensionnelle est d'offrir aux utilisateurs la possibilité d'analyser les données selon différents axes d'analyse (différentes dimensions) afin de confirmer une tendance ou suivre les performances de l'entreprise.

Cette analyse se fait selon le principe OLAP offrant, de ce fait, les possibilités de recourir à différentes opérations facilitant la navigation dans les données. La mise en place de ces outils est une option très intéressante dans la mesure où les données seront accessibles en analyses instantanées. Plusieurs fournisseurs de solution OLAP existent sur le marché et offrent des solutions construites sur des méthodes et technologies différentes.

C'est d'ailleurs pour cela que le choix de la solution doit se faire au préalable, selon les besoins en utilisation, la taille de l'entrepôt et les moyens techniques disponibles.

6.4 Tableaux de bord

« Le tableau de bord est un ensemble d'indicateurs peu nombreux conçus pour permettre aux gestionnaires de prendre connaissance de l'état et de l'évolution des systèmes qu'ils pilotent

et d'identifier les tendances qui les influenceront sur un horizon cohérent avec la nature de leurs fonctions » [9].

Les tableaux de bord sont un outil de pilotage qui donne une vision sur l'évolution d'un processus, afin de permettre aux responsables de mettre en place des actions correctives un exemple d'un tableau de bord est présentée dans la figure 5.

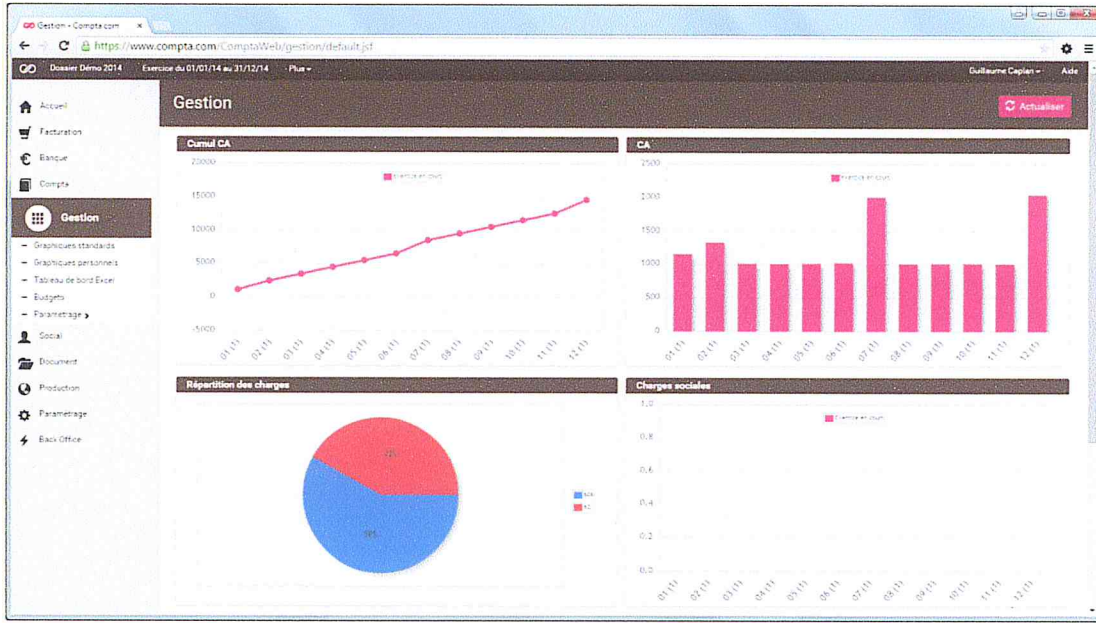


Figure 5 : Exemple d'un tableau de bord [24]

7. Conclusion :

Dans ce chapitre nous avons présenté de manière générale les systèmes d'information décisionnel, leur vocation par rapport aux systèmes SIO, structures de données et les approches de modélisation des entrepôts de données et enfin l'exploitation des données.

Chapitre II

Processus d'intégration des données

(ETL)

1. Introduction

Les technologies réseau actuelles (Intranet, Internet, etc.) permettent l'accès à une variété de source de données stockées dans différents emplacements physiques. Selon [10], le principal intérêt de ces technologies pour l'utilisateur est la possibilité d'accéder à une multitude de sources de données entretenant un certain rapport les unes avec les autres, et de combiner ces données afin d'obtenir l'information désirée. L'intégration de données a pour but, selon [11], de combiner les données réparties dans différentes sources et de fournir à l'utilisateur une vue unifiée de celles-ci grâce à un processus d'intégration chargé de l'Extraction, de Transformation et de Chargement (ETL : *Extract-Transform-Load*).

2. Définition du processus ETL

ETL, acronyme d'*Extraction-Transformation-Loading*, est un système de chargement de données depuis les différentes sources d'information de l'entreprise (hétérogènes) jusqu'à l'entrepôt de données (modèles multidimensionnels) comme le montre la figure 06. Ce système ne se contente pas de charger les données, il doit les faire passer par un tas de moulinettes pour les normaliser, les nettoyer, les contextualiser, puis de les charger de la façon adéquate. Chacune des étapes E, T et L du processus ETL sera définie de façon détaillée dans la suite de ce chapitre.

Il est important de savoir que la réalisation de l'ETL constitue 70% d'un projet décisionnel en moyenne Et ce n'est pas pour rien, ce système est complexe et ne doit rien laisser s'échapper, sous peine d'avoir une mauvaise information dans l'entrepôt pouvant induire en erreur les décideurs dans l'analyse de l'activité de l'entreprise. [19]

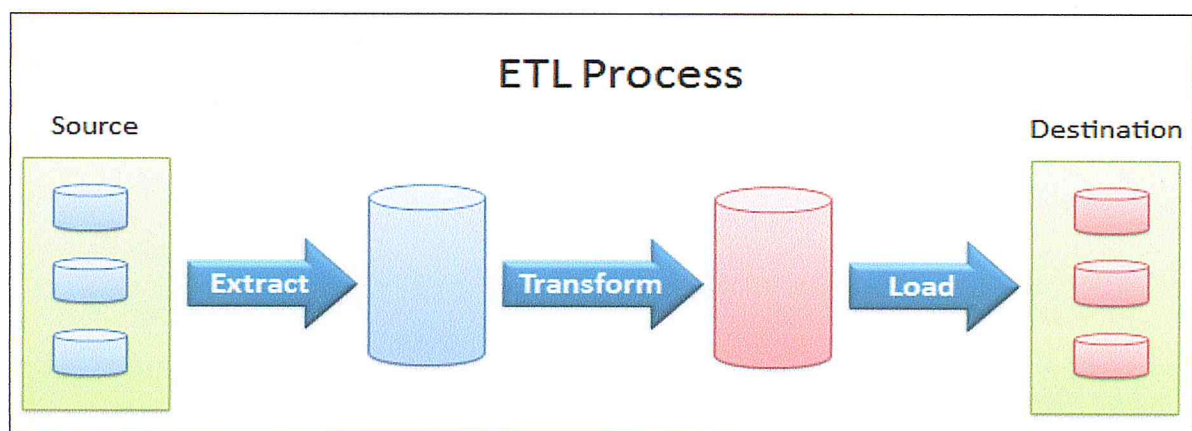


Figure 06 : Les trois phases ETL. [25]

2.1 Extraction

Le but de cette phase est de récupérer les données de production. Ces données ne sont pas forcément stockées dans une seule base de données. Les données pourront être issues de structures propriétaires, de logiciels, de systèmes de fichier, De plus, elles ne sont pas obligatoirement stockées au même endroit géographiquement (ex : siège social à Alger et succursale à Blida). En partant de ce constat, il est important de ne pas minimiser cette étape. Elle implique une très bonne connaissance des sources de données, afin de connaître la structure et la sémantique de chaque information.

Toutes les données sources n'ont pas systématiquement de l'intérêt pour la base de données décisionnelle. Le processus d'extraction aura également pour mission de filtrer les données utiles. [17]

Identification des sources :

- ❖ Énumérer les items cibles (métriques et attributs de dimension) nécessaires à l'entrepôt de données.
- ❖ Pour chaque item cible, trouver la source et l'item correspondant de cette source.
- ❖ Lorsque des données peuvent être extraites à partir de plusieurs sources, il faut choisir la source la plus pertinente.
- ❖ Si l'item cible exige des données de plusieurs sources, former des règles de consolidation.
- ❖ Si l'item source renferme plusieurs items cibles (Ex : un seul champ pour le nom et l'adresse du client), définir des règles de découpage.
- ❖ Inspecter les sources pour des valeurs creuses (manquantes). [17]

2.1.1 Extraction en temps-réel :

S'effectue au moment où les transactions surviennent dans les systèmes sources, la figure (07) représenté l'extraction en temps réel.

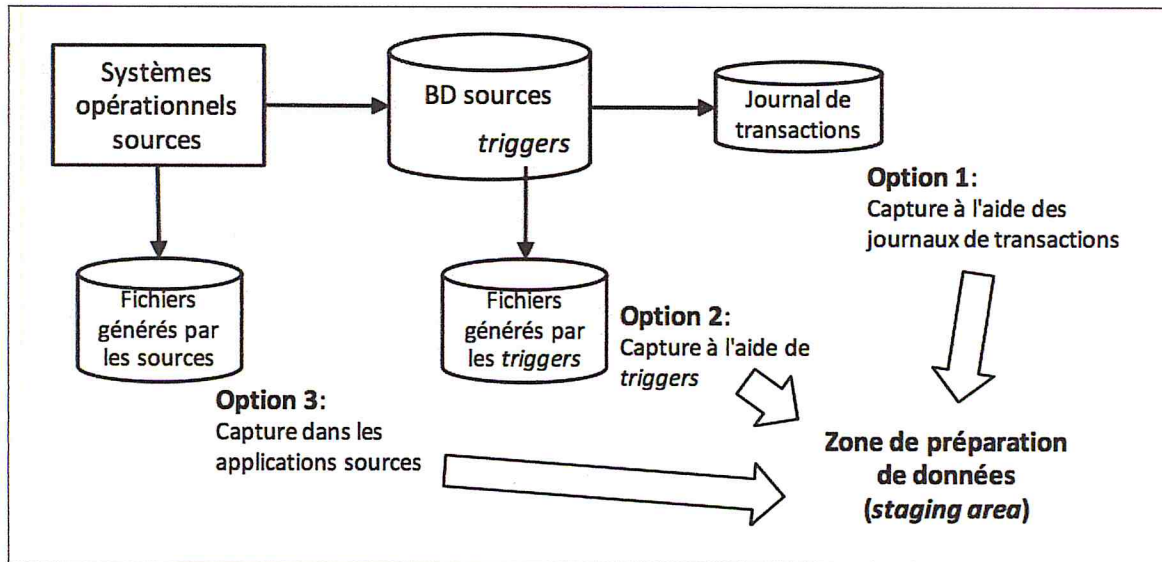


Figure 07 : Extraction temps-réel. [17]

Option 1 : Capture à l'aide du journal des transactions

- Utilise les logs de transactions de la BD servant à la récupération en cas de panne.
- Aucune modification requise à la BD ou aux sources.
- Doit être fait avant le rafraîchissement périodique du journal.
- Pas possible avec les systèmes legacy ou les sources à base de fichiers (il faut une BD journalisée). [17]

Option 2 : Capture à l'aide de triggers

- Des procédures déclenchées (triggers) sont définies dans la BD pour recopier les données à extraire dans un fichier de sortie.
- Meilleur contrôle de la capture d'évènements.
- Exige de modifier les BD sources.
- Pas possible avec les systèmes legacy ou les sources à base de fichiers. [17]

Option 3 : Capture à l'aide des applications sources

- Les applications sources sont modifiées pour écrire chaque ajout et modification de données dans un fichier d'extraction.
- Exige des modifications aux applications existantes.
- Entraîne des coûts additionnels de développement et de maintenance.
- Peut être employé sur des systèmes legacy et les systèmes à base de fichiers. [17]

2.1.2 Extraction différée :

Extrait tous les changements survenus durant une période donnée (ex : heure, jour, semaine, mois), comme le montre la figure 08.

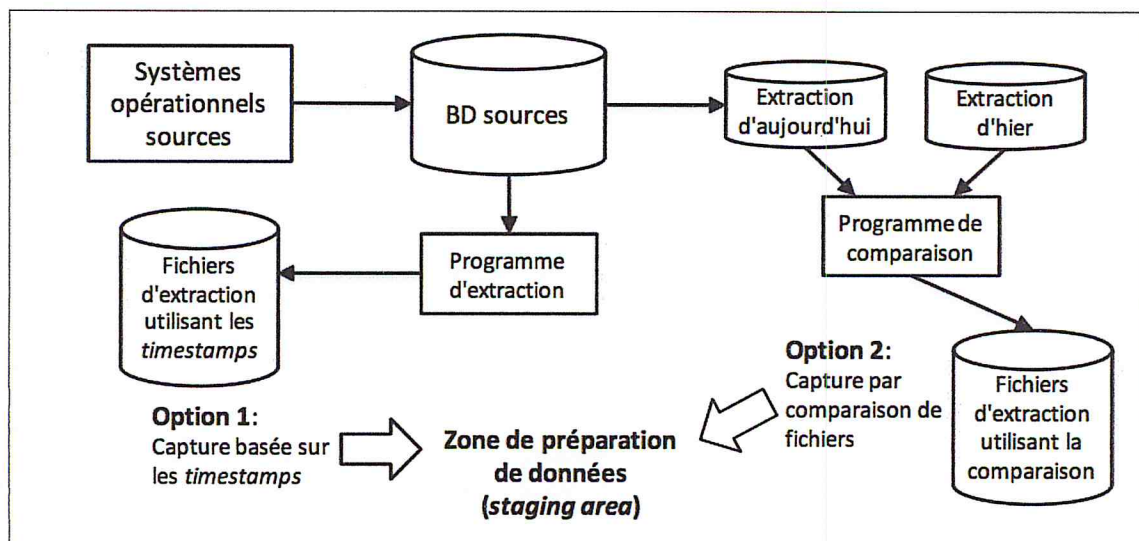


Figure 08 : Extraction différée. [17]

Option 1 : Capture basée sur les timestamps

- Une estampille (timestamp) d'écriture est ajoutée à chaque ligne des systèmes sources.
- L'extraction se fait uniquement sur les données dont le timestamp est plus récent que la dernière extraction.
- Fonctionne avec les systèmes legacy et les fichiers plats, mais peut exiger des modifications aux systèmes sources.

- Gestion compliquée des suppressions. [17]

Option 2 : Capture par comparaison de fichiers

- Compare deux snapshots successifs des données sources.
- Extrait seulement les différences (ajouts, modifications, suppressions) entre les deux snapshots.
- Peut être employé sur des systèmes propriétaires (legacy systems) et les systèmes à base de fichiers, sans aucune modification.
- Exige de conserver une copie de l'état des données sources.
- Approche relativement coûteuse. [17]

2.2 Transformation

La phase T de l'ETL travaille sur les données provenant de l'extraction. Elle a pour but de transformer les données afin de répondre à des contraintes d'ordre techniques, à des besoins fonctionnels et particulièrement à des objectifs en termes d'analyse. Les transformations les plus fréquentes sont le changement de monnaie ou la correction de casse sur un libellé. Nous pouvons aussi transformer des informations afin de les faire correspondre à une nomenclature, ce qui aura pour effet d'uniformiser les dimensions dans l'entrepôt de données. Lors de la transformation nous pouvons scinder ou consolider des données afin d'optimiser les futures requêtes. [17]

2.3 Chargement

La troisième phase (L) du processus d'alimentation est responsable de charger les données préparées dans les zones appropriées du schéma cible. Elle contrôle également l'intégrité des données. Elle pourra, le cas échéant, ajouter des données afin de respecter toutes les contraintes d'intégrité du modèle décisionnel. Par exemple si pour une vente nous n'avons pas la dimension géographique nous indiquerons « inconnue ». Le chargement est le garant de l'évolution des données. Lors de celui-ci, deux possibilités s'offrent à nous : soit historiser les changements des données, soit conserver les dernières modifications. [17]

3. Les structures de données ETL :

3.1 Sources de données :

L'ETL peut prendre en charge différentes natures de sources de données, tant en entrée qu'en sortie, les principales étant bien sûr les SGBD relationnels, les flux XML et il peut s'agir également de fichiers à formats fixes ou avec séparateurs (CSV), de fichiers Excel, etc. [27]

3.2 Bases de données relationnelles :

Les Bases de Données (BD) sont des données structurées. Ces dernières sont des informations organisées et classées en vue de faciliter leur lecture et leur traitement grâce aux SGBDs.

- **Fichiers plats :**

CSV : Le format CSV (*Comma Separated Values*) en français, « valeurs séparées par des virgules ». Il s'agit d'un format de fichiers ouvert qui permet de stocker les données d'un tableau. Chaque ligne du fichier correspond à une ligne du tableau, les colonnes sont en général séparées par des virgules. On peut très bien remplacer les virgules par des tabulations ou tout autre caractère. Ce format est utilisé pour échanger de manière interopérable des données de tableurs, bases de données, annuaires, etc. entre logiciels différents et/ou plateformes différentes. [16]

XML : Le langage XML, acronyme de *eXtended Markup Language*, est un format textuel qui permet de créer des documents contenant des données semi-structurées. Les données semi-structurées peuvent se voir comme une relaxation du modèle relationnel classique, un des fondements des bases de données traditionnelles, dans lequel on autorise une structure moins rigide et homogène des « champs de données ». Ce modèle de données c'est révélé très utile dans la représentation de familles de documents variés: multimédia, hypertexte, données scientifiques, etc. [16]

- **Les problèmes des sources de données :**

- 1) Sources diverses et disparates.
- 2) Sources sur différentes plateformes et système et d'exploitation (*OS : Operating Système*).
- 3) Applications basées sur des veilles technologiques (legacy) utilisant des BD et autres technologies obsolètes.
- 4) Historique de changement non-préservé dans les sources.
- 5) Qualité de données douteuses et changeantes dans le temps.

- 6) Structure des systèmes sources changeante dans le temps.
- 7) Incohérence entre les différentes sources.
- 8) Données dans un format difficilement interprétable ou ambigu. [17]

4. Zone de préparation des données « DSA » :

La zone de préparation des données (*DSA : Data Staging Area*) est une zone de stockage intermédiaire utilisée pour le traitement des données après l'extraction. La zone DSA se trouve entre la source de données et la base cible, qui sont souvent des entrepôts de données, datamart, ou d'autres référentiels de données. [28]

Ils peuvent être mises en œuvre sous la forme de tables dans les bases de données relationnelles, fichiers plats à base de texte (ou des fichiers XML) stockés dans les systèmes de fichiers ou des fichiers binaires formatés propriétaires stockés dans les systèmes de fichiers. [12]

L'architecture de la zone DSA varie en complexité d'un ensemble des tables relationnelles simples dans une base de données cible à des instances de base de données autonomes ou des systèmes de fichiers. [18]

❖ Avantages :

Voici quelques raisons pour lesquelles il est préférable d'utiliser la zone de stockage temporaire :

- Extract Once, transform many : une seule extraction des systèmes sources, et autant de transformations et de chargement dans le DSA.
- Ne pas impacter le fonctionnement des systèmes sources, surtout dans les cas des systèmes 24/7. Le temps nécessaire à extraire les données, devrait être donc minimal. Pour se faire, la transformation ne doit pas se faire en même temps que l'opération d'extraction : on extrait les données le plus rapidement possible et on les met ensuite dans un DSA où s'effectuent les transformations.
- En cas de problèmes de transformation (plantage lors de la transformation, erreur BD...), on ne sera pas obligé de refaire l'extraction, du moment où la source de la transformation est le DSA.

5. Conclusion

Dans ce chapitre, nous avons présenté les différentes approches d'intégration de données, en particulier, celle basée sur l'entrepôt de données (l'approche matérialisée). Dans cette approche nous avons exposé de manière détaillée le processus ETL.

Chapitre III

Analyse de données en ligne (OLAP)

1. Introduction

En informatique, et plus particulièrement dans le domaine des bases de données, le traitement analytique en ligne (OLAP : On Line Analytical Processing) est un type d'applications informatiques orientées vers l'analyse sur-le-champ d'informations selon plusieurs axes, dans le but d'obtenir des rapports de synthèse tels que ceux utilisés en analyse financière. Les applications de type OLAP sont couramment utilisées en informatique décisionnelle, dans le but d'aider la direction à avoir une vue transversale de l'activité d'une entreprise. [28]

2. Définition

On Line Analytical Processing est un environnement d'accès aux données en lecture uniquement. Les programmes fonctionnant sous de tels environnements travaillent sur de très grandes quantités de données, ce qui rend les analyses très complexes. Le système OLAP regroupe l'information provenant de diverses sources. Il les regroupe, les intègre, les stocke, tout ceci afin de donner une vue métier à l'utilisateur. Cette vue métier va l'aider à retrouver l'information rapidement. Une notion importante est l'historisation des données au sein des bases OLAP. Ceci entraîne, avec une architecture différente, une grandeur de base de données supérieure aux bases de données classiques.

En 1993 E.F. Codd a défini dans « Providing OLAP to user-analyst » le concept d'OLAP. Il a mis en évidence 12 règles [5] que doivent respecter les bases pour être OLAP :

1. **Multidimensional Conceptual View** (Vue conceptuelle multidimensionnelle) Permet d'avoir une vision multidimensionnelle des données. L'inverse se nomme les tables unidimensionnelles.
2. **Transparency** (Transparence) L'utilisateur ne doit pas se rendre compte de la provenance des données si celles-ci proviennent de sources hétérogènes. Ces sources peuvent provenir des bases de données de production, de fichiers à plats,
3. **Accessibility** (Accessibilité) L'utilisateur doit disposer d'un accès aux données provenant de sources multiples en faisant abstraction des conversions et extractions de celles-ci.

4. **Consistence Reporting Performance** (Performance continue dans les rapports) Les performances ne doivent pas être diminuées lors de l'augmentation du nombre de dimensions ou lors de l'augmentation la taille de la base de données, mais doivent être proportionnelles à la taille des réponses retournées.
5. **Client-Server Architecture** (Architecture client-serveur) Il est essentiel que le produit soit client-serveur. Le serveur stocke les données et le client les restitue.
6. **Generic Dimensionality** (Dimensionnement générique) Chaque dimension doit être équivalente par rapport à sa structure et à ses capacités opérationnelles pour ne pas fausser les analyses.
7. **Dynamic Sparse Matrix Handling** (Gestion dynamique des matrices creuses) Certaines cellules de l'hypercube peuvent être vides. Elles doivent être stockées de manière à ne pas détériorer les temps d'accès.
8. **Multi-User support** (Support multi-utilisateurs) Les outils OLAP doivent fournir des accès concurrents, l'intégrité, la sécurité et la gestion des mises à jour.
9. **Unrestricted Cross-dimensional Operations** (Opération non restrictive entre les dimensions) Les calculs doivent être possibles à travers toutes les dimensions qui sont régies par les règles de gestion. Toutes les tranches de cube doivent être visualisées.
10. **Intuitive Data Manipulation** (Manipulation intuitive des données) La manipulation des données se fait directement à travers les cellules d'une feuille de calcul, sans recourir aux menus ou aux actions multiples. Au final, il doit permettre l'analyse intuitive dans plusieurs dimensions.
11. **Flexible Reporting** (Flexibilité dans la création des rapports) La création des rapports ou des graphiques se doit d'être simple et efficace pour les utilisateurs.
12. **Unlimited Dimensions & Aggregation Levels** (Nombre illimité de niveaux d'agrégation et de dimensions) Dimensions et niveaux d'agrégation illimités, afin d'autoriser les analyses les plus pointues. [5]

Ces 12 règles ont pour but de normaliser une base de données décisionnelle. Cette base de données peut être un système de gestion de base de données relationnelle ou multidimensionnelle, respectivement SGBDR et SGBDM.

Dans les SGBDM, le stockage des données se base sur le principe des hypercubes. Un hypercube est une matrice décisionnelle avec au minimum quatre dimensions d'analyse. Nous parlons également du cube, qui est une matrice décisionnelle avec trois dimensions.

Le concept d'OLAP est décliné en plusieurs « sous concepts » qui orientent la structure physique des données ou les techniques de traitements. [29]

L'agrégation des données est une notion fondamentale de l'OLAP : afin d'optimiser les temps de calcul, les membres d'une même dimension peuvent être organisés en une hiérarchie de relations 1-n, arborescence comparable aux arbres logiques. Les opérations d'agrégation les plus fréquentes sont la somme et la moyenne, elles permettent de calculer rapidement les valeurs associées aux positions parents. Par exemple, on peut décomposer le temps suivant une hiérarchie à trois niveaux : les années au niveau 1, les trimestres au niveau 2, les mois au niveau 3 (voir Figure 09). A noter qu'une dimension peut se décomposer suivant plusieurs hiérarchies : l'année pourrait se décomposer en mois puis en jours. L'agrégation des données multidimensionnelles permet de mémoriser le résultat des calculs dans la base et d'augmenter ainsi les performances lors de l'interrogation. En contrepartie, l'agrégation nécessite un espace de stockage conséquent. Par opposition, la notion de formule correspond à un calcul effectué à la volée, le résultat n'est pas stocké en base. [29]

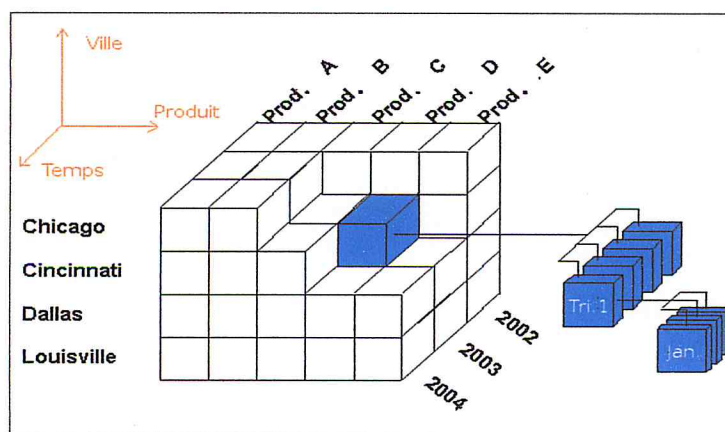


Figure 09 : un cube OLAP et le principe d'agrégation. [29]

3. Comparaison entre OLAP et OLTP

3.1 Définition de OLTP

OLTP (On Line Transaction Processing) est le modèle utilisé par les SGBD. Le mode de travail est transactionnel. L'objectif est de pouvoir insérer, modifier et interroger rapidement et en sécurité la base. Ces actions doivent pouvoir être effectuées très rapidement par de nombreux utilisateurs simultanément. Chaque transaction travail sur de faibles quantités d'informations, et toujours sur les versions les plus récentes des données. [32]

3.2 Tableau comparatif

Caractéristiques	OLTP	OLAP
Utilisation	SGBD (base de production)	Datawarehouse
Opération typique	Mise à jour	Analyse
Type d'accès	Lecture écriture	Lecture
Niveau d'analyse	Elémentaire	Global
Quantité d'information échangées	Faible	Importante
Orientation	Ligne	Multidimension
Taille BD	Faible (max qq GB)	Importante (pouvant aller à plusieurs TB).
Ancienneté des données	Récente	Historique

Tableau 04 : un tableau récapitulatif des différences entre OLTP et OLAP. [32]

4. Les différents outils OLAP

OLAP sera l'outil à privilégier pour les données quantitatives si leur structuration a priori est naturelle (cas rencontré fréquemment pour les applications financières ou commerciales), alors que le requêteur sera idéal pour les données qualitatives et pour toute

analyse impromptue nécessitant l'autonomie de l'utilisateur (cas rencontré fréquemment pour le marketing ou la gestion du personnel). Si les besoins sont à combiner, il faudra choisir entre la richesse fonctionnelle apportée par plusieurs outils interfacés ou l'homogénéité des outils intégrés.

4.1 M-OLAP

MOLAP est conçue exclusivement pour l'analyse multidimensionnelle, avec un mode de stockage optimisé par rapport aux chemins d'accès prédéfinis. Ainsi, toute valeur d'indicateur associée à l'axe temps sera pré-calculée au chargement pour toutes ses valeurs hebdomadaires, mensuelles, etc. [14]

La figure 10, présente l'architecture d'un produit MOLAP :

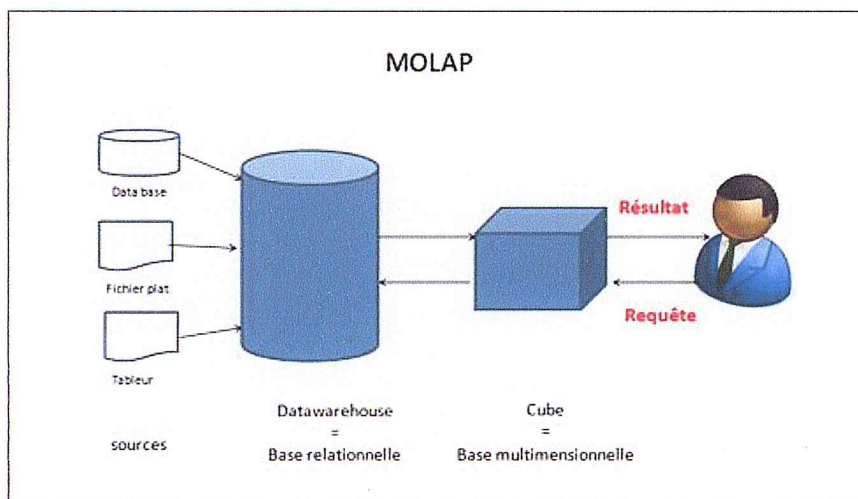


Figure 10 : Architecture d'un produit MOLAP.

MOLAP agrège tout par défaut. Plus le volume de données à gérer est important, plus les principes d'agrégations implicites proposés par MOLAP sont pénalisants dans la phase de chargement de la base, tant en terme de performances que de volume. La limite fréquemment évoquée pour MOLAP étant de quelques giga octets. MOLAP surpasse ROLAP pour des fonctionnalités avancées comme la prévision ou la mise à jour des données pour la simulation. Cependant, ces différences s'expliquent par une plus grande maturité en faveur de MOLAP, concept qui date de près de vingt ans. MOLAP est incompatible avec d'autres modes d'accès aux

données. Si MOLAP doit cohabiter avec d'autres techniques d'accès aux données (par requêteur, par data mining, etc.), deux bases de données doivent cohabiter. En effet, MOLAP repose sur un moteur spécialisé, qui stocke les données dans un format tabulaire propriétaire (cube). Pour accéder aux données de ce cube, on ne peut pas utiliser le langage de requête standard SQL, il faut utiliser une API spécifique. Le marché des bases MOLAP étant plus réduit, il est plus difficile pour les éditeurs qui le représentent d'investir sur de telles évolutions. [14]

4.2 R-OLAP

Les outils ROLAP superposent au-dessus des SGBD/R bidimensionnels un modèle qui représente les données dans un format multidimensionnel. Ces produits diminuent sensiblement le coût lié à la mise en œuvre d'un serveur de base de données multidimensionnelle supplémentaire. Au travers des métadonnées, ils permettent de transformer l'analyse multidimensionnelle demandée par l'utilisateur en requêtes SQL. Pour cela, ces outils s'appuient pour la plupart sur une modélisation particulière des données, distinguant les axes d'analyse et les faits à observer. On parlera notamment de modèle en étoile et de modèle en flocon ou encore des techniques de définition physique d'agrégations. Ceci oblige à définir le modèle en fonction de l'outil à utiliser et des analyses à mener mais est un gage de performance et de cohérence lors de l'utilisation de ce type de produits. Cette contrainte exige un travail important des équipes informatiques et donc enlève beaucoup à l'intérêt d'utiliser un SGBD Relationnel comme support de stockage pour l'analyse multidimensionnelle. [13]

La figure 11 : présente une architecture du produit R-OLAP :

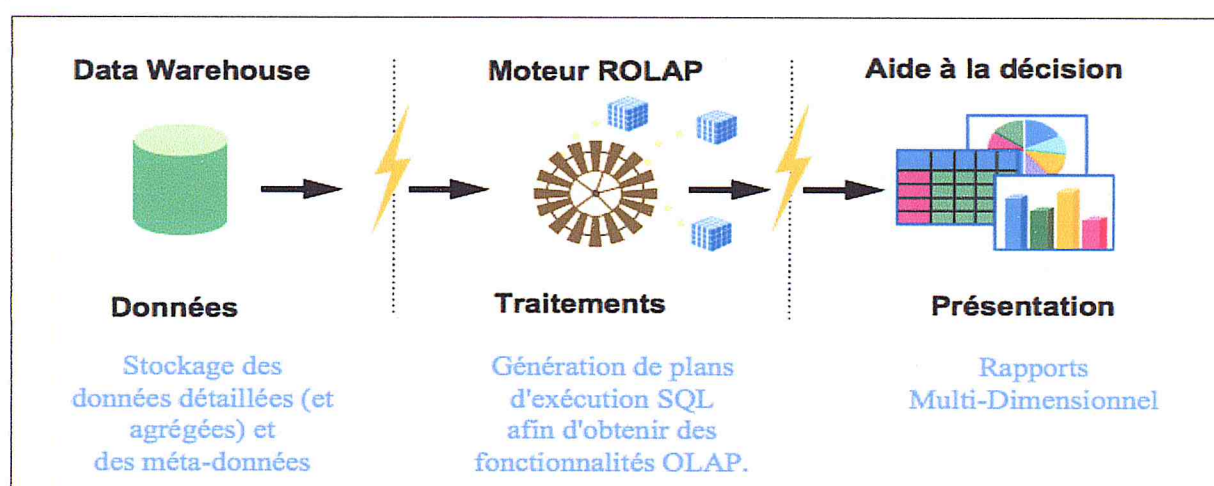


Figure 11 : Architecture d'un produit ROLAP. [14]

Les outils ROLAP, proposent le plus souvent un composant serveur, pour optimiser les performances lors de la navigation dans les données ou pour les calculs complexes. Avec ROLAP, il est déconseillé d'accéder en direct à des bases de données de production pour faire des analyses sérieuses, pour des raisons de performances. ROLAP n'agrège rien, mais tire parti des agrégats s'ils existent. De ce fait ROLAP est plus lourd à administrer que MOLAP, puisqu'il demande de créer explicitement certains agrégats. Certains éditeurs, comme Informai avec Méta cube ou Oracle avec Discovere 2000, pallient cependant à cette faiblesse avec des outils d'administration aptes à conseiller pour une politique d'agrégation adéquate. ROLAP est donc mieux adapté aux gros volumes. En s'appuyant sur les bases relationnelles, référence du marché, ROLAP tire parti des évolutions de celles-ci (adaptation aux architectures hardware sophistiquées, extensions objets, etc.). [13]

4.3 H-OLAP

Hybrid on line analytical processing. H-OLAP utilise R-OLAP et M-OLAP en fonction des données qu'il traite. Sur les données agrégées il utilise M-OLAP (voir figure 12), par contre sur les données plus détaillées, il utilise R-OLAP.

Un Exemple de restitution des données : Les données sont stockées dans un CUBE puis on fait la restitution via un outil de reporting comme BO (Business Objects) par exemple. L'utilisateur pourra donc avoir accès à un rapport contenant les données issues du CUBE ainsi qu'à un autre rapport détaillé contenant les données en provenance des tables. [30]

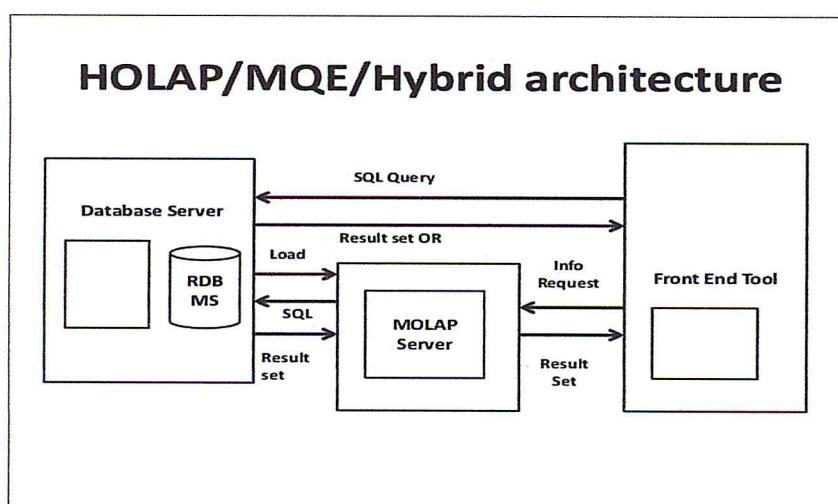


Figure 12 : Architecture d'un produit H-OLAP. [31]

4.4 S-OLAP

Spatial On Line Analytical Processing. S-OLAP est une plateforme visuelle pour l'exploration et l'analyse spatio-temporelle. Ceci dans le but de présenter les données sous une autre forme que celle tabulaire. [31]

4.5 D-OLAP

D-OLAP se définit par Desktop OLAP. La base DOLAP est une base OLAP très limitée en taille, hébergée sur le poste client. Elle est bien entendu très rapide.

L'architecture d'un produit D-OLAP, est présentée dans la figure 13 :

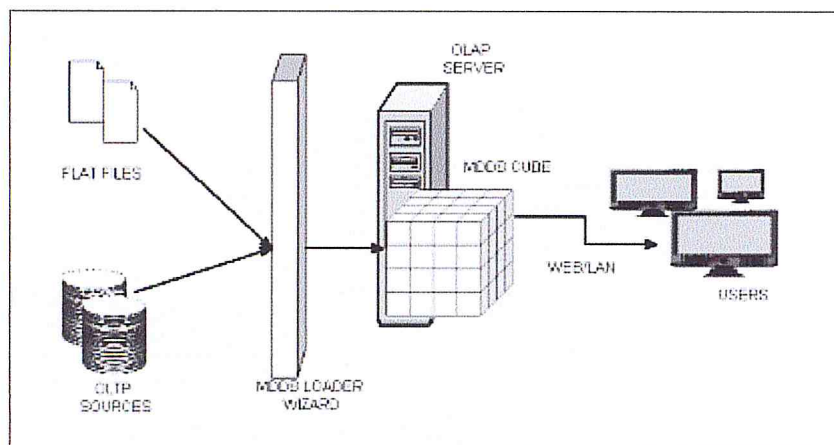


Figure 13 : Architecture d'un produit D-OLAP. [31]

5. Conclusion

Dans ce chapitre, nous avons présenté les technologies OLAP qui sont des outils de Business Intelligence BI. Ils sont facilement intégrables dans les entreprises, ils facilitent le Reporting et le traitement de la donnée. Ils peuvent être utilisés à petite échelle ou à moyenne échelle contrairement aux outils de Big data qui sont réservés pour le traitement d'une masse d'informations. Lorsque l'on fonctionne avec les outils OLAP on ne se sent pas à la ramasse car les requêtes conservent surtout pour le ROLAP une logique SQL.

Partie 2

Conception de l'entrepôt de données DW_CLIENT

Chapitre IV

Présentation de l'organisme d'accueil

1. Introduction

Nous entamons le premier chapitre de cette partie qui est la présentation de l'organisme d'accueil. Nous avons été accueillis par le groupe SONELGAZ, au niveau de la filiale El Djazaïr Information Technology (ELIT), Dans ce chapitre, nous allons commencer par un aperçu global du groupe SONELGAZ, son historique ainsi que son organisation. Puis, nous présenterons ELIT, la filiale informatique du groupe en mettant l'accent sur la Direction Système d'Information Activités Distribution / Gestion de Réseaux (DSID) (où nous avons effectué notre stage.

1.1 Le groupe SONELGAZ

1.1.1 Présentation du groupe

SONELGAZ, (Société nationale de l'électricité et du gaz) est le groupe leader de la production et de distribution de l'électricité et du gaz en Algérie.

Crée en 1947 sous le nom 'Électricité et gaz d'Algérie (EGA)' et rebaptisée SONELGAZ. En 1969, cette compagnie était toujours le monopole de la production et de la distribution de l'électricité et du gaz naturel en Algérie.

Le groupe occupe une position privilégiée dans l'économie du pays en tant que responsable de l'approvisionnement de plus de sept millions de ménages en électricité et de trois millions en gaz naturel, soit une couverture géographique de près de 0.99 en taux d'électrification et 0.55 pour la pénétration gaz.

1.1.2 Organisation du groupe

1.1.2.1 Holding SONELGAZ

En 2012, SONELGAZ spa a été érigée en «société holding» sous la dénomination de «SONELGAZ» qui, avec ses sociétés filiales formeront un ensemble dénommé «Groupe SONELGAZ».

Les statuts révisés de SONELGAZ ont pour objectif de permettre une meilleure clarification des responsabilités entre SONELGAZ et ses entités en ce sens qu'ils délèguent à SONELGAZ sa véritable vocation de société holding alors que, les activités de production, de transport et de distribution de l'électricité et du gaz seront dévolues légalement et statutairement aux filiales.

D'autre part, SONELGAZ conserve le rôle de détenteur du portefeuille des actions

constituant le capital social de ses filiales.

Sur un autre registre, les conseils d'administration des filiales, constitueront les relais permettant à la société holding de suivre et d'orienter le pilotage des filiales.

1.1.2.2 Structure du groupe

Le Groupe SONELGAZ a mis en place, depuis plusieurs années, une nouvelle organisation en optant pour une approche industrielle dans le domaine de l'énergie. C'est ainsi que le Groupe a soutenu les filiales qui ont fait de lui un acteur majeur dans le développement du pays, et de créer les conditions nécessaires pour l'impulsion d'une intégration industrielle.

Actuellement, le groupe SONELGAZ est constitué de la holding SONELGAZ et de 40 filiales et 7 sociétés en participation (voir figure 14)

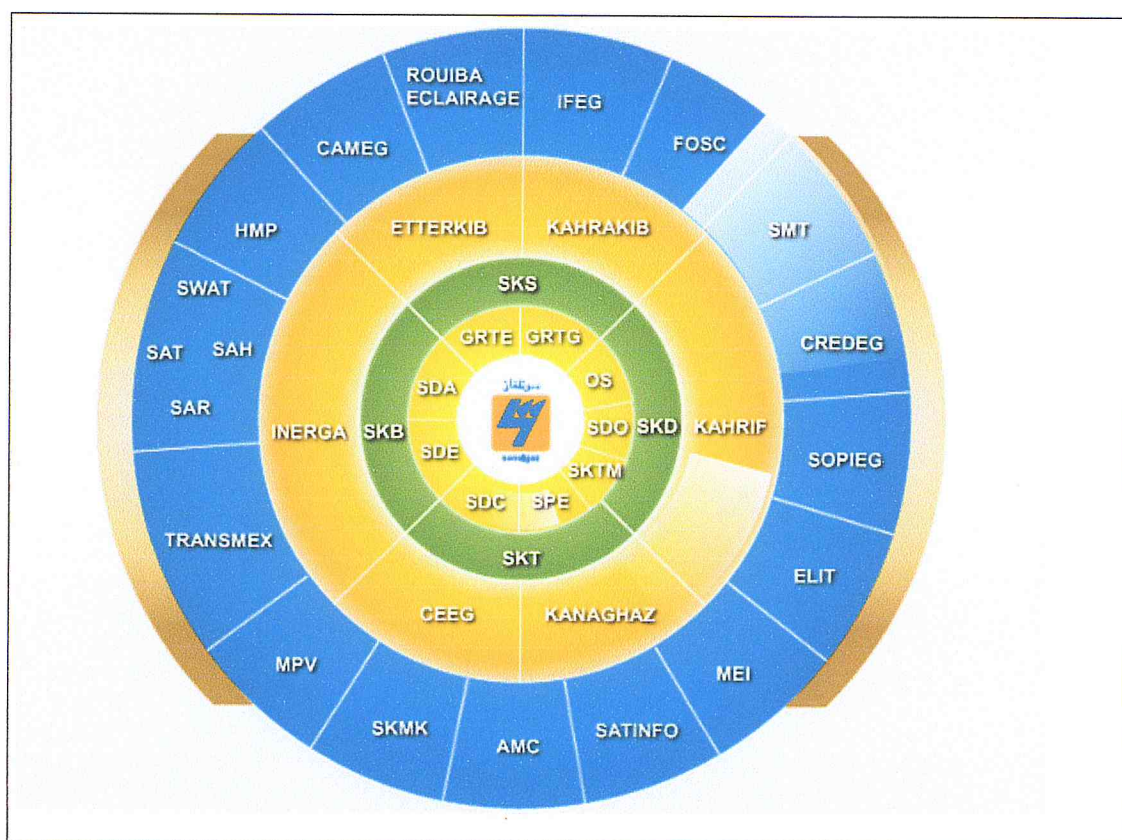


Figure 14 : Structure du groupe SONELGAZ

➤ **Maison mère (La holding Sonelgaz) :**

Dans le cadre des nouvelles orientations, la holding assure le pilotage du groupe, à travers l'exercice des missions de politique et stratégie.

- **Filiales métiers de bases :** Au nombre de Huit (08), ces filiales assurent la production, le transport et la distribution de l'électricité ainsi que le transport et la distribution du gaz par canalisation.
- **Filiales Travaux :** Ce sont des filiales spécialisées dans le domaine de la réalisation des infrastructures énergétiques (engineering, montage industriel, réalisation de réseaux...) et c'est grâce à elles que l'Algérie dispose aujourd'hui d'infrastructures électriques et gazières répondant aux besoins du développement économique et social du pays.
- **Filiales métiers périphériques :** Elles ont en charge d'activités annexes, telles que la maintenance d'équipements énergétiques, la distribution de matériel électrique et gazier, le transport, les manutentions exceptionnelles, la recherche et développement, la formation, etc. Les sociétés de participation : SONELGAZ détient également des participations dans des sociétés, dont le métier est en rapport avec le domaine de l'électricité et du gaz.

1.2 El Djazaïr Information Technology (ELIT)

1.2.1 Présentation d'ELIT

Le 1er janvier 2009, l'activité Systèmes d'Information, confiée à la direction générale au niveau de la Maison Mère du Groupe SONELGAZ, a été érigée en Société par actions, dénommée "EL Djazaïr Information Technology", par abréviation "ELIT SPA". Cette filiale a été créée pour répondre à :

- La stratégie du Groupe SONELGAZ de développer des moyens propres de maîtrise d'œuvre dans le domaine des Systèmes d'Information, et de disposer d'un pôle de compétences technologiques au service de ses sociétés.
- La volonté du Groupe SONELGAZ de confier la propriété des Systèmes d'Information à une entité spécialisée et de focaliser les capacités de ses sociétés sur leurs métiers de base respectifs.

1.2.2 Les missions d'ELIT

La Société El Djazaïr Information Technology – ELIT. Spa, est chargée de définir et de mettre en œuvre la politique générale du Groupe SONELGAZ concernant les systèmes d'information et les technologies de l'information et de la communication. Les missions assignées à ELIT se déclinent à deux niveaux, stratégique et opérationnel. Sur le plan stratégique, ELIT contribue à la stratégie du Groupe SONELGAZ par :

- L'élaboration de la politique des Systèmes d'Information et des technologies de l'information et de la communication du Groupe SONELGAZ.
- La prise en charge des besoins des sociétés du Groupe SONELGAZ en matière d'informatique et de télécommunications. Quant au niveau opérationnel, ELIT s'emploie à :
- Elaborer et mettre en œuvre les Systèmes d'Information destinés au pilotage et à la gestion des différentes activités des sociétés du Groupe SONELGAZ.
- Mettre à la disposition des sociétés du Groupe SONELGAZ les moyens informatiques et de télécommunications (logiciels, matériels, infrastructures, etc.) nécessaires pour assurer le niveau de service attendu.
- Assurer la maintenance et l'administration des Systèmes d'Information, des plateformes et des équipements mis à la disposition des utilisateurs.
- Assurer l'accès à l'information et aux applications et en garantir la sécurité, l'intégrité et la fiabilité.
- Mettre à la disposition des utilisateurs l'expertise technique indispensable à la satisfaction de leurs besoins.
- Proposer, à terme, tous les services IT construits pour les sociétés du groupe aux clients externes.

1.2.3 Structure d'ELIT

Comme toute autre filiale de groupe SONELGAZ, ELIT a sa propre mono-structure relativement indépendante des autres filiales. Cette structure est décrite par l'organigramme

de la figure 15, présentée ci-dessous :

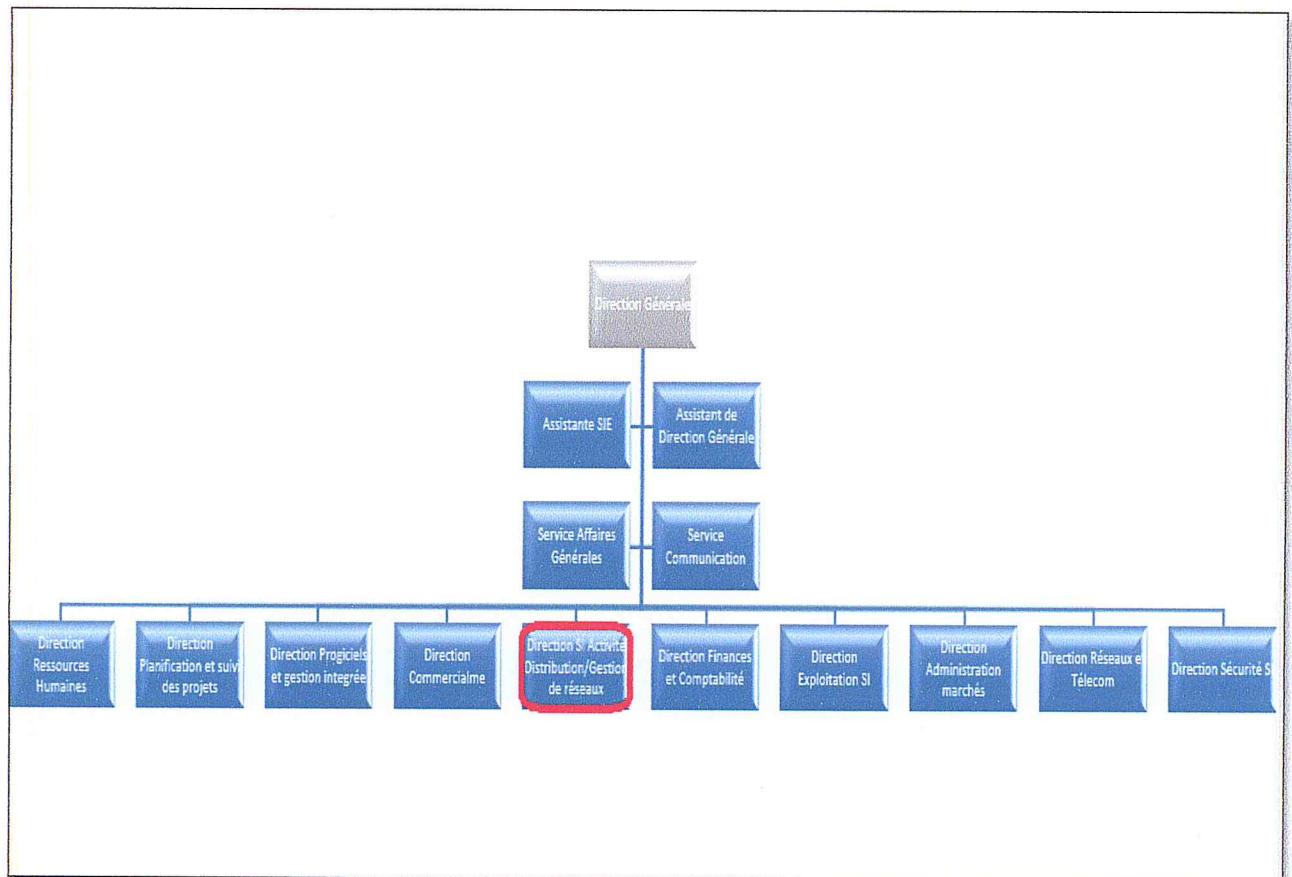


Figure 15 : Structure d'ELIT

ELIT dispose donc d'une direction générale, d'assistants, de secrétariats, de structures métier et de structures support.

1.2.4 Structure d'accueil (la DSID)

Notre stage fin d'étude a été effectué au niveau de la direction SI activités distribution/gestion de réseaux abrégée en DSID. Plus précisément, au niveau de département développement SI distribution.

1.2.4.1 Organisation de la DSID

La DSID possède trois départements comme le montre l'organigramme présenté ci-dessous (figure 16) :

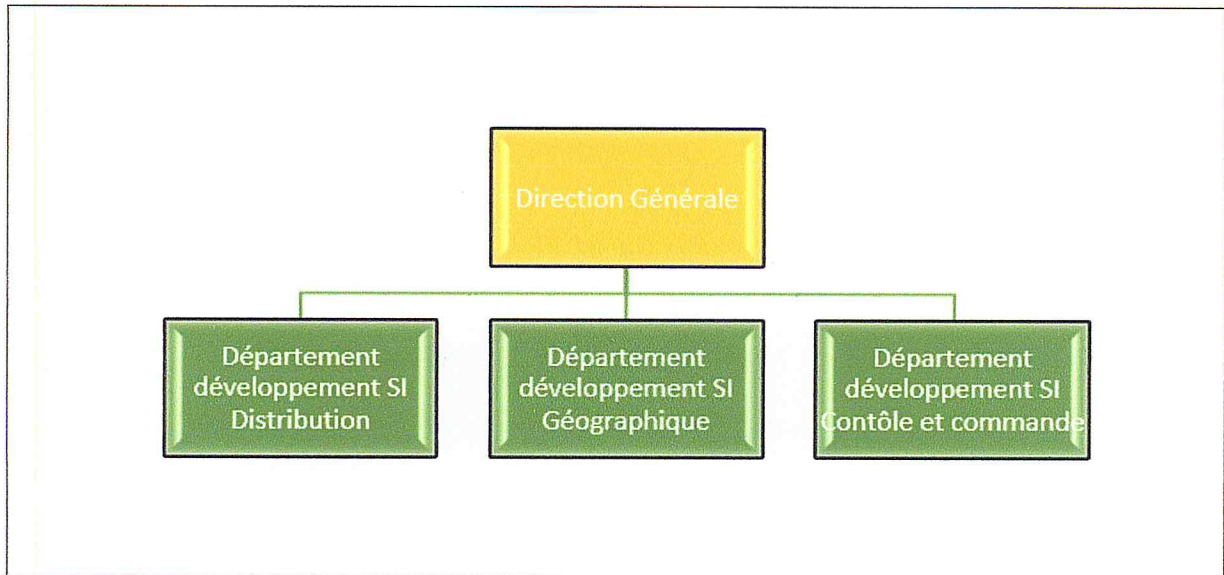


Figure 16 : Structure de la DSID

1.2.4.2 Missions et attributions de la DSID

La Direction SI Activités Distribution / Gestion de Réseaux est chargée de la production, du déploiement et de la maintenance des systèmes d'information, d'abord des métiers de la Distribution, mais également ceux des activités qui supposent l'utilisation d'un réseau ou reposant sur des infrastructures en réseaux et pouvant intégrer des systèmes informatiques embarqués et/ou itinérants.

Il s'agit, entre autres, des activités, nécessitant la mise en place de systèmes intelligents (Smart Grids) liées au transport et à la distribution de l'électricité et du gaz. Ses attributions se résument dans les points suivants :

- Étudier les besoins en systèmes d'information des Sociétés chargées du transport et de la distribution de l'électricité et du gaz, d'un site de production jusqu'aux consommateurs finals.
- Organiser et planifier la réalisation des projets SI, depuis leur conception jusqu'à leur achèvement, (Conception, développements, tests, intégration, migration de données, etc.), en s'appuyant sur des compétences internes ou externes.

- Assurer la maintenance corrective et évolutive des SI développés.
- Assurer la veille technologique.

1.3 Les sociétés de distribution SONELGAZ

1.3.1 Présentation des sociétés de distribution

L'un des métiers les plus importants du groupe, et dans lequel s'inscrit notre projet, est la fourniture et la distribution de l'énergie électrique et gazière. Ce métier, vu l'organisation du groupe, est assuré par quatre filiales qui sont les « Sociétés de Distribution ».

Les sociétés sont :

- Sociétés de Distribution d'Electricité et du Gaz de l'Ouest (SDO).
- Sociétés de Distribution d'Electricité et du Gaz du Centre (SDC).
- Sociétés de Distribution d'Electricité et du Gaz d'Alger (SDA).
- Sociétés de Distribution d'Electricité et du Gaz de l'Est (SDE).

1.3.2 Missions des sociétés de distribution

Les Sociétés de Distribution d'Electricité et du Gaz sont chargées de :

- Exploiter et maintenir les réseaux de distribution de l'électricité et du gaz sur le périmètre de la société en question.
- Développer et moderniser les réseaux électricité et gaz.
- Garantir un accès non discriminatoire aux réseaux électricité et gaz.
- Garantir la qualité et la continuité de service dans la distribution de l'électricité et du gaz.
- Garantir la sécurité des réseaux électricité et gaz.
- Commercialiser l'électricité et du gaz.

1.3.3 Organisation des sociétés de distribution

Une société de distribution possède la structure décrite par l'organigramme présenté dans ce qui suit (figure 17) :

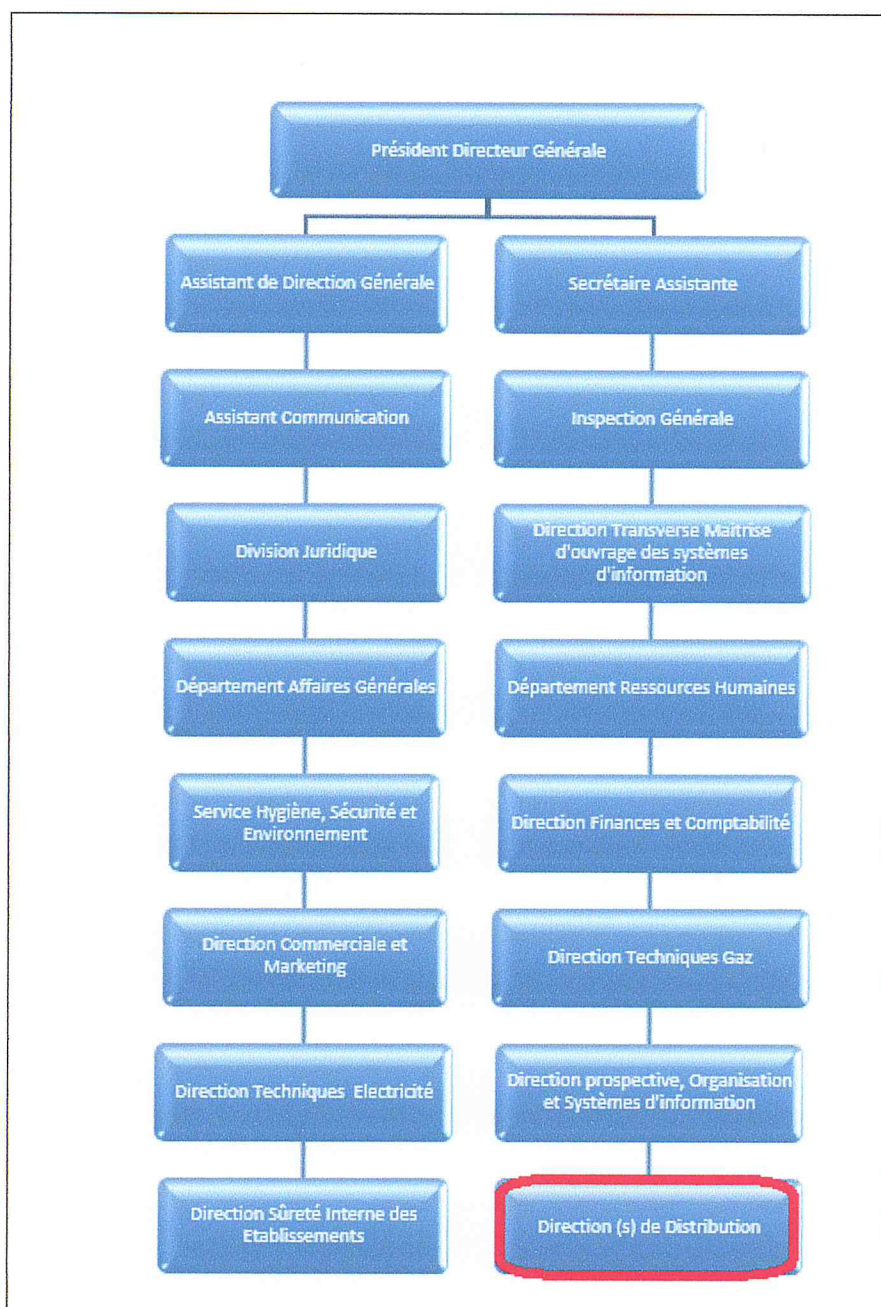


Figure 17 : Organisations des SD

Chaque société de distribution compte huit directions centrales (DC) situés au niveau de son siège ainsi qu'un nombre de directions de distribution (DD) Afin de comprendre la

hiérarchie des structures métier associées à une SD, nous avons conçu l'organigramme ci-dessous, (voir figure 18).

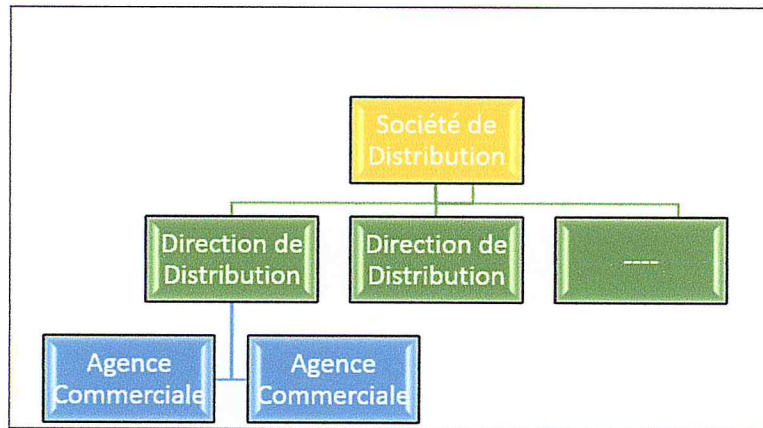


Figure 18 : Hiérarchie de la SD

Les directions de distribution sont des directions régionales qui assurent les fonctions métier de la distribution d'électricité et du gaz. Chaque DD possède un nombre d'agences commerciales (AC) qui présentent le point de contact direct avec le client.

1.3.4 La clientèle de la distribution

Les clients sont classifiés selon le type de l'énergie et la puissance d'alimentation comme suit:

En électricité on distingue trois types de clients selon la tension :

- Client « Basse Tension-BT » : jusqu'à une puissance de 40 KVA. L'abonné est considéré comme un client BT. La tension délivrée est 220 V ou 380 V.
- Client « Haute Tension A-HTA » : d'une puissance de 50 KVA jusqu'à 360 KVA. La tension d'alimentation est de 10 KV ou 30KV.
- Client « Haute Tension B-HTB » : la tension d'alimentation dépasse 60KV En gaz, les clients sont classifiés selon la pression.
- Client « Basse Pression-BP » : tout client alimenté sous une pression de 21 bars à travers un détendeur.
- Client « Haute Pression-HP » : tout client alimenté sous une pression supérieure à

4bars avec un poste de détente gaz. En BT/BP (basse tension et basse pression), on distingue plusieurs types de clients.

- Clients ordinaires.
- Clients FMS (Facturation Multi-sites).

2. Conclusion

A travers ce chapitre, nous avons abordé l'aspect organisationnel du groupe SONELGAZ, de la filiale ELIT et des sociétés de distribution du groupe. Nous avons vu leurs organisations, structures et missions. Le projet qui nous a été confié, est la conception et la réalisation d'un système décisionnel pour le système de la gestion de la clientèle des sociétés de distribution du groupe.

Chapitre V

ÉTUDE DE L'EXISTANT

1. Introduction

L'étude de l'existant est une étape importante pour aborder tout projet. Elle nous permet de prendre connaissance du système opérationnel existant d'une manière apprise. Cette étude s'étend sur le périmètre du projet, à savoir les processus facturation, raccordement et relève pour les clients (BT-BP-HTA-HP-HTB-MP) des sociétés de distribution de Sonelgaz.

Nous présenterons le Système de Gestion de la Clientèle (SGC) de façon générale avant de mettre l'accent sur le module facturation, raccordement et relève. Nous nous intéresserons, dans une deuxième étape, à l'existant décisionnel relatif à la gestion de la clientèle et étudier, ensuite, la procédure de Reporting adoptée actuellement. Nous clôturons ce chapitre par bilan du système actuel.

2. Existant opérationnel

Il existe plusieurs logiciels opérationnels au niveau des agences pour la gestion de la clientèle développée par ELIT ou par les ingénieurs des sociétés de distributions (SD) : ancienne version de SGC, Portails Web des SD, Plateforme Saturne dédiée pour la gestion de l'auto-relève, Consulter-Facture, Réclamation-Client et enfin la Gestion des Signalés.

L'ancienne version du SGC présente quelques limites que nous résumons de la manière suivante :

- Le projet actuel de la refonte de SGC vise à regrouper toutes les fonctionnalités des applications opérationnelles qui sont éparpillés ce qui provoque une duplication des données.
- Problème d'homogénéité et de normalisation de données d'où la refonte du SCC.
- Centralisation La date de début de la refonte SGC est 2015.

Dans ce qui suit, nous présentons la nouvelle version du SGC.

2.1.Système de Gestion de la Relation Client (CRM)

Les deux dimensions collaborative et opérationnelle du CRM ont été complètement développées. Il reste, cependant, à finaliser la partie consolidation des données qui est presque achevée (90%).

2.2. Présentation du système opérationnel actuel (SGC)

La dimension collaborative du CRM est assurée à travers le module « Communication avec le client ». La dimension opérationnelle du CRM, quant à elle, comprend les modules suivants :

- Achat
- Raccordement
- Relève
- Gestion des événements
- Facturation
- Recouvrement
- Réclamation

Le périmètre fonctionnel de la dimension analytique de ce projet couvre trois processus à savoir la gestion des facturations, raccordements et la relève.

2.2.1. Processus Facturation

Le module Facturation est chargé d'établir une facture d'énergie au client afin de lui permettre de régler sa consommation en énergie électrique et/ou gazière.

Sonelgaz distingue deux types de factures d'énergie :

- Factures cyclique : ce sont les factures établies périodiquement en fonction du calendrier de relève.
- Factures hors cycle : elles sont générées suite à un évènement entrainant une facturation.

2.2.1.1. Facturation cyclique

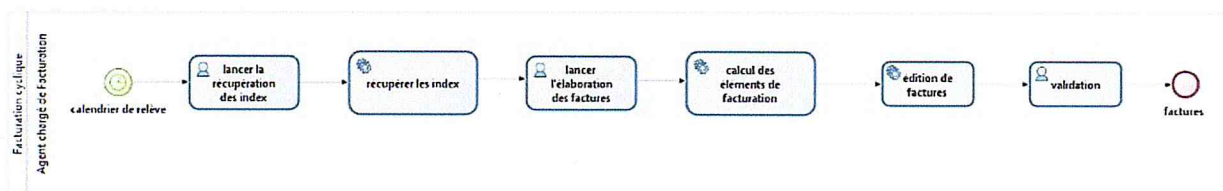


Figure 19 : Facture cyclique

Le processus Facturation cyclique est décrit par le schéma de la figure 19. Il peut être décomposé en trois étapes :

1) Récupération des index :

- La facturation d'énergie chez Sonelgaz se fait de façon trimestrielle.
- Les lieux de consommation à facturer pour une date précise sont extraits en fonction du calendrier de relève-facturation.
- La facturation d'un groupe peut intervenir durant cinq (05) jours ouvrables soit le lendemain de sa relève soit après la relève spéciale.

2) Détermination des éléments de la facturation :

Après la récupération des index, l'utilisateur lance l'élaboration des factures. Le système procède automatiquement au calcul de différents éléments de facturation :

➤ Les éléments se rapportant à la consommation :

- ❖ Période de consommation.
- ❖ La quantité consommée en électricité (kWh) et/ou gaz (thermies).
- ❖ Le montant d'énergie consommée.
- ❖ Le montant des primes fixes en hors taxe.

➤ Les éléments se rapportant aux prestations :

- ❖ Frais occasionnels.
- ❖ RGPE, REPE.
- ❖ Frais de coupure et de rétablissement.
- ❖ Devis.

➤ Les éléments se rapportant aux taxes :

- ❖ Droit fixe sur consommation.
- ❖ Taxe d'habitation et Taux TVA.

3) Edition de factures :

Une fois la phase de calcul terminée, les factures sont générées selon un format défini. L'agent chargé de la facturation procède à une validation à la fin de processus. Cette validation envoie automatiquement les factures générées au module recouvrement pour le règlement.

2.2.1.2. Facturation Hors cycle

Une facture hors cycle est élaborée suite à la demande d'un agent chargé des évènements BT/BP. Les étapes de facturation restent les mêmes sauf que les éléments de la facturation à prendre en considération dépendent de la nature de l'évènement déclencheur de la facturation. Cette différence est gérée automatiquement par le module facturation.

2.2.2. Processus de raccordement :

Le raccordement est un processus propre au métier du (des) gestionnaire(s) de réseaux (GRD). Il prend en charge toute création ou modification d'ouvrages sur le réseau de Distribution.

Ce processus est généralement composé de sous-processus suivants :

- Demande de raccordement
- Recevabilité et Qualification de la demande
- Etude
- Proposition de raccordement technique et financière (devis de raccordement).
- Programmation : Préparation du dossier de l'appel d'offre.
- Réalisation
- Facturation (travaux)
- Mise en service

Le nouveau système préconisé doit prendre en charge tout type de demande, à savoir :

- Raccordement BT, BP, HTA, MP.
- Autres travaux (déplacement, renforcement, ...).

2.2.2.1. Périmètre de raccordement :

Le périmètre de raccordement est limité à la gestion commerciale de toute demande de raccordement ou de modifications du réseau électricité et ou gaz des clients.

Ces demandes comme suit :

- Branchements sans extension (simples) électricité et gaz.
- De compteur électricité et ou gaz.
- Raccordements de lieux de consommation aux réseaux électricité et gaz nécessitant une extension de réseau.

2.2.3. Processus de relève :

La relève est un acte qui consiste en la lecture des index sur les cadrans ou registres des compteurs d'électricité et du gaz suivie d'une transcription de ces mêmes index sur le bordereau de relève ou un TSP. Cette opération peut aussi être effectuée par télé relève (compteur électricité et/ou gaz).

Toute anomalie constatée sur le lieu de consommation (terrain) est également consignée sur un TSP ou sur le bordereau de relève pour être exploitée.

Ce module traite :

- Le calendrier de relève.
- L'organisation de la relève et la répartition des clients (commune, tournée, circuit).
- Les supports de relève (bordereau, TSP, plate-forme télé-relève).
- La saisie des index en intégrant les différents contrôles nécessaires (bordereau ou TSP).
- Le téléchargement des index (télé-relève).
- Les signalés (compte rendu du contrôle des index).

Les types de client concerné sont :

- BT/BP/HTA/MP.

3. Existant décisionnel

Le groupe SONELGAZ veut, par le biais de ce projet, palier à un manque important en matière décisionnelle au niveau de ses sociétés de distribution. Ce manque se caractérise par l'indisponibilité d'un support efficace à l'aide de décision, en mesure de fournir des informations adéquates en temps voulu.

Partant de ce constat, nous présenterons, dans ce qui suit, les procédures et les méthodes de Reporting et de prise de décision, ainsi que les éventuelles lacunes qui peuvent exister.

3.1 Processus de reporting

Afin d'assurer le suivi et l'analyse de son activité de distribution, le groupe Sonelgaz se base, essentiellement, sur des rapports de statistiques périodiques et d'autres générés à la demande. Ces rapports sont élaborés par la filiale ELIT ou par les différentes SD à base des données extraites et consolidées à partir des systèmes opérationnels.

La procédure d'élaboration des rapports dépend de façon directe de l'entité demandeuse. Cette dernière peut être une Agence, une DD, une SD, le CREG, l'une des directions de la maison mère (holding Sonelgaz) ou le ministère de l'énergie.

Après l'analyse des processus reporting utilisés au sein de groupe, nous avons pu recenser deux procédures qui se distinguent par le niveau de la hiérarchie et la structure chargée de leur élaboration (ELIT ou SD).

3.2 Procédure de reporting au niveau du groupe

A ce niveau hiérarchique, les utilisateurs ont besoin de chiffres qui concernent l'ensemble des SD pour suivre l'état de la fonction de distribution sur tout le territoire national. Ce type de reporting est souvent demandé par le ministère d'énergie, les analystes du CREG ou les dirigeants de la société holding.

Généralement, la demande est adressée à ELIT et aux différentes SD. L'objectif de cette action est de vérifier les écarts entre les résultats d'analyse issues d'ELIT et ceux des SD, chercher des interprétations, apporter des corrections si nécessaires, élaborer les états finaux demandés et enfin les transférer aux demandeurs.

Dans ce qui suit, nous présenterons les deux cas selon la structure du rapport.

3.3 Reporting à travers ELIT

La seule structure susceptible à contacter ELIT pour établir les rapports est la DGSP. Cette dernière joue le rôle d'intermédiaire entre les demandeurs de rapports et ELIT. Pour faciliter la compréhension de cette procédure en partant de la demande d'un état donné jusqu'à son élaboration, nous avons établi le schéma suivant (figure 20) :

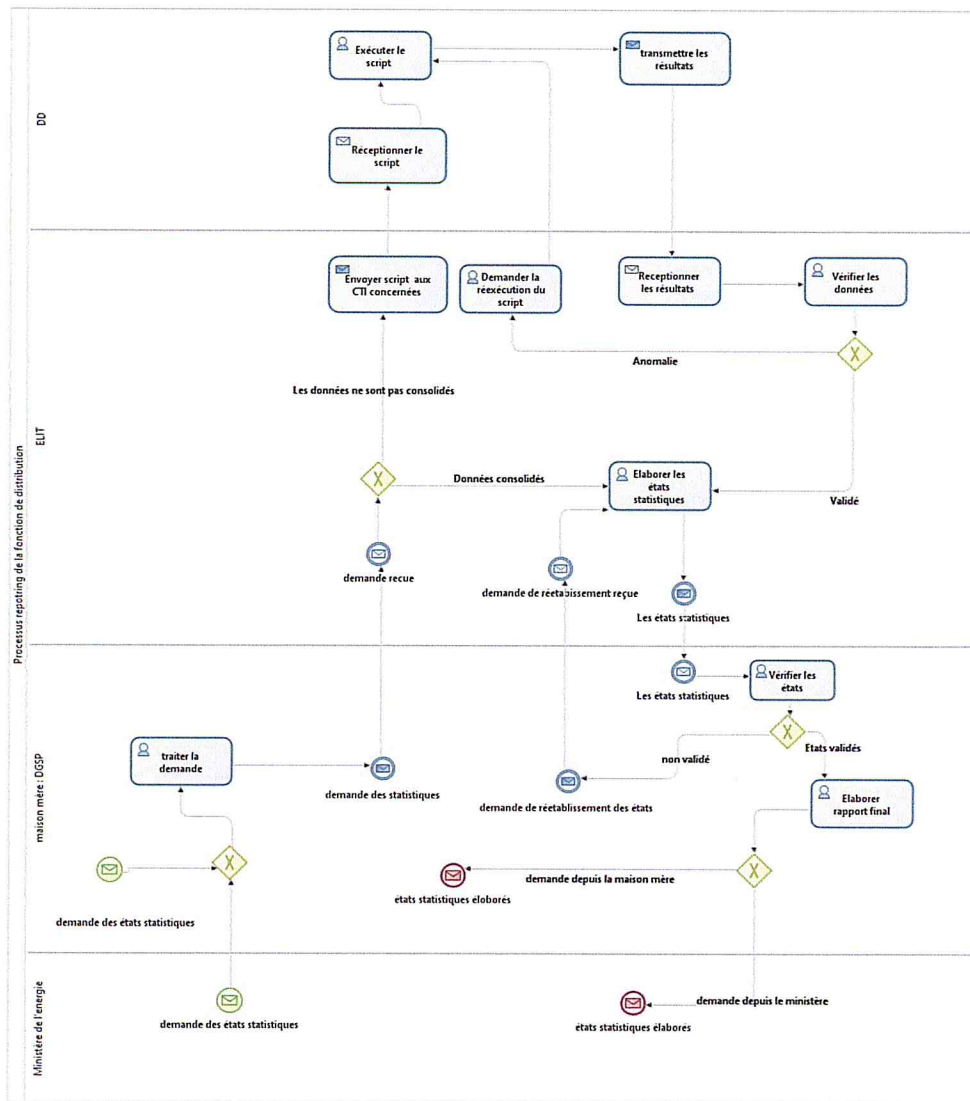


Figure 20 : Processus de reporting à travers ELIT

Comme le montre le diagramme de la figure 20, les étapes du processus de reporting se présentent de la manière suivante :

1. Afin de répondre à leurs besoins d'analyse, les décideurs et les analystes du niveau supérieur du groupe Sonelgaz et du ministère de l'énergie lancent une demande à la DGSP en précisant les états dont ils ont besoin.

2. La DGSP étudie les états demandés et dégage une liste des états intermédiaires, qui seront présentés sous forme des statistiques et des rapports, nécessaires pour l'élaboration des états demandés. Elle envoie par la suite, une demande à la Direction SI Activités Distribution / Gestion de Réseaux (DSID) de la filiale ELIT pour élaborer des états intermédiaires.
3. A la réception de la demande de la DGSP, les ingénieurs de la DSID d'ELIT vérifient si les données nécessaires pour l'élaboration des états intermédiaires sont déjà consolidées.
 - Si des rapports nécessitant les mêmes données ont été déjà préparé par ELIT alors les données sont déjà consolidées donc on passe directement à la sixième étape.
 - Sinon l'équipe chargée de l'édition de ces états prépare un script pour l'extraction de données à partir des bases de données opérationnelles régionale du SGC ainsi qu'un fichier appelé « fichier contexte ». Ce dernier doit être complété par les paramètres de la BDD de la DD par ses propres ingénieurs pour des raisons de confidentialité. ELIT envoie le script accompagné du fichier contexte par courrier électronique au CTI des DD concernés.
4. Les administrateurs de la BDD du SGC au niveau de CTI des DD doivent insérer les paramètres d'accès au BDD dans le fichier contexte puis exécuter le script d'extraction. Le résultat d'extraction sera enregistré sur des fichiers sous format Excel et doit être envoyé en réponse au courrier reçu de la part d'ELIT.
5. Une fois les données consolidées sont reçues, l'équipe d'ELIT doit vérifier si les données sont bonnes. En cas d'anomalie (erreur dans l'intervalle de temps des données extraites, altération ou endommagement du fichier durant le transfert, etc.), elle envoie une demande de réexécution de script par mail en mentionnant l'anomalie pour la corriger.
6. Par la suite, l'équipe d'analystes d'ELIT élabore les états intermédiaires et les envoie à la DGSP.
7. A la réception des états intérimaires, la DGSP procède à une vérification et demandent à ELIT d'effectuer des rectifications si nécessaires.
8. La DGSP, en se basant sur les états intermédiaires, élabore les états demandés et les envoient à la partie demandeuse qu'elle soit interne au groupe ou externe (ministère d'énergie).

Remarque : Le processus reporting ci-dessus concerne l'ancien SGC. Tandis que pour le nouveau SGC qui a été développé et qui n'est pas encore déployé (vu que la consolidation de

données n'est pas encore achevée) ne dispose pas d'un système de reporting qui lui est associé.

3.4 Bilan

A l'issue du diagnostic que nous avons effectué sur les procédures actuelles du reporting, nous avons pu soulever les lacunes et les limites suivantes :

- Absence d'un moyen permettant aux décideurs d'éditer leurs rapports en toute autonomie.
- Indisponibilité des rapports en temps voulu.
- Absence d'un moyen de navigation permettant aux décideurs de visualiser les analyses et les résultats selon le niveau de détail voulu.
- Le niveau élevé de détail des données qui rend l'analyse et la prise de décisions plus difficile.
- La lenteur du processus décisionnel, vu le passage obligatoire par ELIT ou les SD.
- Les erreurs et les écarts des statistiques entre ELIT et les SD.

4. Conclusion

Au niveau de ce chapitre, nous avons présenté les principales fonctionnalités du système opérationnel existant pour la gestion des clients SGC, nous avons décrit par la suite les procédures de reporting actuelles et nous avons fini par l'identification des limites et déficits du système décisionnel actuel.

Dans le chapitre suivant, nous allons collecter les différents besoins des décideurs de distribution du groupe Sonelgaz.

Chapitre VI

Besoins d'analyse des données et aide à la décision



1. Introduction

Nous allons dans ce chapitre présenter la partie « besoins d'analyse des données » qui constitue la partie essentielle de la conception.

2. Identification des besoins

L'identification des besoins est l'étape cruciale de tout système d'information, notamment le SID, durant laquelle les besoins analytiques des utilisateurs sont identifiés.

Dans ce chapitre nous allons commencer par présenter notre démarche suivie pour la collecte des besoins, les utilisateurs finaux du système, les besoins fonctionnels et techniques, ainsi que les difficultés rencontrées. Nous concluons par la présentation de notre solution décisionnelle.

2.1. Démarche de collecte des besoins

Il n'existe pas une démarche standard pour la collecte de besoins. Cependant, elle dépend des compétences de l'entreprise, des enjeux et de ce que nous avons déjà évoqué avec nos utilisateurs.

Notre démarche d'identification des besoins s'est déroulée en trois (03) étapes :

- 1) L'utilisation des techniques d'entretien avec les personnes du métier du groupe Sonelgaz à travers tous les niveaux (Ministère d'énergie, La maison mère, ELIT, Sociétés de distribution, les directions de distribution et les agences).
- 2) L'analyse de la documentation.
- 3) L'étude des sources de données.

2.1.1. Entretiens et réunions

Afin de faire une étude complète et d'obtenir le maximum d'informations, nous avons concentré en premier lieu sur les besoins exprimés par les managers, puis pour les enrichir, nous avons étudié les besoins relatifs aux données réelles après les points de décision.

Une fois le Cahier des charges en notre possession, nous avons noté les noms des personnes l'ayant élaboré et c'est avec eux que nous avons programmé les premières interviews.

Chaque entretien a été sanctionné par un procès-verbal qui synthétise les points débattus ainsi que les résultats acquis.

2.1.2. Documentation

Durant notre étude, nous avons consulté plusieurs documents internes à Sonelgaz tels que :

- Des rapports déjà effectués : Ils nous ont permis de comprendre l'organisation du groupe Sonelgaz et les activités métiers.
- Les différents sites web des SD : Nous les avons consultés pour récupérer les organigrammes, comprendre le métier, désigner les acteurs de notre système décisionnel.
- Le cahier des charges du projet opérationnel et Repoting: Ils nous ont été remis par l'équipe de développement du nouveau système SGC au niveau Étude des sources de données.

Les sources de données qui alimentent notre système décisionnel est la base de données du système opérationnel CRM.

ELIT a opté pour une architecture de base de données centralisée au niveau du groupe Sonelgaz. En effet, les données des différentes sociétés de distribution issues de la CRM sont stockées dans la base de données hébergée au niveau du Data center d'ELIT.

2.2. Identification des utilisateurs finaux du système décisionnel

En se basant sur les résultats de l'analyse de l'existant et des besoins identifiés, il est nécessaire d'identifier les utilisateurs finaux de la solution à mettre en place. Les utilisateurs susceptibles à utiliser dans le système décisionnel sont décrits dans le tableau ci-dessous :

(Voir tableau 05)

Structure	Niveau Organisationnel	Poste de l'utilisateur
Le groupe de Sonelgaz	Siège de Sonelgaz (la maison mère)	<ul style="list-style-type: none"> ➤ PDG du groupe ➤ DGSP ➤ DGDS ➤ CREG
		<ul style="list-style-type: none"> • Responsable de la DC • Les chefs des départements. • Chefs de services : <ul style="list-style-type: none"> - Analyses statistiques et prévisions. - Transactions d'énergies et gestion des contrats. • Les administrateurs de service de contrôle et assistance. • Les analystes de service SGC.
Les sociétés de distribution (SD)	Direction générale de la SD.	PDG de la SD
	Direction commerciale et Marketing (DCM)	<ul style="list-style-type: none"> • Responsables régionales des DD. • Chefs des quatre divisions : <ul style="list-style-type: none"> - Division relations commerciales - Division gestion des systèmes informatiques - Division finances et comptabilité. - Division d'étude d'exécution des travaux. • Les administrateurs de la division gestion des systèmes informatiques. • Les analystes de division relations commerciales. • Les responsables des statistiques de la division finances et comptabilité.
	Agences commerciales (AC)	Responsables des agences

Tableau 05 : Catégories d'utilisateurs du système de Reporting à travers ELIT

2.3.Récapitulatif des besoins

2.3.1. Besoins fonctionnels (Analytique)

Après plusieurs interviews et réunions avec les dirigeants et les utilisateurs métiers, nous avons pu recenser les exigences de ces derniers en termes de rapports et d'indicateurs.

Le besoin primordial des décideurs était de leur fournir un moyen qui leur permettra de générer des états.

Ces derniers sont des techniques qui consistent à collecter, traiter et représenter les données historiques pour qu'elles puissent être correctement interprétées pour des besoins d'études, d'analyses et de prospectives en interne comme en externe (niveaux opérationnels, niveaux décisionnels, Directions Générales Holding Groupe, Ministère de l'Energie GREG – autres), en vue d'obtenir une idée qualitative et quantitative à utiliser souvent comme composante d'aide à la décision.

- **Rapports à la demande :**

Les rapports à la demande sont des états conçus suite à une demande d'un décideur.

Ces rapports peuvent être imprimés ou enregistrés sous différents formats (Excel, Word, PDF, XML etc....).

Les responsables Sonelgaz ont réclamé les rapports suivants :

- **Processus de facturation :**

- ✓ Etats des Facturation pour tous les types de clients.

- **Processus de raccordement :**

- ✓ Etat de Réalisation des branchements (BT /BP).
- ✓ Etat des Instances de réalisation des branchements (BT/BP).
- ✓ Etat des Délais de réalisation (BT/BP).
- ✓ Etat des Délais de réalisation des branchements par phase/Global (BT/BP).
- ✓ Etat de Réalisation des extensions de réseaux Electricité BT et raccordements HTA
- ✓ Etat de Réalisation des extensions de réseaux Gaz BP et raccordements MP
- ✓ Etat des Extensions électricité en instance (BT/HTA).

- ✓ Etat des Raccordements électricité en instance (BT/HTA).
- ✓ Etat des Extensions gaz en instance (BP/MP).
- ✓ Etat des Raccordements gaz en instance (BP/MP).
- ✓ Etat des Délais de réalisation des extensions électricité par phase/Global (BT/HTA).
- ✓ Etat des Délais de réalisation des raccordements électricité par phase/Global (BT/HTA).
- ✓ Etat des Délais de réalisation des extensions gaz par phase/Global (BP/MP).
- ✓ Etat des Délais de réalisation des raccordements gaz par phase/Global (BP/MP).

➤ **Processus de la relève :**

- ✓ Etat de Compte rendu de relève (BT/BP/HTA/MP).
- ✓ Etat de Compte rendu des anomalies signalées en instance de traitement (BT/BP/HTA/MP).
- ✓ Synthèse du traitement des signalées de la relève (BT/BP/HTA/MP).

❖ **Les Indicateurs :** pour chaque processus il y a des indicateurs pour faciliter de traiter les informations pour construire une bonne décision :

+ **La gestion de facturation :**

1) Indicateur sur la qualité de la facturation :

Le nombre facture émises sur index, estimées – ayant au moins une consommation nulle, ayant au moins un index pas atteint ainsi que le nombre de facture annulées et émises manuellement et par groupe, selon la formule :

- **Facture émises sur index :** nombre de factures émises sur index relevés.
- **Factures estimées :** nombre de factures estimées.
- **Forfait :** facture calculées sur la base d'un forfait.
- **Factures ayant au moins une consommation nulle :** nombre Factures ayant au moins un cadran à consommation nulle.
- **Factures ayant au moins un index pas atteint :** nombre de Factures ayant au moins

un cadran dont Index relevé < dernier index facturé.

- **Factures annulées** : nombre de factures annulées par DCA et par motif.
- **Factures manuelles** : nombre de factures redressées ou émises en hors cycles par DCA.

✚ **La gestion de raccordement :**

Tout raccordement doit passer par une demande de raccordement et suivra les différentes phases énumérées ci-dessous jusqu'à l'intégration du demandeur (client) dans le cycle de facturation ou à défaut son annulation. Avec ces différentes phases, le nombre d'affaires et les délais par phase constituent des indicateurs de mesure pour ce processus :

1) Nombre d'instance par phase (Branchement-Extension-Raccordement) :

Le nombre d'instance est calculé par énergie pour les demandes de branchement BT/BP – extension BT/BP et raccordement HTA/MP HTB/HP à la fin du mois selon la formule :

- **Etude** : nombre de demandes reçues non encore étudiées : date d'étude non renseignée.
- **Paiement** : nombre de devis envoyés aux clients non encore payés : date paiement non renseignée.
- **Mise en service** : nombre de branchements dont les devis sont payés non encore mis en service : date MES non renseignée.
- **Intégration** : nombre de branchements mis en service no encore intégrés en facturation : date intégration non renseignée.

2) Nombre de réalisation par phase (Branchement-Extension-Raccordement) :

Le nombre de réalisation est calculé par énergie pour les demandes de branchement BT/BP – extension BT/BP et raccordement HTA/MP HTB/HP à la fin du mois selon la formule :

- **Etude** : nombre de demandes étudiées.
- **Paiement** : nombre de devis envoyés aux clients payés.
- **Mise en service** : nombre de branchements mis en service.
- **Intégration** : nombre de branchements intégrés en facturation.

3) Délais moyens de satisfaction (Branchement-Extension-Raccordement) :

Les délais moyens (jour) de réalisations par phase et global sont calculés sur la base des demandes de branchement BT/BP / extension BT/BP / raccordement HTA/MP HTB/HP, mis en service et intégrés durant la période par énergie selon la formule :

- **Délais moyen étude** : $\sum (\text{date établissement devis} - \text{date réception demande}) / \text{nombre demandes étudiées}$.
- **Délais moyen paiement devis** : $\sum (\text{date paiement devis} - \text{date réception étude}) / \text{nombre devis payés}$.
- **Délais moyen obtention Autorisation de construire** : $\sum (\text{date réception Autorisation de voirie} - \text{date paiement devis}) / \text{nombre Autorisation demandées}$.
- **Délais moyen d'attribution des affaires** : $\sum (\text{date établissement ODS} - \text{date établissement de l'OET}) / \text{nombre devis payés}$.
- **Délais moyen réalisation des affaires** : $\sum (\text{date fin travaux} - \text{date début travaux}) / \text{nombre affaires réalisées}$.
- **Délais moyen de mise en service** : $\sum (\text{date mise en service} - \text{date fin travaux}) / \text{nombre mise en service}$.
- **Délais moyen intégration clients** : $\sum (\text{date intégration} - \text{date mise en service étude}) / \text{nombre intégration}$.
- **Délais objectif** : nombre de jours fixés.

+ La gestion de la relève :

1) Indicateur sur la qualité de la relève :

Le nombre de contrat - nombre des index relevés – non relevés – absents – égalités d'index – consommation faible – évolution de la consommation, selon la formule :

- **Nombre de contrat** : correspond au nombre de compteurs.
- **Index relevés** : nombre de tous les index relevés ou télé relevés.
- **Index non relevés** : nombre de tous les index non relevés et non télé relevés.
- **Absents** : nombre d'index non relevés pour absence clients.

- **Egalités d'index** : nombre d'index dont la consommation est nulle.
- **Consommation faible** : nombre d'index dont la consommation est inférieures à 50% de celle de la même période de l'année.
- **Evolution de la consommation** : consommation année n - consommation année $n-1$.
- **Evolution contrat** : nombre contrat année n - nombre contrat année $n-1$.

2) Indicateur sur les anomalies signalées à la relève :

Le nombre des anomalies signalées à la relève, en instance de prise en charge par motif et par groupe, selon la formule :

- **Compteurs défectueux ou bloqués** : nombre de défectueux ou bloqués.
- **N° de compteur différent** : nombre de compteurs dont le n° ne correspond pas.
- **Clients non intégrés** : nombre de clients non encore intégré dans le système.
- **Clients déclassés** : nombre de clients dont le folio est déclassé.
- **Clients en fraude** : nombre de clients en situation de fraude.
- **Compteurs à vérifier** : nombre de compteurs A vérifier.
- **Index non atteint** : nombre de compteur dont l'index n'est pas atteint.
- **Compteurs inaccessible** : nombre de compteur inaccessible.
- **Double emploi** : le n° de compteur en double.

2.3.2. Besoins techniques :

La direction SI Activités Distribution / Gestion de réseaux avait tracé certains critères que le futur système doit respecter. Ces critères sont les suivants :

- ❖ Les outils utilisés dans le système sont Open Source pour qu'elle soit en adéquation avec la politique Open Source du Groupe Sonelgaz.
- ❖ Serveur de base de données PostgreSQL (PostgreSQL pour la gestion des bases de données).
- ❖ Le système doit être sécurisé.

- ❖ Les accès au système doivent être gérés avec les authentifications.
- ❖ Le système doit permettre un accès facile aux données.
- ❖ Edition automatique des rapports.

2.4. Difficultés rencontrées

Lors de la collecte des besoins nous avons rencontrés plusieurs difficultés :

- ❖ Les procédures métiers décrites dans les cahiers de charges des différents modules opérationnels du CRM ne sont pas exactement les mêmes suivis par le personnel de la distribution alors qu'ils étaient rédigés par des ingénieurs des SD et des DD.
- ❖ Certains documents et sources d'informations sont confidentiels.
- ❖ La complexité de la codification des bases de données transactionnelles.
- ❖ Les personnes interrogées se lamentent et donnent des réponses rapides.

2.5. Proposition de la solution

2.5.1. Le système cible

En se basant sur les besoins recensés dans la phase d'identification de besoins, nous proposons de mettre en place un système d'information décisionnel permettant de :

- Automatiser l'édition des rapports d'analyse et permettre une certaine flexibilité dans la construction de ces derniers.
- Offrir une vision synthétisée sur l'activité commerciale des SD à travers un ensemble d'indicateurs.
- Paramétrer les rapports selon le niveau d'hierarchie de l'utilisateur. Autrement dit, les rapports accessibles par un niveau de ne concernent que son niveau ou des niveaux inférieurs.
- Garantir un accès rapide et sécurisé aux rapports.

2.6. Les solutions proposées

La solution dépend de façon directe des sources de données utilisés pour alimenter notre système de reporting, nous proposons les deux solutions suivantes :

2.6.1. Reporting en utilisant un entrepôt de données

Cette proposition est basée sur la mise en place d'un entrepôt de données qui sera alimenté par les données opérationnelles du SGC et qui sera dédiée au système du reporting.

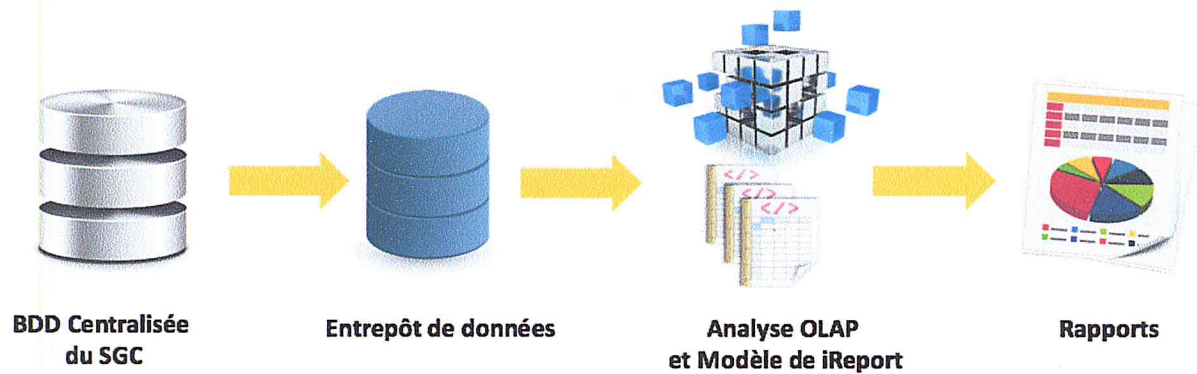


Figure 21 : Reporting en utilisant un entrepôt de données

La solution proposée ci-dessus (voir figure 21), consiste à construire un entrepôt de données à base des données extraites de la BDD du SGC en fonction des besoins analytiques exprimés par les utilisateurs du système (décideurs et analystes). L'étape suivante est l'analyse OLAP avec les modèles de iReport qui consiste à faire des traitements multidimensionnels qui permettent de faire le calcul des indicateurs afin de générer les rapports demandés.

Avantage :

- L'analyse OLAP assure la préparation du rapport ne prend que quelques secondes.
- L'opération de rafraîchissement de l'entrepôt de données qui se fait chaque nuit assure aux utilisateurs d'avoir l'image la plus proche à celle qui pourrait être fournie par une solution en temps réel toute en éliminant les inconvénients de cette dernière.
- Le processus ETL qui doit s'exécuter avant le chargement des données au niveau de l'entrepôt garantit la justesse et la fiabilité des résultats fournis par le système de reporting.
- La construction de l'entrepôt de données supporte l'aspect évolutif de la solution en offrant la possibilité de mettre en place d'autres mécanismes de restitution permettant d'autres types d'analyse tel que le datamining pour l'analyse prédictive.

- L'entrepôt de données assure l'historisation des données servant à la décision pour l'entreprise.
- L'entrepôt de données présente la meilleure solution pour stocker les informations qui servent à l'aide à la décision de façon qui facilite leur manipulation en vue de fournir l'information voulue en temps voulu.

Inconvénient :

- Nombre très important d'états (jusqu'à 1000 états voire plus) et ils prennent trop de temps pour être paramétrés et générés par iReport.
- Il n'y a pas de possibilité pour générer des rapports selon les besoins des utilisateurs non informaticiens.
- La solution nécessite un grand espace mémoire.
- La duplication des données déjà disponibles dans la base de données centralisée du SGC.

2.6.2. Reporting en utilisant un entrepôt de données avec un générateur d'Etat :

Dans ce cas, le système de reporting est le même avec la première solution proposée mais dans ce cas il y a l'existence d'un générateur d'état automatique représentée dans la figure 22 trouvée ci-dessous

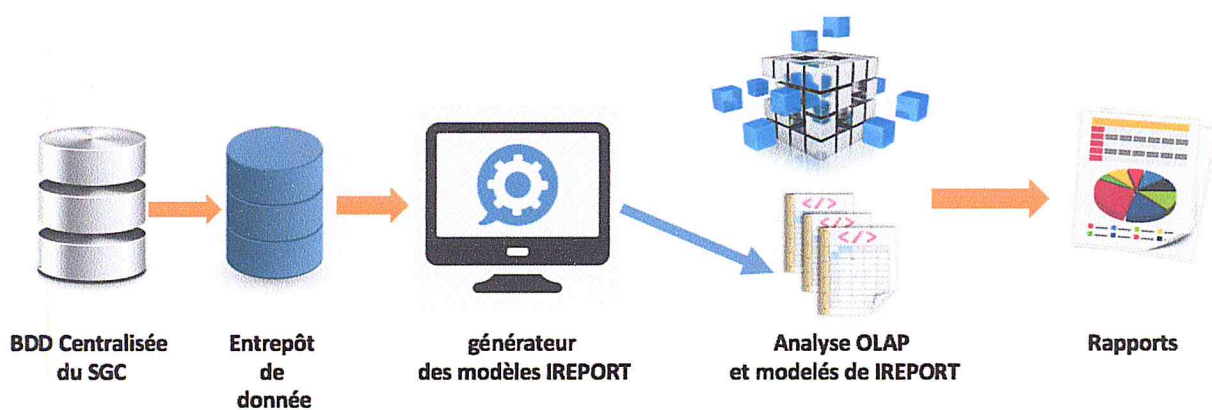


Figure 22 : Reporting en utilisant un entrepôt de données avec un générateur d'état

Avantage :

- Rapidité et souplesse dans la création des états (environ 500 états par heure).
- La solution garantit la génération des rapports demandés en quelques instants (quelques secondes à quelques minutes selon la nature de rapport).
- Possibilité de générer n'importe quel état selon les besoins des utilisateurs non informaticiens.

2.7.La solution sélectionnée

En se basant sur les besoins exprimés par les futurs utilisateurs du système de reporting, et en prenant en considération les avantages et les inconvénients de chaque solution, nous avons décidé, après discussions avec mon promoteur et l'équipe BI de la DSID d'ELIT, d'adopter la deuxième solution présentée dans la figure 23.

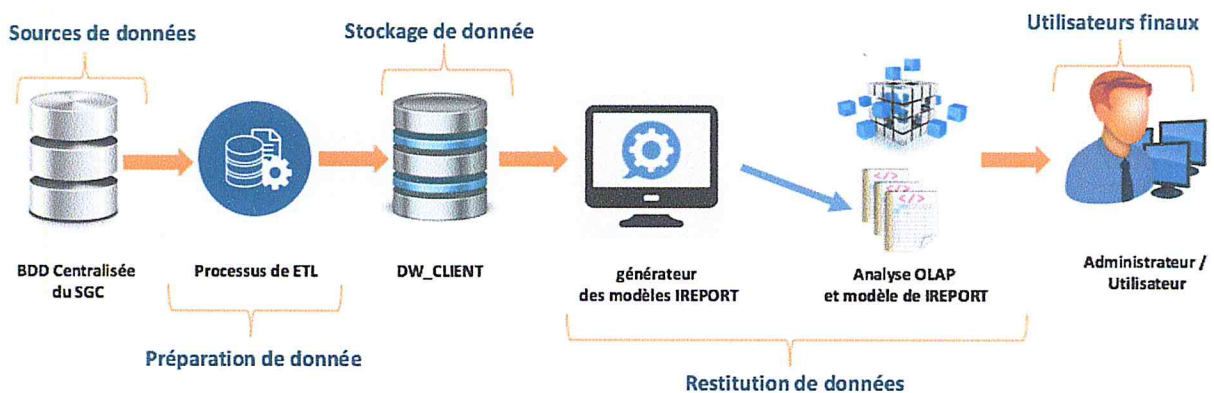


Figure 23 : l'architecture de la solution proposée

2.7.1. Source de données

Vu que le SGC regroupe toutes les fonctionnalités essentielles du métier de la distribution, sa base de données est tellement riche en information qu'elle présente une source suffisante pour alimenter notre système décisionnel.

2.7.2. Préparation de données

Les données subissent de diverses transformations avant d'être stockées au niveau de l'entrepôt de données du système décisionnel.

2.7.3. Stockage de données

L'entrepôt de données (DW_CLIENT) est le cœur de notre solution. Il est constitué de plusieurs magasins de données qui sont construits selon les besoins analytiques constatés dans la phase d'identification des besoins.

2.7.4. Restitution de données

Dans le but de répondre aux besoins des décideurs de différents niveaux, nous proposons de construire un outil du générateur des états basés sur les analyses OLAP permettre au administrateur de générer des rapports décisionnels.

3. Conclusion

L'étude des besoins est une étape plus importante dans notre projet. C'est en effet, à partir de cette étude qu'on a décidé le déroulement de notre solution.

Il nous a permis de connaître d'avantage de détails sur les rouages de l'entreprise et d'identifier les besoins analytiques de l'entreprise. Les besoins étant recensés, la conception du DW_CLIENT peut alors commencer.

Cette conception fera l'objet du chapitre suivant.

Chapitre VII

Modélisation dimensionnelle de

L'entrepôt DW_CLIENT

1. Introduction

Une fois les besoins des utilisateurs identifiés, nous pouvons aborder la conception de DW_CLIENT. Pour cela, nous avons eu recours à la modélisation dimensionnelle qui est souvent associée aux entrepôts de données compte tenu de ses avantages.

2. La modélisation dimensionnelle en étoile :

Considérant que le contenu de l'entrepôt de données traite sur trois (03) métiers différents à savoir Facturation, Relève, Raccordement, chacun représente un magasin de données. Pour une question d'organisation, nous modélisons le contenu de l'entrepôt DW_CLIENT en trois magasins de données.

2.1.Data-mart « Facturation » :

2.1.1. Fait « FACTURATION_BTBP » :

Le modèle « FACTURATION_BTBP » sera comme suit :

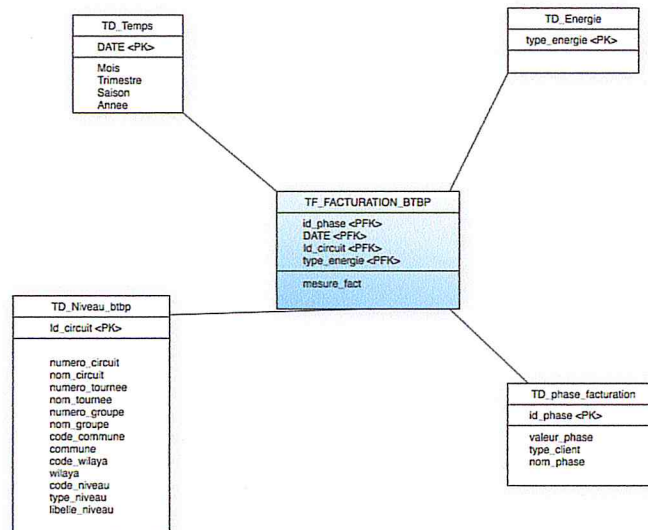


Figure 24 : Fait "FACTURATION_BTBP"

Les mesures de fait « FACTURATION_BTBP » est décrit dans le tableau suivant :

Mesure	Description
Mesure_fact	Le nombre facture émises sur index, estimées – ayant au moins une consommation nulle, ayant au moins un index pas atteint ainsi que le nombre de facture annulées et émises.

Tableau 06 : Mesures de Fait « FACTURATION_BTBP »

2.1.2. Fait « FACTURATION » :

Le modèle « FACTURATION » sera comme suit :

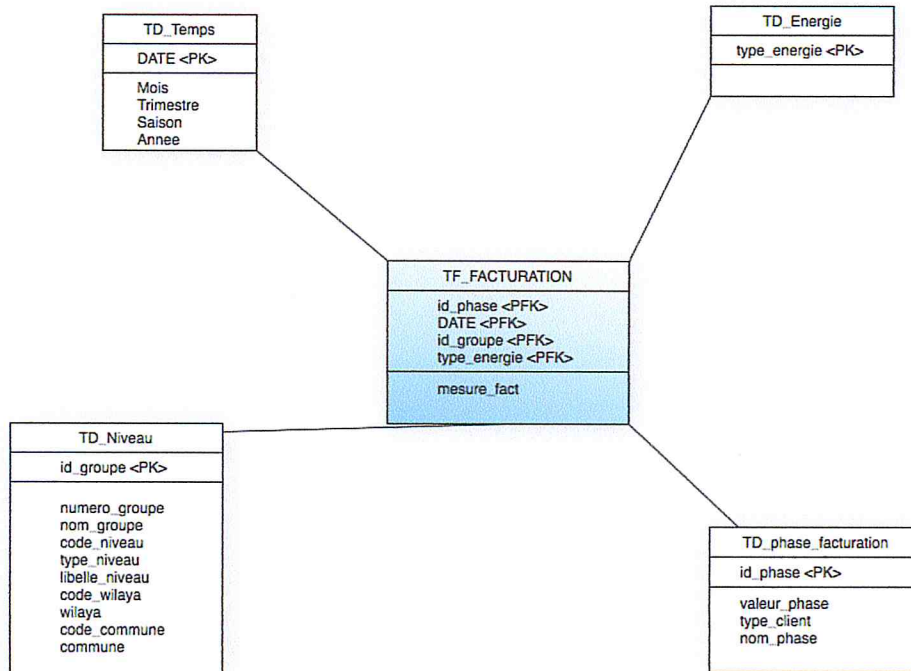


Figure 25 : Fait "FACTURATION "

Les mesures de fait « FACTURATION » est décrit dans le tableau suivant :

Mesure	Description
Mesure_fact	Le nombre facture émises sur index, estimées – ayant au moins une consommation nulle, ayant au moins un index pas atteint ainsi que le nombre de facture annulées et émises.

Tableau 07 : Mesures de fait « FACTURATION »

Remarque : la différence entre la table de fait facturation_btbp et la table de fait facturation c'est la dimension (Niveau) qui n'est pas la même, parce que les clients BT/BP ont un id_circuit qui est le plus petite dans la table C'est lui qui va être la clé primaire par rapport aux clients HTA/MP/HTB/HP ils n'ont pas de circuit.

2.1.3. Schéma global du DATA-MART « Facturation » :

La figure 26 représente le schéma global du Data-mart « facturation ».

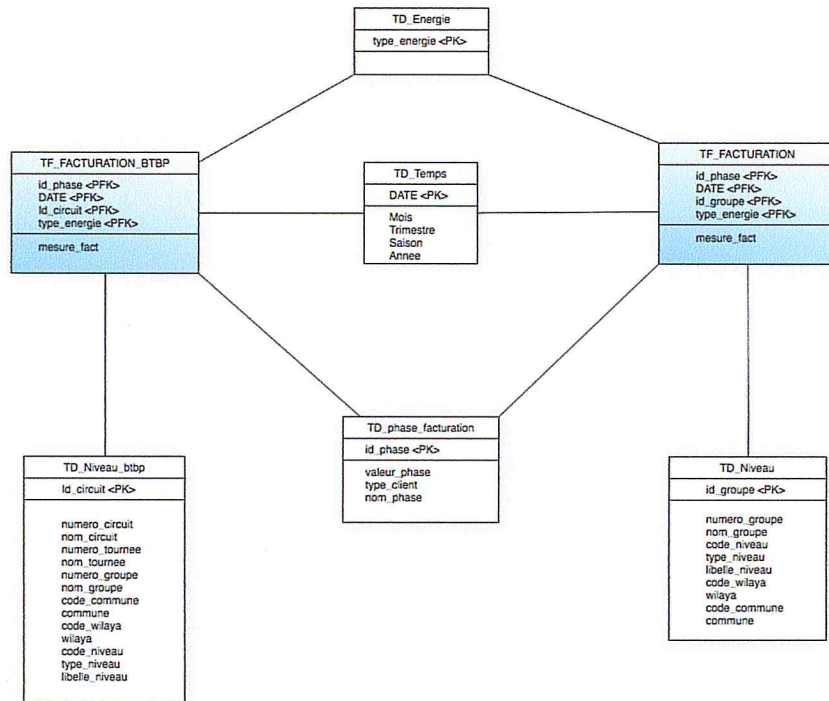


Figure 26 : DATA-MART "FACTURATION "

2.2.Data-mart « Relève » :

Intéressons-nous maintenant au data-mart « Relève » qui sera composé de plusieurs faits :

2.2.1. Fait « RELEVÉ_BTBP » :

Le modèle « RELEVÉ_BTBP » sera comme suit :

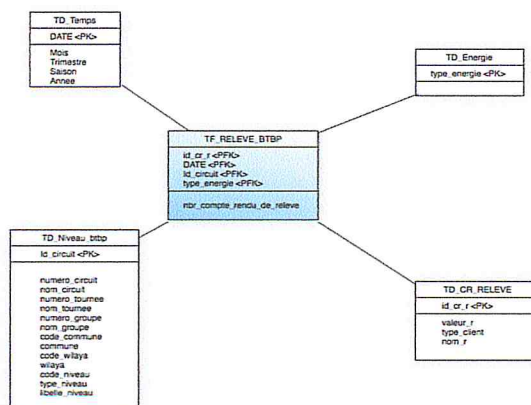


Figure 27 : Fait "RELEVÉ_BTBP"

Les mesures de fait « RELEVÉ_BTBP » est décrit dans le tableau suivant :

Mesure	Description
Nbr_compte_rendu_de_releve	Nombre des clients : année précédente , année en cours , télé relève, relève manuelle ,non relevé ,absent ,consommation nulle ,consommation faible ,consommation année en cours ,consommation année précédente.

Tableau 08 : Mesures de fait « RELEVÉ_BTBP »

2.2.2. Fait « RELEVÉ » :

Le modèle « RELEVÉ » sera comme suit :

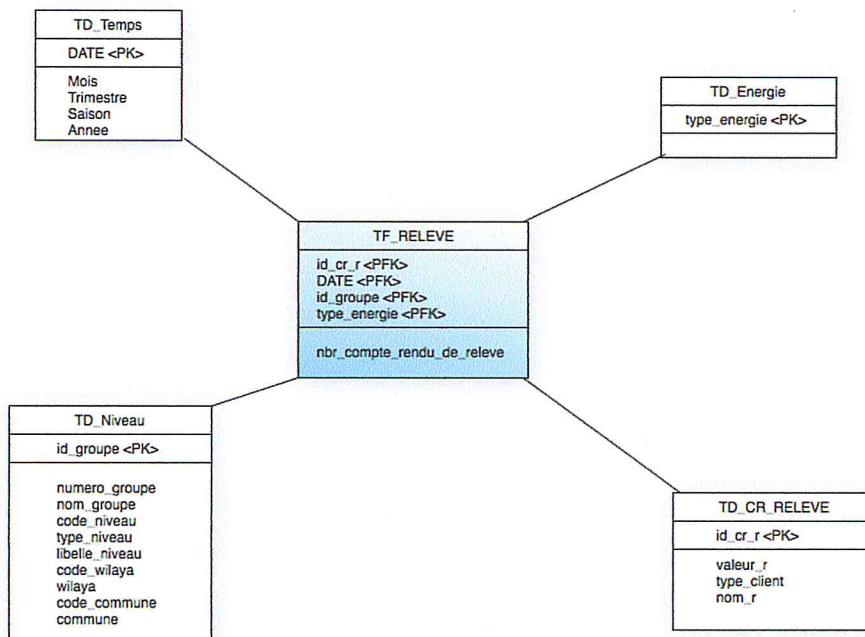


Figure 28 : Fait "RELEVÉ "

Les mesures de fait « RELEVÉ » est décrit dans le tableau suivant :

Mesure	Description
Nbr_compte_rendu_de_releve	Nombre des compteurs : à relever, relevés, a consommation nulle, absence a la relève.

Tableau 09 : Mesures de fait « RELEVÉ »

2.2.3. Fait « RELEVÉ_ANOMALIE_BTBP » :

Le modèle « RELEVÉ_ANOMALIE_BTBP » sera comme suit :

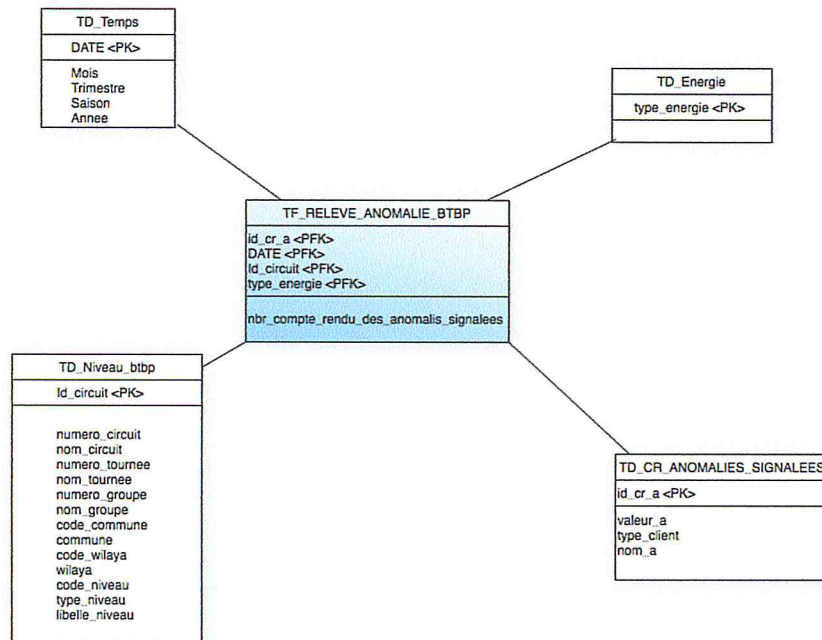


Figure 29 : Fait " RELEVÉ_ANOMALIE_BTBP "

Les mesures de fait « RELEVÉ_ANOMALIE_BTBP » est décrit dans le tableau suivant :

Mesure	Description
Nbr_compte_rendu_des_anomalies_signalees	Nombre des clients : compteur défectueux, compteur inaccessibles, compteur bloques, compteur n'affiche pas, non intègres, en fraudes, compteur index non atteint, compteur résilie consomme, compteur différents, compteur à vérifier, déclassés, consommation nulle réelles, compteur retirés, tarifs incorrects, double emplois.

Tableau 10 : Mesures de fait « RELEVÉ ANOMALIE_BTBP »

2.2.4. Fait « RELEVÉ_ANOMALIE » :

Le modèle « RELEVÉ_ANOMALIE » sera comme suit :

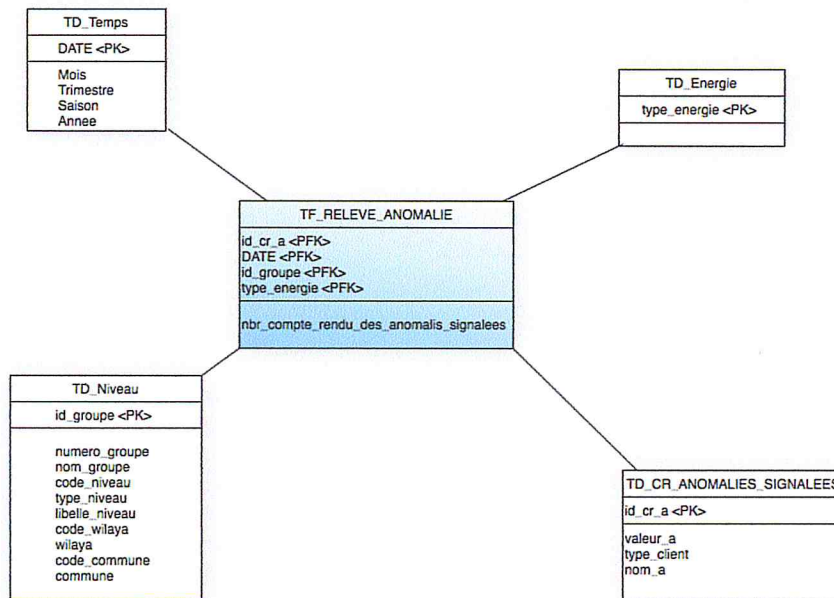


Figure 30 : Fait " RELEVÉ_ANOMALIE "

Les mesures de fait « RELEVÉ_ANOMALIE » est décrit dans le tableau suivant :

Mesure	Description
Nbr_compte_rendu_des_anomalies_signalees	Le nombre des clients : compteur non intégré, compteur a bloquer, compteur a vérifier, compteur fraude, compteur différent.

Tableau 11 : Mesures de fait « RELEVÉ ANOMALIE »

2.2.5. Schéma global du Data-mart « Releve » :

La figure suivant représente le schéma global du Data-mart « Releve ».

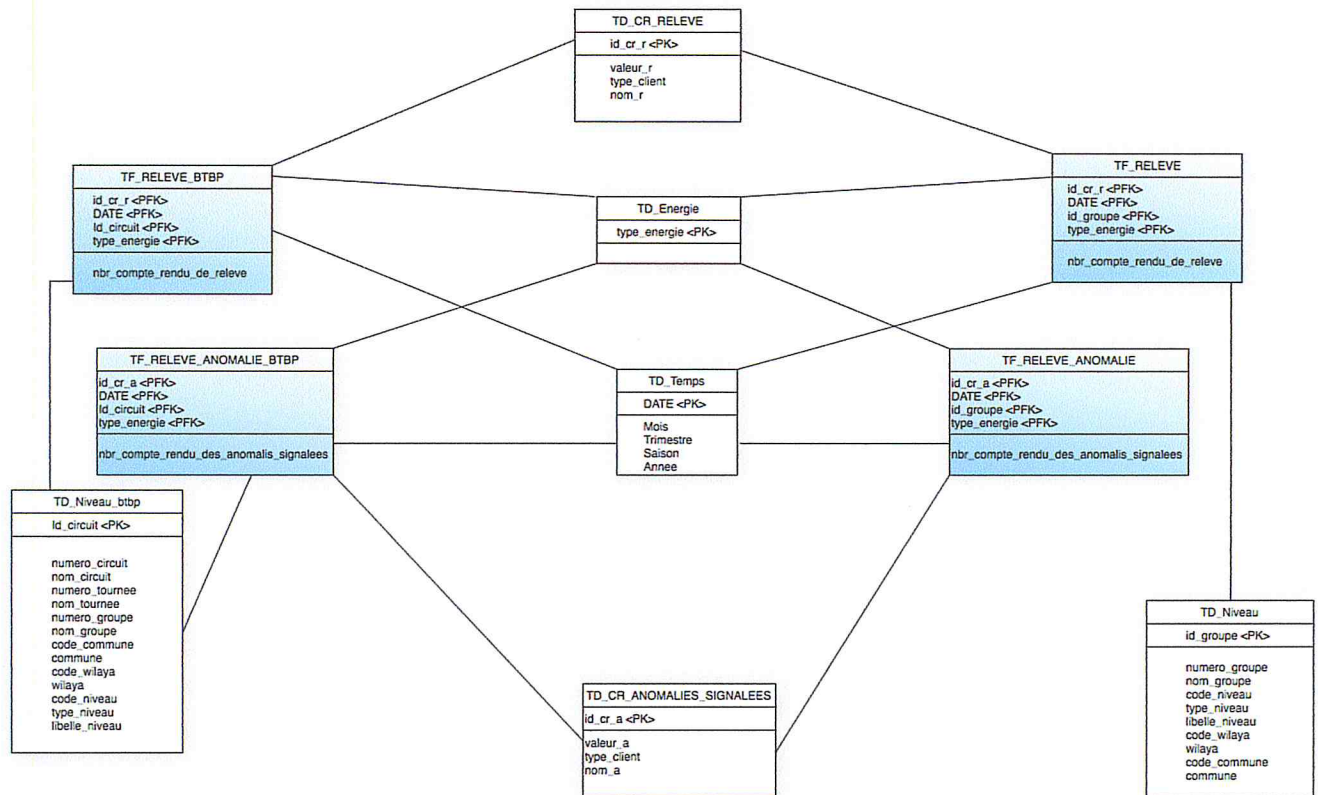


Figure 31 : DATA-MART " RELEVÉ "

2.3.DATA-MART « Raccordement » :

Intéressons-nous maintenant au volet « Raccordement » qui sera composé de plusieurs faits :

2.3.1. Fait « Rcn_branchement » :

Le modèle « Rcn_branchement » sera comme suit :

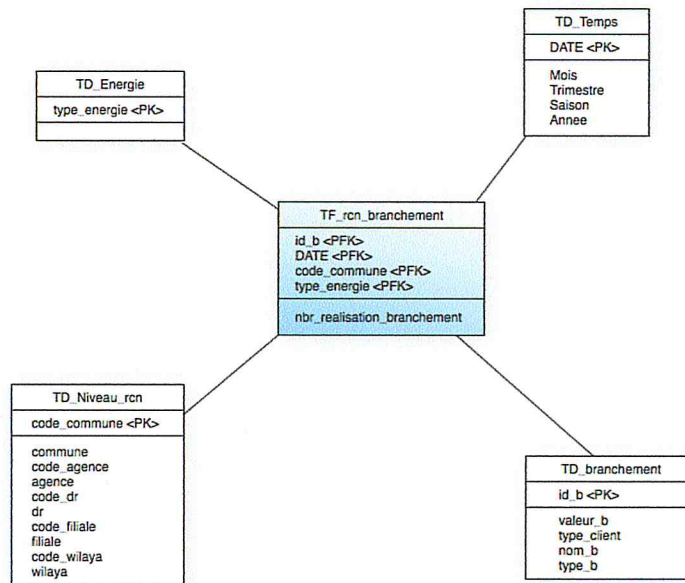


Figure 32 : Fait " RCN_BRANCHEMENT "

Les mesures de fait « RCN_BRANCHEMENT » est décrit dans le tableau suivant :

Mesure	Description
Nbr_realisation_branchement	Le nombre des : portefeuilles début période ,demandes reçues, branchements réalisés , affaires annulées, portefeuille fin période, taux évolution ,étude en instance, paiement en instance, consultation en instance, attribution en instance, réalisation en instance, mise en service en instance, intégration en instance ,étude ,paiement ,réalisation ,mise en service, intégration.

Tableau 12 : Mesures de fait « RCN_BRANCHEMENT »

2.3.2. Fait « Rcn _branchement » :

Le modèle « Rcn_branchement » sera comme suit :

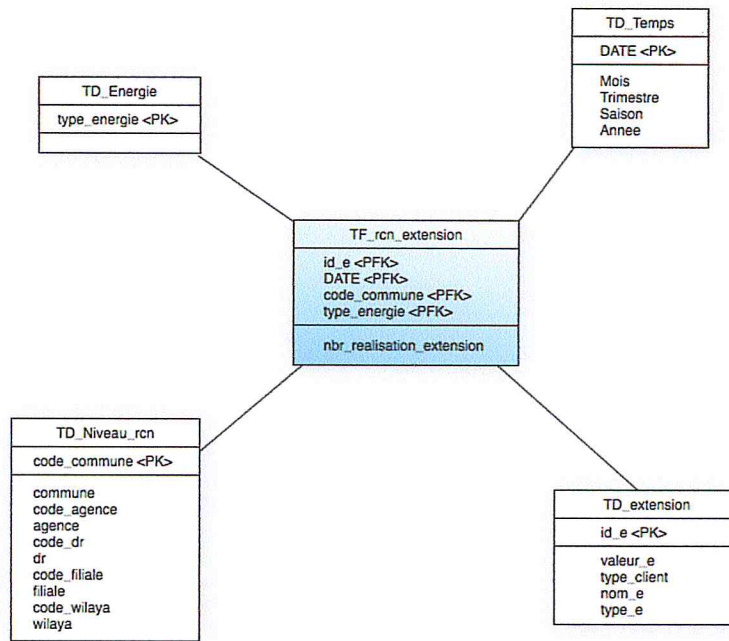


Figure 33 : Fait " RCN _ EXTENSION "

Les mesures de fait « RCN_EXTENSION » est décrit dans le tableau suivant :

Mesure	Description
Nbr_realisation_extension	Le nombre des : portefeuille début période, demandes reçues, branchements réalisés, affaire MES, demandes étudiées, devis paye, extension/raccordement terminées, extension/raccordement MES, extension/raccordement intégration, étude en instance, paiement en instance, ADC en instance, réalisation en instance, MES en instance, intégration en instance.

Tableau 13 : Mesures de fait « RCN _ EXTENSION »

2.3.3. Fait « Rcn_délais_moyen » :

Le modèle « Rcn_délais_moyen » sera comme suit :

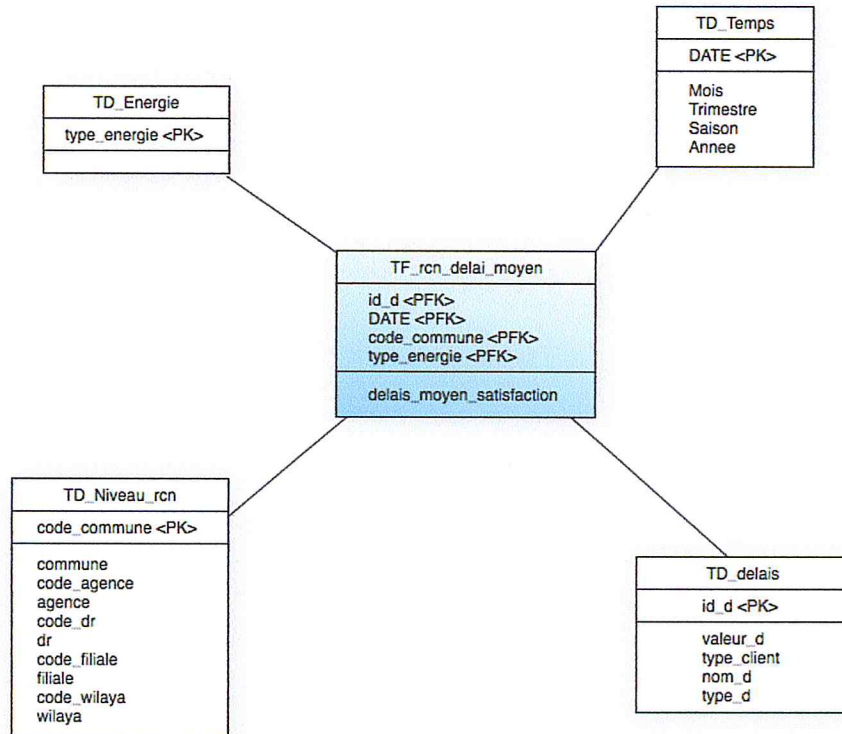


Figure 34 : Fait " RCN_DELAIS_MOYEN "

Les mesures de fait « RCN_DELAIS_MOYEN » est décrit dans le tableau suivant :

Mesure	Description
Delais_moyen_satisfaction	Le délai moyen de satisfaction (jour) est calculé sur la base des délais par : étude, paiement ,attribution ,autorisation ,réalisation ,MES, intégration ,moyen global, branchement réalisé hors délai ,demandes étudiées ,devis payes ,ADC ,extension terminé ,extension intégration ,satisfaction extension ,réalisation extension hors délai ,demandes étudiées ,devis payes.

Tableau 14 : Mesures de fait « RCN_DELAIS_MOYEN »

2.3.4. Schéma global du DATA-MART « Raccordement » :

La figure 35 représente le schéma global du Data-mart « Raccordement ».

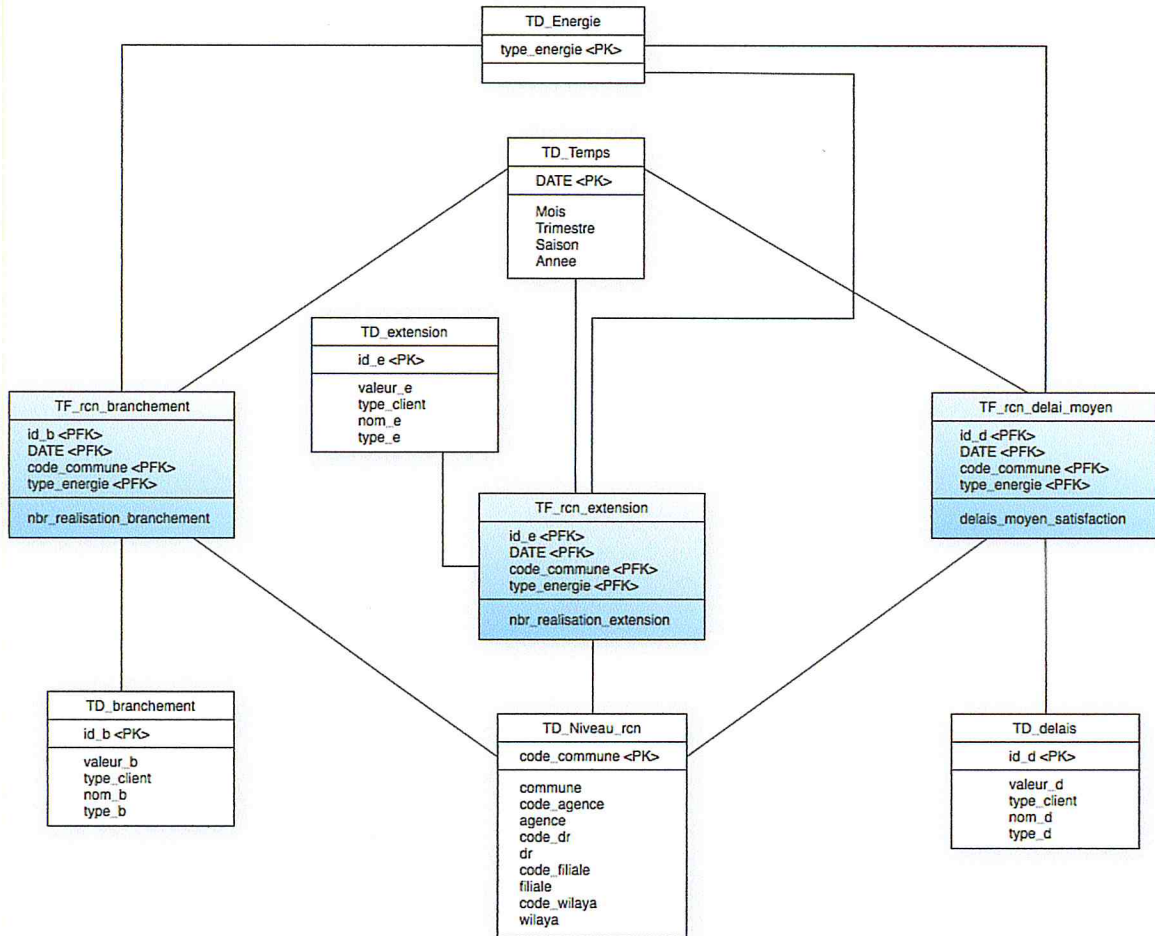


Figure 35 : DATA-MART " RACCORDEMENT "

2.4. Schéma en constellation final de l'entrepôt de donnée « DW_CLIENT » :

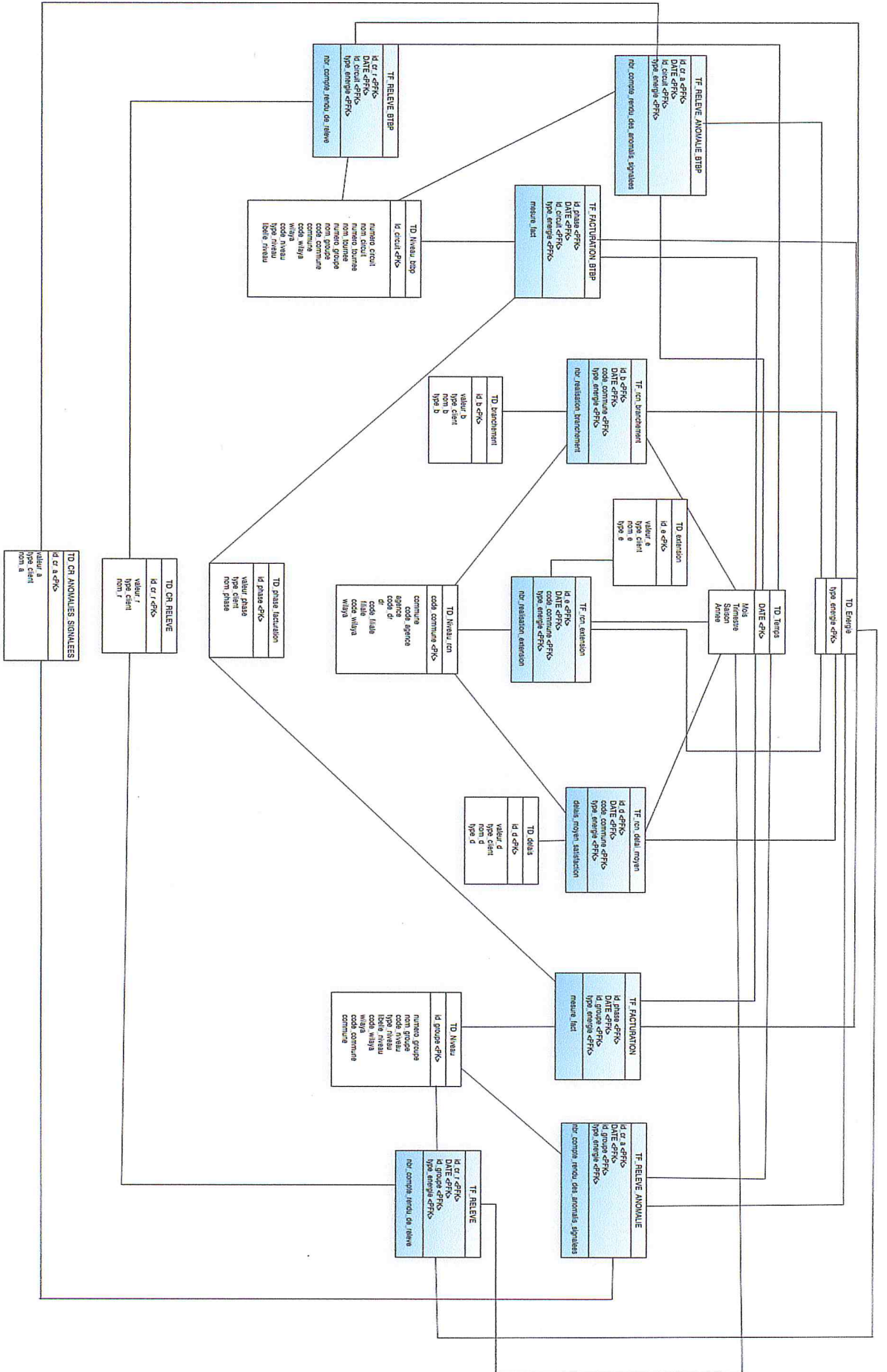


Figure 36 : Schéma en constellation final de l'entrepôt de donnée « DW_CLIENT »

3. Définition des niveaux et hiérarchies des dimensions

Les hiérarchies de chaque dimension ainsi que les niveaux correspondant à chaque hiérarchie vont définir une structure hiérarchique. Cette structure permettra la navigation à l'intérieur de la dimension à travers les opérations de « forage vers le bas » (accès à un niveau de détail supérieur) ou de « forage vers le haut » (accès à un niveau de détail inférieur).

Les tableaux suivants regroupent les dimensions de l'ensemble des cubes de chaque Data-mart, avec la définition des hiérarchies, niveaux et colonnes correspondantes.

3.1.Data-Mart « *FACTURATION* » :

Dimension Hiérarchie		Niveau Colonne		Définition	Type
TD_TEMPS	ALL→N1→N2→N3→N4→N5	Niveau1=N1	- Année	Année de la date.	INT
		Niveau2=N2	- Saison	Saison à laquelle appartient une date.	INT
		Niveau3=N3	- Trimestre	Trimestre de la date.	INT
		Niveau4=N4	- mois	Numéro du mois.	INT
		Niveau5=N5	- date	La date au format complet.	DATE
TD_ENERGIE	ALL→N1	Niveau1=N1	- Type_energie	Type de l'énergie (gaz ou électricité).	CHAR
TD_PHASE_FACTURATION	ALL→N1	Niveau1=N1	- id_phase	Clé artificielle.	CHAR
			- nom_phase	Le nom de la phase de facturation.	CHAR
			- valeur_phase	La valeur de la phase de facturation	CHAR
			- type_client	Type du client (BT/BP/HTB/HTA/HP/MP).	CHAR
TD_NIVEAU	ALL→N1→N2→N3→N4→N5→N6	Niveau1=N1	- code_filiale	Code de filiale (sda,sdc,sdo,sde)	CHAR
			- filiale	La filiale à laquelle le lieu de consommation est affilié.	CHAR
		Niveau2=N2	- code_wilaya	Code de la wilaya.	INT
			- wilaya	Nom de la wilaya.	CHAR
		Niveau3=N3	- code_dr	Code de la direction de distribution.	INT
			- dr	Nom de la direction de distribution.	CHAR
		Niveau4=N4	- code_agence	Code de l'agence (une agence regroupe plusieurs communes).	INT

TD_NIVEAU_BT BP	ALL→N1→N2→N3→N4→N5→N6→N7→N8	Niveau5=N5	- agence	Nom de l'agence.	CHAR
			- code_commune	Code de la commune.	INT
			- commune	Nom de la commune.	CHAR
		Niveau6=N6	- id_groupe	Clé artificielle.	CHAR
			- numero_groupe	Code de groupe.	INT
			- nom_groupe	Nom de groupe	CHAR
		Niveau1=N1	- code_filiale	Code de filiale (sda,sdc,sdo,sde)	CHAR
			- filiale	La filiale à laquelle le lieu de consommation est affilié.	CHAR
		Niveau2=N2	- code_wilaya	Code de la wilaya.	INT
			- wilaya	Nom de la wilaya.	CHAR
		Niveau3=N3	- code_dr	Code de la direction de distribution.	INT
			- dr	Nom de la direction de distribution.	CHAR
		Niveau4=N4	- code_agence	Code de l'agence (une agence regroupe plusieurs communes).	INT
			- agence	Nom de l'agence.	CHAR
		Niveau5=N5	- code_commune	Code de la commune.	INT
			- commune	Nom de la commune.	CHAR
		Niveau6=N6	- numero_groupe	Code de groupe.	INT
			- nom_groupe	Nom de groupe	CHAR
		Niveau7=N7	- numero_tournee	Code de tournée.	INT
			- nom_tournee	Nom de la tournée.	CHAR
Niveau8=N8	- id_circuit	Clé artificielle.	CHAR		
	- nom_circuit	Nom de circuit.	CHAR		
	- numero_circuit	Code de circuit.	INT		

Tableau 15 : Définition des niveaux et hiérarchies des dimensions « FACTURATION »

3.2.Data-Mart « RELEVE » :

Dimension Hiérarchie		Niveau Colonne		Définition	Type
TD_TEMPS	ALL→N1→N2→N3→N4→N5	Niveau1=N1	- Année	Année de la date.	INT
		Niveau2=N2	- Saison	Saison à laquelle appartient une date.	INT
		Niveau3=N3	- Trimestre	Trimestre de la date.	INT
		Niveau4=N4	- mois	Numéro du mois.	INT
		Niveau5=N5	- date	La date au format complet.	DATE

TD_ENERGIE	ALL→N1	Niveau1=N1	- Type_energie	Type de l'énergie (gaz ou électricité).	CHAR
TD_CR_RELEVE	ALL→N1	Niveau1=N1	- id_cr_r	Clé artificielle.	CHAR
			- valeur_r	La valeur du compte rendu de la relève.	INT
			- type_client	Type du client (BT/BP/HTB/HTA/HP/MP).	CHAR
			- nom_r	Le nom du compte rendu de la relève.	CHAR
TD_CR_ANOMALIE S_SINGALEES	ALL→N1	Niveau1=N1	- id_cr_a	Clé artificielle.	CHAR
			- valeur_a	La valeur de l'anomalie signalée de la relève.	INT
			- type_client	Type du client (BT/BP/HTB/HTA/HP/MP).	CHAR
			- nom_a	Le nom de l'anomalie signalée de la relève.	CHAR
TD_NIVEAU	ALL→N1→N2→N3→N4→N5→N6	Niveau1=N1	- code_filiale	Code de filiale (sda,sdc,sdo,sde)	CHAR
			- filiale	La filiale à laquelle le lieu de consommation est affilié.	CHAR
		Niveau2=N2	- code_wilaya	Code de la wilaya.	INT
			- wilaya	Nom de la wilaya.	CHAR
		Niveau3=N3	- code_dr	Code de la direction de distribution.	INT
			- dr	Nom de la direction de distribution.	CHAR
		Niveau4=N4	- code_agence	Code de l'agence (une agence regroupe plusieurs communes).	INT
			- agence	Nom de l'agence.	CHAR
		Niveau5=N5	- code_commune	Code de la commune.	INT
			- commune	Nom de la commune.	CHAR

Tableau 16 : Définition des niveaux et hiérarchies des dimensions « RELEVE »

3.3. Data-Mart « RACCORDEMENT » :

Dimension	Hiérarchie	Niveau	Colonne	Définition	Type
TD_TEMPS	ALL→N1→N2→N3→N4→N5	Niveau1=N1	- Année	Année de la date.	INT
		Niveau2=N2	- Saison	Saison à laquelle appartient une date.	INT
		Niveau3=N3	- Trimestre	Trimestre de la date.	INT
		Niveau4=N4	- mois	Numéro du mois.	INT
		Niveau5=N5	- date	La date au format complet.	DATE

TD_ENERGIE	ALL→N1	Niveau1=N1	- Type_energie	Type de l'énergie (gaz ou électricité).	CHAR		
TD_BRANCHEMENT	ALL→N1	Niveau1=N1	- id_b	Clé artificielle.	CHAR		
			- valeur_b	La valeur de branchement.	INT		
			- type_client	Type du client (BT/BP/HTB/HTA/HP/MP).	CHAR		
			- nom_b	Le nom de branchement.	CHAR		
			- type_b	Type de branchement.			
TD_EXTENSION	ALL→N1	Niveau1=N1	- id_e	Clé artificielle.	CHAR		
			- valeur_e	La valeur de l'extension.	INT		
			- type_client	Type du client (BT/BP/HTB/HTA/HP/MP).	CHAR		
			- nom_e	Le nom de l'extension.	CHAR		
			- type_e	Le type de l'extension	CHAR		
TD_DELAIS	ALL→N1	Niveau1=N1	- id_d	Clé artificielle.	CHAR		
			- valeur_d	La valeur de délais	FLOAT		
			- type_client	Type du client (BT/BP/HTB/HTA/HP/MP).	CHAR		
			- nom_d	Nom de délais	CHAR		
			- type_d	Type de délais	CHAR		
TD_NIVEAU_RCN	ALL→N1→N2 →N3→N4→N5 →N6	Niveau1=N1	- code_filiale	Code de filiale (sda,sdc,sdo,sde)	CHAR		
			- filiale	La filiale à laquelle le lieu de consommation est affilié.	CHAR		
		Niveau2=N2	- code_wilaya	Code de la wilaya.	INT		
			- wilaya	Nom de la wilaya.	CHAR		
		Niveau3=N3	- code_dr	Code de la direction de distribution.	INT		
			- dr	Nom de la direction de distribution.	CHAR		
				Niveau4=N4	- code_agence	Code de l'agence (une agence regroupe plusieurs communes).	INT
					- agence	Nom de l'agence.	CHAR
		Niveau5=N5	- code_commune	Code de la commune.	INT		
			- commune	Nom de la commune.	CHAR		

Tableau 17 : Définition des niveaux et hiérarchies des dimensions « RACCORDEMENT »

4. Conclusion

Au niveau de ce chapitre, nous avons identifié les mesures, les faits qui leur sont associés et les dimensions et les hiérarchies selon les lesquelles seront analysés ces faits, et cela en tenant compte des besoins des utilisateurs définis précédemment.

Nous avons obtenu les schémas en étoiles qui vont constituer chaque magasin de donnée. On a regroupé les trois data-mart dans un schéma en constellation qui est notre objectif DW_CLIENT. L'entrepôt de données que nous avons conçu va être exploité par les utilisateurs.

Nous passons donc à l'étape suivante qui est la Conception et mise en œuvre du système d'analyse.

Partie 3

Conception et mise en œuvre du système d'analyse

Chapitre VIII

Conception du générateur de rapports décisionnels

1. Introduction

Ce chapitre constitue la finalité de notre travail. Nous allons parler, tout d'abord, de la conception du générateur de rapports décisionnels ainsi que des métadonnées ; nous exposerons, ensuite, le mécanisme de fonctionnement du générateur.

L'objectif principal du générateur est de permettre la génération de rapports selon les besoins des utilisateurs basés sur des modèles multidimensionnels des données. En effet, la création, la modification et la génération des rapports se fait de façon très simple et intuitive. L'analyse de ces rapports, particulièrement des indicateurs, permet l'extraction d'informations pertinentes par rapport à des dimensions choisies au moment du paramétrage du rapport afin d'améliorer le processus de prise de décisions.

2. Démarche suivie pour la conception

Nous allons effectuer la modélisation de notre système en nous basant sur le formalisme UML. Pour commencer, nous présenterons les différents acteurs et les cas d'utilisation de notre système. La deuxième étape consistera à définir les diagrammes de séquences, ce qui permettra de visualiser les cas d'utilisation d'une manière plus détaillée. Enfin, le diagramme de classes permettra de visualiser l'ensemble des métadonnées nécessaires au fonctionnement du générateur des rapports décisionnels.

2.1. Présentation des acteurs et des cas d'utilisation

Un acteur représente un rôle joué par une entité externe (utilisateur humain, dispositif matériel ou autre système) qui interagit directement avec le système étudié. Un acteur peut consulter et/ou modifier directement l'état du système. Nous avons présenté les acteurs de notre système de la manière suivante :

- **Décideur** : Un responsable SONELGAZ autorisé à consulter les rapports décisionnels et les statistiques relatifs à son niveau hiérarchique.
- **Administrateur** : Personne désignée pour l'administration du générateur de rapports et pour la gestion des utilisateurs (décideurs).

Nous considérons deux classes de cas d'utilisation, la première concerne le générateur de rapports décisionnels, la seconde concerne la consultation des rapports décisionnels. Les figures ci-dessous en sont une illustration :

2.1.1 Diagramme de cas d'utilisation général :

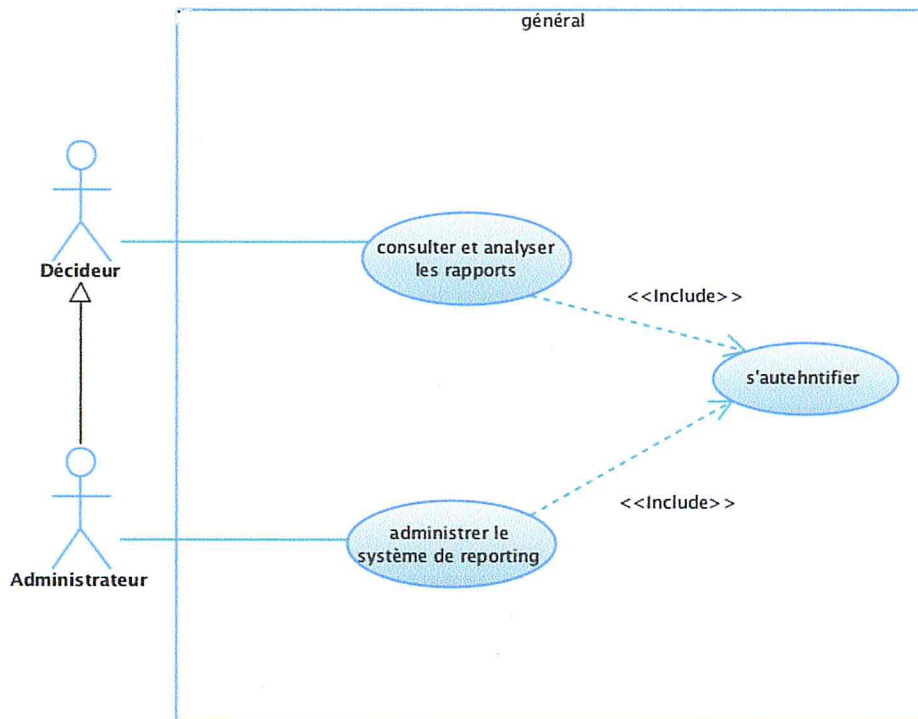


Figure 37 : Diagramme de cas d'utilisation général

2.1.2 Diagramme de cas d'utilisation gestion d'administration détaillé :

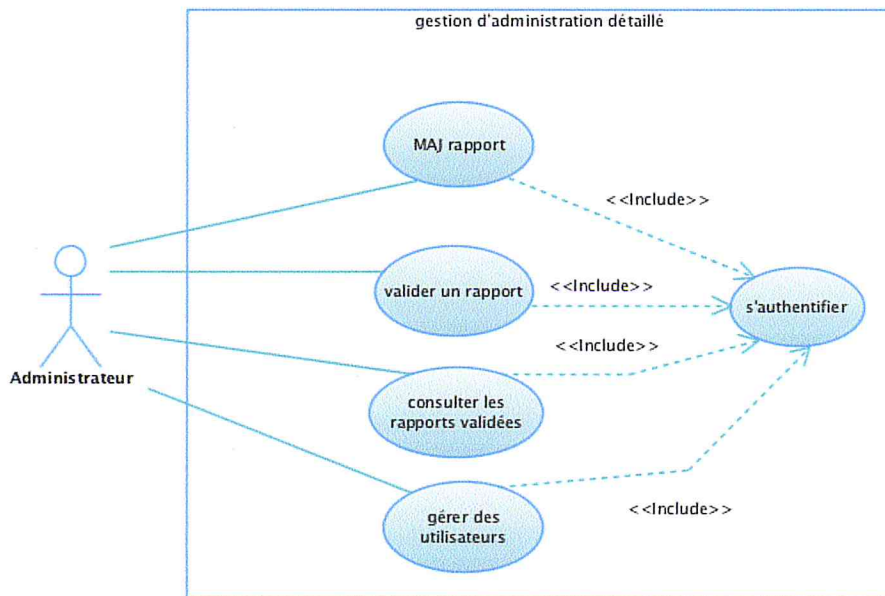


Figure 38 : Diagramme de cas d'utilisation gestion d'administration détaillé

2.1.3 Diagramme de cas d'utilisation MAJ d'un rapport :

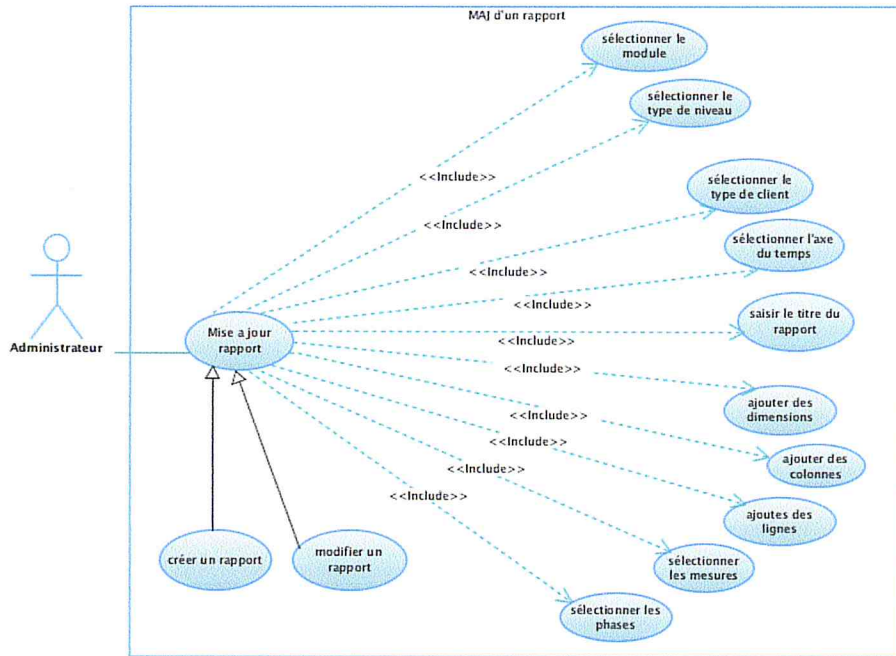


Figure 39 : Diagramme de cas d'utilisation MAJ d'un rapport

2.1.4 Diagramme de cas d'utilisation validation des rapports :

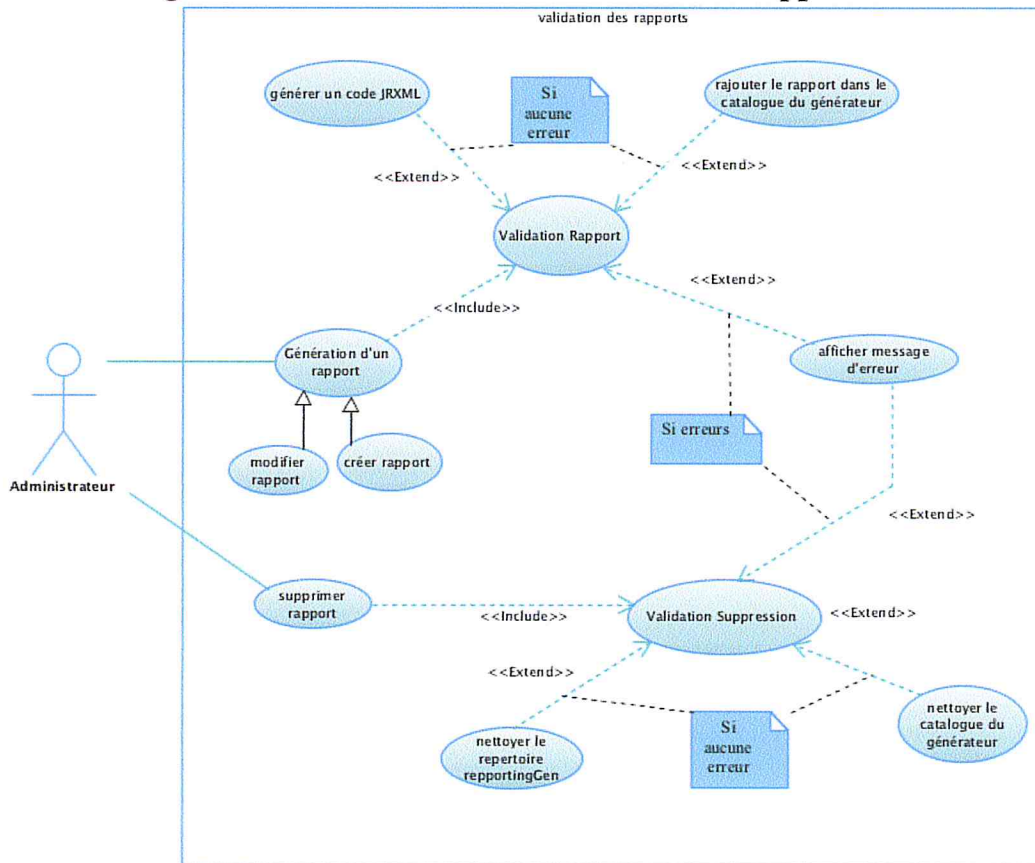


Figure 40 : Diagramme de cas d'utilisation validation des rapports

2.1.5 Diagramme de cas d'utilisation gestion des utilisateurs :

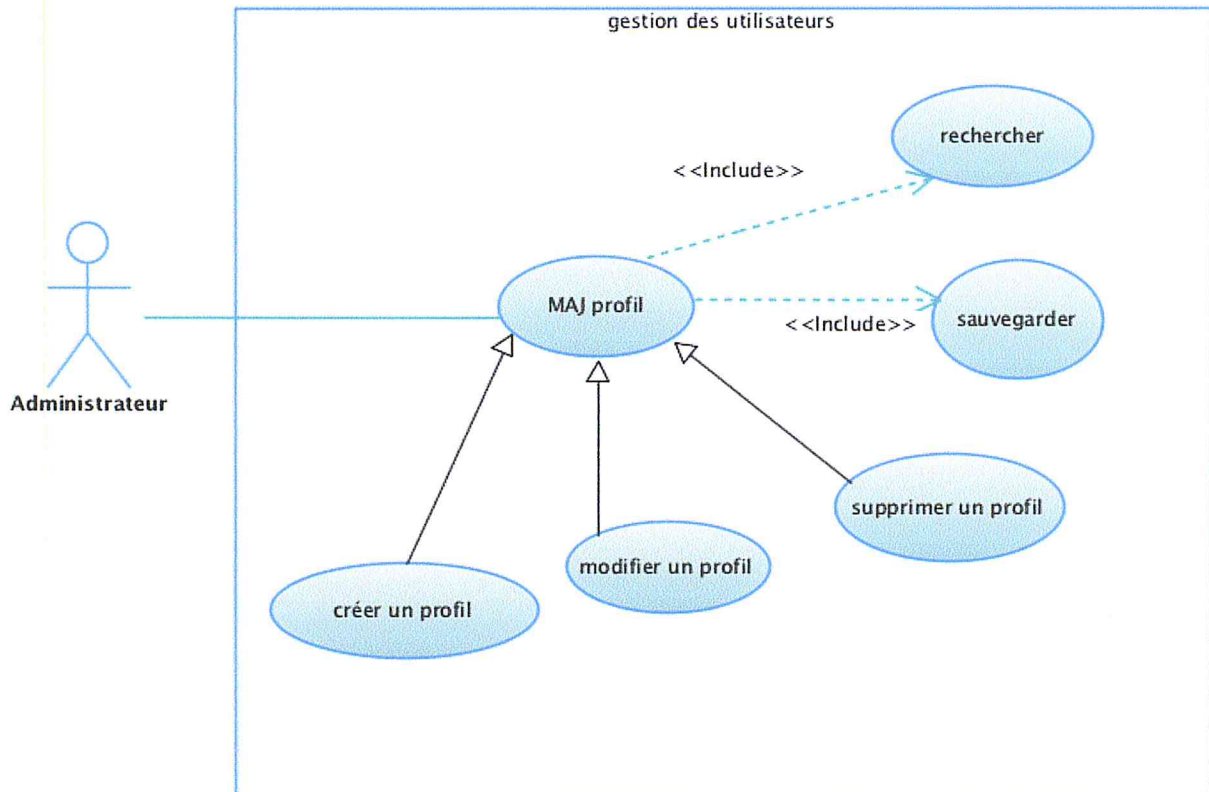


Figure 41 : Diagramme de cas d'utilisation gestion des utilisateurs

2.1.6 Diagramme de cas d'utilisation consultation des rapports décisionnels :

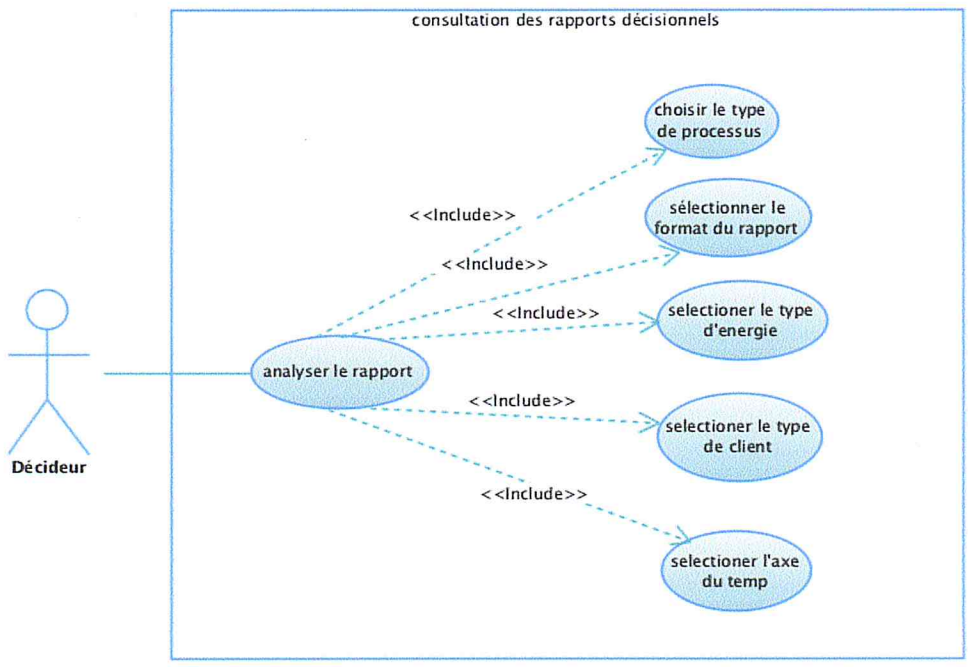


Figure 42 : Diagramme de cas d'utilisation consultation des rapports décisionnels

2.2 Diagrammes de séquence

Dans cette phase, et après identification des cas d'utilisation, et des scénarios associés à chaque cas, nous les représentons à l'aide des diagrammes de séquence, à titre d'exemple, la figure 43 nous montre le diagramme de séquence associé au cas d'utilisation "créer un rapport" :

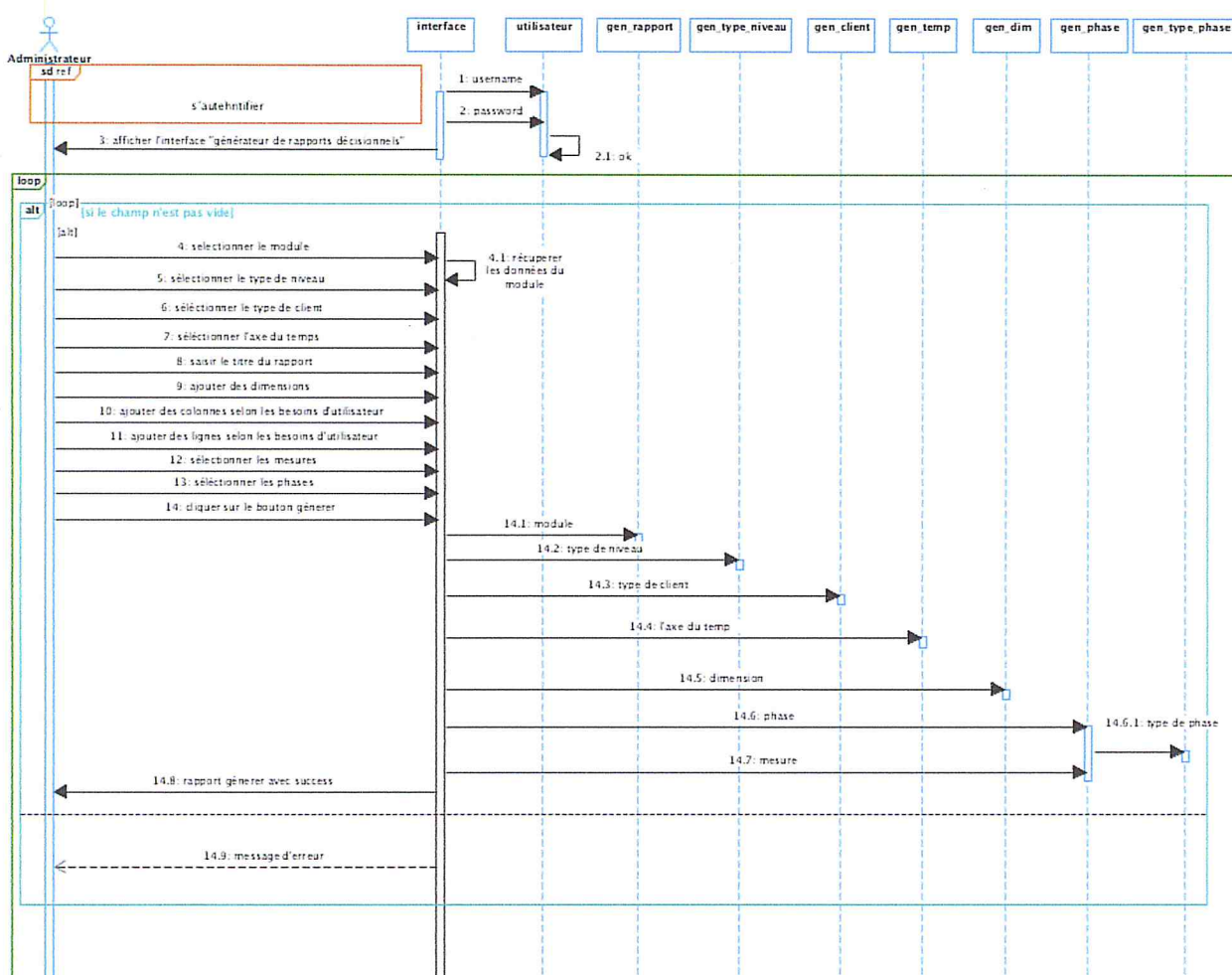


Figure 43 : Diagramme de séquence « créer un rapport »

Scénario :

- L'administrateur s'authentifie au site.
- L'administrateur sélectionne le type de niveau et client et l'axe du temps.
- Saisir le titre de rapport
- Ajouter des dimensions

- Ajouter des colonnes et des lignes selon les besoins
- Sélectionner les mesures et les phases (les indicateurs).
- Vérification du champ
- Le rapport est généré.
- **L'enchaînement d'erreurs :**
- E1 : un champ vide.
- L'interface affiche message d'erreur.

2.3 Diagramme de classe

La dernière étape de notre démarche de conception consiste à définir le diagramme de classe qui décrit le catalogue de notre système. Le catalogue associé au générateur de rapports contient toutes les métadonnées nécessaires pour le fonctionnement de celui-ci. Le diagramme de classes est le point central dans un développement orienté objet. Il a pour objectif de décrire la structure des entités manipulées par les utilisateurs. La figure 44 représente notre diagramme de classe.

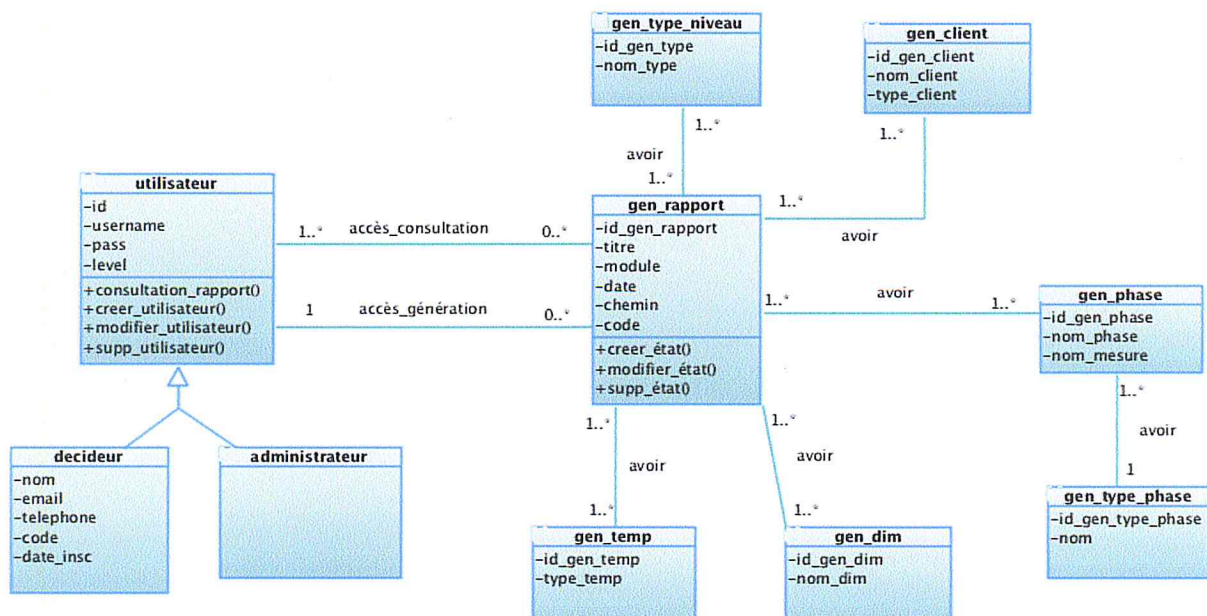


Figure 44 : diagramme de classe

2.3.1 Description des classes de générateur de rapports décisionnels :

Le tableau suivant décrit la description des classes de générateur de rapports décisionnels.

Classe : Nom de classe			
Description de la classe.			
Désignation	Description	Type	Observation
Classe : Utilisateur			
« Elle contient les informations nécessaires d'un compte utilisateur. »			
Id	Identificateur de l'utilisateur	INT(9)	Identifiant (PK, not null)
Username	Nom d'utilisateur	Varchar(50)	
Pass	Mot de passe	Varchar(90)	
Level	Numero du Niveau de l'utilisateur	INT(1)	
Classe : Décideur			
« Est une classe qui contient les informations nécessaires d'un compte décideur. Décideur est une classe héritée de la classe mère Utilisateur »			
Nom	Nom du décideur	Varchar(50)	
Email	L'Email du décideur	Varchar(100)	
Téléphone	Téléphone du décideur	Varchar(100)	
Code	Code de niveau	Varchar(50)	
Date-insc	Date d'inscription de décideur	DATE	
Classe : Administrateur			
« Est une classe qui permet de créer les comptes des utilisateurs et de générer des rapports décisionnels." Administrateur " est une classe héritée de la classe mère " Utilisateur " »			
Classe : Gen_report			
« Elle contient toutes les informations concernant un rapport. »			
Id_gen_report	Identificateur du rapport	INT(9)	Identifiant (PK, not null)

Conception du générateur de rapports décisionnels

Titre	Titre du rapport	Varchar(100)	
Module	Le nom du module du rapport	Varchar(100)	
Date	Date générer le rapport	DATE	
Chemin	Le chemin de fichier Jrxml	Varchar(100)	
Code	Code validation de rapport	Varchar(50)	
Classe : Gen_client			
« Est une classe qui contient les informations des types de client »			
Id_gen_client	Identificateur de type client	INT(9)	Identifiant (PK, not null)
Nom_client	Nom du type client	Varchar(100)	
Type_client	Type d'énergie de client	Varchar(100)	
Classe : Gen_type_niveau			
« Elle contient les information des types de niveau »			
Id_gen_type	Identificateur de type de niveau	INT(9)	Identifiant (PK, not null)
Nom_type	Nom du niveau	Varchar(100)	
Classe : Gen_phase			
« Est une classe qui contient les information des phases d'un rapport »			
Id_gen_phase	Identificateur de phase	INT(9)	Identifiant (PK, not null)
Nom_phase	Nom de la phase	Varchar(100)	
Nom_mesure	Nome de la mesure	Varchar(100)	
Classe : Gen_type_phase			
« Est une classe qui contient les information de type des phases »			
Id_gen_type_phase	Identificateur de type phase	INT(9)	Identifiant (PK, not null)
Nom	Nom de type de la phase	Varchar(100)	
Classe : Gen_dim			
« Elle contient les informations des dimension d'un rapport »			
Id_gen_dim	Identificateur de dimension	INT(9)	Identifiant (PK, not null)

Nom_dim	Nom de la dimension	Varchar(100)	
Classe : Gen_temp			
« Elle contient les information des types de temps d'un rapport »			
Id_gen_temp	Identificateur du temps	INT(9)	Identifiant (PK, not null)
Type_temp	Type du temps	Varchar(100)	

Tableau 18 : Description des classes de générateur de rapports décisionnels

3. Mécanisme du générateur de rapports décisionnels

L'objectif est ici de présenter les concepts généraux du générateur.

3.1 Architecture du générateur de rapports

La figure 45 donne un aperçu de l'architecture du générateur de rapports décisionnels :

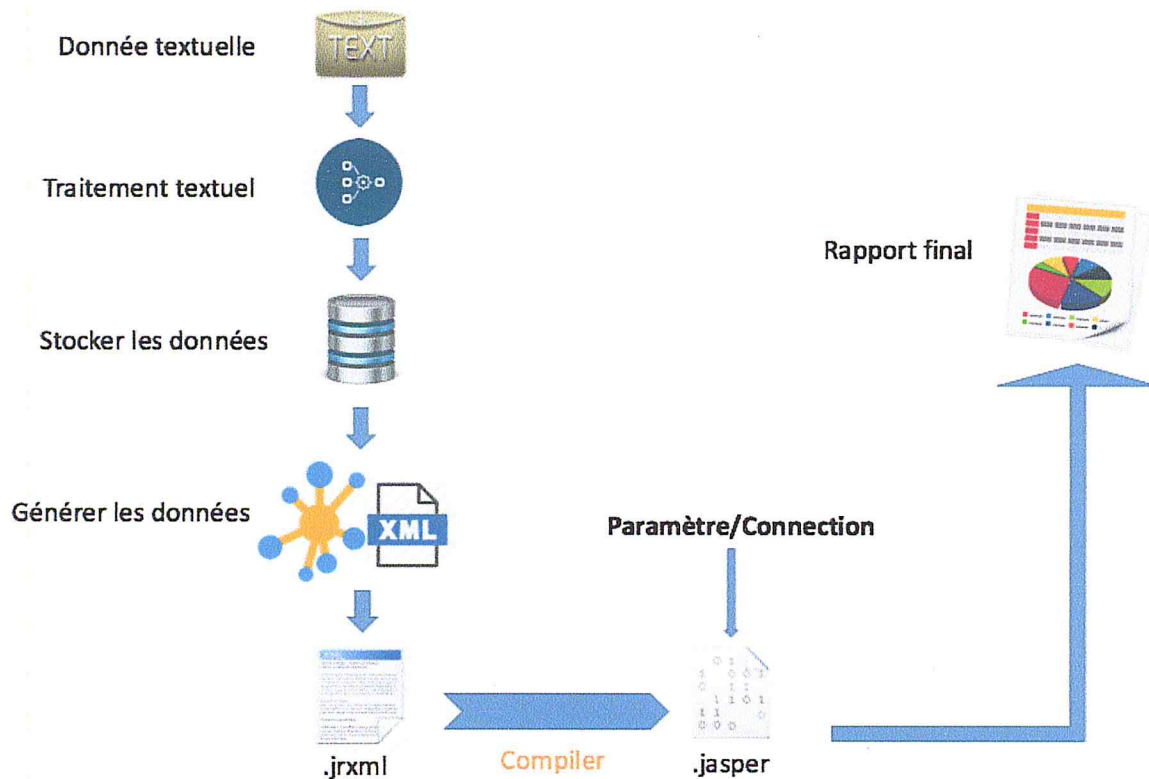


Figure 45 : Architecture de générateur de rapport décisionnels

1) **Données textuelles :**

Ces données représentent des phases, des dimensions et des mesures étalées sur une palette grâce à laquelle l'utilisateur pourra effectuer son choix.

1) **Traitement textuel :**

Après avoir choisi les données nécessaires, cette partie permet de partitionner et arranger ces derniers selon leurs types.

2) **Stocker les données :**

La meilleure solution pour stocker les ensembles des données partitionnés précédemment et les placer dans leurs tables associées, est d'utiliser le catalogue du système, i.e., la base de données associée au générateur

3) **Générer les données :**

Cette partie constitue la phase essentielle du mécanisme du générateur. Ce dernier génère une syntaxe XML (XML de Jasper iReport) et crée une requête SQL à partir de la base de données pour chaque état sélectionné et génère, enfin, une structure d'un rapport sous format JRXML.

Le fichier JRXML du rapport est composé de 10 parties, que nous détaillons ci-après :

- **Background** : l'arrière-plan du rapport (logo, etc ...)
- **Title** : Titre
- **PageHeader** : en-tête de page
- **ColumnHeader** : en-tête de colonne
- **Detail** : corps du rapport
- **ColumnFooter** : pied de colonne
- **PageFooter** : pied de page
- **LastPageFooter** : pied de la dernière page
- **Summary**.
- **No-Data** : sera affiché lorsque la requête SQL (par exemple) ne renvoie aucune donnée.

Chaque paramètre, variable, champ est typé. Les types utilisés sont les classes enveloppées de Java (String,Integer, BigDecimal....), et non les types primitifs,

Une fois que ces quatre (4) étapes sont achevées, nous obtenons un fichier JRXML, qui sera compilé par la suite afin de pouvoir l'utiliser dans un fichier sous format Jasper, cette compilation se réalise dans le code de l'application. Il est extrêmement inutile et coûteux de compiler à chaque fois le fichier XML, lors des appels du rapport. Pour cette raison, le fichier Jasper sera utilisé en combinaison avec les paramètres et la source identifiée dans le catalogue du générateur, nous obtenons alors le rapport final.

4. Conclusion

Nous avons présenté dans ce chapitre la conception et le mécanisme du générateur de rapports décisionnels avec l'analyse multidimensionnelle à partir d'un entrepôt de données DW_CLIENT. Le résultat obtenu est un outil qui offre à la fois l'aspect instrument de contrôle et permettra d'améliorer la qualité, l'intégrité et la transparence des informations utilisées pour éclairer d'une manière efficace, propre aux rapport.

A ce niveau, nous disposons de tous les éléments nécessaires pour entamer la phase description des applications et mise en œuvre, celle-ci est présentée en détail dans le chapitre suivant.

Chapitre IX

Description des applications et mise en
œuvre

1. Introduction

La conception de notre système étant achevée, nous passons à la phase de réalisation. Ce chapitre a pour objectif de décrire l'environnement de développement de l'application.

Nous allons aussi présenter l'implémentation de la solution proposée pour la réalisation de l'application en citant ses différentes interfaces et fonctionnalités.

Nous obtenons à la fin de cette étape un système fonctionnel et nous donnons un aperçu visuel de ce dernier avec ces quelques captures d'écran.

2. Environnement du travail

Dans cette partie, nous allons étudier le choix des outils matériels et surtout les outils logiciels du développement web.

2.1 Environnement matériel

Nous mentionnons les caractéristiques de l'ordinateur sur lesquelles nous avons développé l'application parce qu'elles peuvent donner une idée sur les conditions du travail.

L'application a été développée sur un ordinateur portable MacBook pro qui se caractérise par :

- **Processeur** : Processeur bicœur Intel Core i5 cadencé à 2,5 GHz.
- **Mémoire** : 2,5 GHz4 Go de mémoire DDR3 à 1 600 MHz.

Nous avons énuméré au cours de cette partie les différents outils utilisés tout au long de ce projet pour l'étude et la mise en place de notre application.

- **Système d'exploitation**

Nous avons utilisé comme système d'exploitation :

- MacOS Sierra.

2.2. Outil de modélisation UML



❖ Visual Paradigm V 14.1

Nous avons exploité pour la modélisation UML de l'application l'outil Enterprise

Architect qui est flexible, complet et puissant, conçu pour les plateformes Windows. C'est un outil de création de modèles dont le langage est UML.

2.3.ETL :

❖ **Talend Open Studio (TOS) V5.0.1**

Est une plate-forme d'intégration de données Open Source, basée sur le langage Java. TOS permet de répondre à toutes les problématiques liées au traitement des données dans la chaîne décisionnelle :

- ETL : Extraction, Transformation, et Chargement des données
- EAI : Echange de données Inter-Application
- Synchronisation des données

L'une des grandes forces de TOS réside dans la fait de pouvoir se connecter à quasiment toutes les sources de données, applications métier et type de fichier existant. Et ce grâce à plus de 250 composants utilisables par les développeurs. Parmi ses composants, on trouve différentes familles. Voici quelques exemples de familles de composants :

- **Applications Métier (Mode Ecriture, Lecture) :** Microsoft CRM, SAP, Sage CRM, Salesforce, SugarCRM, ...
- **Base de données (Mode Ecriture, Lecture) :** AS400, MS SQL, Oracle, DB2, MySQL, PostgreSQL, Access, ODBC, ...
- **Fichier (Mode Ecriture, Lecture) :** Excel, CSV, TXT, ...
- **Internet :** FTP, WebServices, HTTP, SSH, ...
- **Orchestration :** Fusion des flux, Réplication des flux, mise en attente de l'exécution, itération sur l'ensemble du contenu d'un répertoire, ...
- **Qualité de données :** Unicité des données, remplacement de caractère dans une chaîne, changement de l'encodage d'un fichier, ...
- **Transformation:** Agrégation, Conversion de type, Filtre, Tri, Mappage
- **XML :** Grâce à sa communauté active de développeurs, TOS s'enrichit de jour en jour en nouveaux composants, jobs, routines, ... Ces éléments sont alors mis à disposition des utilisateurs pour ensuite être utilisés. [33]

2.4. Environnement de développement intégré :



❖ NetBeans V8.1

Nous avons choisi NetBeans comme Environnement de Développement Intégré(EDI) open source lancé par SUN en juin 2009 qui permet de développer des applications Java, PHP, C, C++ et Ruby. Il comprend toutes les caractéristiques d'un IDE moderne (éditeur en couleur, projets multi-langage, refactoring, éditeur graphique d'interfaces et de pages Web). Netbeans est notamment spécialisé pour le développement J2EE. Il permet de gérer un serveur d'application, de générer rapidement un Webservice avec l'API JAX-WS, de générer la structure d'une application pour poste de travail et permet également le développement d'applications Web. [34]

2.5. Serveur d'application



❖ Glassfish V 4.1

Glassfish est un serveur d'application compatible Java EE créé par Sun Microsystems en 2005 en tant que Sun Application Server. Ce serveur permet le développement d'applications distribuées en utilisant les technologies comme EJB3, JPA, JSF, JAX-WS et d'autres.

2.6. Outil d'administration de la base de données



❖ PostgreSQL V 9.4

PostgreSQL est un système de gestion de base de données relationnelle et objet (SGBDRO). C'est un outil libre disponible selon les termes d'une licence de type BSD. Ce système est concurrent d'autres systèmes de gestion de base de données, qu'ils soient libres (comme MariaDB, MySQL et Firebird), ou propriétaires (comme Oracle, Sybase, DB2, Informix et Microsoft SQL Server). Comme les projets libres Apache et Linux, PostgreSQL n'est pas contrôlé par une seule entreprise, mais est fondé sur une communauté mondiale de développeurs et d'entreprises. [35]

2.7. Outil de reporting



❖ JasperReports / iReport V5.6.0

JasperReports est un outil de Reporting Open Source, offert sous forme d'une bibliothèque qui peut être embarquée dans tous types d'applications Java. JasperReports se base sur des fichiers JRXML (qui ont la structure d'un fichier XML) pour la présentation des états. Il permet d'exporter des rapports aux formats : PDF, HTML, XLS, CSV, XML, RTF, TXT. [36]

2.8. Plate-forme de développement « JEE »

L'application a été développée sur la base de la plate-forme de développement JEE (Java Enterprise Edition). Le choix de cette plate-forme a été justement imposé par le client, c'est d'ailleurs la plate-forme de développement utilisée dans la majorité des projets de ELIT.

Cette plate-forme représente en fait un ensemble de technologies et d'APIs pour la plate-forme Java conçus pour soutenir les applications d'entreprise. En général, il s'agit d'applications complexes, à grande échelle, distribuées, transactionnelles et hautement disponibles conçues pour répondre aux besoins d'affaires critiques.

Framework de développement « JSF »

JSF est un framework implémentant le modèle MVC (figure ci-dessous), pour le développement d'applications web JEE. Il permet la mise en place d'un mapping entre HTML et les objets, fournissant ainsi un environnement de développement qui permet la construction d'une interface de type web.

Ce Framework est basé sur une approche de composants rappelant les interfaces graphiques natives, on réagit aux actions de l'utilisateur en mettant en place des composants (boutons, champs de texte, etc.). On ne traite donc pas des requêtes HTTP mais des actions de l'utilisateur. [15]

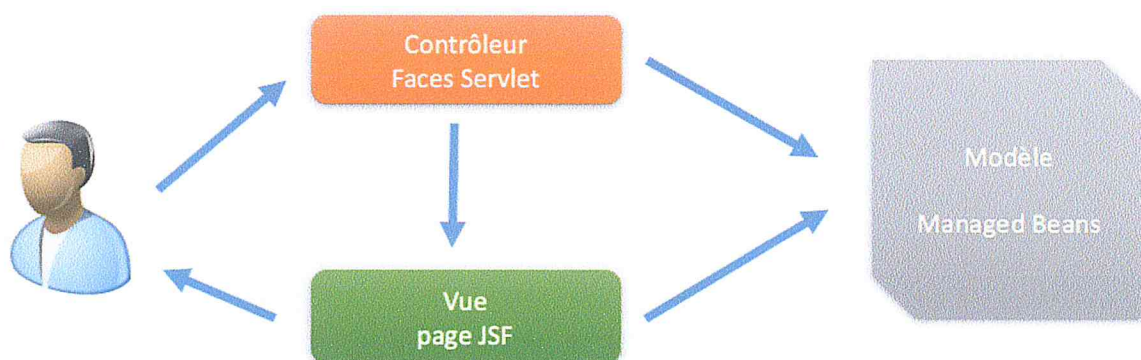


Figure 46 : Le modèle MVC du Framework JSF

3. Architecture technique de la solution

La figure suivant illustre la structure et l'architecture technique de la solution proposée :

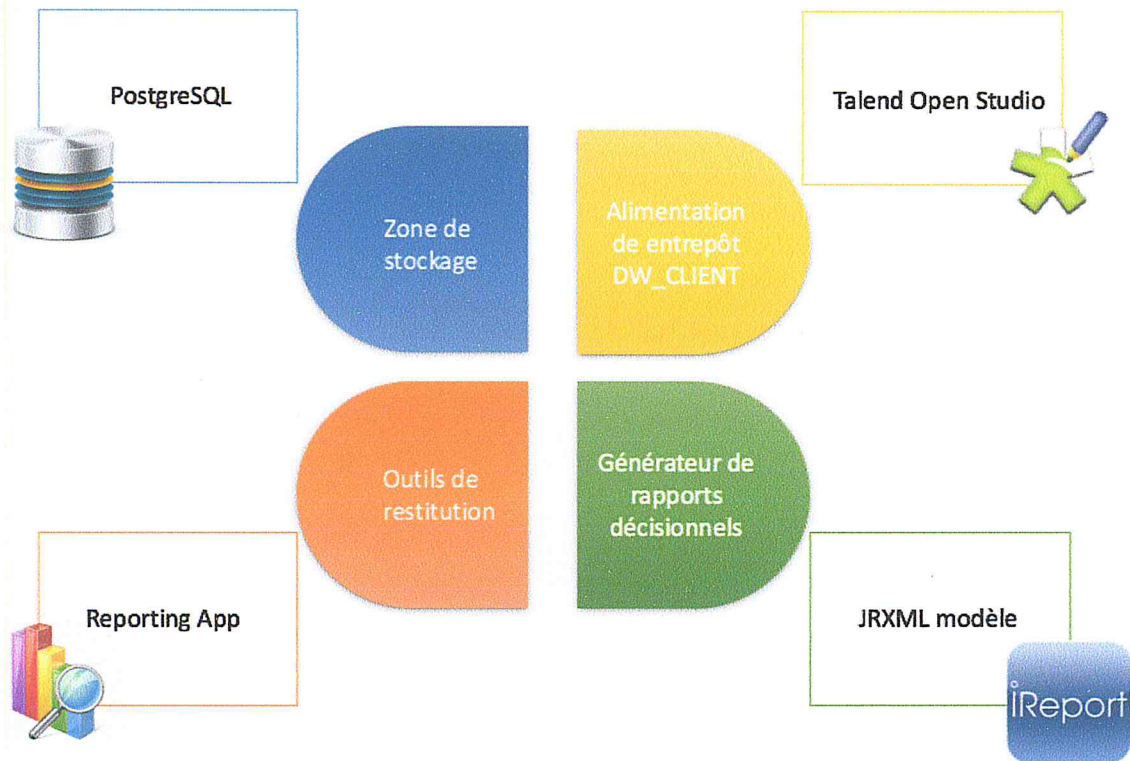


Figure 47 : l'architecture technique de la solution proposée

4. Réalisation du projet

4.1 Construction de l'entrepôt de données

Nous avons créé une base de données relationnelle nommée « DW_CLIENT » sous le SGBD PostgreSQL en suivant le modèle ROLAP. Ce dernier consiste à :

- Les dimensions sont des tables avec des clés primaires créées en utilisant un code spécifique ou générées à l'aide de concaténation,
- Les tables de faits sont des tables avec des clés séquentielles et des clés étrangères de toutes les tables de dimensions associées au fait.

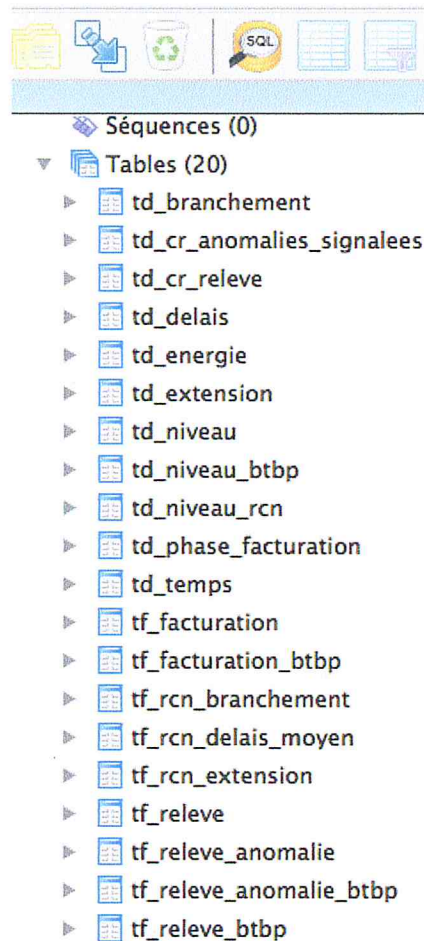


Figure 48 : PostgreSQL

4.2 Extraction, transformation et chargement

Une fois la base de données créée, nous procédons à l'ETL à l'aide de l'outil Talend Open Studio.

La première étape consiste à définir les contextes, chaque contexte correspondant à une base de données précise. Nous appliquerons des processus de manipulation et de transformation des données (appelés jobs) en faisant varier le contexte.

Nous importons le schéma des tables (correspondant aux connexions définies) dont nous avons besoin, ces tables seront utilisées dans les jobs. Ainsi les tables des bases de données sources seront utilisées dans les composants d'entrée et les tables de l'entrepôt dans les composants de sortie, un composant de type « tMap » sera utilisé pour faire les correspondances entre les deux composants.

Nous créons les jobs de remplissage des tables de dimensions et des tables de faits. Par la suite, nous exécuterons les jobs de remplissage des tables de dimensions avant ceux des tables

de faits, car les premiers référencent les seconds à l'aide de clés étrangères. Le remplissage de certaines tables de faits nécessite une synchronisation particulière (charger à partir de nombreuses tables dans un ordre précis). La figure suivante illustre un exemple du job de chargement processus de la relève :

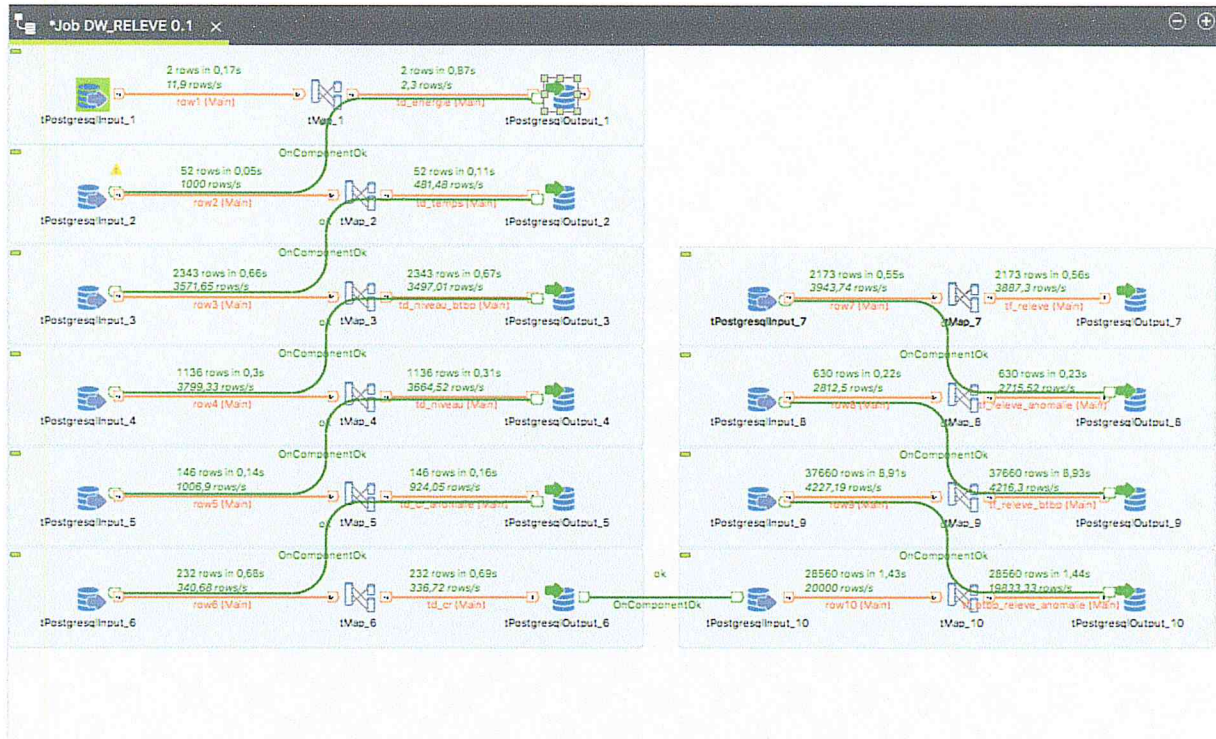


Figure 49 : job de chargement processus de la RELEVE

5. Implémentation des interfaces du système

Nous présentons dans cette section le prototype d'application que nous avons réalisé. Nous allons pour cela prendre comme scénario un utilisateur ayant tous les droits afin que l'on puisse exposer l'ensemble des fonctionnalités de notre système.

5.1. Authentification

Afin de pouvoir accéder à l'application, l'utilisateur étant un 'administrateur' ou un 'décideur', doit d'abord s'authentifier en utilisant les paramètres de connexion qui lui ont été fournis (Nom d'utilisateur et mot de passe).

La figure ci-dessous montre l'interface qui permet à l'utilisateur de s'authentifier.

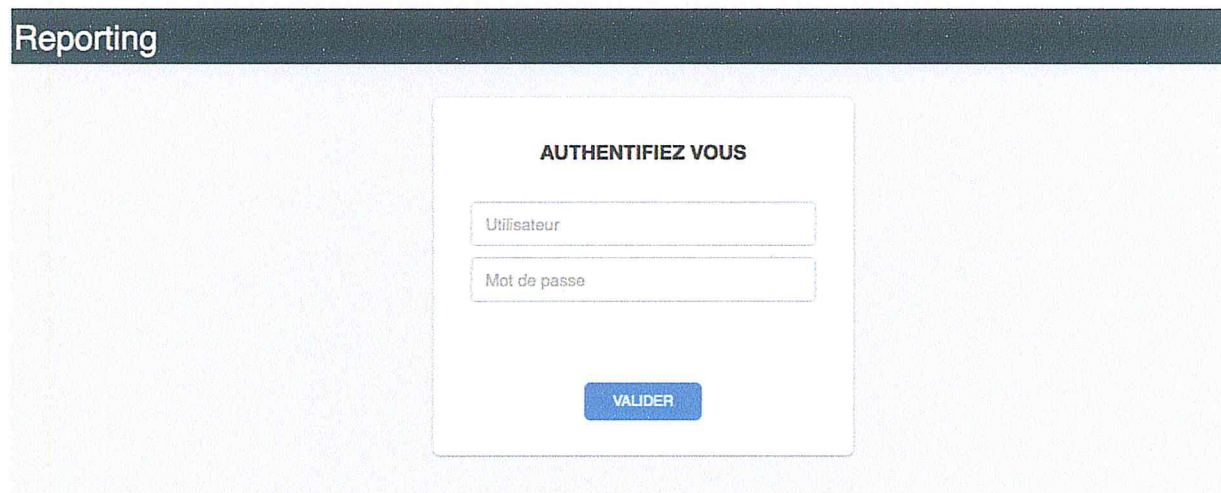


Figure 50 : l'interface qui permet à l'utilisateur de s'authentifier

5.2. Page d'accueil :

Suite à l'authentification, l'utilisateur se dirige selon son type vers des modules différents, tel que l'administrateur accède au « Générateur de rapport décisionnel », ou il pourra gérer les rapports ainsi que les utilisateurs. Quant au décideur, ce dernier accède à la « Consultation des rapports de SONELGAZ ».

5.2.1. Gestion des rapports :

5.2.1.1. Création des rapports :

Afin de pouvoir créer un rapport, l'administrateur doit suivre les étapes suivantes :

- Sélectionner le module (Facturation, Relève ou Raccordement).

- Sélectionner le type de niveau (sd, dd ou agences).
- Sélectionner type client, ce dernier change en fonction du module choisi.
- Sélectionner l'axe du temps.
- Écrire un titre du rapport.
- Cliquer sur le bouton 'ajouter dimension', et glisser les différentes dimensions : temps, niveau et énergie.
- Former les lignes et les colonnes par rapport au 'rapport' voulu.
- Glisser les mesures.
- Glisser les phases (les indicateurs).
- Cliquer sur le bouton 'générer'.

La figure ci-dessous, représente l'interface permettant la création des rapports.

The screenshot shows the 'générateur de rapports décisionnels' interface. The main area is titled 'Créer un rapport'. It features a search bar, a dropdown menu for 'Facturation', and a section for 'Sélectionner le type de niveau' with options 'sd', 'dd', and 'agence'. There are also buttons for 'Ajouter dimension', 'Supprimer', and 'Générer'. A table at the bottom shows various data fields with checkboxes. A red circular stamp is visible on the right side of the interface.

Figure 51 : l'interface permettant la création des rapports

5.2.1.2. Générer le code des Rapports :

Après la création d'un rapport, et afin de pouvoir générer son code l'administrateur accède à la liste des rapports générés, Sélectionne le rapport voulu et clique sur le bouton « Générer code ».



Figure 52 : l'interface permettant la génération du rapport

Une fois le code généré, deux fichiers seront créés l'un sous format Jasper et l'autre sous format Jrxml :

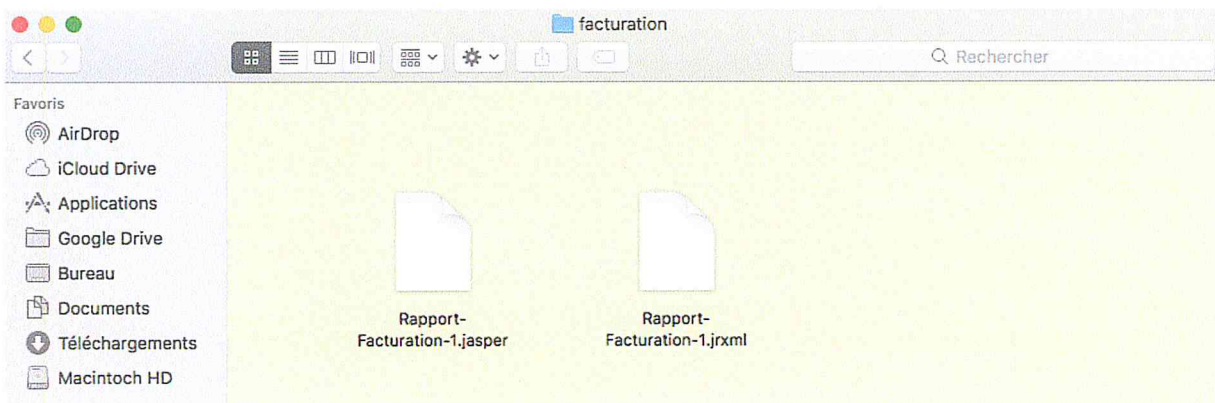


Figure 53 : deux fichiers seront créés l'un sous format Jasper et l'autre sous format Jrxml

5.2.2. Gestion des utilisateurs :

5.2.2.1. Création des utilisateurs :

L'administrateur est responsable sur la création des décideurs, pour ce faire, il doit cliquer sur le bouton 'gérer des utilisateurs', puis sur 'Ajouter un utilisateur', ensuite remplir les différents champs du formulaire, et en fin cliquer sur « submit »

La figure ci-dessous, montre l'interface permettant la création des utilisateurs :

The screenshot shows the 'gestion des utilisateurs' interface. On the left, there is a sidebar with navigation options: 'Créer un rapport', 'Liste des rapports générés', 'Liste des rapports validés', and 'Gérer des utilisateurs'. The main area is titled 'Ajouter un utilisateur' and contains the following form fields:

- Nom d'utilisateur: sda
- Mot de passe:
- Level: SD
- Code: SDA
- Nom et prénom: Nacer Tadbirt
- Email: tadbirt.n@gmail.com
- Numéro de téléphone: 0555679927

A 'Submit' button is located at the bottom of the form. A search bar with the text 'Rechercher ...' is visible in the top right corner.

Figure 54 : l'interface permettant la création des utilisateurs

5.2.2.2. Liste des utilisateurs :

Une fois qu'un décideur est créé, la liste des utilisateurs s'affiche, afin de permettre à l'administrateur de vérifier, modifier ou supprimer les différents utilisateurs existants.

The screenshot shows the 'gestion des utilisateurs' interface displaying a list of users. The table has the following columns: ID, Level, Nom de désignation, Nom d'utilisateur, and Nom et Prénom. The data is as follows:

ID	Level	Nom de désignation	Nom d'utilisateur	Nom et Prénom
1	Administrateur	admin	admin	admin
2	SD	SDA	sda	sda
3	DD	El Harache	dd	harach
4	Agence	Agence de El Harache	agence	agencehara

At the bottom of the table, there is a pagination control showing '10' and navigation arrows. A search bar with the text 'Rechercher ...' is visible in the top right corner.

Figure 55 : l'interface de la liste des utilisateurs

5.2.3. Consultation des rapports :

Dès qu'un utilisateur soit créer, il pourra accéder au système tout en s'authentifiant :

Figure 56 : l'interface de l'authentification de décideur SDA

Une fois l'authentification vérifiée, la page « Consultation des rapports Sonelgaz » sera affichée, l'accueil de cette dernière contient les différentes statistiques de facturation, relève et raccordement, montrées dans la figure qui suit :

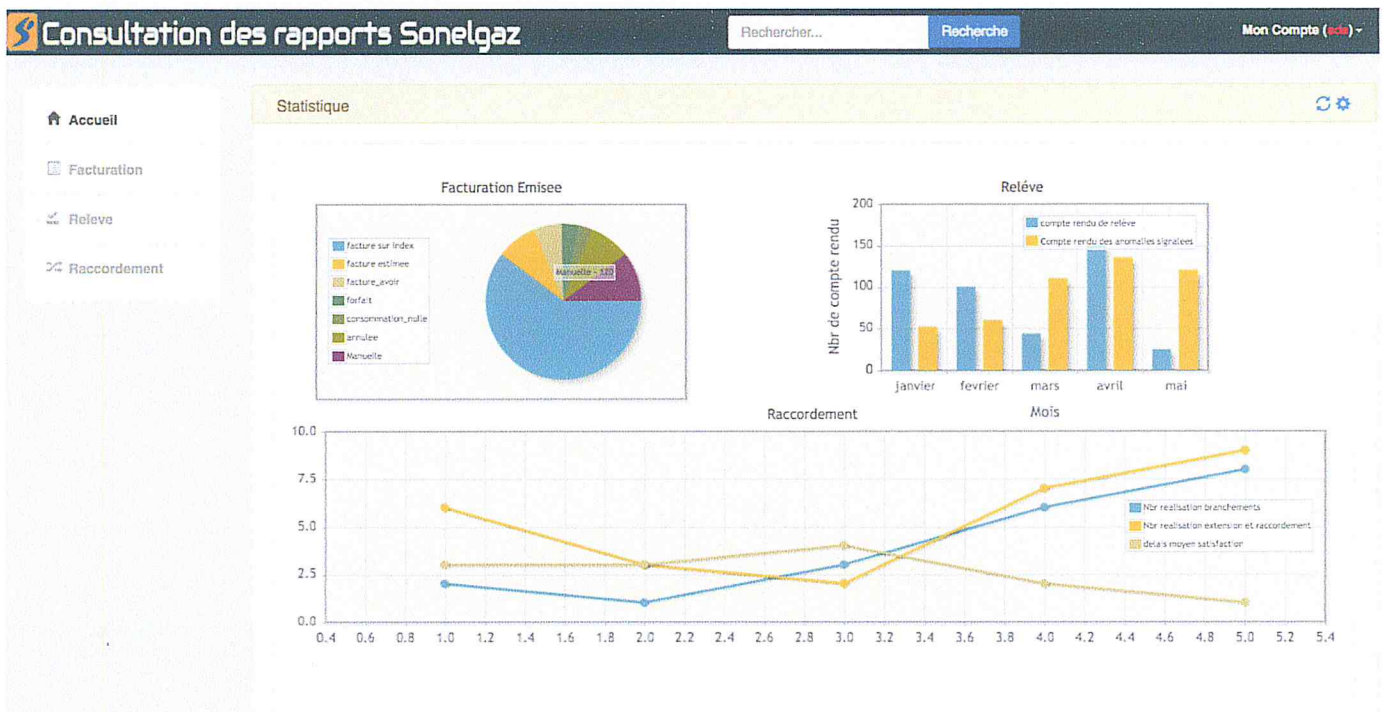


Figure 57 : l'interface des statistiques de facturation, relève et raccordement

En cliquant sur facturation, le décideur pourra consulter les différents rapports existants, pour cela il doit : sélectionner le typer d'énergie, le type de client, l'axe temps et cliquer sur « afficher les rapports »

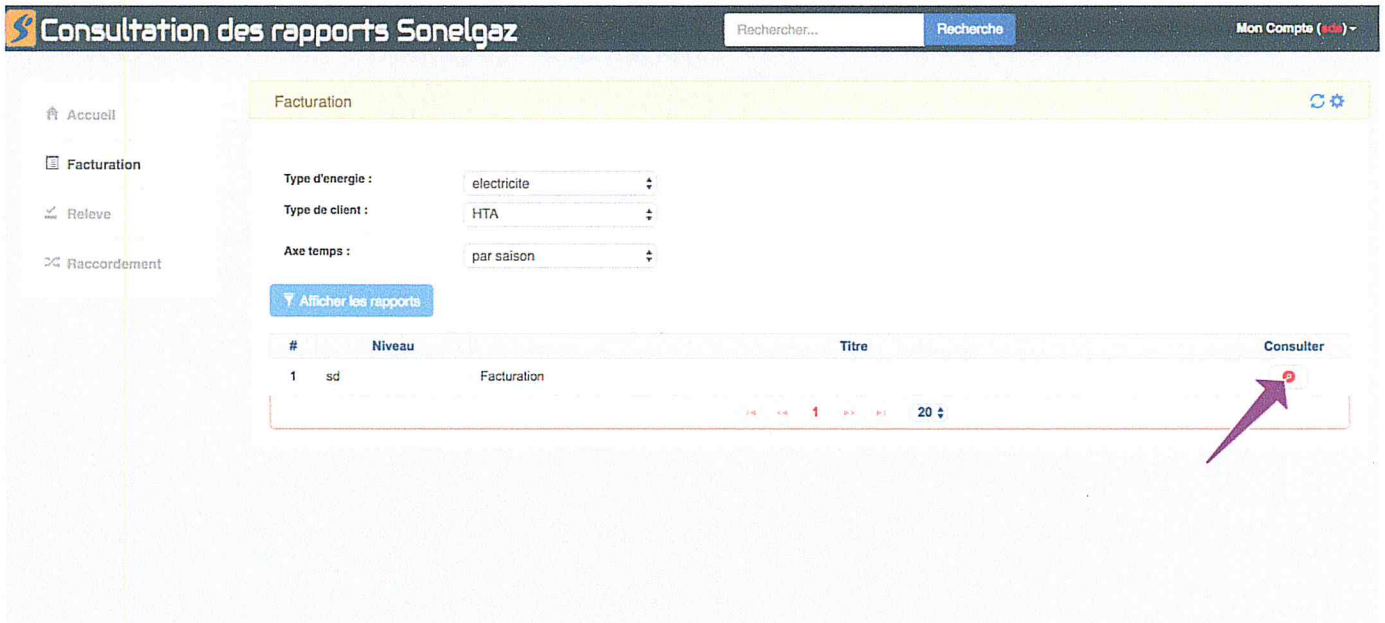


Figure 58 : l'interface permettant la consultation de rapport

Dès que le rapport apparaît dans le tableau, l'utilisateur clique sur « consulter » pour accéder à l'interface suivante :

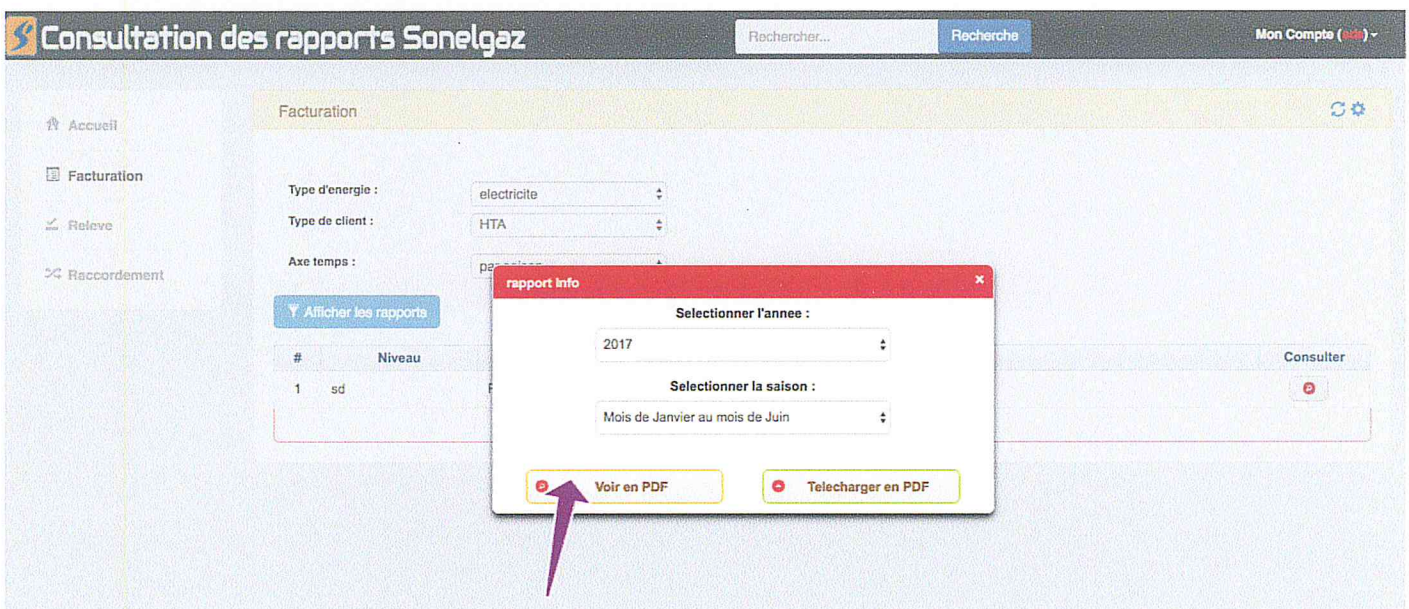


Figure 59 : l'interface de décideur sélectionne l'année et la saison désirées

Le décideur sélectionne l'année et la saison désirées, il aura le choix entre voir ou télécharger le résultat sous format PDF, dans ce qui suit nous montrons le cas où l'utilisateur clique sur « voir en PDF ».

DD	Facture emise				DCA		
	Facture sur index	Facture avoir	forfait	Facture ayant au moins 01 consommation	Facture estimée	Manuelle	annulée
EL MAROUJ	36	0	16	36	0	0	0
EL BIR	66	0	11	66	0	0	0
EL BOUJADJ	138	0	23	138	0	0	0
EL BOUJADJ	120	0	23	120	0	0	0
Total (nombres)	420	0	73	420	0	0	0

Edition: dimanche 04 juin 2017 Page 1/1

Figure 60 : l'interface de rapport en format PDF

6. Conclusion

Dans ce chapitre, nous nous sommes intéressés à l'implémentation de notre application. Nous avons d'abord cité divers outils et environnement utilisés pour la réalisation de notre système. Ensuite, nous avons présenté l'architecture technique de la solution proposée, nous avons décrit par la suite les différentes étapes de réalisation du projet, y compris la construction de l'entrepôt de données et l'ETL, et en fin nous avons présenté différentes interfaces et scénarios d'utilisation.

Conclusion Générale

Au terme de ce travail, élaboré dans le cadre de notre projet de fin d'étude, nous nous sommes amenés à développer un système efficace de consolidation et de présentation des informations sous forme de rapports, en se basant sur la réalisation d'un Data-Warehouse, pour les données du système de gestion de la clientèle. Ce système est destiné aux sociétés de distribution du groupe Sonelgaz, afin d'améliorer leur processus d'analyse et de prise de décision.

En effet, la dispersion des données entre les différents niveaux des sociétés de distribution du groupe Sonelgaz ainsi que l'hétérogénéité des sources de données (bases de données et fichiers Excel) rendent l'analyse une tâche ardue et son résultat désormais incertain ; ce qui influe, évidemment, sur l'efficacité des décisions à prendre. Pour cela, nous avons opté pour l'élaboration d'un système d'information décisionnel basé sur la technologie d'entrepôt de données.

Tout au long de notre travail de conception et de réalisation, nous avons suivi une démarche mixte, alliant de ce fait entre deux approches connues dans le domaine de l'entreposage de données, à savoir les démarches « Besoins d'analyse » et « Sources de données ». Cette démarche a permis de répondre aux attentes et besoins des utilisateurs tout en exploitant au mieux les données générées par les systèmes opérationnels de manière à anticiper sur des besoins non exprimés.

Dans un premier temps, nous avons fait une étude des besoins en se basant sur le cahier des charges « Reporting Commercial » et sur les rapports d'activité actuels du SGC.

Dans un deuxième temps, la modélisation de l'entrepôt DW_CLIENT s'est faite grâce aux principes de la modélisation dimensionnelle. Cette modélisation offre une vision claire et une compréhension intuitive des modèles proposés. Nous avons, de ce fait, proposé des modèles en étoile des trois processus recensés (facturation, relève et raccordement...)

Concernant l'entreposage de données, DW_CLIENT a été implémenté sur un SGBD relationnel à savoir Postgresql. Une fois l'entrepôt implémenté, celui-ci doit être alimenté et rafraîchi périodiquement grâce à un processus d'intégration ETL. Nous avons mis en œuvre la phase d'intégration des données sources avec la plateforme TALEND Open Studio. Dans cette partie, nous avons traité les données sources, considérées brutes et diffuses, afin de les

adapter aux objectifs d'analyse. Pour ce faire, plusieurs fonctions de nettoyage et d'agrégation ont été intégrées dans le processus telles que : la conversion, filtrage, formatage des dates, calcul des mesures, etc.

En fin, l'application a été achevée avec le générateur d'état automatique qui permet la conception des rapports selon les besoins des décideurs. Cet outil offre, à la fois, l'aspect instrument de contrôle et permet d'améliorer la qualité, l'intégrité et la transparence des informations utilisées. Les rapports conçus avec notre générateur permettent d'éclairer d'une manière efficace et sans délai le processus décisionnel du système SGC par catégorie organisationnelle et à différents niveaux de privilèges.

Afin que les utilisateurs finaux puissent exploiter l'ensemble des données de DW_CLIENT, nous avons mis en place une application web permettant aux utilisateurs (décideurs) au niveau des Agences commerciales, Directions régionales, Sociétés de Distribution et de la Maison mère, d'accéder à un ensemble de fonctionnalités dédiées à l'analyse et l'aide à la décision : rapports, exportations et publications vers les outils bureautiques, etc. Avec la mise en place de cette solution, la finalité de notre travail a été atteinte.

Perspectives

Comme tout travail, le système décisionnel que nous avons mis en œuvre mérite d'être complété et amélioré afin d'assurer sa pérennité et une utilisation professionnelle au sein d'une organisation de la taille de l'entreprise Sonelgaz. Après étude des aspects de qualité de notre système, nous résumons les perspectives de notre système de la manière suivante :

- ❖ Suivre le déploiement actuel et recueillir les correctifs et remarques des utilisateurs.
- ❖ Etendre le déploiement de manière à couvrir, à terme la, totalité du territoire national.
- ❖ Utilisation des méthodes et algorithmes de Data Mining pour une meilleure exploitation des données.
- ❖ Enrichir le générateur d'états par des fonctionnalités graphiques et intelligentes permettant plus de souplesse et une réelle aide à la décision.

BIBLIOGRAPHIE

- [1] : Chaudhuri, S., Dayal, U., Narasayya, V. R., « An overview of business intelligence technology ». Communications of the ACM, Vol.54, n°8, p.88-98, 2011.
- [2] : W. H. Inmon; « Building the Data Warehouse Third Edition »; Wiley Computer Publishing 2002.
- [3] : Emmanuel Ferragu, Modélisation des Systèmes d'Information Décisionnels : Techniques de modélisation conceptuelle et relationnelle des entrepôts de données, (2013).
- [4] : R. Kimball and K . Strehlo . Why decision support fails and how to fix it. SIGMOD Record, 24 (3) :92 – 97 , 1995.
- [5] : E. F. Codd ; « Providing OLAP (On-Line Analytical Processing) to User- Analysts : an IT mandate. » ; Technical report ; E.F. Codd & Associates; 1993.
- [6] : A. Agrawal, A. Gupta, and S. Sarawagi. « Modeling multidimensional data bases. Technical report research », I B M, 1997
- [7] : Dr. Kamel Boukhalfa , La conception physique aux outils d'administration et de tuning des entrepôts de données (2009).
- [8] : KHOURI S. Modélisation conceptuelle à base ontologique d'un entrepôt de données. Thèse de magistère en informatique. Alger : Institut National de Formation en Informatique (I.N.I), 2008
- [9] : Bouquin Henry ; « Le contrôle de gestion » ; P.U.F ; 2003.
- [10] : C. Convey, O. Karpenko, N. Tatbul: “Data Integration Services”, Technical Report, Brown University Computer Science Department, (2001).
- [11] : M. Lenzerini : “Data Integration: A Theoretical Perspective”, (2002).
- [12] : Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals p. 137-138, Paulraj Ponniah, 2001.
- [13] : Georges El Helou and Charbel Abou khalil, «Data Mining Techniques d'extraction des connaissances», Management and NTIC project 16 February 2004, PANTHEONASSAS PARIS II University.
- [14] : Didier Nakache ;« *Data Warehouse et Data Mining* »; ConservatoireNational des Arts et Métiers de Lille; Version 1.1; 15 juin 1998.
- [15] : GARDEUX Vincent. Jsf : un frameworkmvc, 2015.

WEB

[16] : CANTALOUBE. J Manipuler les données : XML et CSV, Open data : séquence n°2, 2013.

URL:<http://eduscol.education.fr/sti/sites/eduscol.education.fr/sti/files/ressources/pedagogiques/693/693-isn-opendata-seq2-eleve.pdf> .(Consulté le 29 avril 2017).

[17] : S. Chafki, C. Desrosiers, Entrepôts de données et intelligence d'affaires, Intégration des données et ETL, Département de génie logiciel et des TI, 2011.

URL: https://cours.etsmtl.ca/mti820/public_docs/acetates/MTI820-Acetates-ETL_1pp.pdf (Consulté le 28 avril 2017).

[18] : The Data Warehousing Institute, Phillip Russom, 2012.

URL: <https://tdwi.org/articles/2012/07/10/big-data-staging-area.aspx> (Consulté le 29 avril 2017).

[19] : GRIM Y. Définition de ETL du dictionnaire Business Intelligence 2013
[En ligne]. Disponible sur : <www.developpez.com > (Consulté le 28 avril 2017).

[20] : GRIM Y. Passez en mode BI [en ligne]. Disponible sur : <www.developpez.com > (Consulté le 25 avril 2017).

[21] : Anne Doucet. Intégration de données hétérogènes et réparties, 2007. URL: <http://www-poleia.lip6.fr/~doucet/CoursBDIA/Cours4.pdf> (Consulté le 26 avril 2017).

[22] : J.-F. Desnos. Entrepôt de données [en ligne] Disponible sur : <<http://imss-www.upmf-grenoble.fr>> (Consulté le 26 avril 2017).

[23] : URL : <https://www.digicap.fr/> (Consulté le 26 avril 2017).

[24] : URL : <http://perso.univ-lyon1.fr/haytham.elghazel/BI/presentation.html> . (Consulté le 26 avril 2017).

[25] : URL : <https://www.renaud-dhoker.fr/pages/les-etl-pour-les-entreprises> (Consulté le 28 avril 2017).

[26] : URL : <http://decisionnel-open-source.smile.fr/Les-principes/Les-principes-de-l-etl> (Consulté le 28 avril 2017).

[27] : URL : https://docs.oracle.com/cd/B10501_01/server.920/a96520/concept.htm (Consulté le 28 avril 2017).

[28] : URL : https://fr.wikipedia.org/wiki/Traitement_analytique_en_ligne (Consulté le 03

mai 2017).

[29] : URL : http://blerubrus.free.fr/cnam/ueeng111/solap_html/sectOlap.html (Consulté le 03 mai 2017).

[30] : URL : <http://www.expert-only.com/business-intelligence/concepts/molap-rolap-et-holap-bases-de-donnees-multidimensionnelles> (Consulté le 04 mai 2017).

[31] : URL : <http://www.supinfo.com/articles/single/3522-cube-olap> (Consulté le 04 mai 2017).

[32] : URL : <http://igm.univ-mlv.fr/~dr/XPOSE2005/entrepot/sghd.html> (Consulté le 04 mai 2017).

[33] : URL : <http://www.next-decision.fr/les-editeurs/etl/talend-open-studio> (Consulté le 27 mai 2017).

[34] : URL : <https://fr.wikipedia.org/wiki/NetBeans> (Consulté le 27 mai 2017).

[35] : URL : <https://fr.wikipedia.org/wiki/PostgreSQL> (Consulté le 27 mai 2017).

[36] : URL : <https://business-intelligence.developpez.com/faq/jasper-ireport/?page=Generalite> (Consulté le 28 mai 2017).

