

Université SAAD DAHLEB – Blida 1
Faculté des Sciences
Département d'Informatique



PROJET DE FIN D'ETUDES

Pour obtenir le titre de Master en Informatique
Option : Ingénierie des Logiciels

Par : SADOK Zahra et MEKINEZINaila

Sujet

**Mise en œuvre d'une solution *Business Intelligence*
basée sur un entrepôt de données de la SONELGAZ**

Soutenu le 24/09/2022 devant le jury composé de :

Mme L. OUKID	Maître de conférences (B)	Département d'Informatique, Blida 1	Présidente
Mr A. KHEDIRI	Maître de conférences (B)	Département d'Informatique, Blida 1	Examineur
Mr M. BALA	Maître de conférences (B)	Département d'Informatique, Blida 1	Promoteur
Mr B. BENBARA	Ingénieur conception et développement (niveau 2)	ELIT, SONELGAZ	Encadreur

Remerciements

Nous remercions, tout d'abord, Dieu le tout puissant et le miséricordieux de nous avoir donné la santé et la volonté d'entamer ce projet et de terminer ce mémoire dans les meilleures conditions.

Nous tenons à remercier et exprimer nos profondes gratitude à notre promoteur MR. MAHFOUD BALA pour avoir accepté de mener notre travail et de nous orienter dans le bon chemin. Pour ces précieux conseils qui nous ont aidé à améliorer la réalisation de ce mémoire, surtout pour sa responsabilité et sa disponibilité durant toute la période de notre projet et pour le temps qu'il nous a consacré.

Notre reconnaissance et nos remerciements vont à notre encadreur MR. BENBARA BEDREDDINE pour nous avoir encadré et encouragé, merci pour sa patience et ses conseils ainsi que pour l'expérience enrichissante et l'intérêt qu'il a manifesté pour ce projet.

Nos remerciements s'adressent, particulièrement, aux membres du jury, qui nous ont honorés par leur présence et pour avoir accepté d'évaluer notre travail.

Nous ne saurons terminer cette page sans présenter nos remerciements à toute l'équipe de développement de « SONELGAZ - ELIT- département 09- » pour leur aide précieuse ainsi qu'à tous ceux qui, de près ou de loin, ont contribué à la réalisation de ce travail.

Merci encore une fois.

Dédicaces



*J'ai le grand plaisir de dédier ce modeste travail, comme
preuve de respect, de gratitude, et de reconnaissance :*

*À celui qui s'est toujours sacrifié et a attendu patiemment pour me voir
réussir à ma chère mère. Puisse Dieu, tout puissant vous combler de
santé, de bonheur et vous procurer une longue vie.*

*À celui qui s'est changé la nuit en jour pour m'assurer les bonnes
conditions, mon cher père vous avez toujours été une école de patience,
de confiance et surtout d'espoir d'amour. Vous êtes et vous resterez pour
moi ma référence, la lumière qui illumine mon chemin. Ce travail est le
résultat de l'esprit de sacrifice dont vous avez fait preuve. J'implore
Dieu, tout puissant de vous accorder une bonne santé, une longue vie et
beaucoup de bonheur.*

*À mes frères OUSSAMA et MOHAMED et ma chère petite sœur
SALSABIL.*

*À Mon cher oncle REDOUAN qui m'a toujours soutenue et
encouragée mes efforts.*

*À ma très chère cousine Ikram qui a été toujours à mes côtés dans les
moments difficiles.*

À mes chères cousines Amina, Khadija, Hanna, Meriem.

*À mes grands-mères FATIHA et ZOBIDA ma tante FOUZIA et
À toute ma famille.*

*À mon fiancé Omar qui m'as toujours encouragé, incité à faire de mon
mieux, ton soutien m'a permis de réaliser le rêve tant attendu. Aucun
mot ne pourrait exprimer ma gratitude, mon amour et mon respect.*

*À mon binôme ZAHRA qui a toujours été présente dans les moments
difficiles.*

*Une spéciale dédicace à mon promoteur MR. MAHFOUD BALA qui
nous a très bien orientés. Il nous a appris beaucoup de choses, sans lui
le travail n'aurait jamais pris fin.*

MEKINEZI NAILA

Dédicaces

Je dédie ce modeste travail

accompagné d'un profond amour :

*À ma famille, elle qui m'a doté d'une éducation digne et qui m'a porté
aussi haut que je suis.*

*Particulièrement à mon tendre père, pour le gout à l'effort qu'il a suscité
en moi.*

*À ma mère, celle qui m'as tant encouragé. Quel que soit les termes
embrassés je n'arriverai jamais à lui exprimer mon amour sincère pour
son soutien.*

*À mes adorables sœurs qui savent toujours me procurer sourire
d'espoirs et motivation.*

*À tous ceux qui me sens proche, mes amis d'enfance et du long parcours
scolaire et universitaire.*

Enfin, je remercie mon binôme qui a contribué à la réussite de ce travail.

Zahra SADOK

ملخص

سونلغاز في إطار مشروعها، الذي جعلتنا جزءا منه، لتطوير نظام ذكاء الأعمال لتلبية احتياجات التحليل لأنشطتها المختلفة . حددنا جميع احتياجات اتخاذ القرار وكذلك وظائف التحويل للعمل على البيانات، من أجل تصميم عملية دمج البيانات المسؤولة عن وضعها في مخطط مناسب للتحليل . سمحت لنا مرحلة التنفيذ بتطوير وتهيئة عملية التكامل على منصة *Talend Open Studio (TOS)* . كما تم تطوير جزء *OLAP* وتكوينه على منصة *KNOWAGE* .

الكلمات المفتاحية: نظام صنع القرار، مستودع البيانات، اتصالات البيانات، ذكاء الأعمال.

Résumé

SONELGAZ nous a confié, dans le cadre de son projet, la mise en œuvre d'un système de *Business Intelligence* pour les besoins d'analyse de ses différentes activités. Après avoir recensé tous les besoins décisionnels, ils ont été formalisés sous forme de modèles dimensionnels exprimant les sujets d'analyse (faits) ainsi que les axes selon lesquels se feront les analyses (dimensions). La fusion des différents modèles dans une constellation, nous a servi à identifier, les fonctions de transformation à opérer sur les données, afin de concevoir le processus d'intégration des données responsable de les mettre dans un schéma adapté aux analyses. La phase d'implémentation, quant à elle, nous a permis de développer et paramétrer le processus d'intégration sur une plateforme *ETL* de référence à savoir *Talend Open Studio (TOS)*. La partie *OLAP*, consistant en la création des cubes de données et la mise en œuvre de l'application d'analyse, a été développée et paramétrée sur la plateforme *KNOWAGE*.

Mots Clés : système décisionnel, entrepôt de données, communication de donnée, *BI*.

Abstract

SONELGAZ within the framework of its project, of which it has made us a part, to develop a business intelligence system to meet the analysis needs of its various activities. We identified all the decision-making needs as well as the transformation functions to operate on the data, in order to design the data integration process responsible for putting them into an appropriate schema for analysis. The implementation phase allowed us to develop and configure the integration process on the *Talend Open Studio (TOS)* platform, While The *OLAP* part is developed and configured on the *KNOWAGE* platform.

Keywords : decision-making system, datawarehouse, reporting, *BI*.

Table des matières

Chapitre 1 : Introduction Générale	12
1. Contexte du projet.....	13
2. Problématique.....	13
3. Présentation de l'organisme d'accueil.....	14
3.1. Groupe SONEGAS.....	14
3.2. Présentation d' <i>ELIT</i>	15
3.3. Structure et organisation d'accueil de la « DSID ».....	16
4. Objectifs.....	16
5. Organisation du mémoire.....	17
Chapitre 2 : Généralités sur les systèmes décisionnels	18
1. Introduction.....	19
2. Définition du <i>reporting</i>	19
3. <i>Business Intelligence (BI)</i>	19
3.1. Définition de la <i>BI</i>	19
3.2. Opérationnel vs décisionnel.....	20
4. Entrepôt de données (<i>DW : Data Warehouse</i>).....	21
4.1. Définition d'un <i>DW</i>	21
4.2. Modélisation d'un <i>DW</i>	22
4.3. Structure des données d'un <i>DW</i>	24
4.4. Architecture décisionnelle.....	25
4.5. Alimentation du <i>Data Warehouse</i>	27
5. Concept d' <i>OLAP (OnLine Analytical Processing)</i>	31
4.1. Définition d'un cube <i>OLAP</i>	31
4.2. Architectures des serveurs.....	32
4.3. Algèbre <i>OLAP</i>	33
6. Conclusion.....	35
Chapitre 3 : Analyse des besoins et Conception	36
1. Introduction.....	37
2. Les entretiens.....	37
3. Les documents de référence.....	37
4. Modélisation dimensionnelle.....	38
4.1 Les dimensions partagées.....	38
4.2 Les modèles dimensionnels.....	49
5. Transformation et Intégration des données.....	59

5.1	Processus de l'entreposage de données	59
5.2	Planification de chargement des données	60
6.	Conception des cubes multidimensionnels.....	61
7.	Diagramme de cas d'utilisation	62
8.	Conclusion	63
Chapitre 4 : Mise en œuvre du système de reporting commercial de SONELGAZ		64
1.	Introduction.....	65
2.	Environnement et outils de développement.....	65
3.	Transformation et Intégration des données.....	67
3.1	Développement sous la plateforme <i>Talend</i>	67
3.2	Aperçu sur nos <i>Job</i>	69
3.3	Planification des <i>Jobs ETL</i>	76
4.	Architecture du système de <i>reporting</i>	76
5.	Description des différents modules du système	78
5.1	Les cubes multidimensionnels.....	78
5.2	Tableau de bord stratégique (<i>Cockpit</i>).....	80
6.	Tests et Résultats	81
6.1	Les cubes multidimensionnels.....	81
6.2	Les <i>cockpits</i>	83
7.	Conclusion	84
Conclusion Générale		86

Liste des abréviations

BD	Base de Données
BI	<i>Business Intelligence</i>
BP	Basse Pression
BT	Basse Tension
CDC	<i>Change Data Capture</i>
DD	Direction Distribution
DSID	Direction Système d'Informations activités Distribution/gestion de réseaux
DW	<i>Data Warehouse</i>
ETL	<i>Extract-Transform-Load</i>
FSM	Facture Sur Mémoire
HP	Haute Pression
HTA	Haute Tension Type A
HTB	Haute Tension Type B
KPI	<i>Key Performance Indicator</i>
MDX	<i>MultiDimensional eXpressions</i>
MP	Moyenne Pression
OLAP	<i>OnLine Analytical Processing</i>
RD	Région de Distribution
SGBD	Système de Gestion de Base de Données
SGBDR	Système de Gestion de Base de Données Relationnelles
SGC	Système de Gestion de la Clientèle
SONELGAZ	SOciété Nationale de l'Electricité et du GAZ
SQL	<i>Structured Query Language</i>
TOS	<i>Talend Open Studio</i>
TTC	Toutes Taxes Comprises
TVA	Taxes sur la Valeur Ajouté
XML	<i>eXtensible Markup Langages</i>

Liste des tableaux

Tableau 1. Système opérationnel vs Système décisionnel.....	21
Tableau 2. Description détaillée des dimensions.	39
Tableau 3. Description détaillée des dimensions (suite).....	40
Tableau 4. Description détaillée des dimensions (suite).....	41
Tableau 5. Description détaillée des dimensions (suite).....	42
Tableau 6. Description détaillée des dimensions (suite).....	43
Tableau 7. Description détaillée des mesures.	44
Tableau 8. Description détaillée des mesures (suite).....	45
Tableau 9. Description détaillée des attributs de chaque dimension.	46
Tableau 10. Description détaillée des attributs de chaque dimension (suite).	47
Tableau 11. Description détaillée des attributs de chaque dimension (suite).	48
Tableau 12. Périodicité de chargement des modèles dimensionnels.	61
Tableau 13. Détails sur les cas d'utilisation.....	62
Tableau 14. Détails sur les cas d'utilisation (suite).....	63
Tableau 15. Liste des composants TOS.	68

Liste des figures

Figure 1. Structure de la filiale ELIT.	15
Figure 2. Structure de la DSID.	16
Figure 3. Exemple table de fait « Ventes ».....	23
Figure 4. Exemple de table de dimension « Temps ».....	23
Figure 5. Structure des données d'un Data Warehouse [10].	24
Figure 6. Architecture décisionnelle [29].	25
Figure 7. Étapes d'alimentation d'un DW.	27
Figure 8. Méthode <i>push</i>	30
Figure 9. Méthode pull.....	30
Figure 10. Méthode Push-Pull.....	31
Figure 11. Exemple d'un cube OLAP « SONELGAZ ».....	32
Figure 12. Résumé des opérations OLAP.....	35
Figure 13. Modèle en étoile « ventes ».....	50
Figure 14. Modèle en étoile « Achats ».....	51
Figure 15. Modèle en étoile « Objectifs ».....	52
Figure 16. Modèle en étoile « chiffres d'affaires ».....	53
Figure 17. Modèle en flocon de neige « Achats ».....	54
Figure 18. Modèle en flocons de neige « ventes ».....	55
Figure 19. Modèle en flocon de neige « Chiffres d'affaires ».....	56
Figure 20. Modèle en flocon de neige « Objectifs ».....	57
Figure 21. Modèle en constellation.....	58
Figure 22. Processus d'extraction et transformation de données « achats ».....	59
Figure 23. Processus de chargement de données « achats ».....	60
Figure 24. Cube multidimensionnel « achats ».....	61
Figure 25. Diagramme de cas d'utilisation.....	62
Figure 26. Liste des Job.	69
Figure 27. Le Job « dim_type_facture ».....	70
Figure 28. tMap_1 du Job « dim_type_facture ».....	71
Figure 29. Résultat de l'exécution avec un CDC.....	72
Figure 30. Job « dim_objctf ».....	72

Figure 31. Job « tf_vente ».....	73
Figure 32. tMap_1.....	73
Figure 33. Job « tf_objectif ».....	74
Figure 34. « tAggregateRow_1 ».....	74
Figure 35. Lors de la construction/vérification du Job « tf_objectif ».....	75
Figure 36. La console après exécution avec tLogRow.	75
Figure 37. Job globale.....	76
Figure 38. Architecture globale du système reporting décisionnel.....	77
Figure 39. Cube « chiffres d'affaires » en format XML.....	79
Figure 40. Visualisation du cube « chiffres d'affaires ».....	80
Figure 41. Tableau de bord « chiffres d'affaires ».....	81
Figure 42. Exemple de création d'un cube multidimensionnels « achats ».....	82
Figure 43. Visualisation d'un sous-cube « achats ».....	83
Figure 44. Différentes opérations sur les cockpits.....	84

Chapitre 1 : Introduction Générale

1. Contexte du projet

La mondialisation, la concurrence et l'évolution technologique rendent les marchés très instables et poussent les entreprises à repenser sans cesse la stratégie de leur business. En effet, beaucoup de grandes entreprises ont eu recours aux techniques du *Business Intelligence (BI)* pour comprendre et contrôler efficacement leurs affaires afin de gagner une meilleure part du marché.

La *BI* est utilisée, principalement, pour améliorer l'actualité et la qualité de l'information, et permettre aux gestionnaires de mieux comprendre la position de leurs *business* par rapport à la concurrence. Les applications et les technologies du *BI* aident les entreprises à analyser les tendances en matière de part du marché, des changements dans le comportement, en outre à étudier les préférences et les dépenses des clients ainsi que les capacités de l'entreprise et les conditions du marché.

En raison de sa dimension économique et de sa position sur le marché énergétique algérien, les activités quotidiennes de SONELGAZ génèrent des données complexes et volumineuses. Malgré brutes et complexes, ces données constituent une source précieuse capable de produire des informations pertinentes ayant des valeurs statistiques et analytiques permettant aux gestionnaires de transformer et d'améliorer les processus métier du groupe. Cependant, SONELGAZ n'exploite pas pleinement ses sources de données, ce qui compromet complètement ses processus d'analyse et d'aide à la décision. C'est dans ce contexte d'informatique décisionnelle que se situe notre projet de fin d'études qui ambitionne de trouver une solution pour le déficit que connaissent les processus d'entreposage et d'aide à la décision de la SONELGAZ.

2. Problématique

Les systèmes *BI* génèrent une image synthétique sur les différentes activités de la SONELGAZ, ceci donne une vision aux décideurs et à leurs collaborateurs sur l'état et les performances des processus métier. Dans un schéma plus avancé, les systèmes *BI* sont capables de donner une vision sur la réussite ou l'échec d'une stratégie qui a été adoptée. Les systèmes *BI* permettent de produire des rapports d'activités afin de rendre compte à une hiérarchie des performances de ces activités.

SONELGAZ a vécu une période très difficile caractérisée par l'absence d'outils décisionnels. Durant cette période, les responsables de la SONELGAZ ont opté pour des solutions traditionnelles pour élaborer leurs rapports et leurs indicateurs de performances sous forme de documents Word, classeurs Excel, etc. C'est pour cette raison que la plus simple des opérations d'analyse devient une tâche ardue. En effet, la société de distribution « DD » se trouve dans l'incapacité de faire des analyses fiables, efficaces et dans des délais raisonnables. Ainsi, les principales difficultés rencontrées peuvent être résumées de la manière suivante :

- La plupart des rapports et des tableaux récapitulatifs sont, très souvent, élaborés de façon manuelle.
- La difficulté de gestion est liée à une surcharge informationnelle intense au niveau des directions métiers.
- La lenteur de la procédure de *reporting* provoque l'indisponibilité des rapports en temps opportun.
- La procédure de *reporting* est très coûteuse pour l'entreprise, dont la production des résultats est trop tardive pour un pilotage efficient.
- Le niveau élevé de détail des données qui rend l'analyse et la prise de décisions plus difficile.

3. Présentation de l'organisme d'accueil

Nous présentons, dans cette section, l'organisme ayant accueilli notre projet de fin d'études à savoir la filiale « *EL Djazair Information Technology* » (*ELIT*), filiale informatique du groupe SONELGAZ.

3.1. Groupe SONELGAZ

Créé en 1947 sous le nom Electricité et Gaz d'Algérie (EGA) puis rebaptisée SONELGAZ, SOciété Nationale de l'Electricité et du GAZ, via l'ordonnance N°69-59 du 28 juillet 1969 [1], SONELGAZ a détenu 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 qu'acteur industriel responsable de l'approvisionnement en énergie avec un taux de couverture en électricité à 98%

pour 10 494 000 clients et un taux de pénétration du gaz à 65% pour 6 451 000 clients [2].

3.2. Présentation d'ELIT

Spécialisée dans les technologies de l'information et de la communication, attachée au début à la Direction Générale des Systèmes d'Information au niveau de la SONELGAZ, *ELIT* a été érigée en société par actions le 1er janvier 2009 [3], dénommée « *EL Djazair Information Technology* », par abréviation "*ELIT SPA*".

Et comme toute autre filiale du groupe SONELGAZ, *ELIT* a sa propre mono-structure relativement indépendante des autres filiales, illustrée par l'organigramme dans la figure 1.

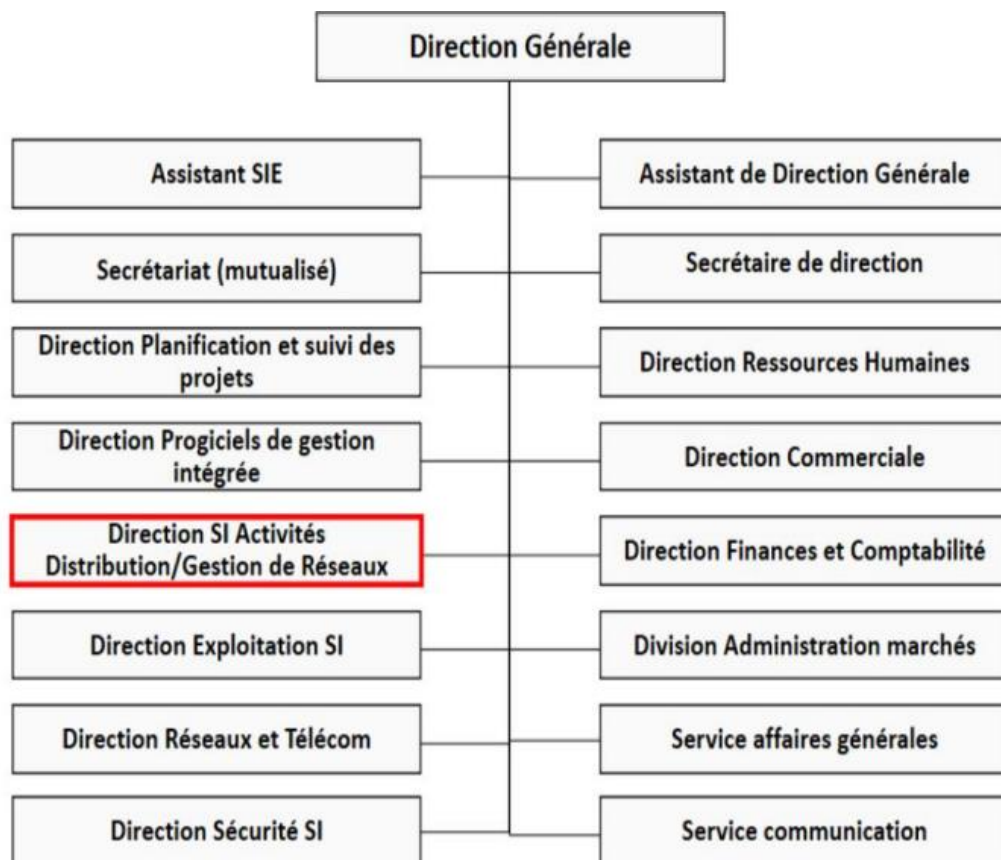


Figure 1. Structure de la filiale *ELIT*.

3.3. Structure et organisation d'accueil de la « DSID »

Comme nous réalisons ce projet au sein de « la Direction Système d'Informations activités Distribution/gestion de réseaux » abrégée en DSID. Plus précisément, au niveau de département « développement SI distribution ». Nous présentons, dans la figure 2, les trois départements de la DSID.

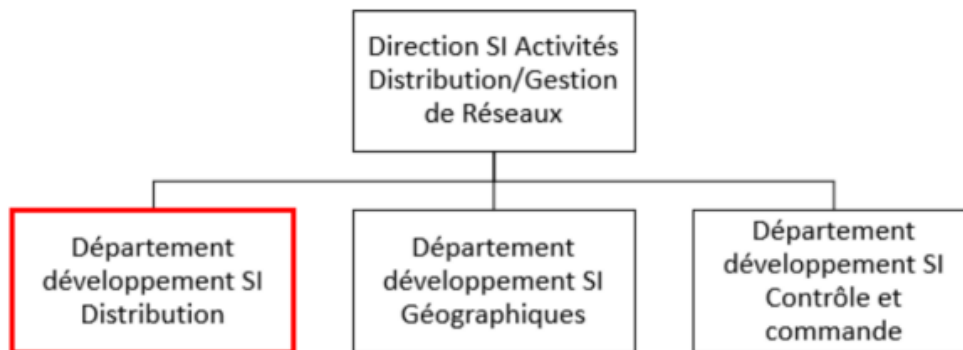


Figure 2. Structure de la DSID.

4. Objectifs

Afin d'évaluer l'imprécision des besoins décisionnels et pour améliorer la qualité de la conception de l'entrepôt de données et la qualité de l'analyse *OLAP* et suite aux problèmes mentionnés auparavant, ce mémoire vise à proposer des solutions (naïve et rigoureuse) pour évaluer l'imprécision des besoins.

Les objectifs atteindre de cette solution sont :

- Une Interface graphique intuitive et simple pour chaque utilisateur avec l'outil d'analyse et de business intelligence « *KNOWAGE* » pour l'analyse des données à partir de l'entrepôt de données.
- Construire les tableaux, les rapports, les graphes, etc. Suite aux résultats générés à partir d'une requête de l'analyse.
- Construire des cubes multidimensionnels à partir des fichiers *XML* pour manipuler les différents opérateurs algébrique *OLAP*.
- L'exportation des résultats d'analyse.
- Sauvegarde de l'analyse créée, pour pouvoir les réutiliser ultérieurement.
- La publication de l'analyse.

Ce projet consiste à élaborer un cahier de charge afin de mettre en place un système décisionnel pour la production des tableaux de bord, (*KPI : Key Performance Indicator*) et les états de sorties alimentés par les données commerciales de la distribution à travers le système de gestion clientèle (*SGC*) et mis à jour par un rafraîchissement automatique des données.

5. Organisation du mémoire

Le présent document constitue une synthèse des aspects les plus pertinents traités lors de notre projet de fin d'études. Pour ce faire, nous l'avons structuré autour des axes suivants.

Le premier chapitre intitulé « Introduction générale » est l'entrée de notre sujet où nous présentons l'organisation d'accueil, la problématique et les objectifs à atteindre.

Afin de permettre une lecture aisée du mémoire, le deuxième chapitre présente les concepts fondamentaux sur lesquels traite notre travail à savoir la *Business Intelligence*, les entrepôts de données, le processus d'intégration des données *ETL*, l'environnement d'analyse des données en ligne, etc.

Le chapitre 3 est organisé en deux grandes sections à savoir (1) le recueil des besoins de la SONELGAZ en termes d'analyse et d'indicateurs de performance et (2) la conception du système futur.

Le chapitre 4, quant à lui, détaille la mise en place du système décisionnel en décrivant son architecture, les sources de données, les différents *Jobs* du processus d'entreposage de données ainsi que l'environnement d'analyse et de *reporting*.

Nous clôturons ce document par une conclusion générale qui récapitule les éléments essentiels du travail effectué. En outre, des idées pour une éventuelle continuité du projet sont développées sous forme perspectives d'évolution.

Chapitre 2 : Généralités sur les systèmes décisionnels

1. Introduction

La *BI* est l'une des principales préoccupations des grandes entreprises, elle fournit une visibilité totale de l'activité de l'entreprise, ce qui favorise l'efficacité opérationnelle. Dans le contexte actuel de crise et de concurrence excessive, la *BI* constitue une opportunité pour les entreprises pour accélérer et améliorer la prise de décisions commerciales et d'augmenter les chiffres d'affaires de l'entreprise. Le gain de temps apporté entraîne un accroissement de la productivité. La *BI* permet notamment d'identifier des manières pour accroître les bénéfices, analyser le comportement des clients et pour anticiper sur les tendances du marché.

2. Définition du *reporting*

Le *reporting* ou rapport d'activité en français, est une famille d'outils permettre aux décideurs et aux dirigeants des entreprises de produire des rapports d'activité et de mettre en place des tableaux de bord. Afin de faciliter la communication des résultats chiffrés et d'acquérir une vue d'ensemble de l'entreprise.

3. *Business Intelligence (BI)*

3.1. Définition de la *BI*

La *Business Intelligence* (Intelligence économique en français) est le processus technique d'analyse des données et de fourniture d'informations dans le but d'aider les dirigeants d'entreprises à prendre des décisions judicieuses en préparant les données à analyser pour créer des rapports, des tableaux de bord et d'autres outils de visualisation de données [4].

Ces outils fournissent des informations telles que la veille concurrentielle et les tendances du marché, ainsi que des informations internes telles que la recherche des raisons des opportunités perdues [5].

Ces dernières années, la *Business Intelligence* a été largement vulgarisée et est passée d'un domaine appartenant principalement à l'informatique à une fonction clé dans toutes les entreprises, comme la recherche d'opportunités commerciales et l'optimisation de leurs activités marketing ou financières [6].

3.2. Opérationnel vs décisionnel

Les systèmes opérationnels dits transactionnels (*OLTP : On-Line Transaction Processing*) sont des environnements de traitement transactionnel [7] dédiés aux métiers de l'entreprise pour les assister dans leurs tâches de gestion quotidiennes, répétitives et atomiques (Insertion, modification, suppression) qui sont effectuées par les employés de l'entreprise [8].

Les systèmes décisionnels basés sur des processus d'analyse (*OLAP : On-Line Analytical Processing*), quant à eux, sont des environnements d'analyse multidimensionnelle dédiés au management de l'entreprise et au pilotage de l'activité permettant à un utilisateur de consulter et d'extraire facilement les données pour les analyser de différentes facettes [9].

Le tableau 1 présente un comparatif entre les systèmes opérationnels et les systèmes décisionnels par rapport à quelques critères [7] [9]:

Critères	Système opérationnel (<i>OLTP</i>)	VS	Système décisionnel (<i>OLAP</i>)
Utilisateurs	<ul style="list-style-type: none"> Données orientées application. Boîte à outils. Données atomiques. Requêtes simples. 		<ul style="list-style-type: none"> Données orientées utilisateur. Processus et analyses métiers déjà implémentés. Données atomique et agrégées. Requêtes complexes.
Données	<ul style="list-style-type: none"> Sur une période courte. Détaillées. Des données partagées. Mise à jour. Temps de validité de transaction. 		<ul style="list-style-type: none"> Historiées. Agrégées. Peuvent être anonymes. Recalculées. Temps de validité de transaction et d'extraction.
Source de données	<ul style="list-style-type: none"> Les données saisies en l'état proviennent d'une source unique. 		<ul style="list-style-type: none"> Les données sont collectées et normalisées à partir de nombreuses sources.

Type de données	<ul style="list-style-type: none"> • Données de gestion (courantes). 	<ul style="list-style-type: none"> • Données d'analyse (résumées, historiées).
Systèmes	<ul style="list-style-type: none"> • Modélisation entités / relations. • Mises à jour fréquentes. • Optimisation du stockage. 	<ul style="list-style-type: none"> • Modélisation multidimensionnelle. • Mises à jour contrôlées. • Optimisation des temps de réponse.
Objectifs	<ul style="list-style-type: none"> • Gestion et production. 	<ul style="list-style-type: none"> • Consultation et analyse.

4. Entrepôt de données (*DW : Data Warehouse*)

4.1. Définition d'un *DW*

Dans son livre « *Building the Data Warehouse* » [10], Bill (william H) Inmon définit un entrepôt de données de la manière suivante : « *Le Data Warehouse* est une collection de données orientées sujet, intégrées, non volatiles et évolutives dans le temps, organisées pour le support d'un processus d'aide à la décision ».

- Orientées sur le sujet : Organisées par thème, il est possible d'utiliser le *Data Warehouse* pour analyser n'importe quel secteur particulier de l'entreprise.

- Intégrées : Avant toute utilisation, les données récupérées de sources hétérogènes internes ou externes sont intégrées au *Data Warehouse*. Pour cela, il est nécessaire de les mettre en forme **Tableau 1.** Système opérationnel vs Système décisionnel. et de les unifier pour garantir une certaine cohérence. Les données proviennent principalement de traitement transactionnel en ligne (*OLTP*).

- Non volatiles : Les données une fois stockées dans l'entrepôt de données ne peuvent jamais être modifiées.

- Évolutives dans le temps : Les données passées sont également conservées dans le *Data Warehouse*, contrairement à certains systèmes transactionnels traditionnels où seules les données les

plus récentes sont stockées. Cela permet de visualiser l'évolution dans le temps des différentes données.

De l'autre côté Ralph Kimball décrit un *DW* comme suit :

« Un Entrepôt de Données (*DW : Data Warehouse*) est un système qui extrait, nettoie, conforme et fournit des données source dans un espace de stockage multidimensionnelles, puis prend en charge et implémente l'interrogation et l'analyse à des fins de prise de décision » [11].

4.2. Modélisation d'un *DW*

Selon Kimball [12] « Le modèle dimensionnel vise à présenter les données sous une forme intuitive dont l'objectif est de se rapprocher de la manière dont les décideurs perçoivent les données d'analyse ».

4.2.1. Modèle en étoile

Le modèle en étoile est un paradigme de modélisation simple et commun, dans lequel un entrepôt de données comprend des dimensions associées à une seule table de fait. La force de ce type de modélisation est sa lisibilité et sa performance [12].

- **Concept de fait**

La table de fait est la table principale d'un modèle dimensionnel qui contient des données observables (les mesures), que l'on possède selon divers axes d'analyse (les dimensions) [12] [13].

La figure 3 décrit un exemple de table de fait *ventes* dédiées à l'analyse des mesures *consommation_vendu*, *consommation_brute*, *nombre_clients*, *nombre_factures* et ce par rapport aux dimensions (*dim_temps*, *dim_type_facture*, *dim_categorie_client*, etc).

The diagram shows a table named 'fait_vente' with two main sections. The first section, labeled 'La clé de la table', contains 13 dimension columns, each with a green key icon and 'int8' data type. The second section, labeled 'Les mesures', contains 4 measure columns, each with a blue document icon and various data types (float8, int8).

fait_vente	
dim_tarifid	int8
dim_categorie_clientid	int8
dim_type_contratid	int8
dim_origine_releveid	int8
dim_type_clientid	int8
dim_energieid	int8
dim_region_metierid	int8
dim_tempsid	int8
dim_region_administrativeid	int8
dim_etat_factureid	int8
dim_type_factureid	int8
dim_activite_economiqueid	int8
dim_type_abonnementid	int8
consommation_vendu	float8
consommation_brute	float8
nombre_client	int8
nombre_facture	int8

Figure 3. Exemple table de fait « Ventes ».

- **Concept de dimension**

Dans un modèle dimensionnel la table de dimension est une table qui fait partie intégrante de la table de fait, qui contiennent des paramètres appelés aussi des colonnes ou attributs, décrivent les lignes de cette table de dimension [12] [13].

La figure 4 présente un exemple de table de dimension *dim_temps* caractérisés par ses attributs : *Id*, *mois*, *libellé_mois*, *année*, *code_trimestre* et *libellé_trimestre*.

The diagram shows a table named 'dim_temps' with one primary key attribute and five other attributes. The primary key 'id' is marked with a yellow key icon and has an 'int8' data type. The other attributes are marked with blue document icons and have various data types (int4, varchar(255)).

dim_temps	
id	int8
mois	int4
libelle_mois	varchar(255)
annee	varchar(255)
code_trimestre	int4
libelle_trimestre	varchar(255)

Figure 4. Exemple de table de dimension « Temps ».

- **Clé de substitution**

La clé de substitution (*SK : Surrogate Key*), aussi appelée la clé non intelligente ou clé technique est une colonne simple, générée par le système avec un identificateur unique pour chaque ligne et indépendante de l'entreprise [12].

4.3. Structure des données d'un *DW*

L'entrepôt de données est structuré en quatre classes de données spécifiques, organisées selon un axe historique est un axe synthétique [14].

La figure 5 illustre le positionnement de ces classes les unes par rapport aux autres.

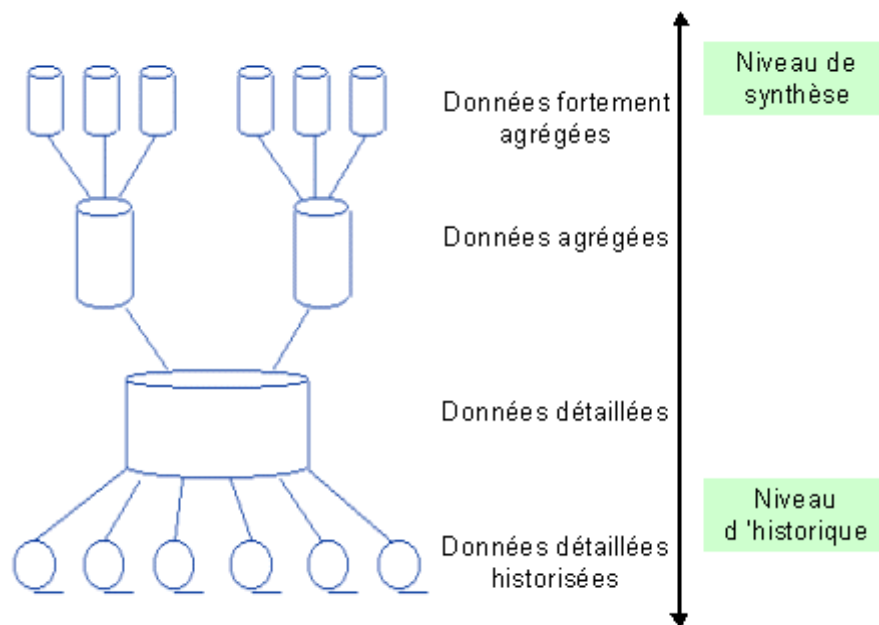


Figure 5. Structure des données d'un *Data Warehouse* [10].

- **Les données fortement agrégées**

Il s'agit des données agrégées traitées par le système à un niveau d'agrégation plus élevé que les données agrégées. Très souvent les données à fédérer dans le *Data Warehouse* proviennent de sources très hétérogènes. Cela nécessite un dictionnaire unique qui gère l'ensemble des fonctions,

Cette cohérence du dictionnaire est décrite au sein des métadonnées du dictionnaire du *DW*.

- **Les données agrégées**

Ce sont les données générées par le système à partir des données détaillées et représentent un premier résultat d'analyse et de synthèse des données contenues dans les systèmes de production.

- **Les données détaillées**

Ce sont les données qui reflètent les événements qui surviennent lors des activités d'une organisation. Elles sont généralement volumineuses puisque détaillées, les données provenant des systèmes de production sont intégrées à ce niveau.

- **Les données détaillées historisées**

Une mise à jour dans l'entrepôt de données (insertion et modification), n'écrase pas les anciennes valeurs mais crée une nouvelle insertion.

4.4. Architecture décisionnelle

Après avoir exposé et défini chacun des éléments constituant l'environnement d'un *Data Warehouse*, il serait intéressant de connaître le positionnement de ces éléments dans une architecture globale d'un *Data Warehouse* (voir figure 6).

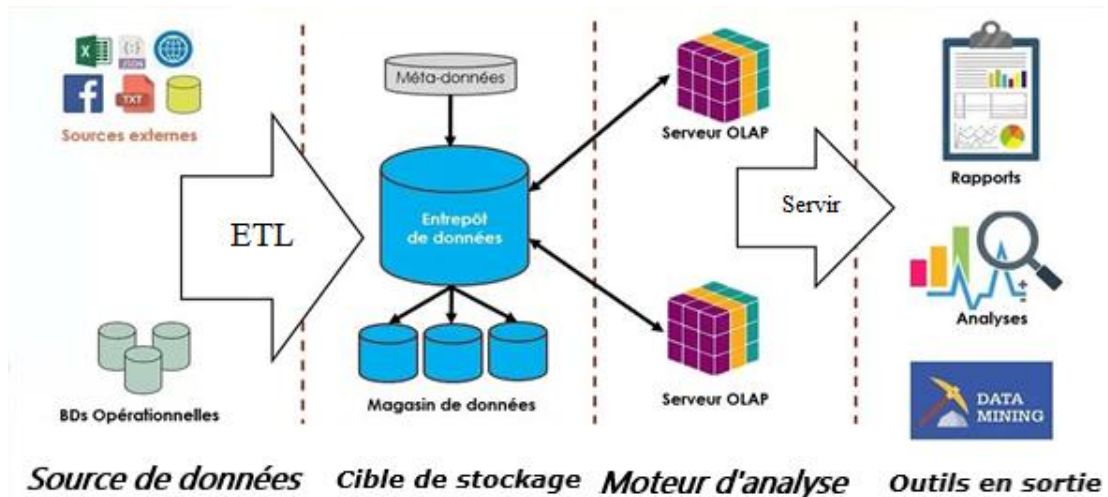


Figure 6. Architecture décisionnelle [29].

- **Sources de données**

Les sources de données sont constituées de *SGBD* relationnels et d'autres fournisseurs et formats de données (classeurs Excel, documents *XML*, fichiers textes, pages web, etc.).

- ***ETL (Extract-Transform-Load)***

ETL est un ensemble de composants (logiciels, scripts, bibliothèques, etc.) permettant d'extraire, de nettoyer et de transformer les données provenant de diverses sources en vue de les filtrer, les corriger, les convertir vers les formats appropriés, les agréger et leur donner une valeur analytique et enfin les charger dans l'entrepôt de données.

- **Cible de stockage**

1. Entrepôt de données (*Data Warehouse*) : ensemble de données thématiques, cohérentes, évoluant dans le temps, fiables, sur lequel les dirigeants fondent leur schéma du processus de décision. C'est l'élément central de l'informatique décisionnelle.
2. Magasin de données (*Data Mart*) : Etant donné que les *Data Warehouses* sont en général, très volumineux et très complexes à concevoir, il a été décidé de les diviser en bouchées plus faciles à créer et à entretenir. On peut faire des divisions par fonction (un *Data Mart* pour les ventes, pour les commandes, pour les ressources humaines) ou par sous-ensemble organisationnel (un *Data Mart* par succursale ou par département) [15].
3. Les métadonnées sont toutes les informations de *DW* de données qui ne sont pas les données elles-mêmes [15], décrivent la structure de l'entrepôt de données et *Data Marts*, à la fois au niveau conceptuel/logique (qui inclut les faits, dimensions, hiérarchies, définitions de données dérivées) et au niveau

physique (tels que les index, les partitions et la réplication). De plus, ces métadonnées contiennent des informations de sécurité (autorisation de l'utilisateur et contrôle d'accès) et surveillent les informations (telles que les statistiques d'utilisation, les rapports d'erreurs et les pistes de vérification) [16].

- **Moteur d'analyse**

La troisième phase dans l'architecture décisionnelle correspond aux moteurs d'analyses et comprend un serveur d'analyse des données en ligne (*OLAP*) qui met en œuvre des règles de gestion et de logique applicative et transforme les données en une structure adéquate pour les requêtes d'analyses appelée cube de données.

- **Outils en sortie**

C'est la partie visible et interactive pour l'utilisateur. Il contient des outils de *data analytics*, de *data mining* et de *reporting*.

4.5. Alimentation du *Data Warehouse*

Après la conception de l'entrepôt de données, on a besoin d'acquérir des données pour la mise à jour régulière de l'entrepôt. L'usage du processus de l'alimentation du *Data Warehouse*, le plus souvent appelé processus *ETL* « *Extract-Transform-Load* » offrant de nombreux avantages pour les organisations. L'*ETL* est un moteur qui extrait, transforme et charge les données et permet aussi d'effectuer des synchronisations massives d'information depuis les différentes sources d'information hétérogènes de l'entreprise vers l'entrepôt de données (modèles multidimensionnels) [17]. La figure 7 décrit les principales phases du processus d'entreposage.



Figure 7. Étapes d'alimentation d'un *DW*.

4.5.1. Les phases d'alimentation « E.T.L »

Les trois phases du processus *ETL* « *Extract-Transform-Load* » représentent la mécanique d'alimentation du *Data Warehouse*, et se déroulent comme suit :

a. Identification des systèmes sources

Sélectionner les systèmes des données, parmi des divers types, pour l'extraction des données. Proviennent des fichiers Excel de différentes autorités dans un domaine spécifique, des factures, des données géographiques exportés des *API* web, des données cartographiques ...

b. L'extraction des données

L'extraction est la première phase du processus *ETL* permettre d'identifier et extraire les données depuis des sources différentes (logiciels, systèmes, fichier, ...), qui ne sont pas forcément stockées dans une seule base de données, et ne sont pas obligatoirement stockées au même endroit géographique, ayant subi une modification depuis la dernière exécution afin de récupérer les données de production. Cette phase a pour but de filtrer et produire des données propre et utiles, faciles d'accès et qui peuvent être exploitées efficacement par l'analytique, la *Business Intelligence* ou les opérations commerciales [12].

c. La transformation des données

La phase transformation est la seconde phase du processus. Sert à appliquer les diverses transformations aux données pour les nettoyer, les intégrer et les agréger. Elle a pour but de transformer les données afin de répondre à des contraintes d'ordre techniques, a des besoins fonctionnels et particulièrement à des objectifs en termes d'analyse [12].

Cette phase assure en réalité plusieurs tâches qui garantissent la fiabilité des données [18]:

- Consolidation des données.
- Correction des données et élimination de toute ambiguïté, effectuer des calculs, des traductions ou des résumés basés sur les

données brutes. Cela peut inclure la modification des en-têtes de ligne et de colonne pour la cohérence, la conversion de devises ou d'autres unités de mesure, la modification de chaînes de texte, etc.

- Élimination des données redondantes, en utilisant un ensemble de modèles de conception pour déterminer et suivre les données qui ont été modifiées afin d'effectuer une action spécifique.
- Compléter et renseigner les valeurs manquantes.

Le processus ETL est parfois utilisé pour générer des tables d'agrégation qui sont ensuite proposées dans des rapports de base ; dans ce cas, les données doivent être triées puis agrégées.

d. Le chargement des données

Le chargement des données est la dernière phase du processus et consiste à insérer les données et les transformées dans l'entrepôt et gérer les changements aux données existantes [12].

Il existe deux méthodes principales pour charger les données dans un entrepôt [19]:

Chargement complet

Le chargement complet des données qui a lieu la première fois que la source est chargée dans l'entrepôt.

Chargement incrémentale

Par contre, a lieu à intervalles réguliers. Ces intervalles peuvent être soit des incréments de flux (petits volumes de données) ou des incréments de lots (plus grands volumes de données).

4.5.2. Politiques d'alimentation

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

Le processus de l'alimentation du *DW* se fait de différentes manières. Ces politiques sont :

a. *Push*

La méthode *push* consiste à charger les données dans le système décisionnel. Il « pousse » les données vers la zone cible de préparation quand il en a l'occasion. Comme le montre la figure 8.

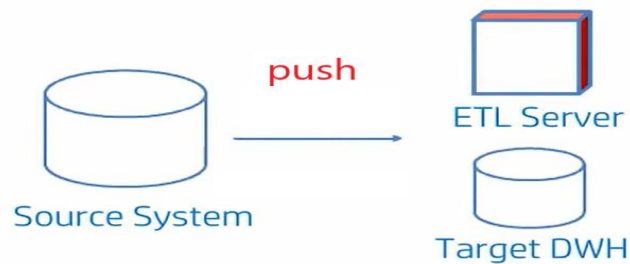


Figure 8. Méthode *push*.

b. *Pull*

Dans cette méthode, Le *Pull* « tire » les données de la source vers la zone de préparation contrairement à la méthode *push*. L'inconvénient de cette méthode est qu'elle peut surcharger le système s'il est en cours d'utilisation. Comme l'explique la figure 9.

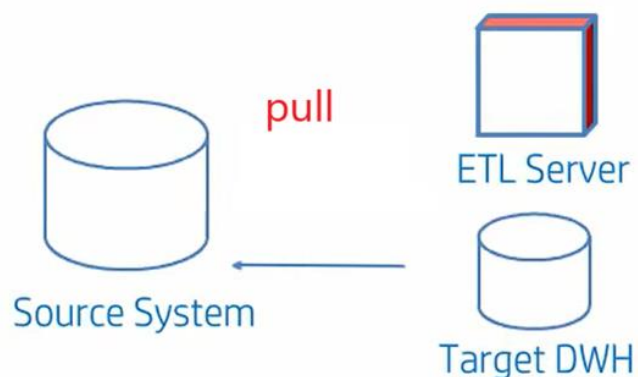


Figure 9. Méthode *pull*.

On cite deux méthodes possibles pour la résolution de ce problème :

1. Charger les données après les heures de travail (généralement la nuit).
2. L'utilisation de la méthode *master-slave*. Tel-que pour chaque table de ses *BD* : Une table *master*, qui a comme opération la lecture et l'écriture ; et une autre *slave*, qui a comme opération la lecture seulement (*read only*).

c. *Push-pull*

C'est la combinaison des deux méthodes *push* et *pull*. La source prépare les données à envoyer et indique à la zone de préparation qu'elle est prête. La zone de préparation va alors récupérer les données. Voir la figure 10.

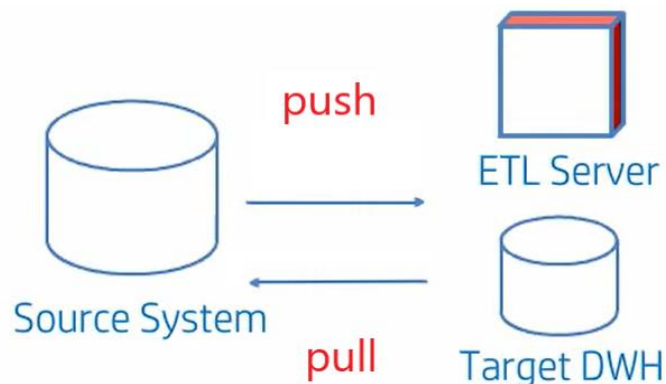


Figure 10. Méthode *Push-Pull*.

5. Concept d'OLAP (*OnLine Analytical Processing*)

4.1. Définition d'un cube OLAP

OLAP (OnLine Analytical Processing) est un logiciel permettant d'extraire les données de plusieurs ensembles de données relationnelles et les réorganise dans un format multidimensionnel, d'effectuer des analyses, et de fournir des vues multidimensionnelles des données à des outils d'aide à la décision. Ce qui permet un traitement très rapide et perspicace [6].

Voici quelques éléments à garder à l'esprit afin de déployer des données dans des cubes *OLAP* [20]:

- Un schéma en étoile hébergé dans une base de données relationnelle est une bonne base physique pour construire un cube *OLAP*, et il est généralement considéré comme une base plus stable pour la sauvegarde et la restauration.
- Les cubes *OLAP* offrent généralement des options de sécurité plus sophistiquées que les *SGBDR*, comme limiter l'accès aux données détaillées mais fournir un accès plus ouvert aux données récapitulatives.
- Les cubes *OLAP* offrent des capacités d'analyse nettement plus riches que les *SGBDR*, qui sont aux prises avec les contraintes de *SQL*. Cela peut être la principale justification de l'utilisation d'un produit *OLAP*.
- Les cubes *OLAP* peuvent imposer des contraintes détaillées sur la structure des dimensions clés qui implémentent des hiérarchies *drill-down* par rapport aux bases de données relationnelles.

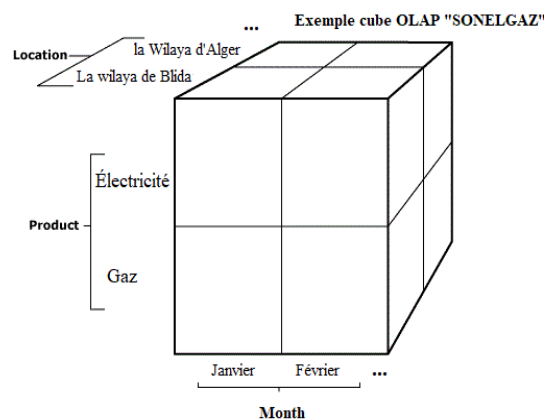


Figure 11. Exemple d'un cube *OLAP* « SONELGAZ ».

4.2. Architectures des serveurs

Le noyau d'un système *OLAP* est son serveur. Ces serveurs sont classés en trois architectures systèmes qui peuvent être distinguées comme suit [7]:

MOLAP « Multidimensional On-Line Analytical Processing »

OLAP qui fonctionne directement avec un cube *OLAP* multidimensionnel est appelé *OLAP* multidimensionnel ou *MOLAP*.

MOLAP est le type d'analyse de données multidimensionnelle le plus rapide et le plus pratique.

Cependant, il existe deux autres types d'*OLAP* qui peuvent être préférables dans certains cas.

ROLAP « Relationnel On-Line Analytical Processing »

ROLAP, ou *OLAP* relationnel, est une analyse de données multidimensionnelle qui opère directement sur les données de tables relationnelles, sans réorganiser au préalable les données dans un cube.

HOLAP « Hybride On-Line Analytical Processing »

HOLAP, ou *OLAP* hybride, tente de créer la division optimale du travail entre les bases de données relationnelles et multidimensionnelles au sein d'une même architecture *OLAP*. Les tables relationnelles contiennent de plus grandes quantités de données, et les cubes *OLAP* sont utilisés pour les agrégations et le traitement spéculatif.

HOLAP nécessite un serveur *OLAP* prenant en charge à la fois *MOLAP* et *ROLAP*.

4.3. Algèbre *OLAP*

L'algèbre *OLAP* est une caractéristique fondamentale du modèle multidimensionnel est qu'il permet de visualiser les données sous plusieurs angles et à plusieurs niveaux de détail. Les opérations *OLAP* permettant à un décideur d'analyser facilement les données de plusieurs systèmes de base de données pour les comparer de différentes façons [21].

Par exemple, une entreprise peut comparer ses ventes d'énergies en janvier avec ses ventes en février, puis comparer ces résultats avec les ventes d'un autre endroit, qui pourraient être stockées dans une base de données différente.

OLAP fournit aux décideurs toutes les informations dont ils ont besoin, il constitue, à ce titre, donc c'est un outil d'aide à la décision très pratique pour les entreprises pour la prise des décisions efficaces sur les orientations stratégiques de l'entreprise [22].

La manipulation du cube *OLAP* se fait en se basant sur des opérateurs algébriques dont (le résultat est également un cube de données appelé cuboïde), ces opérateurs sont les suivants [7]:

4.3.1 *Slice & Dice*

Le « *Slicing* », appelé aussi l'opération de découpage, est une sélection sur une dimension particulière de données à partir d'un cube initial.

Le « *Dicing* », quant à lui, est une sélection sur plusieurs dimensions du cube initial.

La différence entre eux se manifeste dans le fait que : le *Dicing* consiste à faire une sélection de plages de valeurs sur deux ou plusieurs dimensions alors que le *Slicing* est basé sur une sélection d'une valeur sur une seule dimension [7].

4.3.2 *Roll-up & Drill-down*

Le forage vers le haut ou vers le bas (*Roll-up* ou *Drill-down*) permettent au décideur d'analyser les données respectivement de manières agrégées ou détaillées en modifiant les différents niveaux de granularité utilisés pour visualiser les données.

L'opération *Drill-down* permet d'obtenir une vue agrégée ou plus détaillée sur les données et ce en descendant dans la hiérarchie. Ainsi, les décideurs peuvent effectuer leurs analyses en parcourant les cubes vers le bas en ajoutant des dimensions sans écrire ou émettre de nouvelles requêtes. Cela permet aux décideurs de naviguer plus profondément parmi les dimensions.

L'opération *Roll-up* permet d'obtenir une vue synthétique ou moins détaillée sur les données. Elle consiste à représenter les données d'un cube à un niveau de granularité inférieur, en supprimant des dimensions sans écrire ou émettre de nouvelles requêtes [7].

4.3.3 *Pivot*

L'opération de pivotement, permet de faire une rotation sur les axes de données et ce en changeant soit un axe d'analyse par un autre, soit en changeant la hiérarchie sur un même axe afin d'obtenir une nouvelle vue des données [7].

La figure 12 résume le concept des opérateurs de l'algèbre *OLAP*.

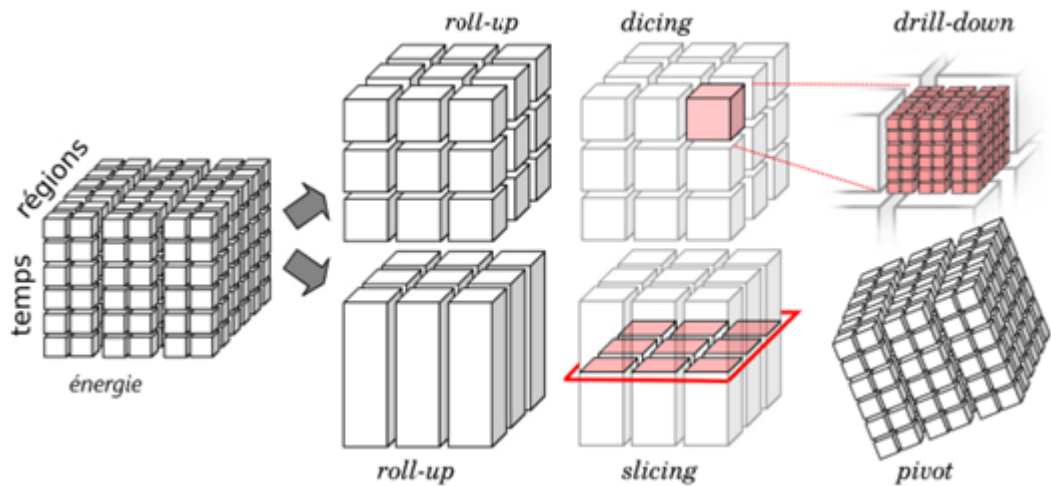


Figure 12. Résumé des opérations *OLAP*.

6. Conclusion

Après avoir appris les concepts de base de la *BI*, tels que les définitions, l'architecture décisionnelle, l'alimentation du *Data Warehouse*, etc. Nous passerons à l'analyse des besoins et la modélisation multidimensionnelle de l'entrepôt de données.

Dans le prochain chapitre on portera sur une étude détaillée de l'existant et la conception de notre projet.

Chapitre 3 : Analyse des besoins et Conception

1. Introduction

L'intérêt des systèmes décisionnels est de répondre aux attentes des utilisateurs. Cela nécessite de faire une étude approfondie de leurs besoins.

L'objectif de ce chapitre est de présenter une conception et une modélisation dimensionnelle qui répond aux besoins cités précédemment et aux exigences techniques.

2. Les entretiens

Dans la phase d'analyse de notre projet, nous avons effectué un certain nombre d'entretiens avec les responsables SONELGAZ, en vue de recenser leurs besoins en termes d'analyse et de *reporting*.

Les principaux objectifs des entretiens peuvent être résumés de la manière suivante :

- Faire une étude complète et obtenir le maximum d'informations.
- Comprendre l'aspect métier de SONELGAZ afin de recenser les exigences et les besoins analytiques.
- Avoir une vision plus claire et précise des besoins des décideurs de l'entreprise.

Cette partie est très importante, car elle nous a permis de comprendre le fonctionnement des différents services et de recenser les besoins implicites et explicites de l'entreprise.

3. Les documents de référence

En plus des entretiens qui nous ont permis de discuter de vive voix avec les différents intervenants dans le système décisionnel de la Sonelgaz, nous avons utilisé aussi un certain nombre de documents considérés comme sources d'informations très pertinentes pour le recueil des besoins dont voici un résumé :

- **Quelques bilans des objectifs**

Les bilans des objectifs, représentent les objectifs fixés annuellement dans le système *Reporting* par les décideurs, Ceci

nous a permis d'analyser les différentes valeurs des objectifs annuelles.

- **Cahier de charge *reporting* commercial**

L'équipe de développement du nouveau système SGC au niveau de la filiale *ELIT*, nous ont fourni le cahier de charge *Reporting* commerciale de leur projet « REFONTE SGC ». Qui nous a aidé à bien comprendre l'aspect métier et identifier les besoins des utilisateurs à travers les tableaux de bord (tableau de bord clients, achats, ventes, etc.). Ainsi que les formules de calcul tel que (taux de pertes, chiffres d'affaires travaux, consommation brute, consommation vendue, etc.).

- **Les sources de données**

Les sources de données qui alimentent notre système décisionnel est la base de données du système opérationnel, tel que les *BD* de la wilaya de Biskra et Tindouf.

Ces documents nous ont permis de mieux comprendre les besoins décisionnels, nous avons aussi consulté les différents sites web de SONELGAZ pour comprendre l'organisation du groupe SONELGAZ et ses activités métiers.

4. Modélisation dimensionnelle

4.1 Les dimensions partagées

Les tableaux 2, 3, 4, 5, 6 décrit les dimensions (ou axes d'analyse) retenus par les responsables de la SONELGAZ pour l'analyse des indicateurs de performances du processus commercial.

Dimensions	Description	Objectifs
Catégorie client	<ul style="list-style-type: none"> La dimension catégorie client représente les différentes catégories des clients définis par SONELGAZ: administration, secteur économique, usages propres, filiales, client ordinaire. 	<ul style="list-style-type: none"> Cette dimension permettra d'identifier l'importance de la catégorie client en termes de consommation, chiffres d'affaires, énergie ... etc. Ceci comportera aux décideurs de considérer le client stratégique ou non stratégique.
Type client	<ul style="list-style-type: none"> Cette dimension représente les deux types de client : physique ou moral. 	<ul style="list-style-type: none"> La dimension « type client » permettra d'identifier la valeur de type client en termes des différentes mesures, ce qui aide les décideurs à analyser les statistiques par rapport aux clients physiques ou morales.
Type contrat	<ul style="list-style-type: none"> Cette dimension représente les différents types de contrat que SONELGAZ offre à ses clients : BT / BP / HTA / MP / HTB /HP. 	<ul style="list-style-type: none"> Cette dimension assure l'aspect évolutif de la conception et permet de supporter les autres types de contrat qui sont hors de notre périmètre, dans toutes les analyses les modèles seront détaillés par type contrat pour séparer les différents types tensions (BT/HTA...) Ou pression (BP, MP...).

Tableau 2. Description détaillée des dimensions.

Dimensions	Description	Objectifs
Énergie	<ul style="list-style-type: none"> • Cette dimension représente l'électricité ou le Gaz. 	<ul style="list-style-type: none"> • La dimension « énergie » permettra de détailler tous les indicateurs, dans toutes les analyses les modèles seront détaillés par énergie (électricité ou gaz).
Origine relève	<ul style="list-style-type: none"> • La relève consiste à collecter les index lus sur les cadrans des compteurs (d'électricité et du gaz) sur le lieu de consommation ou à distances et de constater toute anomalie et la consigner sur un rapport journalier de parmi les origines de relève : Relevé, mouvement, relève estimer, forfaitise. 	<ul style="list-style-type: none"> • La dimension « origine relève » permettra d'identifier l'intérêt des différents types d'origine relève (Relevé, mouvement, relève estimer, forfaitise) en termes de client, consommation, chiffres d'affaires, énergie ... etc.
Région métier	<ul style="list-style-type: none"> • Cette dimension représente les différents régions métier de SONELGAZ: RD, DD, Agence. 	<ul style="list-style-type: none"> • La dimension « région métier » permettra d'identifier l'importance de la région en termes de consommation, chiffres d'affaires, énergie, etc. • Ceci comportera aux décideurs de considérer la région stratégique ou non stratégique.

Tableau 3. Description détaillée des dimensions (suite).

Dimensions	Description	Objectifs
<p style="text-align: center;">Temps</p>	<ul style="list-style-type: none"> Cette dimension représente la Période : année, trimestre, mois. 	<ul style="list-style-type: none"> Cette dimension permettra d'évaluer la consommation, chiffres d'affaires sur l'axe temporelle et permettra aussi de comparer le taux d'évolution des différentes périodes pour identifier une performance ou une régression. Ex : Comparer chiffres d'affaires entre deux années différentes.
<p style="text-align: center;">Région administrative</p>	<ul style="list-style-type: none"> Cette dimension représente les régions administratives de SONELGAZ: Wilaya / Commune. 	<ul style="list-style-type: none"> La dimension « région administrative » permettra d'identifier l'importance de la région administrative en termes de consommation, chiffres d'affaires, énergie, etc. Dans toutes les analyses les modèles seront détaillés par la région administrative suite aux demandes de Ministère / gouvernement ...

Tableau 4. Description détaillée des dimensions (suite).

Dimensions	Description	Objectifs
État facture	<ul style="list-style-type: none"> • Cette dimension représente l'état de la facture : validée ou annulée. 	<ul style="list-style-type: none"> • Cette dimension permettra de définir l'intérêt de l'état de la facture en termes de consommation, chiffres d'affaires... etc. • Cela permettra aux décideurs d'analyser et d'étudier les différents états de factures. Ex : le nombre de facture validée, annulée, etc.
Type facture	<ul style="list-style-type: none"> • Cette dimension représente les types de factures de SONELGAZ, Ex : facture cyclique, facture hors cycle, facture avoir, facture fraude, etc. 	<ul style="list-style-type: none"> • Cette dimension permettra aux décideurs de Distinguer entre les différents types de facture. • Cela aide aussi à enquêter sur les factures fraudes.
Activité économique	<ul style="list-style-type: none"> • Cette dimension représente les différentes activités économiques. <p>Ex : agriculture, Administration, forage, abonnée, ménage, etc.</p>	<ul style="list-style-type: none"> • Cette dimension permettra d'identifier l'importance de l'activité économique par rapport aux consommations, ventes, etc. <p>Ex : La portée de la consommation de l'électricité et le gaz chez boulangeries.</p>

Tableau 5. Description détaillée des dimensions (suite).

Dimensions	Description	Objectifs
Type abonnement	<ul style="list-style-type: none"> Cette dimension représente les différents types d'abonnement définis par SONELGAZ. <p>Ex : abonnement pour les clients ordinaires, abonnement pour les clients FSM.</p>	<ul style="list-style-type: none"> Cette dimension permettra d'identifier l'importance de type d'abonnement en termes des mesures. Ceci permettra aux décideurs de connaître le nombre des clients FSM ou hors FSM (ordinaires).
Tarif	<ul style="list-style-type: none"> Cette dimension représente les différents tarifs que SONELGAZ propose pour les différents types des clients. 	<ul style="list-style-type: none"> La dimension « tarif » permettra aux décideurs d'identifier l'importance des tarifs en termes de consommation, chiffres d'affaires, énergie, etc. Plusieurs statistiques se basent sur les tarifs.
Type objectif	<ul style="list-style-type: none"> Cette dimension représente les différents types de mesure d'objectifs. <p>Ex : Consommation vendu, nombre client, montant achat, etc.</p>	<ul style="list-style-type: none"> La dimension « type objectif » permettra aux décideurs d'analyser les chiffres d'affaires, les ventes, énergie, etc. <p>Ceci permettra de comparer ses données à celles de la concurrence ou par rapport aux <i>Targets</i>.</p>

Tableau 6. Description détaillée des dimensions (suite).

Les tableaux 7 et 8 décrivent les mesures, appelées aussi indicateurs ou faits, retenues par les responsables de la SONELGAZ pour l'analyse et l'évaluation du processus commercial.


Mesures	Formule	Unité
Consommation vendue	Σ des consommations vendues – Σ consommation annulées – Σ consommation avoirs + Σ consommation annulées avoirs.	Float
Consommation Brute	La différence entre les index : Nouveau index - L'ancien index.	Float
Nombre Clients	Nombre d'abonnement + Clients nouveaux + Clients réabonnés – Clients résiliés.	Int
Nombre factures	Σ des identifiants des factures.	Int
Achat Energie	Σ énergie.	Int
Montant achat	Σ totale achat.	Float
Chiffres d'affaires travaux	Σ (montant ventes matérielles + prestations + travaux).	Float
Montant TVA	Σ total valeurs taxes TVA.	Float
Montant TTC	Σ total valeurs taxes TTC.	Float
Taxe habitation	Σ total valeurs taxes habitation.	Float
Taxe TVPE	Σ total valeurs taxes.	Float
Prime fixe	Σ montant d'abonnement.	Float
Droit fixe	Σ (taux *montant HT).	Float
Objectif	Calculé à partir d'une application.	Float


Tableau 7. Description détaillée des mesures.

Mesures	Formule	Unité
Chiffres d'affaires ventes	<p>Première formule : Montant (Primes fixes + Consommation) + Σ (Montant énergie + prime fixe) vendu – (Montant énergie + prime fixe) annulée – (Montant énergie + prime fixe) avoir + (Montant énergie + prime fixe) annulée avoir des émissions ordinatrices + émissions manuelles – les annulations – les avoir + les annulations des avoirs.</p> <p>Montant (Redevances + PMD/DMD + PMA/DMA + Consommation) des émissions ordinatrices + émissions manuelles – les annulations – les avoir + les annulations des avoirs.</p> <p>Deuxième formule : Montant (Redevances + PMD/DMD + PMA/DMA + Consommation) des émissions ordinatrices + émissions manuelles – les annulations – les avoir + les annulations des avoirs.</p>	Float

Tableau 8. Description détaillée des mesures (suite).

Les tableaux 9, 10, 11 présente une description détaillée des attributs de chaque dimension avec le détail tel que :

 L'attribut est la clef primaire dans cette table.

 L'attribut est unique dans cette table.














Dimensions	Attributs	Détails
Catégorie client	 id int8  code varchar(255) U  libelle varchar(255) U	<p>Id : identification.</p> <p>Code : codification des catégories clients.</p> <p>Libelle : Libellée, description des champs.</p> <p>Ex : UP= usage propre.</p>
Type client	 id int8  code varchar(255) U  libelle varchar(255) U	<p>Id : identification.</p> <p>Code : codification</p> <p>Libelle : Libellée, description des champs.</p> <p>Ex : Ph = Physique.</p>
Type contrat	 id int8  code varchar(255) U  libelle varchar(255) U  energie varchar(255)	<p>Id : identification.</p> <p>Code : codification des types contrats.</p> <p>Libelle : Libellée, description des champs.</p> <p>Energie : électricité ou gaz.</p>
Energie	 id int8  code varchar(255) U  libelle varchar(255) U	<p>Id : identification.</p> <p>Code : codification des énergies.</p> <p>Libelle : Libellée, descriptions des champs</p> <p>Ex : E = électricité, G = Gaz.</p>

Tableau 9. Description détaillée des attributs de chaque dimension.






















Dimensions	Attributs	Détails
Tarif	 id int8  code varchar(255) U	<p>Id : identification.</p> <p>Code : codification des tarifs.</p>
Origine relève	 id int8  code varchar(255) U  libelle varchar(255)	<p>Id : identification.</p> <p>Code : codification des origines relèves.</p> <p>Libelle : Libellée, descriptions des champs.</p> <p>Ex : M= Mouvement.</p>
Région métier	 id int8  code_agence varchar(255)  code_rd varchar(255)  libelle_agence varchar(255)  code_dd varchar(255)  libelle_dd varchar(255)  libelle_rd varchar(255)	<p>Id : identification.</p> <p>Code_rd, Code_dd, Code_agence : codification des champs.</p> <p>Libelle_rd, Libelle_dd, Libelle_agence : description du champ.</p>
Temps	 id int8  mois int4  libelle_mois varchar(255)  annee varchar(255)  code_trimestre int4  libelle_trimestre varchar(255)	<p>Id : identification.</p> <p>Mois : numérotation des mois.</p> <p>Libelle_mois : description du champ mois.</p> <p>Annee : Année, description du champ année.</p> <p>Code_trimestre : numérotation des trimestres.</p> <p>Libelle_trimestre : description du champ trimestre.</p>
Activité économique	 id int8  code varchar(255) U  libelle varchar(255) U	<p>Id : identification.</p> <p>Code : codification activité économique.</p> <p>Libelle : description des champs.</p>

Tableau 10. Description détaillée des attributs de chaque dimension (suite).



















Dimensions	Attributs	Détails
Type abonnement	 id int8  code varchar(255) U  libelle varchar(255) U	<p>Id : identification.</p> <p>Code : codification des types d'abonnements.</p> <p>Libelle : description des champs.</p>
Région administrative	 id int8  libelle_comune varchar(255)  code_comune varchar(255) U  libelle_wilaya varchar(255)  code_wilaya varchar(255)	<p>Id : identification.</p> <p>Libelle_commune : description des champs.</p> <p>Code_commune : codification commune.</p> <p>Libelle_wilaya : description des champs. Ex :16 = Alger.</p> <p>Code_wilaya : numérotation des wilayas.</p>
État facture	 id int8  code varchar(255) U  libelle varchar(255) U	<p>Id : identification</p> <p>Code : codification état facture.</p> <p>Libelle : description des champs.</p> <p>Ex : v = validee.</p>
Type facture	 id int8  code varchar(255) U  libelle varchar(255) U	<p>Id : identification.</p> <p>Code : codification des factures.</p> <p>Libelle : description des champs.</p> <p>Ex : HC = HORS_CYCLE.</p>
Type objectif	 id int8  code varchar(255) U  libelle varchar(255)  unite varchar(255)	<p>Id : identification.</p> <p>Code : codification des types d'objectifs.</p> <p>Libelle : noms des objectifs.</p> <p>Unite : Unité, DA / Th / m³.</p>

Tableau 11. Description détaillée des attributs de chaque dimension (suite).

4.2 Les modèles dimensionnels

Une clé de substitution (*Surrogate key*) est utilisée aux niveaux des différentes tables des différents modèles en étoile. Celle-ci sera générée automatiquement par le système et indépendante de la clé naturelle (*Business Key*) qui provient des systèmes opérationnels [23].

Nous avons opté pour l'utilisation des *SK* pour les raisons suivantes :

- L'utilisation d'une clé de substitution simplifiera la jointure entre une table de faits et une table de dimension.
- Historique des changements et granularité infinie, l'historique des changements de dimension selon un certain critère, nécessite la clé de substitution.
- Améliorer la performance, Les clés de substitution présentent des avantages significatifs en termes de performances.

L'utilisation seulement des clés naturelles dans des modèles dimensionnels est généralement une bévue, l'utilisation de clés de substitution doit être considérée comme la norme pour les modèles dimensionnels.

4.2.1 Modèles en étoile

Les figures 13, 14, 15, 16 représentent les quatre modèles en étoiles pour les activités ventes, achats, objectifs et chiffres d'affaires.

- **Modèle en Etoile « Ventes »**

Le modèle en étoile des ventes permet aux décideurs d'avoir une vue sur l'évolution des ventes d'énergie réalisées, la consommation vendue, la consommation brute, nombre de clients, nombre de factures, les chiffres d'affaires des ventes réalisés par rapport aux dimensions associées.

La figure 13 décrit le modèle en étoile des ventes.

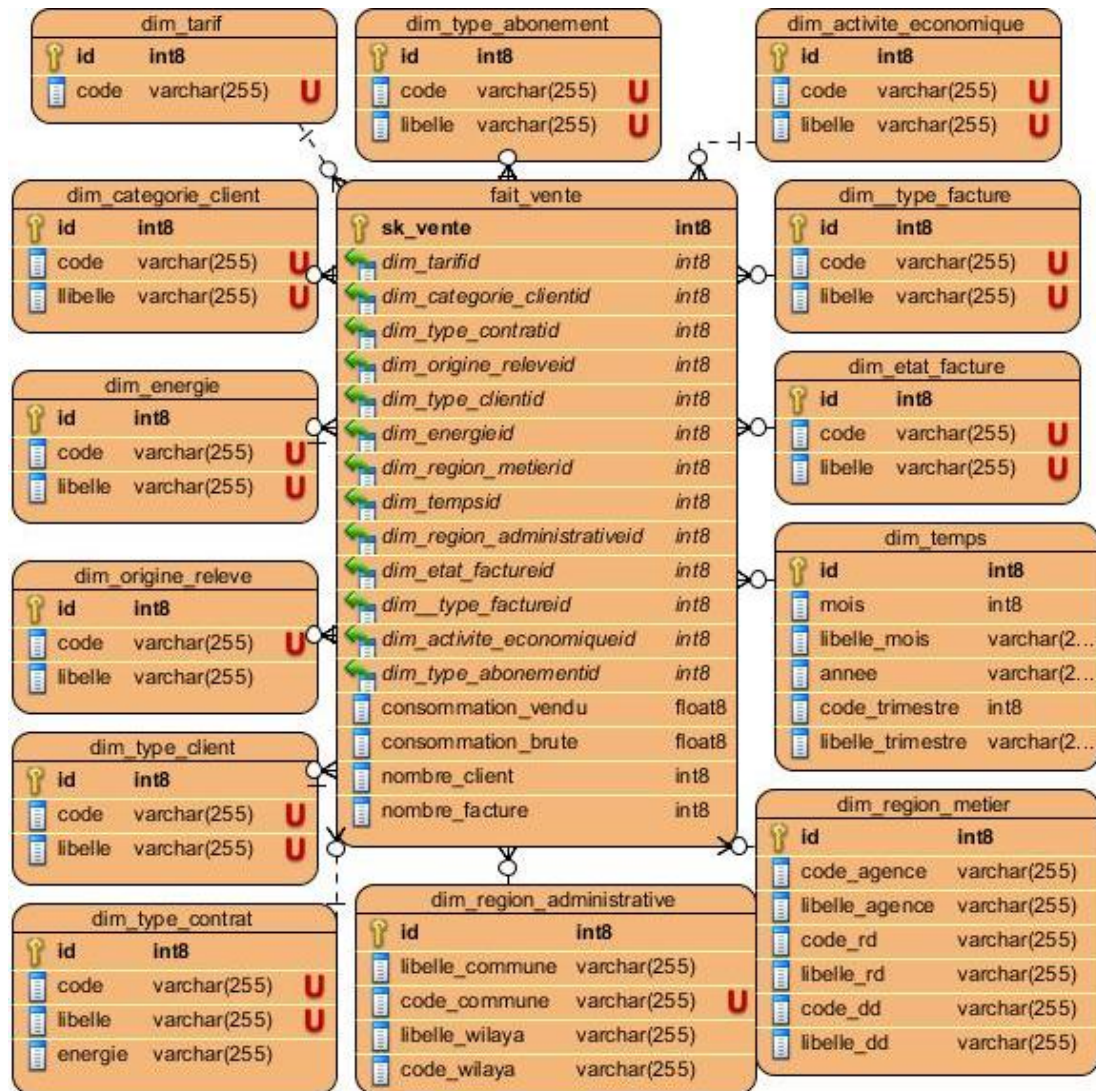


Figure 13. Modèle en étoile « ventes ».

- **Modèle en étoile « Achats »**

La figure 14 représente le modèle en étoile relatif aux achats de la SONELGAZ auprès de ses fournisseurs. Il permet aux décideurs de suivre l'évolution des achats des équipements, d'énergies, montant d'achat par rapport aux dimensions associées.

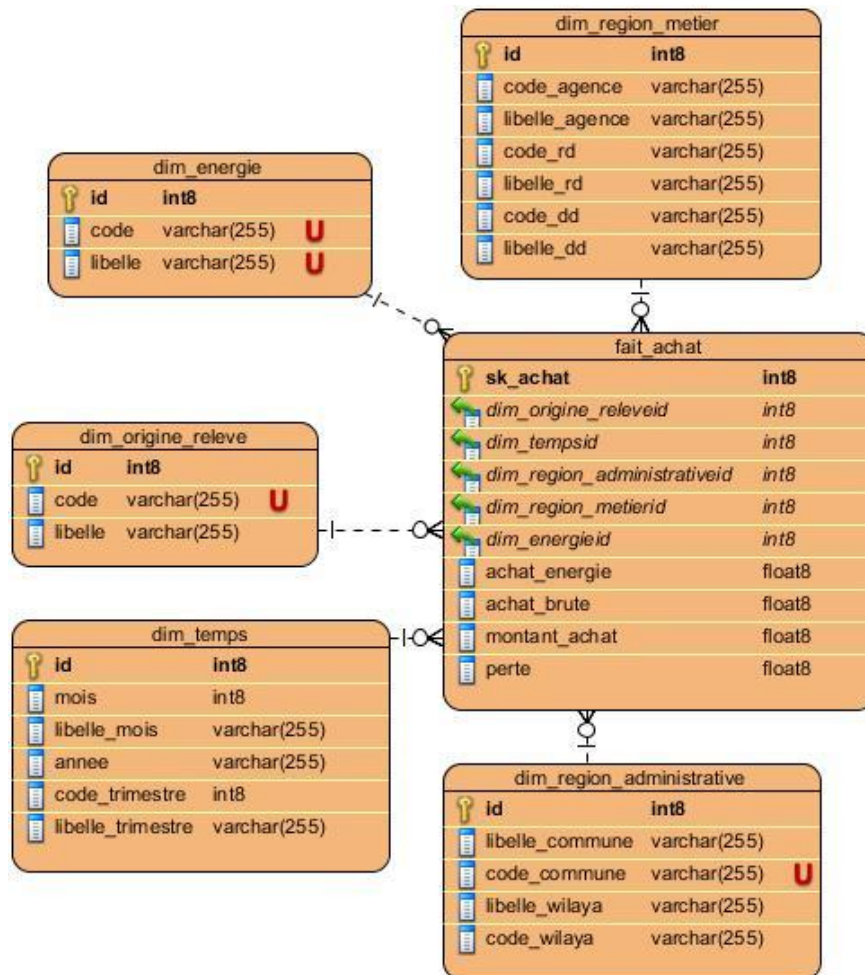


Figure 14. Modèle en étoile « Achats ».

- **Modèle en étoile « Objectifs »**

Permettre aux décideurs de suivre l'évolution des résultats et les écarts par rapport à des valeurs de référence (objectifs fixés, normes internes ou externes, références statistiques, etc.), en temps réel et par rapport aux régions administratives et métiers, type objectifs, type contrats.

La figure 15 représente le modèle en étoile des objectifs.

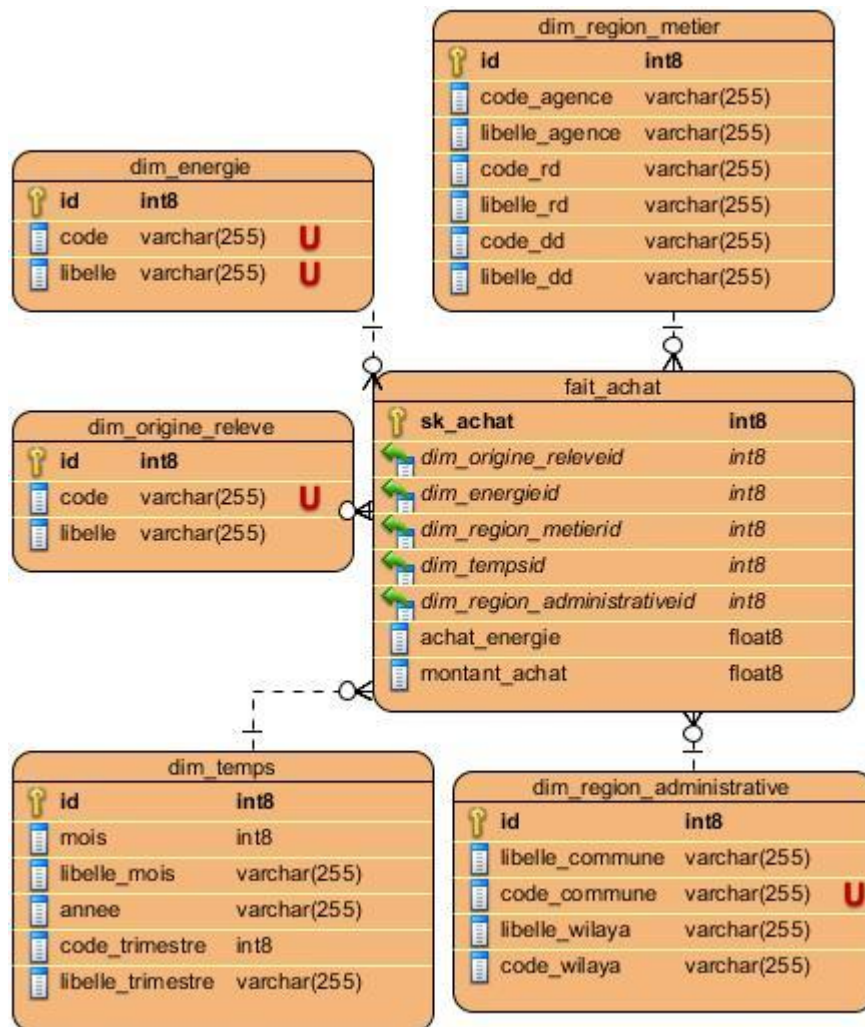


Figure 15. Modèle en étoile « Objectifs ».

- **Modèle en étoile « Chiffres d'affaires »**

Permettre aux décideurs d'observer plusieurs indicateurs, basés sur les chiffres d'affaires, qui leurs permettront d'interpréter la performance de l'entreprise.

Par exemple : les chiffres d'affaires d'énergies par rapport aux activités économiques, régions métier, etc.

La figure 16 représente le modèle en étoile des chiffres d'affaires.

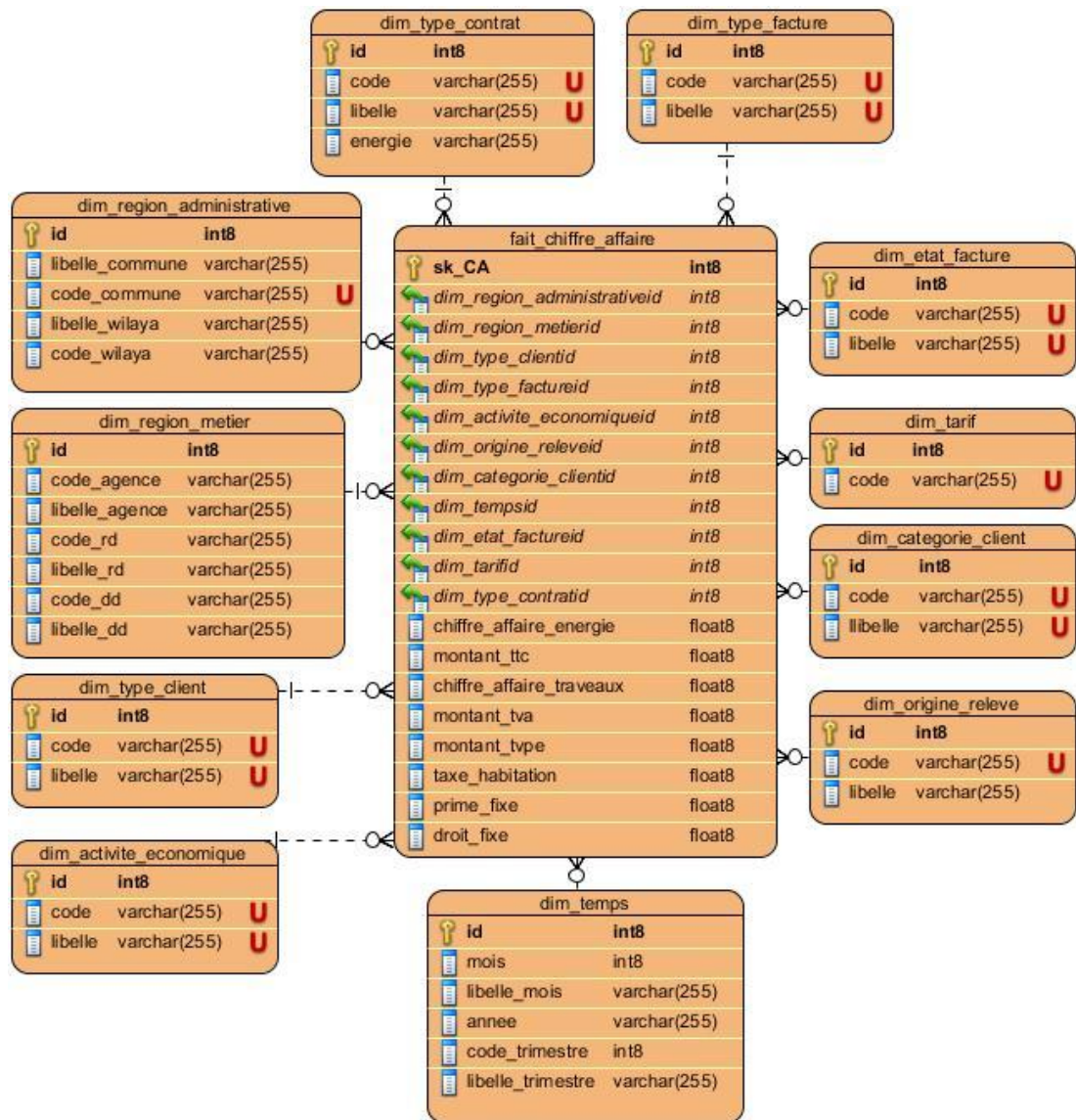


Figure 16. Modèle en étoile « chiffres d'affaires ».

4.2.2 Modèles en flocons de neige

Le modèle en flocon de neige est un modèle pour lequel chaque dimension est représentée avec plusieurs tables, c'est une version normalisée du modèle en étoile, ce qui génère des hiérarchies au niveau des dimensions normalisées [20].

La modélisation d'un modèle en flocon de neige est généralement justifiée par l'économie d'espace de stockage en normalisant les données, cependant elle peut

s'avérer moins compréhensible pour l'utilisateur final, et très couteuse en termes de performances.

Les figures 17, 18, 19, 20 représentent les quatre modèles en flocon de neige pour les activités achats, ventes, chiffres d'affaires et objectifs.

La figure 17 représente le modèle en flocon de neige relatif aux achats selon les dimensions « *dim_energie*, *dim_region_metier*, *dim_temps*, *dim_region_administrative* ».

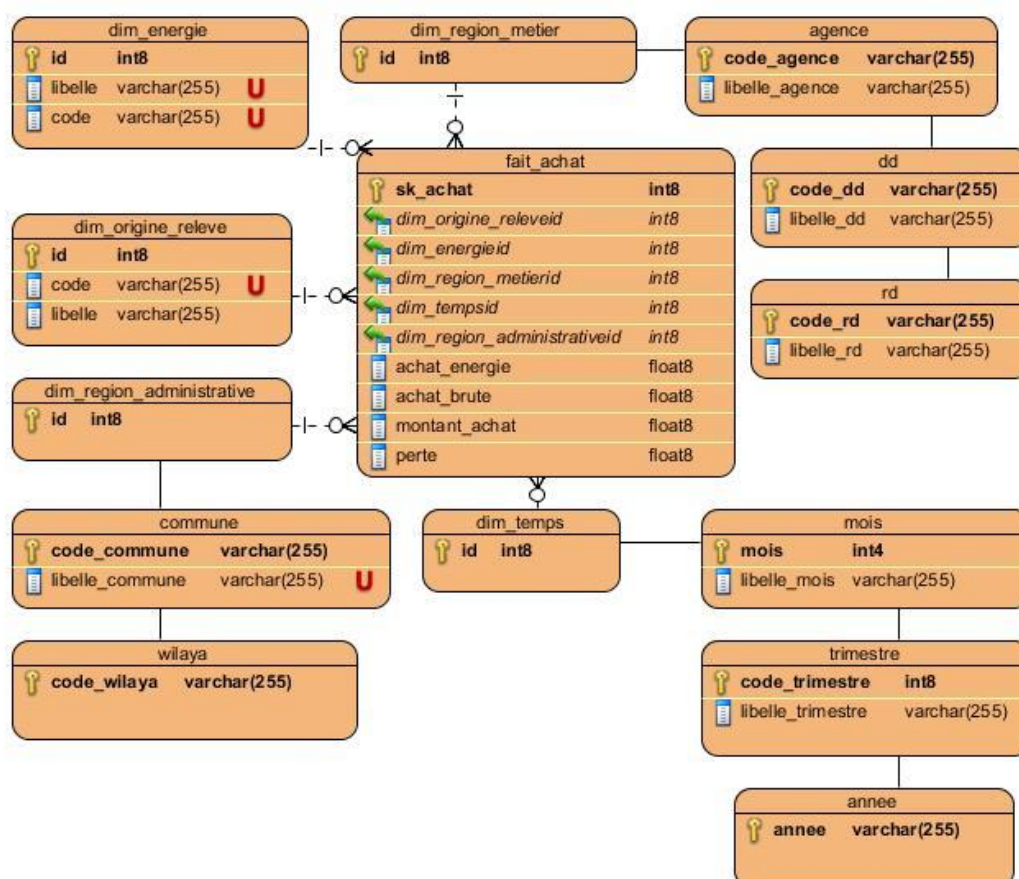


Figure 17. Modèle en flocon de neige « Achats ».

La figure 18 représente le modèle en flocon de neige relatif aux ventes.

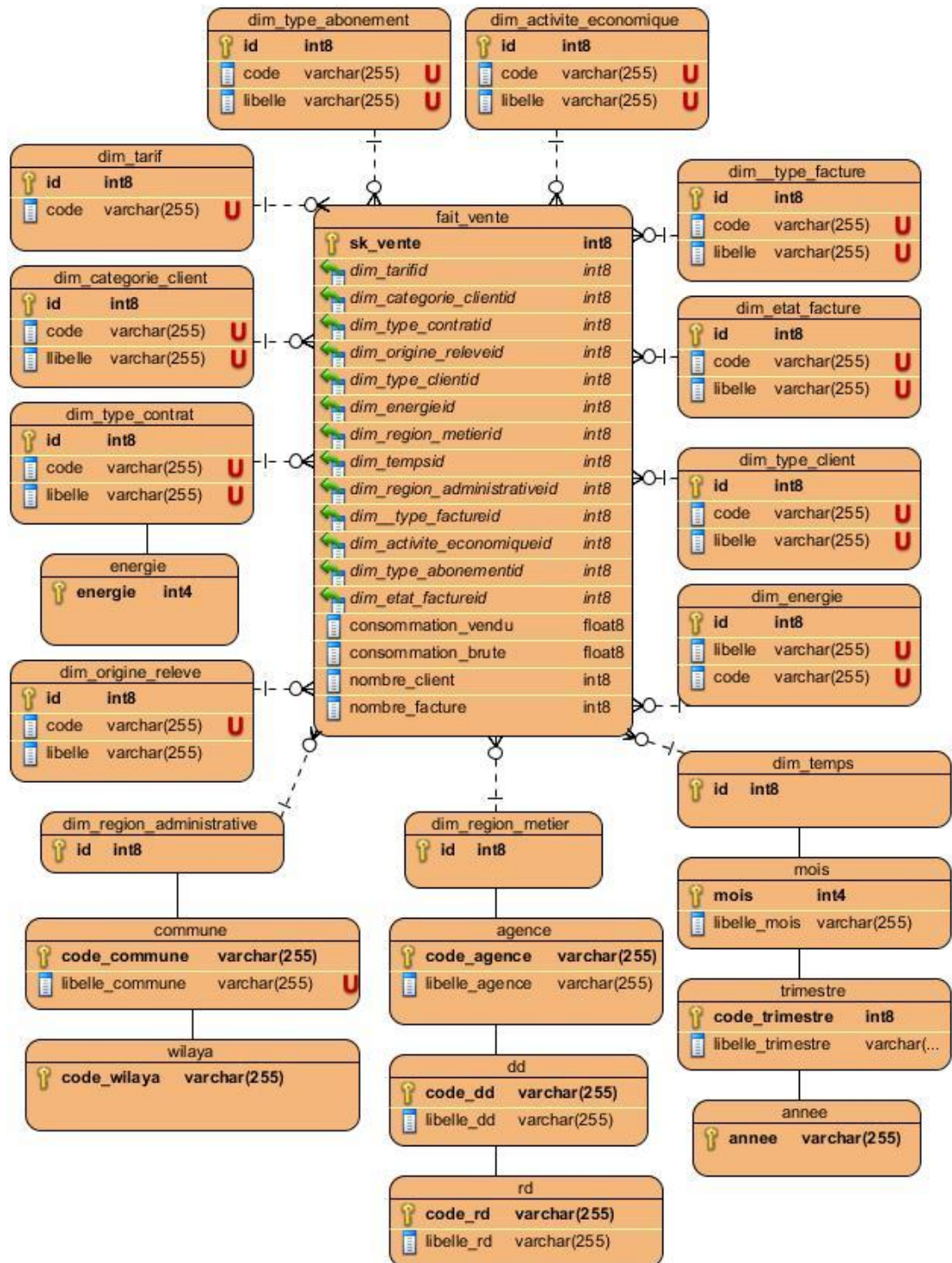


Figure 18. Modèle en flocons de neige « ventes ».

La figure 19 représente le modèle en flocon de neige relatif au chiffre d'affaires selon les dimensions « *dim_region_metier*, *dim_temps*, *dim_region_administrative*, *dim_type_client*, *dim_type_facture*, *dim_etat_facture*, *dim_activite_economique*, *dim_tarif*, *dim_categorie_client*, *dim_type_contrat*, *dim_origine_releve* ».

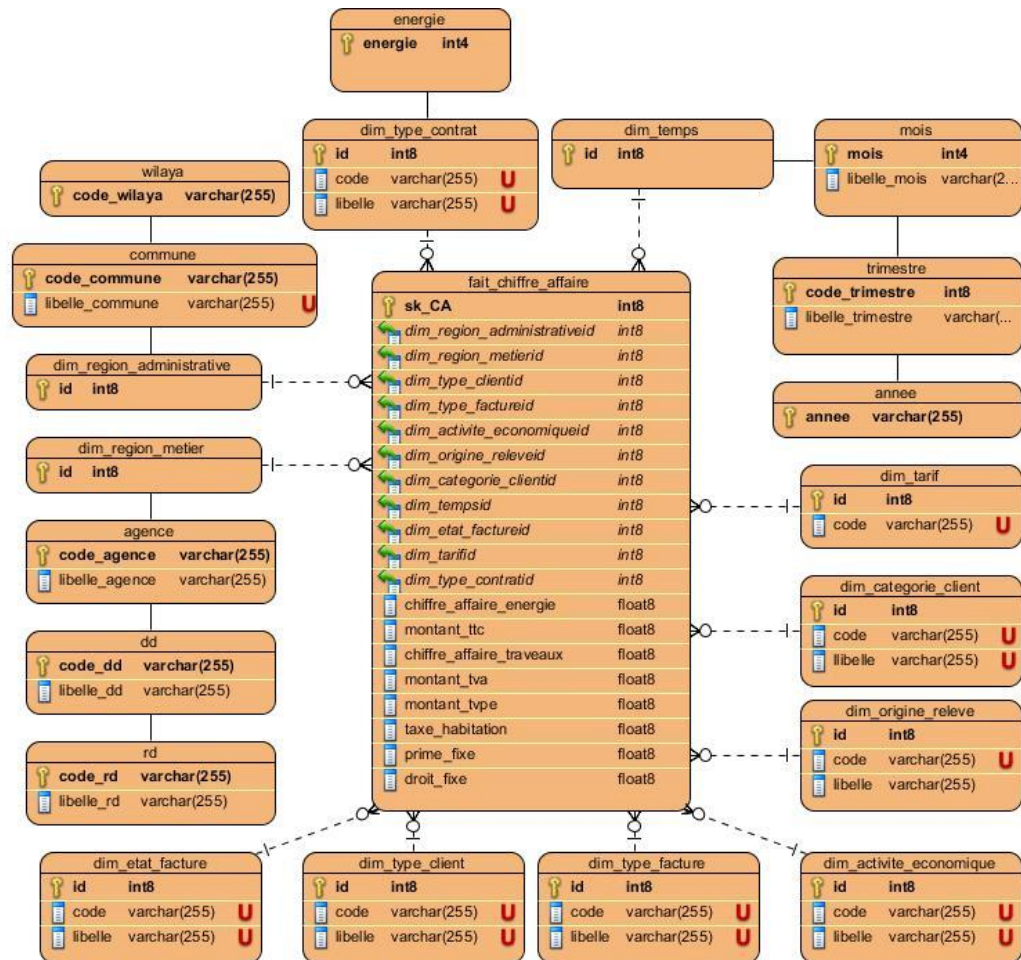


Figure 19. Modèle en flocon de neige « Chiffres d'affaires ».

La figure 20 représente le modèle en flocon de neige relatif aux objectifs selon les dimensions « *dim_region_metier*, *dim_temps*, *dim_region_administrative*, *dim_type_contrat*, *dim_type_objectif* ».

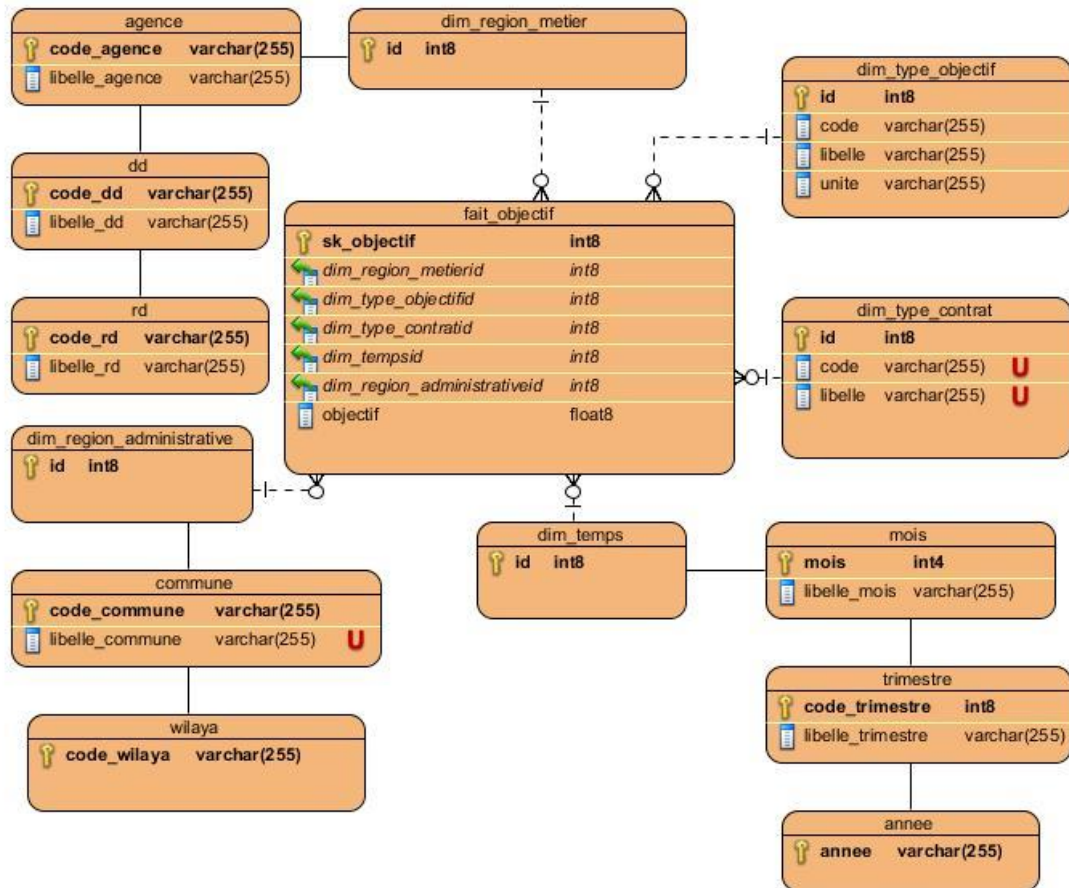


Figure 20. Modèle en flocon de neige « Objectifs ».

4.2.3 Modèle en constellation

Le modèle en constellation consiste à regrouper l'ensemble des modèles en étoile cité précédemment dans les figures (17, 18, 19, 20), sans duplication des tables de dimension [24].

Nous présenterons par la suite, la figure 21 qui illustre notre modèle en constellation avec les quatre tables de fait (*tf_achat*, *tf_vente*, *tf_objectif*, *tf_chiffre_affaire*) et quatorze tables de dimensions (*dim_temps*, *dim_categorie_client*, etc).

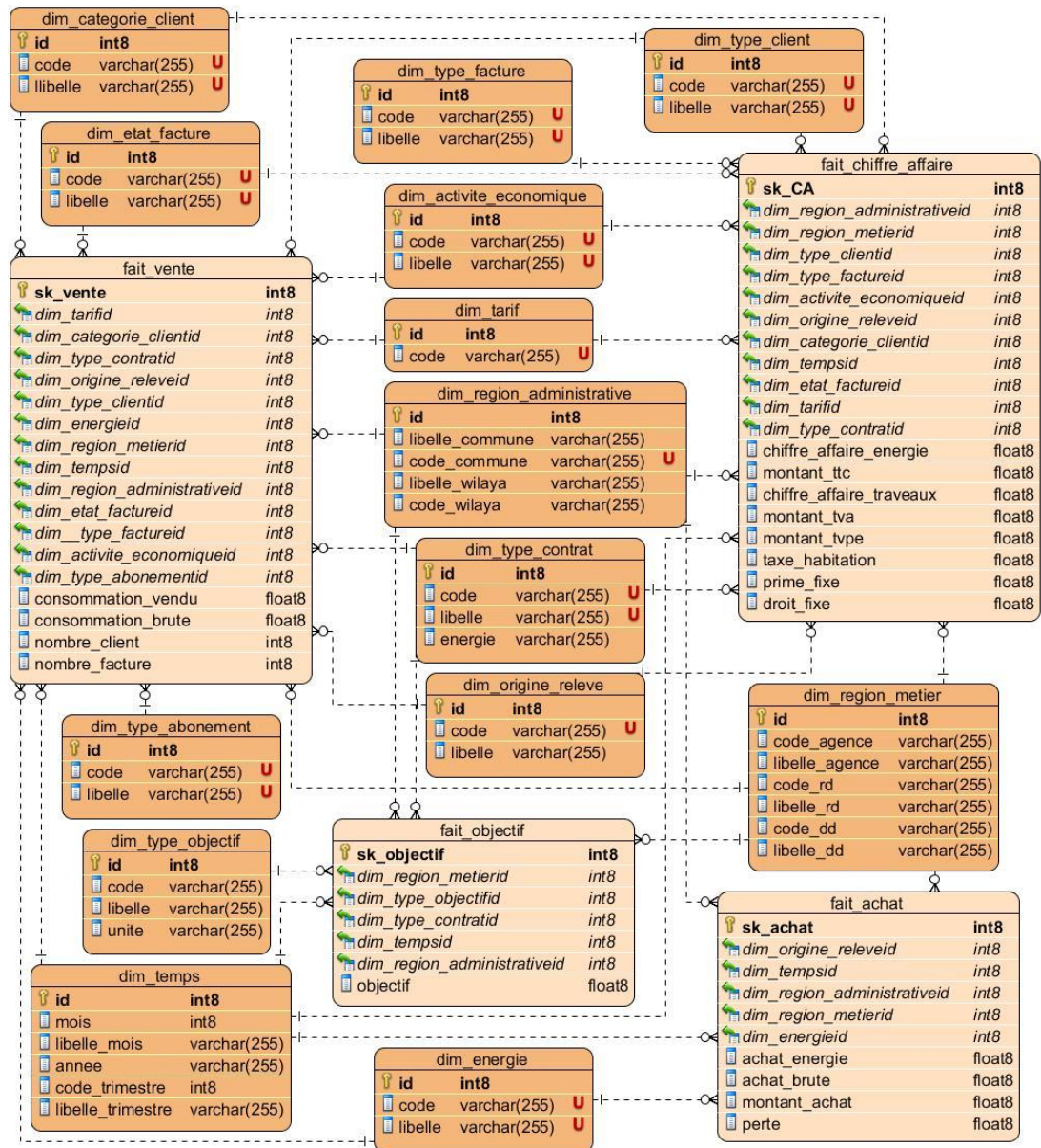


Figure 21. Modèle en constellation.

5. Transformation et Intégration des données

5.1 Processus de l'entreposage de données

Afin de décrire le fonctionnement global du processus *ETL*, nous avons choisi le diagramme d'état transition.

La figure 22 présente les deux phases d'extraction et transformation des données « achats ».

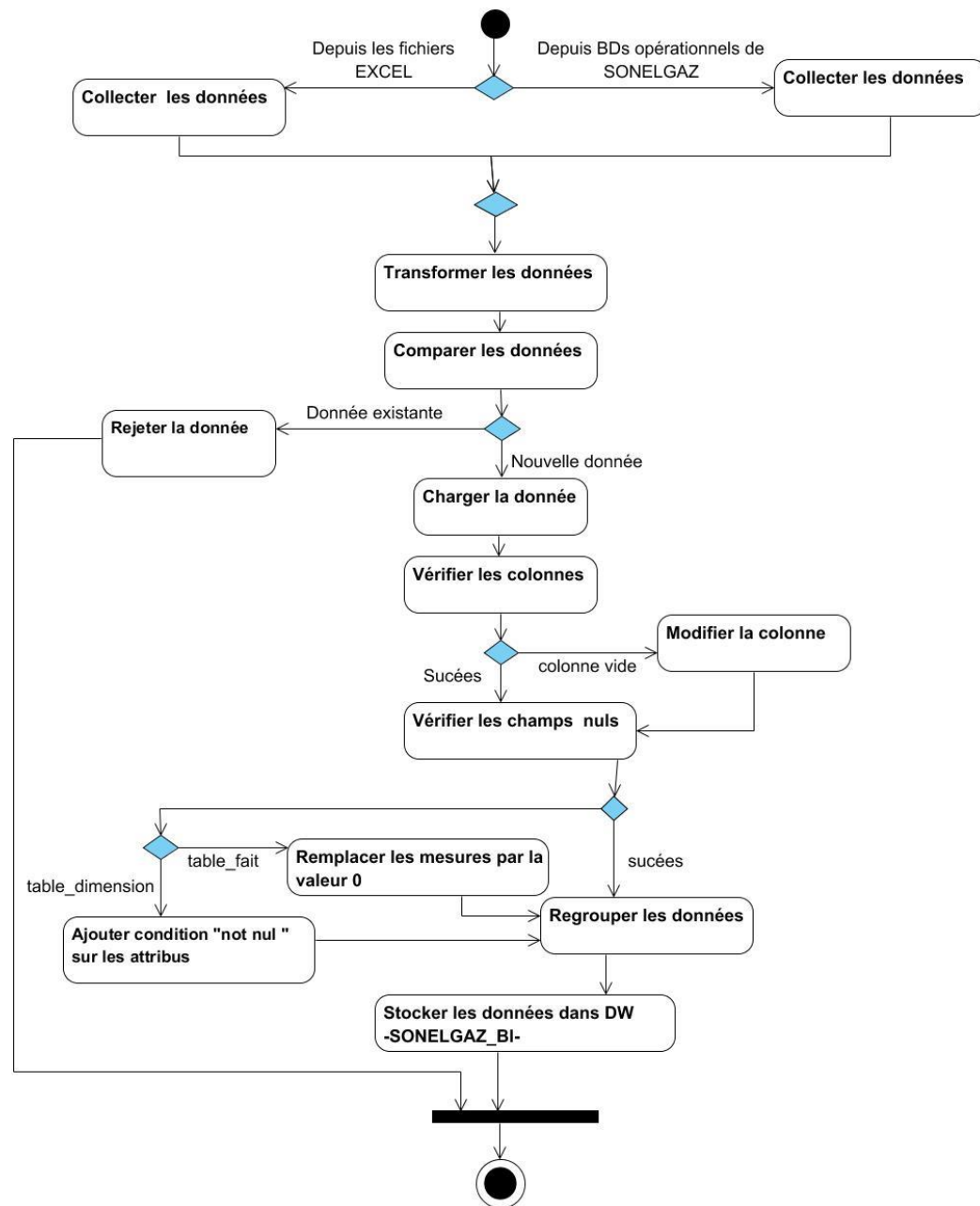


Figure 22. Processus d'extraction et transformation de données « achats ».

La figure 23 présente la phase de chargement des données « achats ».

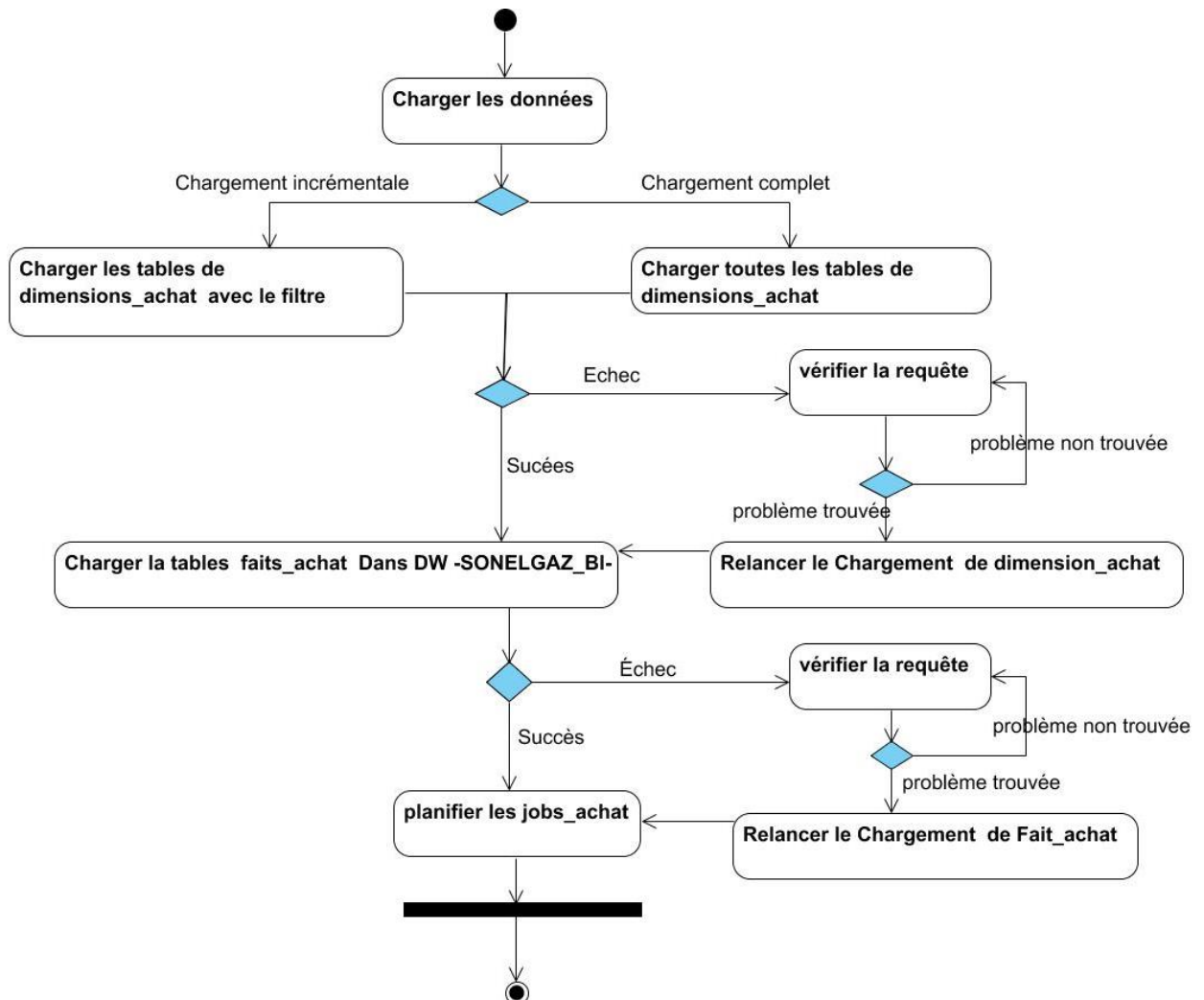


Figure 23. Processus de chargement de données « achats ».

5.2 Planification de chargement des données

SONELGAZ aborde la notion de périodicité de chargement afin d'éviter tout blocage possible de la *BD* du SGC.

La périodicité de l'exécution du chargement de données, varie d'un modèle à un autre. En effet elle dépend de :

- La fréquence de changement des données de la *BD* source.
- La quantité de données à charger.
- La période d'inactivité des systèmes sources.

Supposons que le système de facturation fonctionne de huit heures jusqu'à dix-sept heures. Le chargement du *DW* s'effectuera à partir de dix-sept heures et se déclenchera automatiquement de façon périodique.

Le tableau 12 montre le rafraîchissement des quatre modèles de notre *DW*:

Modèle dimensionnelle	Périodicité de chargement
Ventes	Journalière
Achats	Journalière
Chiffres d'affaires	Journalière
Objectifs	Annuelle

Tableau 12. Périodicité de chargement des modèles dimensionnels.

6. Conception des cubes multidimensionnels

Après avoir identifié nos dimensions et nos mesures et les présenter sous forme d'un modèle multidimensionnel (modèle en étoile), nous procéderons à la construction des cubes multidimensionnels. Afin de faciliter la navigation aux données et leurs visualisations selon plusieurs axes et à plusieurs niveaux de détail.

La figure 24 illustre un exemple graphique d'un cube multidimensionnel.

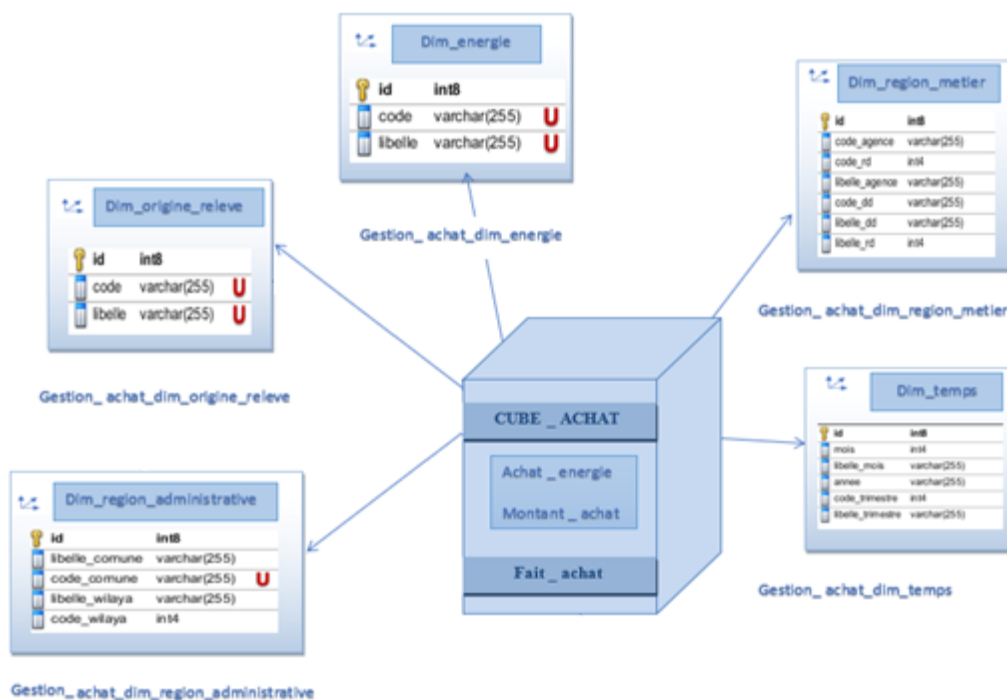


Figure 24. Cube multidimensionnel « achats ».

7. Diagramme de cas d'utilisation

Dans le but d'offrir une vision globale de notre système et pour identifier et décrire tous type d'analyse effectuée par l'administrateur et les utilisateurs (décideurs), nous avons recouru à présenter les besoins à travers le diagramme de cas d'utilisation illustré dans la figure 25.

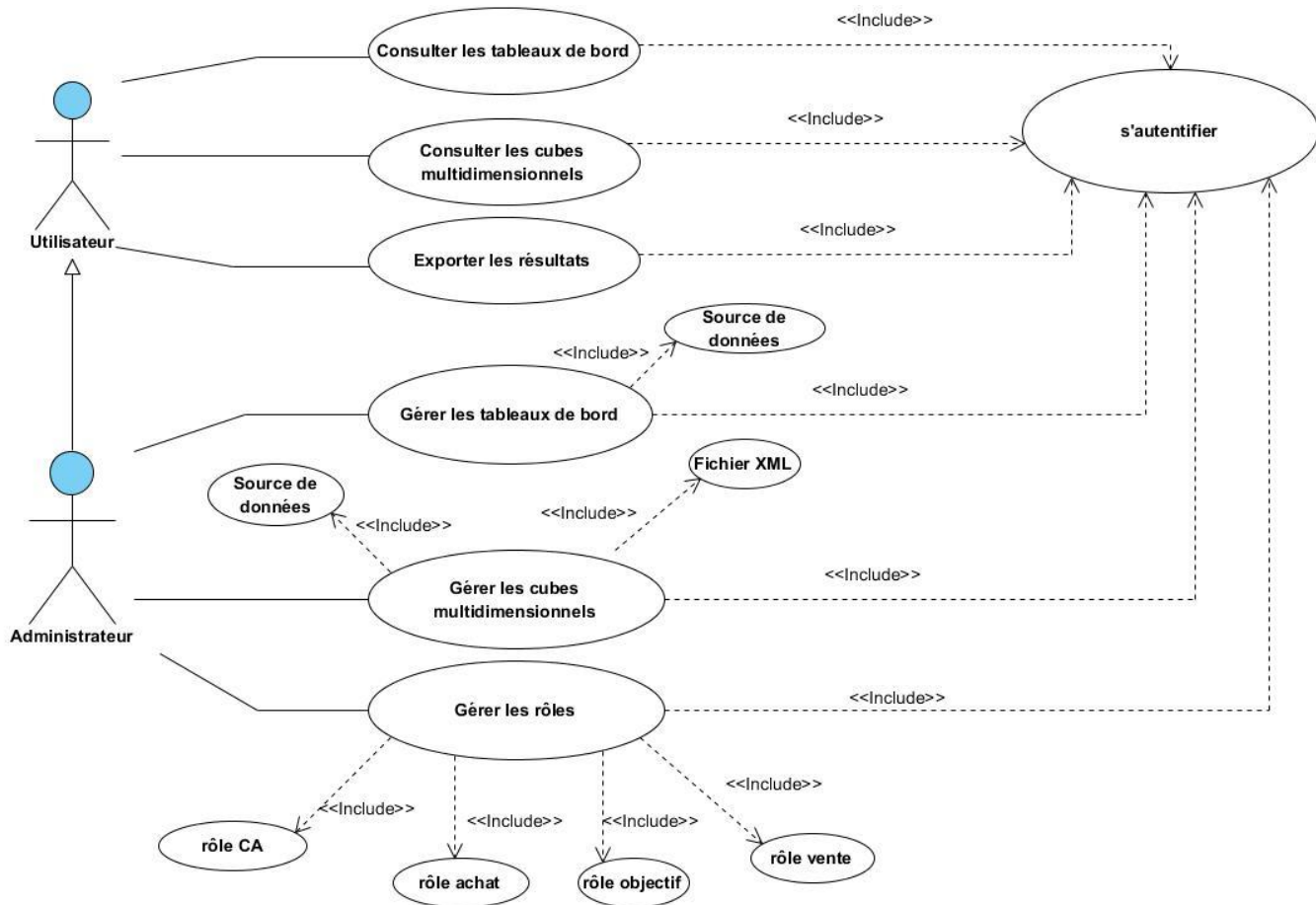


Figure 25. Diagramme de cas d'utilisation.

Le tableau 13 explique les détails des différents cas d'utilisation de notre système par rapport aux acteurs précédents.

Acteur	Cas d'utilisation	Détail
Administrateur	Gérer les tableaux de bord.	L'administrateur est une personne désignée pour l'administration du système et pour la gestion des utilisateurs (ajouter, modifier et supprimer des : tableaux de bord, cubes et rôles), après être s'autentifier pour accéder à l'interface.
	Gérer les cubes multidimensionnels.	
	Gérer les rôles.	

Tableau 13. Détails sur les cas d'utilisation.

Acteur	Cas d'utilisation	Détail
Utilisateur	Consulter les tableaux de bord.	L'utilisateur est un responsable de SONELGAZ, a la possibilité de consulter les tableaux de bord, les cubes multidimensionnelles et exporter les résultats relatifs à son niveau hiérarchique, après être s'authentifier pour accéder à l'interface.
	Consulter les cubes multidimensionnels.	
	Exporter les résultats.	

Tableau 14. Détails sur les cas d'utilisation (suite).

8. Conclusion

À travers ce chapitre nous avons présenté l'une des plus importantes étapes de conception d'un projet *BI*, elle représente plus de 60 % de mise en œuvre. Au début nous avons identifié les dimensions après avoir faire plusieurs entretiens, l'objectif de chaque dimension, les attributs de chaque dimension et la description détaillée de chaque mesure avec sa formule, puis nous avons présenté quelque diagrammes *UML* tels que les diagrammes de cas d'utilisation, diagramme d'état transition. Enfin nous avons montré la modélisation multidimensionnelle de l'entrepôt de donnée qui est le composant essentiel de l'aspect technique de ce projet.

Dans le prochain chapitre on bascule vers l'implémentation technique et la réalisation de la solution retenue.

**Chapitre 4 : Mise en œuvre du
système de *reporting* commercial de
SONELGAZ**

1. Introduction

Nous présentons, dans cette section, les détails importants sur la réalisation et la mise en œuvre du système *BI* de la SONELGAZ, en commençant par décrire les environnements et les outils de développement, les différentes technologies adoptées. Nous présenterons, par la suite, le processus *ETL* mis en œuvre pour l'alimentation de l'entrepôt de données. Nous terminerons ce chapitre par la création des cubes de données dimensionnels ainsi que les tableaux de bord afin de permettre aux décideurs une meilleure exploitation des données contenues dans l'entrepôt de données.

2. Environnement et outils de développement

Les conditions du développement sont assez importantes et ont un grand impact sur la qualité du travail. La réalisation de ce projet a été basée sur les outils suivants :

- **Outils de modélisation**

Pour la modélisation nous avons utilisé l'outil professionnel *Visual Paradigm for UML* de la version 15.2, est un logiciel permettant aux programmeurs de modéliser leurs objets avec des diagrammes *UML*.

Parmi les avantages offerts par *Visual Paradigm* :



- Il offre de nombreux outils pour créer différents types de schémas (diagramme de cas d'utilisation, modèle en étoile, modèle en constellation ... etc.).

- Un bon nombre de navigateurs permettant de personnaliser chaque élément.

- *Visual Paradigm* permet de générer la base de données à partir du modèle créé ou des codes sources en divers langages comme *Java* ou *c++*.

- Capture avec la souris, *Visual Paradigm* permet d'utiliser la souris pour manipuler facilement le diagramme à créer.

- **Langage de modélisation : "UML2.0"**

C'est le standard en matière de langage de modélisation des systèmes informatiques.

- **Système de Gestion des Bases de Données (SGBD)**

PostgreSQL est le type de *SGBD* utilisé vu que les sources de données opérationnelles de la SONELGAZ sont hébergées sur cet *SGBD*. L'entrepôt de données a été, lui aussi, implémenté sur *PostgreSQL*.



PostgreSQL est un système de gestion de base de données, libre *open source* capable de prendre en charge en toute sécurité les charges de travail de données les plus complexes [25].

Pour le haut niveau d'intégrité et de personnalisation des données nous nous sommes réfugiés à *PostgreSQL* la version 14.2.

- **Outils d'administration de base de données**

pgAdmin est un outil de gestion des bases de données relationnelles dérivées telles que le serveur avancé *EDB (Enterprise DataBase)*, et La plateforme



d'administration et de développement *open source* la plus populaire et la plus riche en fonctionnalités pour *PostgreSQL* [26].

- **Outil de développement de processus d'intégration et de rafraichissement d'entrepôt de données (ETL)**

Pour le rafraichissement de l'entrepôt de donnée, nous avons conçu et mis en œuvre un certain nombre de *Job* synchronisés grâce à la plateforme *Talend Open*



Studio (TOS) for Data Integration de la version 17.1.1 considérée comme référence dans ce domaine. C'est un logiciel *open source* dédié au développement de processus *ETL*, basé sur le langage *Java* et développé par la société *Talend*.

Cet outil d'intégration de données permet de répondre à toutes les problématiques liées au traitement des données de manière générale.

Parmi les points fort de *Talend* :

- Documentation et communauté très présente.

- Permet aux développeurs de créer ses propres fonctionnalités répondant plus précisément à ses besoins et à les intégrer à la plateforme.

- **Outil de visualisation des données**

Pour la visualisation des données nous avons recourue à l’outil *KNOWAGE* c’est le nouveau nom de la suite open source pour *Business Analytics*, *SPAGOBI*. Nous avons utilisé la version 6.4 qui a été lancer le 19 mai 2019 [27].

C’est une solution de *Business Intelligence* qui permet de créer des cubes de données et des tableaux de bord d’entreprise simples et modernes.

Parmi les points forts de *KNOWAGE* :



- Licence standard gratuite et complète.
- Offre aux développeurs la possibilité de créer leurs propres éléments d’analyse.

- Offre la possibilité de créer les cubes multidimensionnels et manipuler les différents opérateurs algébriques : *SLICE – DICE – ROLL UP – DRILL DOWN*.

3. Transformation et Intégration des données

Pour la transformation des données et leurs intégrations nous avons besoin de : sources de données (*BD*, fichier Excel ...), et base de données cible, qui est notre *Data Warehouse*.

3.1 Développement sous la plateforme *Talend*

Un *Job* est une conception graphique d’un ou plusieurs composant(s) connecté(s) entre eux. Permet de mettre en place le flux de données, traduire les besoins métier en codes, puis se charge d’exécuter ces derniers [28].

Nos Jobs utilisent les composants cités dans le tableau 15, parmi une grande liste des composants.



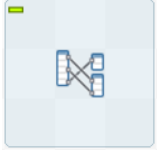





Composant	Nom	Utilisation
	<i>tDBInput</i>	Chargé pour spécifier une table en entrée (source).
	<i>tDBOutput</i>	Pour la sortie de flux (la <i>BD</i> cible donc notre <i>Data Warehouse</i>).
	<i>tMap</i>	Ce composant transforme et route des données à partir d'une ou plusieurs source(s) de données vers une ou plusieurs destination(s).
	<i>tLogRow</i>	Ce composant affiche les données ou les résultats dans la console de la vue « Exécution » afin de monitorer les données traitées.
	<i>tAggregateRow</i>	Ce composant reçoit un flux de données et fait une agrégation basée sur une ou plusieurs colonnes.
	<i>tWarn</i>	Ce composant déclenche un avertissement, En cas d'erreur. Et il n'arrête pas l'exécution du Job.
	<i>tLogCatcher</i>	Capteur des champs et des messages des fonctions <i>Java Exception</i> et des composants <i>tWarn</i> et les transmet vers la sortie définie.
	<i>tFileInputExcel</i>	Ce composant lit un fichier Excel ligne par ligne pour le diviser en champs et envoie les champs comme défini dans le schéma au composant suivant.

Tableau 15. Liste des composants *TOS*.

3.2 Aperçu sur nos *Job*

- La liste de *Job* que nous avons besoins pour l'alimentation de notre *Data Warehouse*, où l'exécution de chaque *Job* alimente une table de fait ou une table de dimension, comme illustre la figure 26.

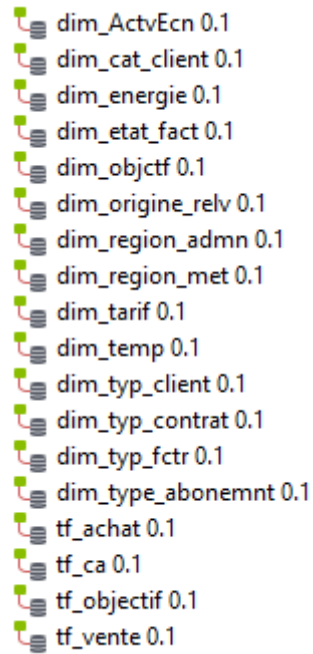


Figure 26. Liste des *Job*.

- Nous avons utilisé comme source de données des bases de données, et un fichier Excel dans le cas des objectifs :

1. Exemple de *Job* d'une dimension qui a une table de la base de données comme *tDBInput*. La dimension « *type facture* » dans la figure 27.

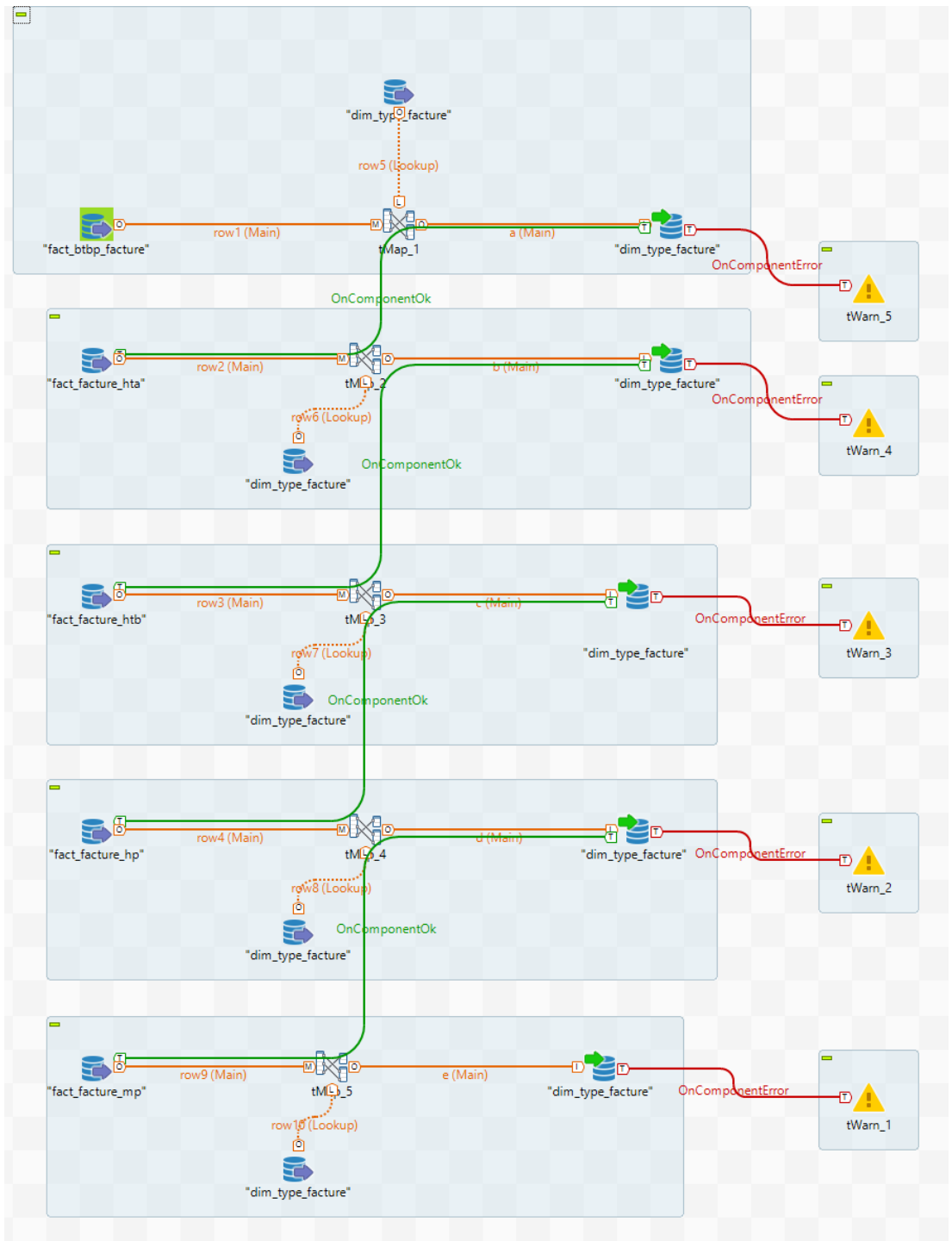


Figure 27. Le Job « dim_type_facture ».

L'exécution du *Job* exécute les requêtes SQL spécifiées dans chaque *tDBInput*. Chaque carré bleu représente un sous-*Job*, et la liaison entre eux est « *OnComponentOk* » qui élabore un lien vers le prochain sous-*Job* lorsque le composant réussit, pour que ce dernier commence à s'exécuter.

Remarquons que dans chaque sous-*Job*, la table cible du *DW* appartient deux fois, une fois comme *tDBInput* et une autre comme *tDBOutput*. Pour une question de performance nous avons utilisé un concept nommé *CDC* (*Change Data Capture*), pour l'identification des modifications apportées aux sources de données.

La figure 28 montre comment le *tMap* nous permis de fournir ces contrôles, en prenant comme exemple *tMap_1* du *Job* actuel (voir figure 27).

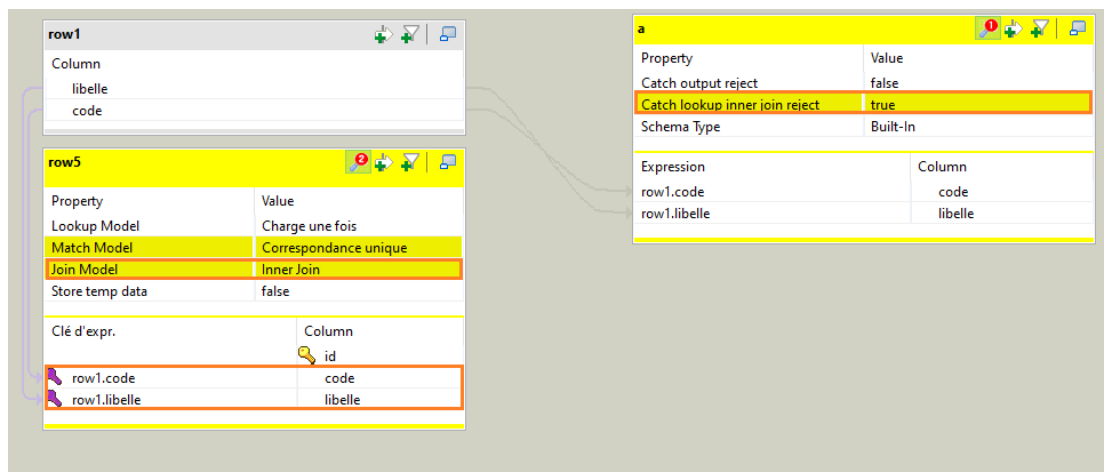


Figure 28. *tMap_1* du *Job* « *dim_type_facture* ».

- Traduisant la partie en rouge en une requête *SQL* donne « *Inner Join row1 on row1.code=code and row1.libelle = libelle* ».

Si on exécute le *Job* (figure 27) une autre fois, sans établir aucune modification, aucune nouvelle ligne ne sera insérée/Mise à jour dans la *BD* cible, comme illustre la figure 29.

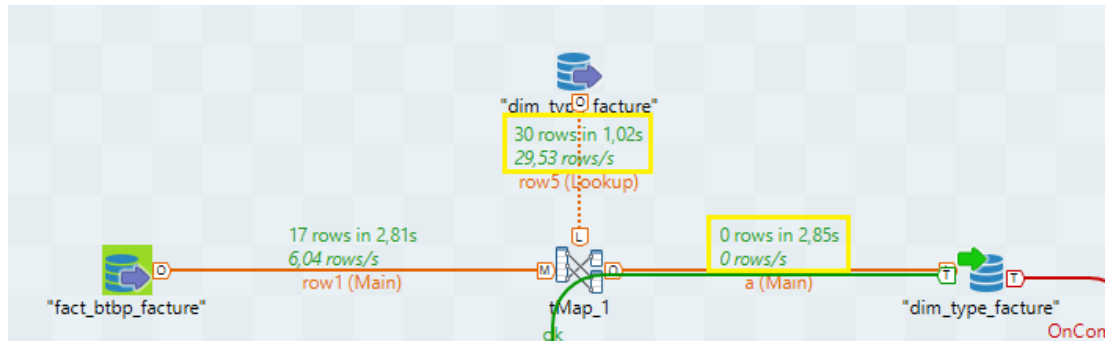


Figure 29. Résultat de l'exécution avec un *CDC*.

2. Exemple de *Job* d'une dimension qui a un fichier Excel comme *tDBInput*. La dimension « *type objectifs* » dans la figure 30.

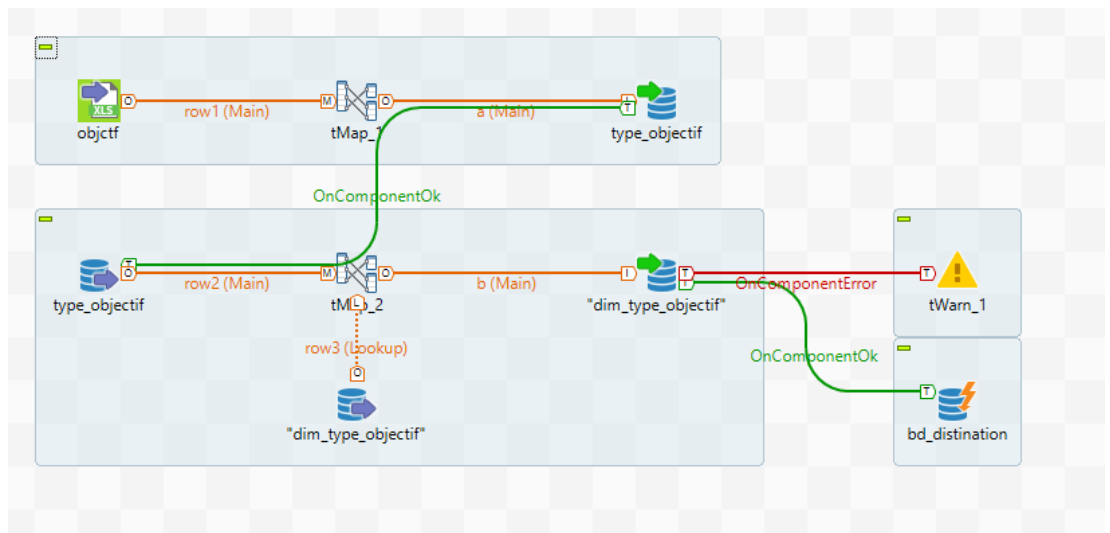


Figure 30. *Job* « *dim_objctf* ».

« *type_objectif* » est une table temporaire, où on va stocker tous les données dont on a besoins à partir du fichier « *objctf* », afin de les manipuler plus facilement avec une requête *SQL*. Dans ce qui suit cette table sera notre *tDBInput*.

3. Exemple de *Job* d'une table de fait avec une *BD* comme *tDBInput* :

La table de fait « *ventes* » est un exemple de récupération des données de deux bases de données différentes, où *tDBInput* de chaque sous-*Job* appartient à une autre base de données (voir figure 31).

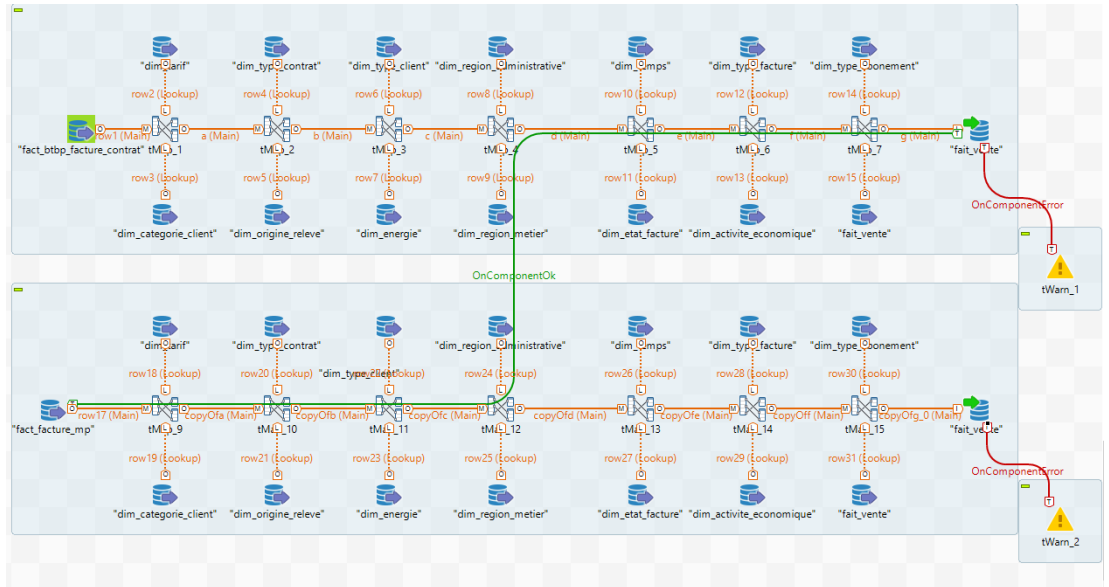


Figure 31. *Job* « *tf_vente* ».

L'idée dans les *Job* des tables de fait est de récupérer l'identifiant de chaque donnée retourner par la requête exécutée par le *tDBInput*, à partir de la dimension concernée.

Un double-clic sur *tMap_1* donne la figure 32.

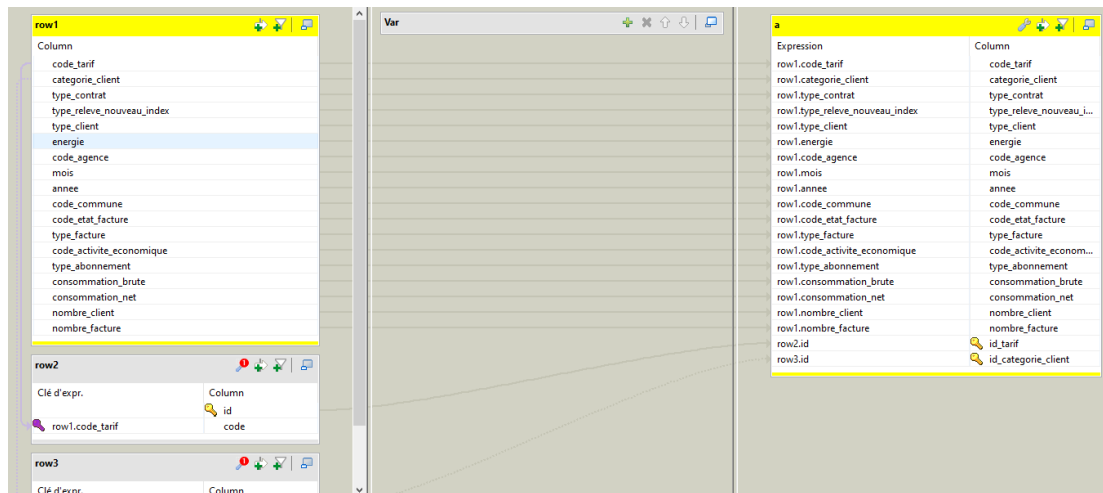


Figure 32. *tMap_1*.

La liaison entre « row1 » et « row2 » consiste de récupérer l'« id » qui représente chaque « code_tarif » retourner par la requête, en faisant cela avec toutes les dimensions concernées, on obtient chaque identifiant dont on a besoin, alors le dernier *tMap* n'a que transmettre les mesures calculées par la requête et les identifiants récupérés précédemment, à la table de fait cible.

4. Exemple de *Job* d'une table de fait avec un fichier Excel comme *tDBInput*. La table de fait « *tf_objectif* » dans la figure 33.

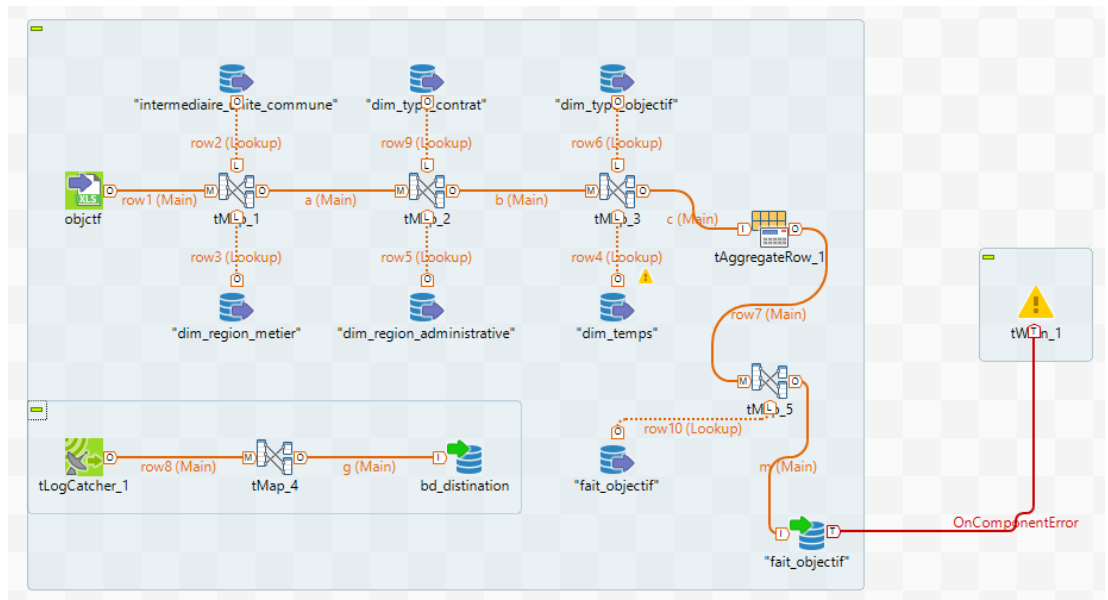


Figure 33. *Job* « *tf_objectif* ».

Pour l'agrégation (le calcul des mesures) nous avons utilisé « *tAggregateRow* ». Un double-clic sur « *tAggregateRow_1* » donne la figure 34.

Group by	Colonne de sortie	Position de la colonne d'entrée
	dim_tempsid	id_temps
	dim_region_administrativeid	id_region_admn
	dim_region_metierid	id_region_met
	dim_type_contratid	id_type_contrat
	dim_type_objectifid	id_type_obj

Opérations	Colonne de sortie	Fonction	Position de la colonne d'entrée	<input type="checkbox"/> Ignorer les valeurs null
	objectif	somme	valeur	<input type="checkbox"/>

Figure 34. « *tAggregateRow_1* ».

Dans la section « Opération » on opte pour la colonne en entrée et celle en sortie, et la fonction d'agrégation.

Ensuite, on choisit les colonnes en entrée et en sortie pour le « *group by* » dans la section du même nom.

Le reste des opérations se fait de la même façon que dans le cas d'une table d'une *BD* comme *tDBInput*.

5. L'utilisation de « *tLogRow* » :

Ce composant a été utilisé lors de la construction de nos Job. Pour avoir une idée sur les données qui circulent dans notre *Job*, afin de régler les exceptions et les erreurs.

- La figure 35 illustre la vérification de la conformité des données en sortie de « *tAggregateRow_1* » avant de les charger dans la table cible.

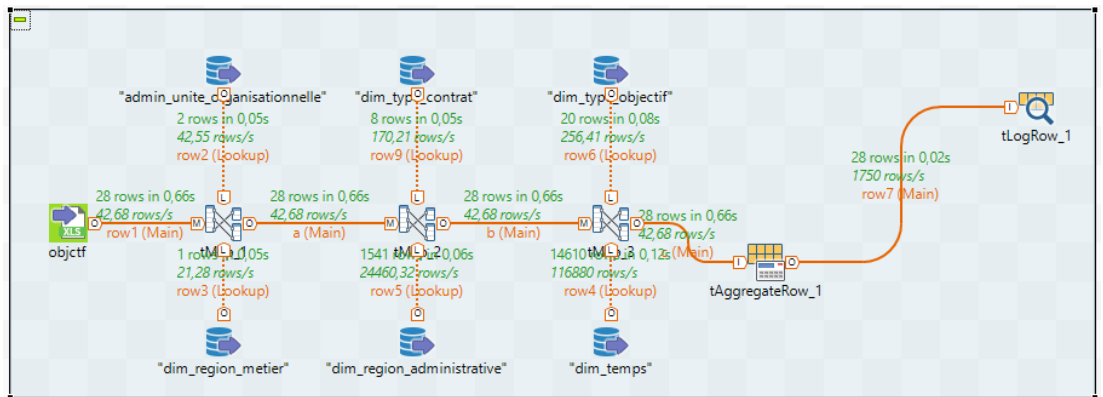


Figure 35. Lors de la construction/vérification du Job « *tf_objectif* ».

L'exécution du *Job* de la figure 35 donne la figure 36.

Exécution

Exécuter Arrêter Effacer

Démarrage du job *tf_objectif* a 17:49 26/05/2023.
 [statistics] connecting to socket on port 4056
 [statistics] connected

tLogRow_1					
dim_tempsid	dim_region_administrativeid	dim_region_metierid	dim_type_contratid	dim_type_objectifid	objectif
4749	1428	1	5	14	200000.0
4749	1428	1	5	18	660000.0
4749	1428	1	5	19	150000.0
4749	1428	1	7	18	150000.0
4749	1428	1	4	2	60.0
4749	1428	1	7	19	650000.0
4749	1428	1	7	17	460000.0
4749	1428	1	5	20	190000.0
4749	1428	1	6	8	600000.0
4749	1428	1	6	12	470000.0
4749	1428	1	2	19	900000.0
4749	1428	1	3	4	360000.0
4749	1428	1	2	18	250000.0
4749	1428	1	6	15	450000.0
4749	1428	1	2	17	710000.0
4749	1428	1	3	6	570000.0
4749	1428	1	3	19	210000.0
4749	1428	1	3	18	390000.0
4749	1428	1	3	21	150000.0
4749	1428	1	3	20	600000.0
4749	1428	1	3	17	790000.0
4749	1428	1	0	17	566000.0
4749	1428	1	1	8	970000.0
4749	1428	1	4	19	800000.0

Nombre limite de lignes 10000 Retour automatique à la ligne

Figure 36. La console après exécution avec *tLogRow*.

3.3 Planification des *Jobs ETL*

Après avoir développé les différents *Jobs ETL* sur *Talend* et dans le but de remettre en état, périodiquement, le chargement des données. Les *Jobs* concernés doivent être exportés en scripts exécutables (.sh sur *Linux*, .bat sur *Windows*).

Afin de simplifier cette phase nous avons résumé l'exécution de ces scripts dans un *Job* globale (figure 37).

Pour garantir la synchronisation des mises à jour, ce *Job* assure la bonne exécution de toutes les *Jobs* des dimensions avant de passer aux tables de fait.

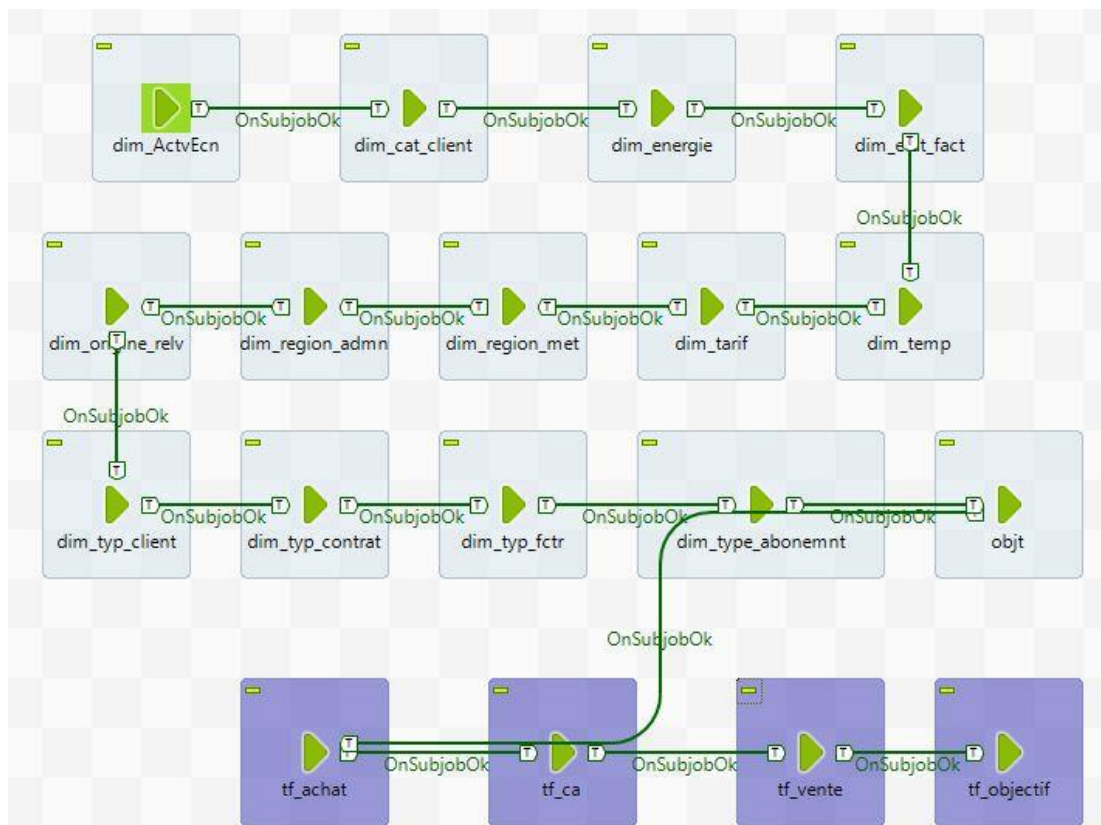


Figure 37. *Job* globale.

4. Architecture du système de reporting

Sur le plan pratique et technique, la *BI* se compose d'une famille d'éléments et d'outils informatiques structurés en quatre catégories correspondantes chacune à une fonction spécifique et à une phase du processus.

Nous présenterons, dans la figure 38, l'architecture globale du système décisionnel et de *reporting* :

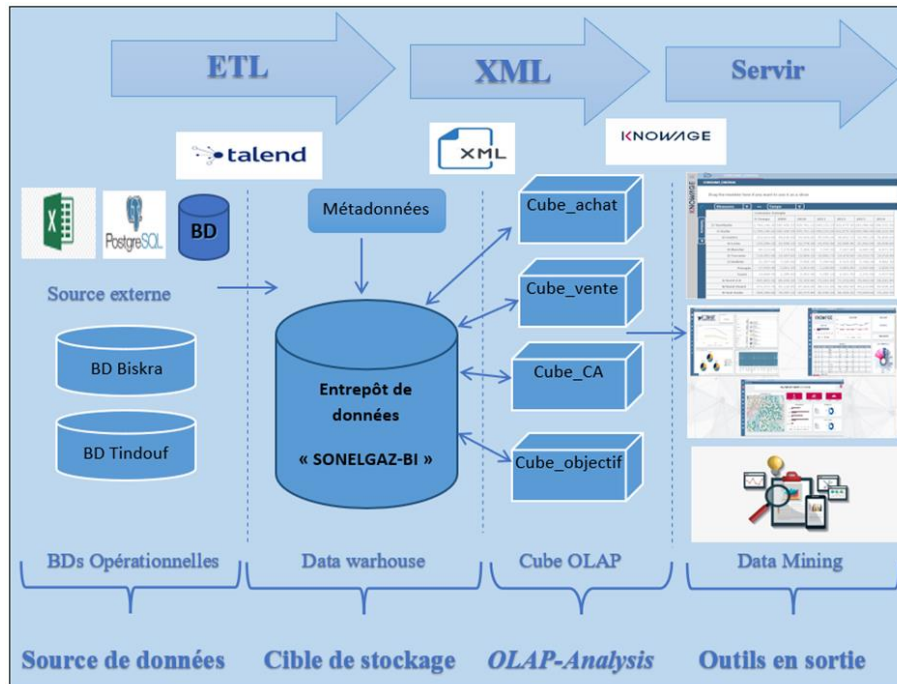


Figure 38. Architecture globale du système *reporting* décisionnel.

- **Source de données**

La première phase consiste à collecter des données importantes dans les systèmes de SONELGAZ, particulièrement, les *BD* opérationnelles. Dans notre projet, nous nous intéressons aux bases de données des wilayas de BISKRA et TINDOUF, à restaurer dans le *SGBD PostgreSQL*. La deuxième source de données retenue dans le cadre de ce projet, est le classeur Excel contenant des données formalisant les objectifs de SONELGAZ.

- **BD cible dédiée pour le stockage de l'entrepôt de données**

Après l'extraction et la transformation des données, l'*ETL* permet de gérer les flux de données en alimentant la base de stockage qui n'est autre que l'entrepôt de données « SONELGAZ-BI ». Afin de stocker les tables de dimensions et les tables de fait relatives à notre conception.

- ***OLAP-Analysis***

Les cubes multidimensionnels sont générés via des fichiers *XML*, selon une structure spécifique, qui recevront des données du *DW*, et qui seront exploités lors des requêtes d'analyse multidimensionnelles.

- **Outils en sortie**

Une fois les données collectées, nettoyées, stockées, et accessibles, elles peuvent être, enfin, analysées pour faire de la prédiction et prendre les bonnes décisions. La plateforme *KNOWAGE* nous a permis de présenter et de visualiser les données à travers des cubes multidimensionnels et des tableaux de bord avec les indicateurs clés de performance (*KPI : Key Performance Indicator*).

5. Description des différents modules du système

5.1 Les cubes multidimensionnels

Après avoir parfaire la phase d'entreposage et chargement des données, nous avons procédé à la création des axes préalablement identifiés lors de la phase d'identification des besoins aux cubes multidimensionnels.

D'abord, on commence par sélectionner les dimensions et identifier leurs hiérarchies, puis spécifier leurs mesures, ce qui donne un cube au format *XML* dans le but de mapper les structures *OLAP*.

La figure 39 décrit le cube « chiffres d'affaires » en format *XML*. Nous remarquons qu'on a spécifié le nom du schéma dont notre *DW* est inclus, dans cette balise même :

- Chaque dimension est représentée comme une balise. La spécification de la table et les colonnes souhaités à présenter dans notre cube se fait sous les balises des niveaux hiérarchiques, en classant ces derniers d'une façon descendante (du niveau le plus grand au plus petit).
- Pour la table fait, à l'intérieur de la balise *<cube>* on peut distinguer en plus de la balise du nom de la table, deux catégories de balises : *<DimensionUsage>* pour la spécification de la table source de chaque clé étrangère de cette table, et *<measure>* pour la description des mesures (nom

de cette colonne dans la table de fait, l'agrégation et le format d'affichage souhaité).

```
<?xml version="1.0"?>
<Schema name="public">
  <Dimension name="dim_temps">
    <Hierarchy name="dim_temps" hasAll="true" allMemberName="Temps" primaryKey="id">
      <Table name="dim_temps"/>
      <Level name="level1" column="annee" uniqueMembers="false"/>
      <Level name="level2" column="libelle_trimestre" uniqueMembers="false" ordinalColumn="id"/>
      <Level name="level3" column="libelle_mois" uniqueMembers="false" ordinalColumn="id"/>
    </Hierarchy>
  </Dimension>
  <Dimension name="dim_categorie_client">
    <Hierarchy name="dim_categorie_client" hasAll="true" allMemberName="All" primaryKey="id">
      <Table name="dim_categorie_client"/>
      <Level name="level1" column="libelle" uniqueMembers="true"/>
    </Hierarchy>
  </Dimension>
  <Dimension name="dim_type_client">
    <Hierarchy name="dim_type_client" hasAll="true" allMemberName="All" primaryKey="id">
      <Table name="dim_type_client"/>
      <Level name="level1" column="libelle" uniqueMembers="true"/>
    </Hierarchy>
  </Dimension>
  <Dimension name="dim_origine_releve">
    <Hierarchy name="dim_origine_releve" hasAll="true" allMemberName="All" primaryKey="id">
      <Table name="dim_origine_releve"/>
      <Level name="level1" column="libelle" uniqueMembers="false"/>
    </Hierarchy>
  </Dimension>
  <Dimension name="dim_etat_facture">
    <Hierarchy name="dim_etat_facture" hasAll="true" allMemberName="All" primaryKey="id">
      <Table name="dim_etat_facture"/>
      <Level name="level1" column="libelle" uniqueMembers="true"/>
    </Hierarchy>
  </Dimension>
  <!--Le reste des dimensions-->
  <Cube name="chiffre_affaire">
    <Table name="fait_chiffre_affaire"/>
    <DimensionUsage name="dim_temps" source="dim_temps" foreignKey="dim_tempsid"/>
    <DimensionUsage name="dim_categorie_client" source="dim_categorie_client" foreignKey="dim_categorie_clientid"/>
    <DimensionUsage name="dim_type_client" source="dim_type_client" foreignKey="dim_categorie_clientid"/>
    <DimensionUsage name="dim_origine_releve" source="dim_origine_releve" foreignKey="dim_origine_releveid"/>
    <DimensionUsage name="dim_etat_facture" source="dim_etat_facture" foreignKey="dim_etat_factureid"/>
    <DimensionUsage name="dim_type_facture" source="dim_type_facture" foreignKey="dim_type_factureid"/>
    <DimensionUsage name="dim_activite_economique" source="dim_activite_economique" foreignKey="dim_activite_economiqueid"/>
    <DimensionUsage name="dim_tarif" source="dim_tarif" foreignKey="dim_tarifid"/>
    <DimensionUsage name="dim_type_contrat" source="dim_type_contrat" foreignKey="dim_type_contratid"/>
    <DimensionUsage name="dim_region_metier" source="dim_region_metier" foreignKey="dim_region_metierid"/>
    <DimensionUsage name="dim_region_administrative" source="dim_region_administrative" foreignKey="dim_region_administrativeid"/>
    <Measure name="chiffre_affaire_energie" column="chiffre_affaire_energie" aggregator="sum" formatString="" />
    <Measure name="montant_ttc" column="montant_ttc" aggregator="sum" formatString="" />
    <Measure name="chiffre_affaire_travaux" column="chiffre_affaire_travaux" aggregator="sum" formatString="" />
    <Measure name="montant_tva" column="montant_tva" aggregator="sum" formatString="" />
    <Measure name="montant_tpe" column="montant_tpe" aggregator="sum" formatString="" />
    <Measure name="taxe_habitation" column="taxe_habitation" aggregator="sum" formatString="" />
    <Measure name="prime_fixe" column="prime_fixe" aggregator="sum" formatString="" />
    <Measure name="droit_fixe" column="droit_fixe" aggregator="sum" formatString="" />
  </Cube>
</Schema>
```

Figure 39. Cube « chiffres d'affaires » en format XML.

En utilisant les spécifications des cubes *OLAP* en format *XML*, le serveur *OLAP* Mondrian, de l'outil *KNOWAGE*, construit des requêtes écrites en langage *MDX* pour la création des cubes multidimensionnels, ce qui facilite la navigation dans ces derniers, directement sur les tables et les colonnes de la base de données relationnelle.

À partir de la visualisation du cube « chiffres d'affaires » sous *KNOWAGE* (voir figure 40). Nous pouvons distinguer : un sous-cube (du cube « chiffres d'affaires »), créé à partir des mesures représentées comme colonnes dans la figure 40, des dimensions représentées comme lignes dans la même figure, ainsi que le volet permettant de naviguer dans l'ensemble des données du cube.

			chiffre_affaire_energie	chiffre_affaire_travaux
⊖ Temps	⊖ All	⊖ All	1,052,603,507.09	19,047,959.44
		Morale	84,731.52	0
		Physique	76,599,503.6	34,188.04
	⊖ Basse pression	⊖ All	168,659,002.76	9,517,313.06
		Morale	0	0
		Physique	1,357,165.62	17,094.02
	Gaz	⊖ All	168,659,002.76	9,517,313.06
		Morale	0	0
		Physique	1,357,165.62	17,094.02
	⊖ Basse tension	⊖ All	702,972,434.45	9,517,313.06
		Morale	84,731.52	0
		Physique	18,171,512.78	17,094.02
	Electricite	⊖ All	702,972,434.45	9,517,313.06
		Morale	84,731.52	0
		Physique	18,171,512.78	17,094.02

Figure 40. Visualisation du cube « chiffres d'affaires ».

5.2 Tableau de bord stratégique (*Cockpit*)

Un tableau de bord stratégique et un outil d'analyse *OLAP*, permet aux utilisateurs finaux d'exploiter les données.

Comme nous pouvons le constater dans les figures 41 et 44, un tableau de bord comprend un ou plusieurs :

- Tableaux d'affichage des données, figure 41 (a).
- Représentation graphique qui synthétise et illustre les résultats des tableaux d'une manière facilement interprétable, voir figure 41 (b).
- « *selector* » pour naviguer entre les catégories des données constituant le centre d'intérêt de l'utilisateur, voir figure 41 (c).

- « KPI_CA » est un indicateur clé de performance permettant de mesurer la performance des services commerciaux, correspond à la mesure « chiffres d'affaires » dans l'exemple figure 41 (d).

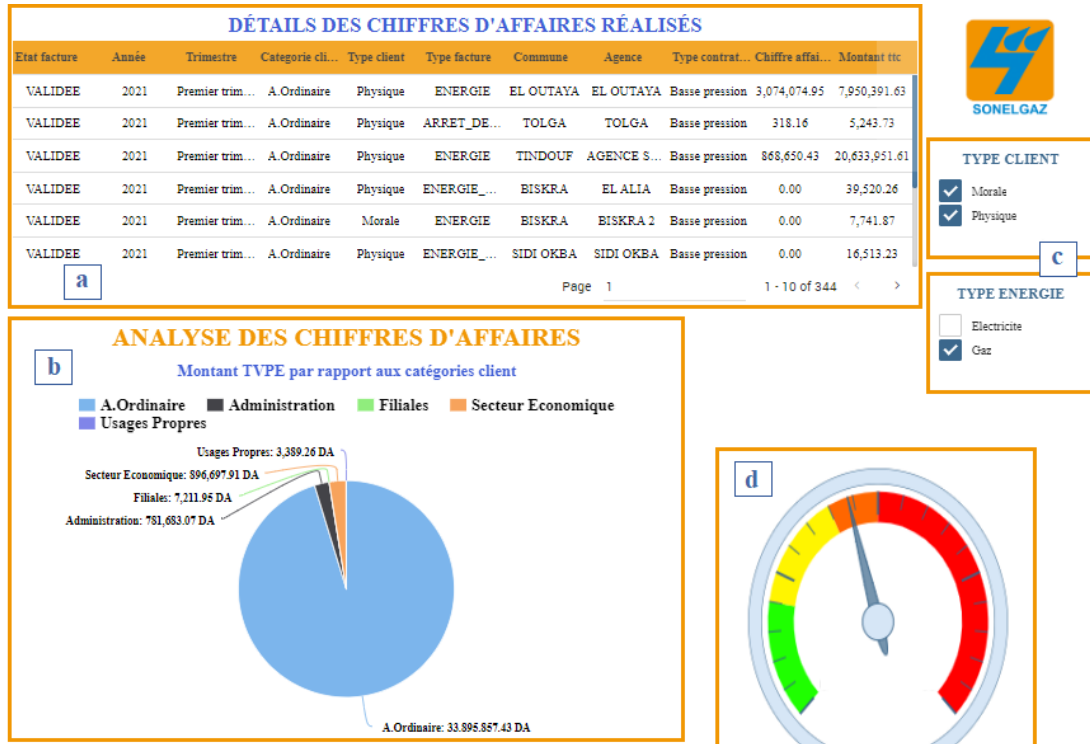


Figure 41. Tableau de bord « chiffres d'affaires ».

6. Tests et Résultats

6.1 Les cubes multidimensionnels

La figure 42 (a) montre la connexion de notre base de données « SONELGAZ_BI » avec l'interface de l'administrateur. Alors que la figure 42 (b) représente le script SQL pour l'ensemble de données de la source de données « achats ». La figure 42 (c) montre un exemple de cube « achats » en format XML puis le téléchargement de ce dernier sur la plateforme *KNOWAGE*, voir figure 42 (d). Enfin, la figure 42 (e) illustre la dernière étape ci celle de la création de cube « achats » sous la plateforme.

a

Label *
achat

Description

Dialect *
PostgreSQL

Read Only
 Read Only
 Read and write
 Use as cache

Type:
 JDBC
 JNDI

URI *
jdbc:postgresql://localhost:5433/SONELGAZ_B|

User
postgres

Password

Driver *
org.postgresql.Driver

1. Connexion avec la base de données

b

DataSet Type *
Query

Data Source *
achat

Query

```

1 select t.libelle_mois mois, t.annee, t.libelle_trimestre trimestre,
2 a.libelle_commune commune, a.libelle_wilaya wilaya,
3 m.libelle_agence agence, m.libelle_rd region_distribution,
4 m.libelle_dd direction_distribution,
5 e.libelle_origine_releve, e.libelle energie,
6 sum(achat_achat) achat_energie,
7 sum(montant_achat) montant_achat
8 from fait_achat
9 join dim_temps t on t.id=dim_tempsid
10 join dim_region_administrative a on a.id=dim_region_administrativeid
11 join dim_region_metier m on m.id=dim_region_metierid
12 join dim_origine_releve r on r.id=dim_origine_releveid
13 join dim_energie e on e.id=dim_energieid
14 group by 1,2,3,4,5,6,7,8,9,10

```

2. Éditation de script pour l'ensemble de données

c

```

<?xml version="1.0"?>
<Schema name="public">
  <Dimension name="dim_energie">
    <Hierarchy name="dim_energie" hasAll="true" allMembers="all" primaryKey="id">
      <Table name="dim_energie"/>
      <Level name="level1" column="libelle" uniqueness="true"/>
    </Hierarchy>
  </Dimension>
  <Dimension name="dim_temps">
    <Hierarchy name="dim_temps" hasAll="true" allMembers="all" primaryKey="id">
      <Table name="dim_temps"/>
      <Level name="level1" column="annee" uniqueness="false"/>
      <Level name="level2" column="libelle_trimestre" uniqueness="false"/>
      <Level name="level3" column="libelle_mois" uniqueness="false"/>
    </Hierarchy>
  </Dimension>
  <Dimension name="dim_region_metier">
    <Hierarchy name="dim_region_metier" hasAll="true" allMembers="all" primaryKey="id">
      <Table name="dim_region_metier"/>
      <Level name="level1" column="libelle_rd" uniqueness="false"/>
      <Level name="level2" column="libelle_dd" uniqueness="false"/>
      <Level name="level3" column="libelle_agence" uniqueness="true"/>
    </Hierarchy>
  </Dimension>
  <Dimension name="dim_region_administrative">
    <Hierarchy name="dim_region_administrative" hasAll="true" allMembers="all" primaryKey="id">
      <Table name="dim_region_administrative"/>
      <Level name="level1" column="libelle_wilaya" uniqueness="false"/>
      <Level name="level2" column="libelle_commune" uniqueness="false"/>
    </Hierarchy>
  </Dimension>
  <Dimension name="dim_origine_releve">
    <Hierarchy name="dim_origine_releve" hasAll="true" allMembers="all" primaryKey="id">
      <Table name="dim_origine_releve"/>
      <Level name="level1" column="libelle" uniqueness="false"/>
    </Hierarchy>
  </Dimension>
  <Cube name="achat">
    <Table name="achat"/>
    <DimensionUsage name="dim_energie" source="dim_energie" foreignKey="dim_energieid"/>
    <DimensionUsage name="dim_region_metier" source="dim_region_metier" foreignKey="dim_region_metierid"/>
    <DimensionUsage name="dim_temps" source="dim_temps" foreignKey="dim_tempsid"/>
    <DimensionUsage name="dim_region_administrative" source="dim_region_administrative" foreignKey="dim_region_administrativeid"/>
    <DimensionUsage name="dim_origine_releve" source="dim_origine_releve" foreignKey="origine_releveid"/>
    <Measure name="achat_energie" column="achat_energie" aggregation="sum" formatString=""/>
    <Measure name="montant_achat" column="montant_achat" aggregation="sum" formatString=""/>
  </Cube>
</Schema>

```

3. Création de fichier XML « achat »

d

Q Search

ACTIVE	FILE NAME	CREATOR	CREATION DATE
<input checked="" type="radio"/>	achat (1).xml	demo_admin	12/06/2022 19:55:34

4. Téléchargement de fichier XML « achat »

e

Label
cube_achat

Name
cube_achat

Description
cube_achat

Type
On-line analytical processing

Engine
OLAP Engine

Data Source
achat

State
Released

Community

Refresh seconds
0

Criptable
 True False

Visible
 True False

Visibility restrictions

Preview file
Choisir un fichier Aucun fichier choisi

Locked by user
 True False

Template
Choisir un fichier Aucun fichier choisi

Template build

5. Création de cube OLAP « achat »

Figure 42. Exemple de création d'un cube multidimensionnels « achats ».

Une illustration graphique d'un sous-cube *OLAP* « achats » avec l'opérateur algébrique *DRIL-DOWN* est donnée dans la figure 43. Nous pouvons remarquer les différentes valeurs des deux mesures *achat_energie* et *montant_achat* par rapport aux dimensions *dim_temps*, *dim_energie*, *dim_region_metier*.

			achat_energie	montant_achat
⊖ Temps	⊖ Energie	⊖ Region metier	5,708,658,617,990.965	19,218,758,839,898.727
		⊖ Region de Distribution centre	5,708,241,807,914.805	19,217,723,052,148.582
		⊖ Biskra	5,708,241,807,914.805	19,217,723,052,148.582
		⊖ Region de Distribution oest	416,810,076.16	1,035,787,750.144
		⊖ DD TINDOUF	416,810,076.16	1,035,787,750.144
		AGENCE SONELGAZ TDF	416,810,076.16	1,035,787,750.144
	Electricite	⊖ Region metier	79,450,192,196	79,450,192,196
		⊖ Region de Distribution centre	79,291,289,484	79,291,289,484
		⊖ Biskra	79,291,289,484	79,291,289,484
		⊖ Region de Distribution oest	158,902,712	158,902,712
		⊖ DD TINDOUF	158,902,712	158,902,712
		AGENCE SONELGAZ TDF	158,902,712	158,902,712
	Gaz	⊖ Region metier	5,629,208,425,794.967	19,139,308,647,702.727
		⊖ Region de Distribution centre	5,628,950,518,430.807	19,138,431,762,664.582
		⊖ Biskra	5,628,950,518,430.807	19,138,431,762,664.582
		⊖ Region de Distribution oest	257,907,364.16	876,885,038.144
		⊖ DD TINDOUF	257,907,364.16	876,885,038.144

Figure 43. Visualisation d'un sous-cube « achats ».

6.2 Les cockpits

La figure 44 (a) Montre les différents paramètres pour la création d'un *cockpit* « achats », Quant à la figure 44 (b) illustre un *cockpit* de visualisation des données « achats » tel que *achats_energie* et *montant_achat* par (*années, trimestre, wilaya, agence, direction de distribution, etc.*) Enfin la figure 44 (c) montre l'export des résultats de *cockpit* en format Excel et PDF.

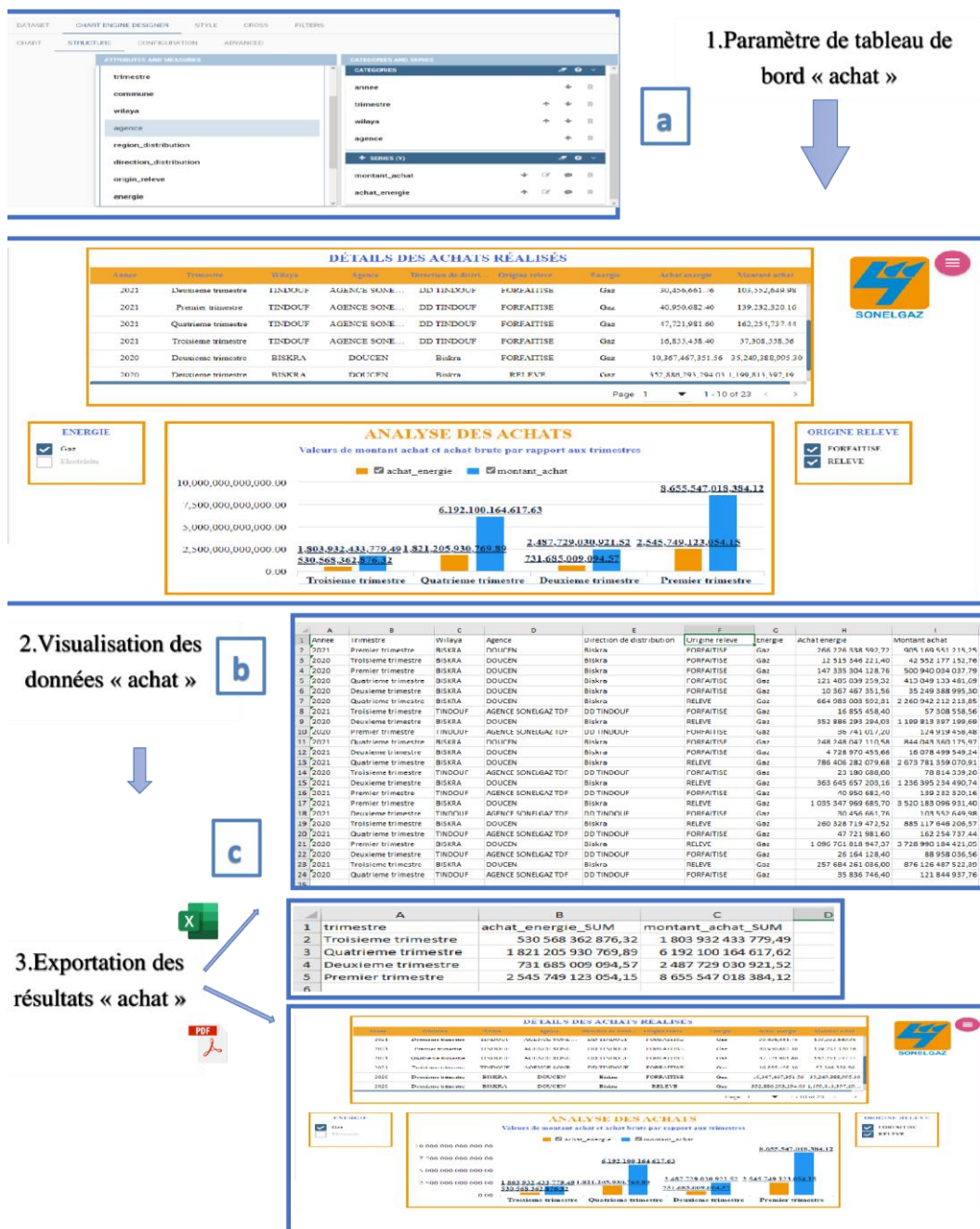


Figure 44. Différentes opérations sur les cockpits.

7. Conclusion

Ce chapitre est dédié pour la réalisation et le déploiement de notre système décisionnel.

Nous avons cité les différents environnements et les outils de développement

que nous avons adopté pour l'implémentation de ce projet à savoir *Visual paradigm*, *Talend*, *PostgreSQL*, *KNOWAGE*, etc.

Ensuite nous avons présenté les détails d'implémentation tel que la transformation et l'intégration des données, la planification des *Jobs ETL* mené grâce au logiciel *Talend*.

Après nous avons montré l'architecture détaillée de notre système *reporting* avec l'explication de chaque phase. Comme nous avons donné un exemple sur les étapes d'exécution des cubes multidimensionnels et les tableaux de bord sur la plateforme *KNOWAGE*.

Enfin nous avons présenté les cubes multidimensionnels ainsi que les différents tableaux de bord liés aux activités des achats, ventes, chiffres d'affaires, objectifs.

Conclusion Générale

Afin d'offrir un système de *reporting* décisionnel fiable et efficace, pour surmonter les problèmes récurrents dans le processus décisionnel, au niveau de la SONELGAZ, la filiale *ELIT* a lancé le projet intitulé « Mise en œuvre d'une solution *Business Intelligence* basée sur un entrepôt de données », pour la mise en place d'un système qui permet aux responsables et décideurs de l'entreprise, d'analyser soigneusement les données commerciales, pour avoir l'information voulue au moment voulu.

Notre travail a commencé par l'étude bibliographique sur les systèmes décisionnels et la collecte progressive des connaissances théoriques et pratiques dans le domaine de la *Business Intelligence*.

La deuxième étape de notre démarche consiste à interroger les décideurs pour recueillir les besoins. Ensuite, nous nous sommes lancés dans l'aspect métier afin d'identifier les différents besoins analytiques et techniques.

Une fois les besoins fixés et validés par l'équipe du projet, nous avons entamé la phase de la conception et la modélisation multidimensionnelle de l'entrepôt de données dans laquelle nous avons proposé des modèles capables de répondre aux besoins exprimés par la SONELGAZ.

Nous avons abordé, ensuite, la phase d'alimentation de l'entrepôt, où nous avons défini le schéma du processus de transformation et d'intégration des données et une démarche de réalisation de ce processus. L'entrepôt de données étant créé et son alimentation assurée grâce au processus *ETL*, nous avons enchaîné par la conception de la zone de restitution correspondant à la création des cubes de données.

Enfin nous sommes passé à l'étape de visualisation des données à travers la plateforme *KNOWAGE*, et le partage des résultats (tableaux de bord, rapports, cubes *OLAP*) afin d'accélérer et maximiser la performance des analyses.

Nous n'ignorons pas les difficultés que nous avons pu surmonter lors de ce parcours, nous avons traversé des moments difficiles et nous avons fait face aux nombreux défis que nous résumons de la manière suivante :

- Puisque la mise en œuvre de tout projet de *reporting* nécessite une parfaite compréhension du métier de l'entreprise, nous avons rencontré des difficultés

à comprendre l'aspect métier et les différentes activités de SONELGAZ, convenable pour la modélisation selon les besoins des utilisateurs.

- La plateforme *KNOWAGE* et une version récente de *SPAGOBI*, dont la plupart des formations sont payantes. En effet, la plupart des ressources ne sont pas accessibles gratuitement sur Internet.
- La découverte de la plateforme *KNOWAGE* et la manipulation des différentes fonctionnalités dans un temps record.
- Les bugs et les erreurs lors du développement de l'*ETL* sous *Talend*, lors de la création des cubes de données et des tableaux de bord sous *KNOWAGE* ont constitué une réelle contrainte en plus de la contrainte de temps que nous avons vécue déjà le long du projet.

Pour finir, et avant de citer les perspectives du projet, nous pouvons dire que ce stage au niveau de SONELGAZ était à la hauteur de nos attentes. En effet, il nous a offert l'occasion de nous intégrer dans un environnement professionnel afin de nous initier aux exigences et aux attentes de cet environnement. Il nous a permis, également, d'acquérir une très bonne expérience et d'avancer dans un domaine qui nous était totalement inconnu à savoir la *Business Intelligence*.

Comme tout projet, un projet dans le cadre de la *Business Intelligence* n'est jamais complètement fini, il mérite toujours d'être amélioré et enrichi par des idées nouvelles et des fonctionnalités avancées et d'actualité. Nous pouvons citer à titre d'exemples, les perspectives et les développements qu'on souhaite réaliser pour une éventuelle continuité de ce travail.

- Suivre le déploiement actuel et recueillir les correctifs et les nouvelles remarques des utilisateurs.
- Utilisation des méthodes et algorithmes de *machine learning* et du *data mining* pour une meilleure exploitation et analyse plus intelligente des données en vue d'améliorer de manière significative la prise de décisions.

- Migration de l'entreposage de données vers un environnement *Big Data* (utilisation de solution *NoSQL*) pour prendre en charge les données non structurées dans le processus décisionnel et garantir un accès plus fiable aux données.

Références

- [1] «Historique,» [En ligne]. Available: <https://www.sonelgaz.dz/fr/category/historique>. [Accès le 10 Janvier 2022].
- [2] «Qui sommes nous ?,» [En ligne]. Available: <https://www.sonelgaz.dz/fr/category/qui-sommes-nous>. [Accès le 10 Janvier 2022].
- [3] Assistante du PDG, «Dossier presse : Rapport d'activités et comptes de gestion consolidés du Groupe Sonelgaz,» Conférence de presse du 21 juin, 2010.
- [4] R. Stackowiak, J. Rayman et R. Greenwald, Oracle Data Warehousing and Business Intelligence Solutions, Indianapolis: Wiley Publishing, Inc., 2007, pp. 4-6.
- [5] W. ECKERSON, Performance Dashboards Measuring, Monitoring, and Managing Your Business, 2e éd., New Jersey: John Wiley & Sons, Inc., 2011, pp. 31-32.
- [6] S. Scheps, Business Intelligence FOR DUMMIES, Indiana: Wiley Publishing, Inc., 2008, pp. 9-11, 68-70.
- [7] A. Vaisman et . E. Zimanyi, Data Warehouse Systems, Design and Implementation, Berlin: Springer, 2014, pp. 59, 63-67, 78-79, 270-271.
- [8] J. Celko, Joe Celko's Analytics and OLAP in SQL, San Francisco: Morgan Kaufmann Publishers, 2006, pp. xix-xx.
- [9] C. Imhoff, N. Galemme et J. Geig, Mastering Data Warehouse Design : Relational and Dimensional Techniques, Indianapolis: Wiley Publishing, 2003, p. 423.
- [10] William-H.-Inmon, Building The Data Warehouse, 4e éd., Indiana: Wiley Publishing, Inc., 2005, pp. 29, 175-176.
- [11] R. Kimball et J. Caserta, The Data warehouse ETL Toolkit, Wiley Publishing, Inc., 2004, p. 23.
- [12] R. Kimball et M. Ross, The Data Warehouse Toolkit Second Edition The Complete Guide to Dimensional Modeling, 2e éd., R. Elliott, Éd., Wiley Computer Publishing, John Wiley and Sons, Inc., 2002, pp. 8, 14-16, 21, 58-59.
- [13] A. Larbi, *Conception des entrepôts de données: Evaluation des besoins flexibles*, mémoire de thèse de doctorat, Faculté des sciences exactes, Université djilali liabes, 2018, p. 27.

- [14] E. h. Ziyati, : *Optimisation de requêtes OLAP en Entrepôts de Données*, Mémoire de thèse de doctorat, Faculté des sciences, UNIVERSITÉ MOHAMMED V –, 2010, p. 20.
- [15] L. Liu et M. Tamer Özsu, *Encyclopedia of database systems*, 2 éd., New York: Springer New York, 2018, pp. 201,793.
- [16] R. Kimball, *The data warehouse lifecycle toolkit*, New Jersey: Wiley Publishing, 2008, pp. 103-104.
- [17] B. Inmon, *Dw 2.0; architecture for the next generation of data warehousing*, Morgan Kaufmann, 2006, p. 12.
- [18] F. Geisler et B. Kettner, *Pro Serverless Data Handling with Microsoft Azure Architecting ETL and Data-Driven Applications in the Cloud*, New York: Apress, 2022, pp. 32-33.
- [19] M. Golfarelli et S. Rizzi, *Data Warehouse Design, Modern Principles and Methodologie*, New Delhi: Tata McGraw Hill Education Private Limited, 2009, p. 18.
- [20] R. Kimball et M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3rd Edition, 3e éd., John Wiley & Sons, Inc., 2013, pp. 45-46, 50.
- [21] F. Silvers, *Building and Maintaining Data Warehouse*, New York: CRC Press Publisher, 2008, pp. 210-211.
- [22] E. Thomsen, *OLAP Solutions Building Multidimensional Information Systems*, 2e éd., Wiley Computer Publishing, 2002, pp. 5, 621.
- [23] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy et B. Becker, *The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence*, Indianapolis: Wiley Publishing, 2010, p. 33.
- [24] M. Bala, *Parallélisation et distribution du processus d'intégration ETL pour le traitement de données massives*, mémoire de thèse de doctorat en sciences, Faculté d'Electronique et Informatique, USTHB, 2017, pp. 36-37.
- [25] «Que-ce que PostgreSQL ?,» [En ligne]. Available: <https://www.oracle.com/fr/database/definition-postgresql.html>. [Accès le 20 Mai 2022].
- [26] «Accueil/FAQ,» [En ligne]. Available: <https://www.pgadmin.org>. [Accès le 27 Avril 2022].
- [27] «Knowage 6.4 release,» [En ligne]. Available: <https://www.knowage-suite.com/site/knowage-6-4->

release/?fbclid=IwAR0HZkOSwqJDYSOxejesbsrmw1k7qnkMYm7Hjw4R4V
FP312H3tdbu8f5vww. [Accès le 28 Juin 2022].

- [28] «Guide d'utilisation de Talend Open Studio for Data Integration,» [En ligne].
[Accès le 13 Mai 2022].
- [29] B. Czernicki, Next-Generation Business Intelligence Software with Silverlight
3, New York: Springer, Apress, 2009, p. 11.