

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

UNIVERSITE SAAD DAHLEB BLIDA



Faculté des sciences

Département d'informatique

Mémoire de fin d'études

Pour l'obtention du diplôme de master en informatique

Thème

***Entrepôt de données de la gestion de
l'information cadastral textuel et
géographique***

Réalisé par:

- FERRACHE fethia
- SEMMANI khadidja

Devant le jury :

- | | |
|------------------------------------|--------------|
| - Mr.DERRAR | président |
| - Mme.OUAHRANI | Examinatrice |
| - Dr.BENBLIDIA Nadja | Promotrice |
| - M ^{elle} . AMIRAT Imene | Encadreur |

Promotion 2015/2016

REMERCIEMENTS

Premièrement et avant toute chose, nous rendons grâce à Allah, le tout puissant, de nous avoir permis de suivre le chemin du savoir, et donné le courage d'achever ce travail.

Un remerciement tout particulier à M^{elle} *IMIRAT Imene*. pour nous avoir encadrées durant cette dernière année du cursus. Nous les remercions profondément pour leurs encouragements continus, disponibilités et leurs critiques constructives qui nous ont permis d'améliorer ce mémoire.

Nous adressons nos vifs remerciements à notre promotrice *BENBLIDEA Nadja* pour son attention et la confiance qu'il nous a accordées ainsi qu'à toute l'équipe de la direction du cadastre, qui nous ont donné l'opportunité de travailler sur un sujet d'une telle envergure.

Nous tenons à remercier M. CHOUDER Mohamed enseignant à l'école nationale d'informatique pour leur aide, leur soutien, leur disponibilité, et leurs conseils, qui nous étaient vraiment bénéfiques.

Un grand Merci au corps enseignant ainsi qu'à l'administration de la faculté Saad Dahleb pour tout le savoir qu'ils ont su nous transmettre durant ces cinq dernières années, et aussi d'être toujours là pour nous guider à retrouver le bon chemin par leur sagesse et leurs précieux conseils.

Enfin, nous remercions les membres du jury d'avoir accepté d'évaluer notre modeste travail, ainsi que toutes les personnes qui ont contribué de près ou de loin à l'élaboration de ce mémoire.

DEDICACES

Je dédie ce travail :

*À l'âme de celui qui a dressé mon chemin sans
Voir le fruit de ses sacrifices,
À mon père ;*

*Aux bougies qui s'éteignent doucement pour
allumer mon obscurité,
À ma mère ;*

*À l'âme de celle qui m'a toujours
accompagné par ses prières,
À ma grande mère ;*

*À ceux qui me soutiennent, et Cultivent
en moi la confiance,
À mes frères ;*

*À toute ma grande famille
et mes très chers amis ;*

Fethia

Je dédie ce travail :

*À ceux qui m'ont fait voir la lumière, qui ont veillé durant mes
nuits pour faire la réussite de mes jours et à qui je dois tout et
que rien ne suffira pour les remercier.
À mes très chers parents ;*

*À ceux qui me soutiennent, et Cultivent
en moi la confiance,
À mes frères ;*

*À celui qui a toujours été là pour moi dans les moments
les plus difficiles, et que je ne remercierai
jamais assez pour toute l'aide
qu'il m'a prodigué.
À toi Mohamed ;*

*À toute ma grande famille
et mes très chers amis ;*

Khadija

Table des matières

Introduction générale	1
Contexte général de l'étude	1
Problématique	1
Objectifs	2
Organisation du mémoire	2
Chapitre 01 : Etude bibliographique	4
Introduction	4
1. système décisionnel.....	5
1.1. Définition du système décisionnel	5
1.2 Objectifs des systèmes décisionnels	5
1.3 Architecture du système décisionnel.....	5
1.4 Processus décisionnel	6
1.5 Comparaison avec les systèmes transactionnels	8
2. Entrepôt de données.....	9
2.1 Définition de l'entrepôt de données	9
2.2 Objectifs d'un entrepôt de données	11
2.3 Structure de données d'un ED	12
2.4 Composants de base d'un entrepôt de données	13
2.4.1 Systèmes Sources	13
2.4.2 Zone de préparation des données	13
2.4.3 Zone de présentation des données	14
2.4.4 Outils d'accès aux données	15
2.4.5 Métadonnées	15
2.5. Approches d'entreposage de données	16
2.5.1. Approche guidée par les sources de données	16
2.5.2. Approche guidée par les besoins d'analyse	16
2.5.3. Approche hybride	16
3. modélisation dimensionnelle.....	17
3.1. Définition de la modélisation dimensionnelle	17
3.2. Modèles et formes de la modélisation dimensionnelle	18
3.3. Architecture des serveurs OLAP.....	20
3.3.1. ROLAP : OLAP Relationnel	20
3.3.2. MOLAP: OLAP Multidimensionnel	20
3.3.3. HOLAP: OLAP hybride	21
4. Conduite et construction d'un entrepôt de données.....	21
Conclusion	22
Chapitre 02 : Analyse des besoins	23
Introduction	23
1.1. Définition du cadastre national(ANC)	23
1.2 Organigramme général de la direction du cadastre (ANC)	24
1.3. Définition de cadastre de wilaya de Blida	25
1.4. Organigramme général de la direction du cadastre Blida	26
1.5. Mission du cadastre	26
2.1. Situation informatique	26
2.1.1. Système de gestion de base de données (GIC)	26
2.1.2. Système d'information graphique(ArcGIC)	28
2.2. Pourquoi utiliser les deux logiciels	28
2.3. Postes de travail	28
2.3.1. Poste externes	28

1.4. Organigramme général de la direction du cadastre Blida	26
1.5. Mission du cadastre	26
2.1. Situation informatique	26
2.1.1. Système de gestion de base de données (GIC)	26
2.1.2. Système d'information graphique(ArcGIC)	28
2.2. Pourquoi utiliser les deux logiciels	28
2.3. Postes de travail	28
2.3.1. Poste externes	28
2.3.2. Poste internes	28
2.4. Description du processus métier	29
2.4.1. Processus de production	30
2.4.2. Processus de mise à jour	33
2.5. Critique de l'existant	35
2.5.1. Le système d'information opérationnel	35
2.5.2. Le système décisionnel.....	36
2.6. Présentation de la démarche d'identification des besoins	42
2.7. Préparation et conduite d'entretiens	42
2.8. Documentation	42
2.9. Etude des sources de données	43
2.10. Analyse des besoins	43
Conclusion	44
Chapitre 03 : Conception de l'entrepôt de données.....	45
Introduction	45
1. Conception de la zone d'entreposage.....	45
1.1 Matrice de l'architecture en bus décisionnel	45
1.2. Modélisation des activités	46
1.2.1. Activité de suivi des sections	46
1.2.2. Activité de suivi des ilots	48
1.2.3. Activité de suivi des parcelles	50
2. Conception de la zone d'alimentation.....	53
2.1. Etude des sources de données	53
2.2. Processus d'alimentation de l'ED	54
2.2.1. Processus d'extraction des données	54
2.2.2. Transformation des données	55
2.2.3. Chargement des données	57
2.2.4. Diagramme d'activité du processus ETL	57
3. Conception des cubes dimensionnels.....	59
3.1. Niveaux et hiérarchies des dimensions	59
3.2. Liste des cubes dimensionnels	60
3.3. Représentation des cubes dimensionnels	60
3.3.1. Représentation du cube CUBE_SUIVI_SECTION	61
3.3.2. Représentation du cube CUBE_SUIVI_ILOT	62
3.3.3. Représentation du cube CUBE_SUIVI_PARCELLE	63
Conclusion	63
Chapitre 04 : Mise en œuvre de notre système décisionnel.....	64
Introduction	64
1. Architecture technique de la solution	64
2. Sélection des outils	65
2.1. Plate-forme du SGBD	65
2.2. Outils de préparation des données	65

Conclusion Générale	74
Références bibliographiques	75
Références web graphies	76
Annexe	77

LISTE DES FIGURES

• Figure 1: Architecture du système décisionnel	7
• Figure 2 : Etapes du processus décisionnel	8
• Figure 3:L'entrepôt de données est orienté sujet	10
• Figure 4:L'entrepôt de données est intégré	11
• Figure 5:L'entrepôt de données est non volatile	11
• Figure 6: Les données d'un entrepôt de données sont historiées	12
• Figure 7: Structure de données d'un entrepôt de données	13
• Figure 8 : Architecture de l'entrepôt de données	14
• Figure 9: Plateforme du processus ETL	15
• Figure 10 : Exemple d'une modélisation en étoile	19
• Figure 11 : Exemple d'une modélisation en flocon	20
• Figure 12 : Exemple d'une modélisation en constellation	20
• Figure 13: Architecture ROLAP.....	21
• Figure 14 : Architecture MOLAP	22
• Figure 15 : Schéma du cycle de vie dimensionnel.....	22
• Figure 16 : Organigramme général de la direction du cadastre(ANC)	24
• Figure 17 : Carte géographique de cadastre de Blida	25
• Figure 18 : Organigramme général de la direction du cadastre de Blida.....	26
• Figure 19 : Processus de Production	30
• Figure 20 : Diagramme de séquence de processus du production.....	32
• Figure 21 : Processus de mise à jour.....	33
• Figure 22 : Diagramme de classe de la base littéral (Base GIC)	37
• Figure 23 : Figure 23 : Table ARCHIVE_APPARTIENT_ILOT.....	38
• Figure 24 : Table section.....	38
• Figure 25 : Diagramme de classe de base graphique(BaseArcGIC)	40
• Figure 26 : Table Cad_IlotArchive.....	41
• Figure 27: Table cad_section	41
• Figure 28 : Schéma en étoile de suivi des sections.....	48
• Figure 29 : Schéma en étoile de suivi des ilots.....	50
• Figure 30: Schéma en étoile de suivi de parcelles.....	53
• Figure 31 : Suppression des champs inutiles dans la table personne.....	58
• Figure 32 : Diagramme d'activité du processus ETL.....	60
• Figure 33 : Cube dimensionnel "CUBE_SUIVI_SECTION".....	61
• Figure 34: Cube dimensionnel "CUBE_SUIVI_ILOT"	62
• Figure 35 : Cube dimensionnel "CUBE_SUIVI_PARCELLE".....	63
• Figure 36 : Architecture technique du système proposé.....	64
• Figure 37: Outils utilisés pour la réalisation de notre système.....	65
• Figure 38 : Mapping du chargement de la dimension parcelle.....	67
• Figure 39: Mapping du chargement de la table de faits suivi_section.....	68

• Figure 40 : Mapping du chargement de la table de faits suivi_ilot.....	68
• Figure 41 : Mapping du chargement de la table de faits suivi_parcell.....	69
• Figure 42: Ensemble de Cubes et Dimensions créés sous Pentaho Schéma WorkBench.	70
• Figure43 : statistiques sur les sections cadastrées par commune.....	71
• Figure44: statistiques sur les ilots cadastrés.....	71
• Figure 45 : statistique sur le nombre de sections réalisées (Ain Romana).....	72
• Figure 46 : statistiques sur le nombre des ilots cadastrés.....	72

LISTE DES TABLEAUX

• Tableau1:Tableau comparatif : (systèmes transactionnels, systèmes décisionnels)	8
• Tableau 2: Postes de travail concernés par le cadastre.....	29
• Tableau 3: les tables utilisées dans la BDD littéral.....	39
• Tableau 4:les tables utilisées dans BDD graphique.....	42
• Tableau 5 : Besoins analytiques classés par sujet d'analyse.....	44
• Tableau 6: Utilisateurs finaux de l'ED.....	44
• Tableau 7: Matrice de l'architecture en bus dimensionnel.....	45
• Tableau 8: Dimension section.....	46
• Tableau 09: Dimension commune.....	47
• Tableau 10: Dimension wilaya.....	47
• Tableau 11: Dimension géographique_section.....	47
• Tableau 12: Dimension ilot.....	48
• Tableau 13: Dimension section_a.....	49
• Tableau 14: Dimension commune_a.....	49
• Tableau 15: Dimension wilaya_a.....	49
• Tableau 16: Dimension géographique_ilot.....	49
• Tableau 17: Dimension parcelle.....	51
• Tableau 18: Dimension ilot_a.....	51
• Tableau 19: Dimension géographique_parcelle.....	51
• Tableau 20: Dimension section_b.....	51
• Tableau 21: Dimension commune_b.....	52
• Tableau 22: Dimension wilaya_b.....	52
• Tableau 23: Tables nécessaires de la source de données.....	54
• Tableau 24: Types d'extraction de données.....	55
• Tableau 25: Liste des niveaux et des hiérarchies des dimensions de l'entrepôt.....	60
• Tableau 26: Liste des cubes.....	60

RESUME

La constitution d'entrepôt de données est une réponse au problème de l'intégration d'une grande quantité de données variées, relatives à un certain domaine d'application.

Notre travail consiste à faire un système décisionnel, qui est un entrepôt de données dans le but de répondre aux besoins du cadastre de Blida, d'avoir des informations nécessaires et fiables, et de disposer d'un système d'aide à la décision.

Mots clés :

Aide à la décision, Informatique décisionnelle, Magasins de données, Processus d'analyse en ligne, Outils ETC, Générateur d'états, Modélisation dimensionnelle.

ABSTRACT

Development of data warehouse is a response to the problem of integrating a large amount of varied data relating to a certain field of application.

Our job is to make a decision-making system, which is a data warehouse in order to meet the needs Maps of Blida, to have the necessary and reliable information, and to have a decision support system.

Keywords:

Decision support, Business Intelligence, Data Mart, OLAP, ETL tools, Reporting, Dimensional modeling.

LISTE DES ABREVIATIONS

- **BDD** : Base De Données
- **ED** : Entrepôt de données
- **ANC**: Agence National du Cadastre
- **DCW** : Direction Cadastral de Wilaya
- **ETL** : Extraction/ Transformation/ Loading
- **HOLAP** : Hybrid On-Line Analytic Processing
- **IDE** : Integrated Development Environment
- **MOLAP** : Multi dimensional On-Line Analytic Processing
- **OLTP** : On-Line Transaction Processing
- **OLAP** : On-Line Analytic Processing
- **ROLAP** : Relational On-Line Analytic Processing
- **SGBD** : Système de Gestion de Base de Données
- **SI** : Système d'Information
- **SID** : Système d'Information Décisionnel
- **TB** : Tableau de bord

Introduction générale

Contexte Général :

De nos jours, les entreprises s'intéressent de plus en plus à l'usage des systèmes d'information, qui contribuent de manière évidente aux performances de l'organisation, en assurant le contrôle, le stockage et la maîtrise de l'information. Ceci relève un nouveau défi, qui est le traitement, en temps réel d'une grande masse de données, provenant de diverses sources hétérogènes, dans le but d'améliorer le degré de connaissance du phénomène étudié et permettre de prendre des décisions efficaces.

Dans un tel contexte, le besoin d'intégration se fait de plus en plus sentir. Cependant, pour répondre à ce besoin, le développement des applications d'intégration se voit contraint de composer avec la répartition des sources et l'hétérogénéité de leurs structures et de gérer l'intégration entre les données en différents formats qu'ils manipulent.

Cette évolution a également conduit aux développements des systèmes d'informations très complexes, gérant des volumes d'informations très importants. Ces informations sont regroupées dans des sources de données hétérogènes, réparties et autonomes.

Problématique :

La constitution d'entrepôt de données est une réponse au problème de l'intégration d'une grande quantité de données variées, relatives à un certain domaine d'application, et stockées physiquement dans différentes sources de données. L'entrepôt de données regroupe certaines informations sous une forme exploitable par des traitements utiles pour l'aide à la décision. Ces informations, qui sont potentiellement pertinentes pour telle ou telle catégorie de décideurs du domaine, doivent être :

- extraites : Accéder à la majorité des systèmes de stockage de données (SGBD, document XML,..) et récupérer les données pertinentes.
- Transformées : Transformer les données avant de le charger dans le datawarehouse.
- Fusionnées : Insérer les données dans le datawarehouse.

L'intégration nécessite alors l'accès aux données des sources et la capacité de les fusionner.

D'une manière générale, la problématique peut se résumer comme suit :

La direction du cadastre de wilaya de Blida (DCW de Blida) national du cadastre(ANC) exerce ses fonctions au but d'établir les documents cadastraux suffisant pour aboutir à livrets fonciers des propriétaires de la wilaya de Blida qui sont produits au niveau de la conservation cadastral de Boufarik. L'établissement de ces documents nécessite l'enregistrement des informations dans les deux logiciels utilisés par la cadastre (Gic et ArcGIS), et cela pose un grand problème de discordances entre les deux logiciels l'ors de la saisit, donc ce problème oblige la cadastre de faire des statistique pour aboutir aux informations fiables. Ces statistiques sont aussi demandées par la direction régionale et nationale, mais malheureusement sont faites d'une manière manuelle.

Objectif :

Afin de répondre à la problématique précédente, la direction du cadastre décide de se lancer dans un projet de développement d'un système d'aide à la décision afin de pouvoir satisfaire les besoins des décideurs de la direction. Dans le cadre de ce projet, notre travail consiste à mettre en place une solution destinée aux décideurs du cadastre, dont le principal objectif est de construire un système décisionnel qui répond aux besoins des décideurs pour :

- Un meilleur contrôle d'accès de l'information cadastral.
- Aboutir à une information compréhensible, utile et rapide.

Organisation du mémoire

Notre mémoire est organisé comme suit :

Chapitre 01 : Etude bibliographique

L'objet de ce chapitre est de donner un panorama du concept de système décisionnel, où nous détaillerons ses objectifs, son architecture, ainsi qu'une comparaison entre ce dernier et le système transactionnel. Nous allons aussi aborder la notion d'entrepôt de données, où nous allons définir et discuter de ses objectifs, présenter les différents aspects relatifs à la structure de l'ED, ses composants de base ainsi que les différentes approches d'entreposage de données. Et aussi nous allons traiter une méthode de conception logique qui vise à présenter les données sous une forme standardisée, qui est la modélisation dimensionnelle.

Chapitre 02 : Analyse des besoins

Dans ce chapitre, nous allons faire une présentation de l'organisme d'accueil, son environnement, sa structure ainsi que ses activités. Nous présentons la situation informatique, les processus métiers, ainsi que les différents critiques des systèmes. Et aussi nous allons identifier la démarche d'identification des besoins, qui consiste en préparation et conduite des entretiens, étude des sources de données de l'organisation, puis nous analysons les besoins fonctionnels du cadastre qui sont classés selon trois sujets d'analyse, ainsi que les utilisateurs concernés par l'entrepôt.

Chapitre 03: Conception de l'entrepôt de données

À travers ce chapitre, nous allons proposer une modélisation dimensionnelle des données (Faits/Dimensions), répondant aux besoins des utilisateurs et permettant d'exploiter au mieux les données stockées dans l'entrepôt. Ensuite, nous présenterons la conception de la zone d'alimentation de notre entrepôt, qui permet de préparer les données de l'entreprise avant leur intégration dans l'ED. Puis nous allons faire les spécifications particulières du modèle OLAP, afin de permettre aux utilisateurs de faire des analyses dimensionnelles complexes sur les données et d'exploiter de manière efficace les informations contenues dans l'entrepôt.

Partie 04 : Mise en œuvre de notre système décisionnel

Nous allons, dans ce chapitre, présenter les détails de la réalisation de notre solution, qui consiste à donner une forme concrète à la phase de conception. Cette partie décrit l'architecture technique de notre solution, les étapes de réalisation, pour terminer avec un plan permettant d'initier les utilisateurs à l'outil.

Chapitre 01 : Etude bibliographique

Introduction

Chaque jour, les entreprises gèrent un très grand volume de données qui proviennent de diverses sources, essentiellement stockées dans les systèmes opérationnels. Et comme l'information est le pilier central du processus de prise de décision, il devient complexe pour un décideur d'avoir une vue synthétique du métier de l'entreprise afin de prendre les bonnes et meilleures décisions.

L'entrepôt de données est devenu une pratique courante pour la plupart des grandes entreprises à travers le monde. Et sa création est un effort qui exige la participation de l'ensemble des assortiments du domaine fonctionnel de la firme [JUKIC, 2006]. De ce concept, est née la notion de **modélisation dimensionnelle** qui facilite l'analyse et garantit la cohérence, la lisibilité et l'accessibilité des informations de l'entrepôt.

Dans ce chapitre, nous allons vous présenter la partie théorique de notre projet. Dans un premier temps, nous allons clarifier la notion d'**informatique décisionnelle**, ensuite nous allons introduire le **concept d'entrepôt de données**.

Ce chapitre qui va porter sur les systèmes décisionnels et les entrepôts de données, est organisé de la manière suivante :

Tout d'abord, la section 1 s'intéresse à expliquer que veut dire un système décisionnel, que peut-il offrir de plus à une organisation ? Nous nous citons quelques-uns de ses objectifs, son architecture, et nous allons faire une comparaison des systèmes décisionnels aux systèmes opérationnels.

La section 2 portera sur la définition des entrepôts de données, ainsi que leurs caractéristiques, Et leurs objectifs. Et aussi nous allons définir la structure de données des ED, les composants architecturaux de l'ED, et les différentes approches d'entreposage de données.

Dans la section 3 nous nous focalisons sur la modélisation dimensionnelle, ainsi que les concepts qui y sont liés (table de faits, dimensions, mesures). Puis, nous présentons les différents modèles et formes d'entreposage de données (en étoile, en flocon, et en constellation).

1. Système décisionnel

1.1. Définition du système décisionnel

La raison d'être d'un entrepôt de données est la mise en place d'une informatique décisionnelle, il est donc nécessaire de donner quelques définitions, les plus répondues à ce concept :

D'après **Polleto** : « *L'informatique décisionnelle (Decision Support System ou Business Intelligence) désigne les méthodes, les outils et les moyens qui permettent de collecter, consolider et modéliser les données d'une entreprise afin d'offrir une aide à la décision et de permettre au corps exécutif de l'entreprise d'avoir une vue d'ensemble de l'activité.* » [POLETTTO, 2012]

Nous pouvons conclure donc que le principal intérêt d'un système décisionnel est d'offrir au décideur une vision transversale de l'entreprise intégrant toutes ses dimensions.

1.2. Objectifs des systèmes décisionnels

L'informatique décisionnelle est un sujet en pleine évolution, c'est un outil qui permet d'avoir une vue d'ensemble des différentes activités de l'entreprise. Les systèmes décisionnels ont comme objectifs de :

- Faciliter la définition et la mise en œuvre des stratégies gagnantes.
- Analyser et interpréter les données complexes de l'environnement économique de l'entreprise et s'adapter continuellement à ce dernier.
- Restituer les informations nécessaires à l'analyse, plus rapidement et à tout moment.
- Aider les dirigeants dans leurs prises de décisions et dans l'analyse de la performance de leur entreprise.

1.3. Architecture du système décisionnel

La notion de l'informatique décisionnelle englobe les solutions informatiques dont le but est de consolider les informations disponibles au sein des bases de données de l'entreprise [POLETTTO, 2012]. Nous pouvons modéliser ce flux informationnel circulant dans la firme de la manière suivante :

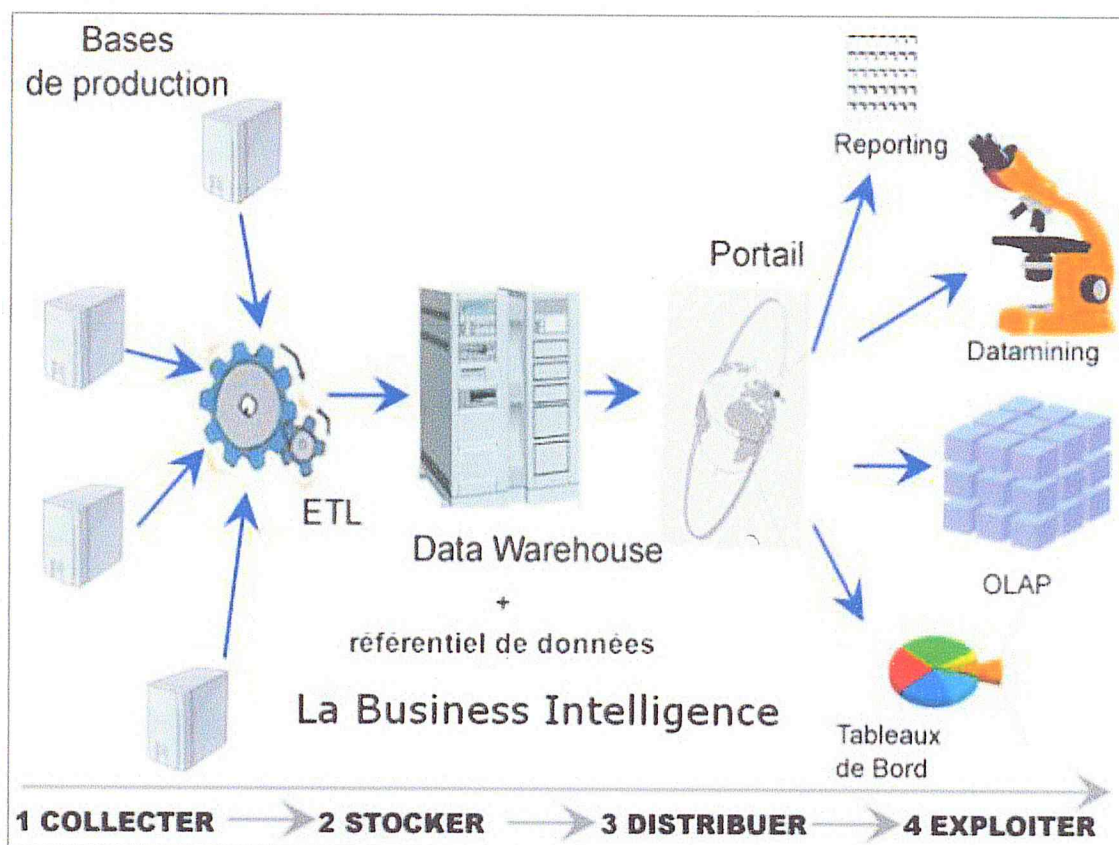


Figure 1: Architecture du système décisionnel [FERNANDEZ, 2013]

Dans cette architecture, on dispose d'un entrepôt de données (généralement, il s'agit plutôt d'un magasin de données¹ qui est plus petit que l'ED et qui concerne un domaine bien particulier) [TASLIMANKA SYLLA, 2007].

L'entrepôt centralise les données issues de plusieurs sources (bases de production de l'entreprise, fichiers textes, documents web, etc.). Ces données sont fusionnées dans l'entrepôt qui est généralement une grosse base de données (SQL Server, Oracle, etc.) [TASLIMANKA SYLLA, 2007].

Ensuite, une fois l'entrepôt confectionné, des données sont extraites dans des serveurs d'analyse ou serveurs OLAP sous forme de cubes de données² afin d'être analysées. Enfin, des générateurs d'états³ sont utilisés afin de présenter l'étude aux utilisateurs finaux ou décideurs.

1.4. Processus décisionnel

Dans la section précédente, nous avons vu les éléments et les outils composant un processus décisionnel, nous allons maintenant expliquer ce processus qui peut se résumer selon [FERNANDEZ, 2013] en quatre parties principales : Collecte de données, Stockage, Distribution et Exploitation.

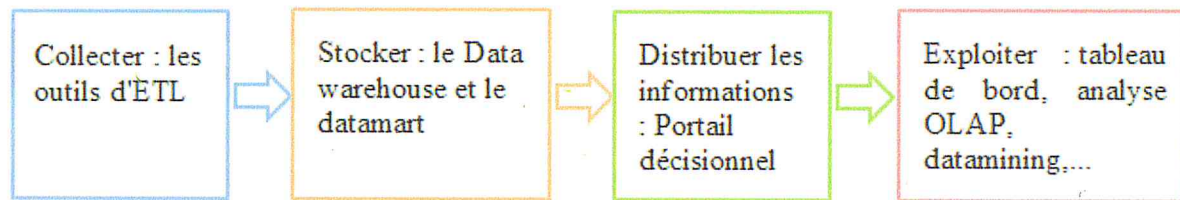


Figure 2 : Etapes du processus décisionnel [FERNANDEZ, 2013]

- **Collecter :**

C'est une étape sans doute très importante mais aussi très délicate. En effet, les sources de données sont souvent diverses et variées, internes ou externes à l'entreprise, il s'agit donc dans cette étape de trouver des outils ETL (Extraction / Transformation / Loading) permettant de : les extraire de leurs emplacements originels, les nettoyer, les transformer et les mettre dans l'entrepôt de données, tout en se focalisant sur les besoins spécifiques de l'utilisateur final afin de pouvoir répondre à toutes ses attentes.

- **Stocker :**

L'étape de stockage consiste à concevoir et à modéliser un entrepôt de données (le cœur du système décisionnel). Cet entrepôt de données permet de centraliser les données nettoyées et consolidées au préalable sous une forme adaptée pour les analyses. Il permet également de simplifier le modèle de données et d'organiser l'information afin de faciliter la restitution, les mises à jour et diminuer le temps de réponse. Son utilité ne sera sentie qu'une fois la quantité d'information à gérer devient vraiment très importante.

- **Distribuer :**

L'information est considérée comme flux et non pas comme unité de stockage, ce qui rend cette étape une clef de la modélisation de l'informatique décisionnelle, car c'est à ce moment que l'information sera livrée à l'utilisateur, et que les différents domaines fonctionnels seront distingués pour les interroger directement dans la phase de restitution.

- **Exploiter :**

Une fois les données sont collectées, stockées et distribuées, elles sont exploitées, c'est-à-dire les informations à valeur ajoutée sont représentées de façon lisible. C'est à ce niveau que les outils de restitution vont intervenir, où nous pouvons citer les tableaux de bord, les générateurs d'états, les outils de navigation dans des cubes, et les outils de statistiques.

1 Le magasin de données (datamart) est une base de données plus réduite que l'ED.

2 Le cube de données est une représentation abstraite d'informations multidimensionnelles, et qui est composé de cellules représentant les mesures.

3Générateur d'état ou reporting.

1.5. Comparaison avec les systèmes transactionnels

Les systèmes « opérationnels » ou « de gestion », également appelés systèmes OLTP (*on-line transaction processing*), sont dédiés aux métiers de l'entreprise pour les assister dans leurs tâches de gestion quotidiennes. [METAIS, 2014]

Bien que le système transactionnel ait des points communs avec le système décisionnel, il est non adapté aux besoins des décideurs. Ces deux systèmes présentent de profondes différences qui peuvent être résumées dans le tableau ci-dessous :

Différence	OLTP	OLAP
Pour les données	<ul style="list-style-type: none"> ▪ Orientées applications ▪ Courantes ▪ Données détaillées, codées et non redondantes ▪ Données Dynamiques ▪ Pas de référentiel commun 	<ul style="list-style-type: none"> ▪ Orientées sujets et thèmes (d'analyse) ▪ Historiques ▪ Informations agrégées cohérentes souvent avec redondance ▪ Informations stables et synchronisées dans le temps (statiques) ▪ Un référentiel unique
Pour les utilisateurs	<ul style="list-style-type: none"> ▪ Accès à peu d'information ▪ Pour les opérationnels ▪ Mise à jour et requêtes simples ▪ Temps de réponse immédiats ▪ Mise à jour et interrogations ▪ Requêtes prédéfinies 	<ul style="list-style-type: none"> ▪ Accès à de nombreuses informations ▪ Pour les décideurs ▪ Lecture unique et requêtes complexes transparentes ▪ Temps de réponse moins rapides ▪ Interrogations ▪ Requêtes imprévisibles et complexes

Tableau 1: Tableau comparatif : (systèmes transactionnels, systèmes décisionnels)

adapté de [TESTE, 2000] et [KHOURI, 2009]

Une base de données transactionnelle, utilisée par les opérationnels permet d'exécuter des **transactions** en temps réel telles que l'ajout, la modification, la suppression ou l'interrogation des données. Ces actions manipulent une faible quantité d'information, effectuées sur les versions les plus récentes de données et peuvent être exécutées par de nombreux utilisateurs simultanément.

Quant à un entrepôt de données, utilisé par les décideurs, est uniquement destiné à l'exécution de **questions statistiques** sur des données agrégées, ou de synthèse, historiées, de grande quantité et en lecture seulement dont les objectifs sont de : regrouper, organiser, intégrer et stocker des informations issues de différentes sources, afin d'assurer à l'utilisateur une vue orientée métier et par conséquent un accès rapide et plus efficace à la connaissance. Un entrepôt de données ne permet pas la modification de données que pour le stockage en blocs de manière régulière.

2. Entrepôt de données

2.1. Définition de l'entrepôt de données

William. H. Inmon et **Ralph Kimball** peuvent aisément et raisonnablement se prétendre les fondateurs d'entreposage de données. Ils ne sont peut-être pas les premiers à avoir posé les définitions relatives à un ED mais leurs travaux dans ce domaine et les principes de conception d'un entrepôt que l'un et l'autre ont posé en font des références incontestables.

Selon **Kimball** : « *L'entrepôt de données est une copie des données de transaction spécifiquement structurées pour la demande et l'analyse.* » [KIMBALL, 2011]

Dans son ouvrage, **Inmon** définit l'ED comme : « *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.* » [INMON, 2002]

À partir de cette définition, nous pouvons ressortir les caractéristiques principales d'un ED :

- **orienté sujet** : les données d'un entrepôt sont organisées autour des thèmes majeurs de l'entreprise tels que : la vente, la production, la comptabilité, ...etc.

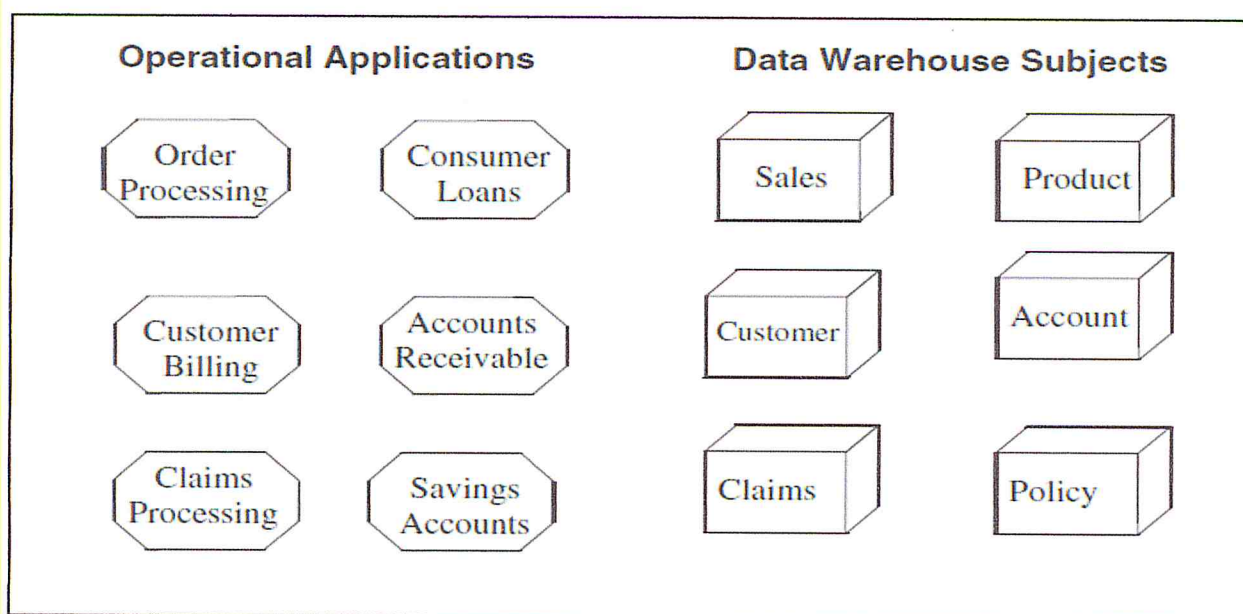


Figure 3: L'entrepôt de données est orienté sujet [PONNIAH, 2004]

- **Intégré** : les données proviennent de divers systèmes opérationnels hétérogènes et des applications de production.

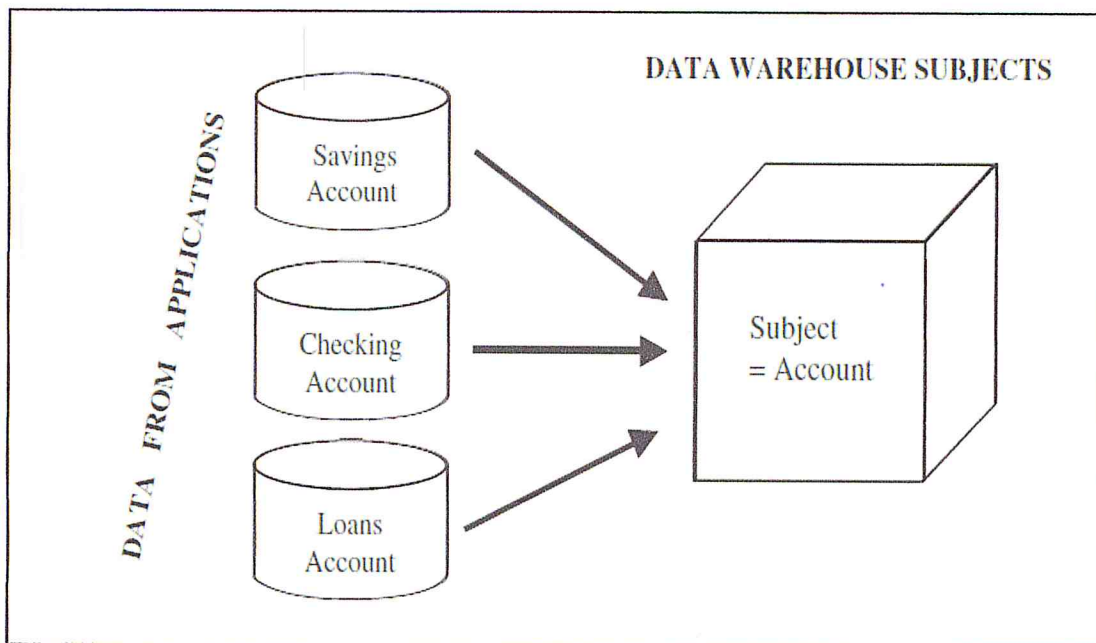


Figure 4: L'entrepôt de données est intégré [PONNIAH, 2004]

- **Non volatile** : les données stockées dans l'entrepôt ne sont pas modifiées par les utilisateurs.

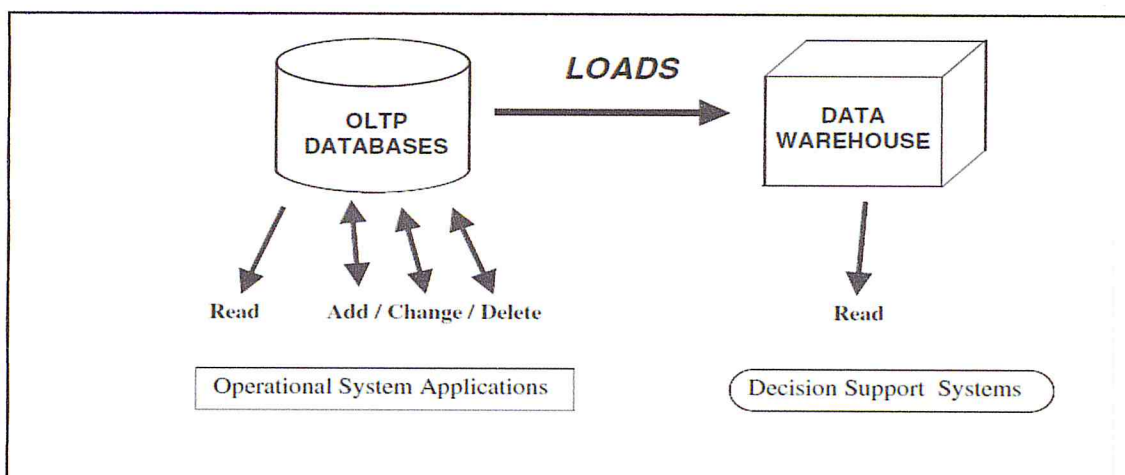


Figure 5: L'entrepôt de données est non volatile [PONNIAH, 2004]

- **Historié** : un référentiel de temps est associé à la donnée.

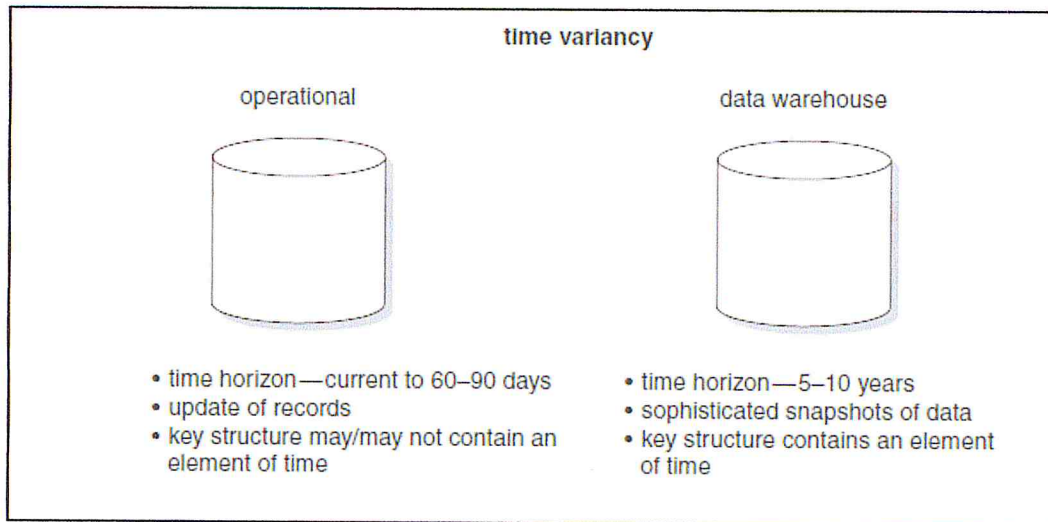


Figure 6: Les données d'un entrepôt de données sont historiées [INMON, 2002]

- **Disponible pour l'interrogation et l'analyse** : l'ED offre à l'utilisateur de manière efficace les informations nécessaires à tout processus de prise de décision [INMON, 2002].

2.2. Objectifs d'un entrepôt de données

L'entreprise est toujours riche par les informations qu'elle possède, et qui peuvent se présenter sous deux formes : système opérationnel enregistrant les données et l'entrepôt de données, les analysant et qui constitue une porte vers ces informations.

Les objectifs d'un entrepôt de données selon Kimball [KIMBALL, 2002] sont :

- **Facilité d'accès aux informations** : Les informations enregistrées dans un ED doivent être accessibles de manière immédiate et directe, elles doivent également être compréhensibles et significatives pour les décideurs,
- **Cohérence des informations** : Cela signifie que l'information se trouve dans l'ED sous une seule forme, ce qui garantit toujours le même résultat pour des requêtes équivalentes,
- **Adaptation et résistance aux changements** : L'ED est conçu de manière à répondre aux nouvelles questions sans altérer, modifier ou invalider les données enregistrées,
- **Sécurité des données** : L'ED permet de contrôler l'accès aux données les plus précieuses de l'entreprise et donc garantit leur protection,
- **Base décisionnelle de l'entreprise** : L'ED stocke et fournit à l'utilisateur les informations propres, purifiées et nécessaires à la prise de décisions,
- **Réutilisation des données** : Les données sont collectées, nettoyées et accumulées dans l'ED et donc elles sont soigneusement rassemblées pour être utilisées et réutilisées selon les besoins des utilisateurs.

2.3. Structure de données d'un ED

Inmon structure l'entrepôt de données selon différents niveaux d'agrégation et de détails des données en deux axes, chaque axe à son tour est constitué de deux classes de données [INMON, 2002]. La figure suivante illustre cette structuration :

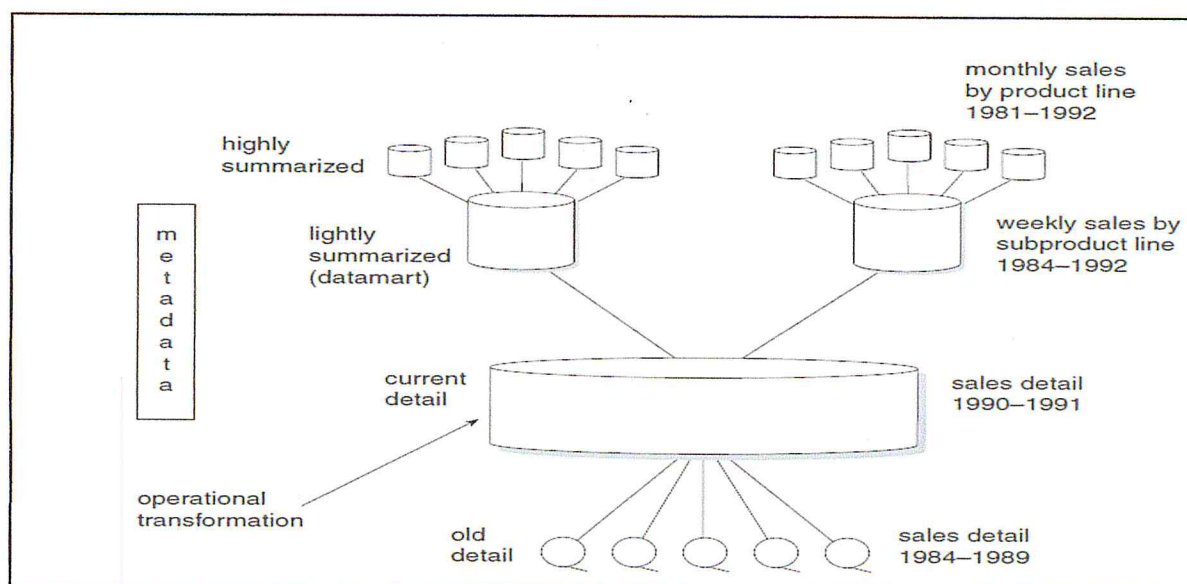


Figure 7: Structure de données d'un entrepôt de données [INMON, 2002]

- **Axe historique :**

Les volumes de données traitées et manipulées à ce niveau sont plus importants que ceux gérés en transactionnel, il est composé des deux classes de données suivantes :

- **Les données détaillées :** À ce niveau, résident les données associées aux événements récents ou issues des systèmes de production et intégrées régulièrement, elles sont volumineuses et fréquemment consultées,

- **Les données détaillées historiques :** Données associées aux transactions historiques, rarement consultées et peu volumineuses.

- **Axe synthétique :**

Ce niveau correspond aux résultats d'analyse et de synthèse de l'information contenue dans les systèmes décisionnels. Il est composé des deux classes de données suivantes :

- **Les données agrégées :** Données agrégées ou composées à partir des données détaillées, très utilisées, elles sont en général des résultats d'analyse et de synthèse de l'information,

- **Les données fortement agrégées :** Données agrégées à partir des données détaillées, à un niveau d'agrégation plus élevé que les données agrégées, à la seule différence de performance, car elles permettent parfois d'exécuter les requêtes 100 fois, voire 1000 fois plus rapidement.

2.4. Composants de base d'un entrepôt de données

L'entrepôt de données est un élément de base qui rentre dans la composition des systèmes décisionnels, c'est l'endroit de stockage des données provenant des différentes sources, traitées et transformées en informations manipulées par les outils d'analyse pour des fins de prise de décision.

Avant d'assembler les composantes d'un entrepôt de données et de le construire, nous allons d'abord expliquer chaque composant séparément. L'ED est composé de quatre zones distinctes :

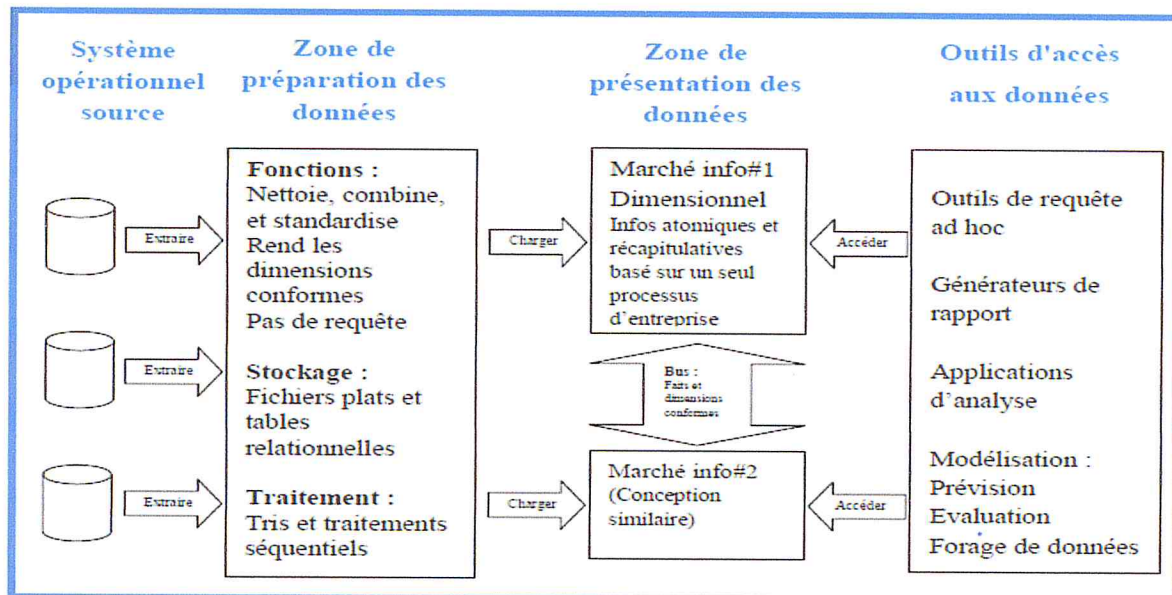


Figure 8 : Architecture de l'entrepôt de données [KIMBALL, 2002]

2.4.1. Systèmes Sources

Afin d'alimenter l'entrepôt de données, il faut en premier lieu un ensemble de sources à partir duquel les données sont extraites. Ces données peuvent être regroupées en quatre grandes catégories (Opérationnelles, internes, archivées et externes) [PONNIAH, 2004].

2.4.2. Zone de préparation des données

Cette zone, inaccessible par les utilisateurs est chargée du traitement des données extraites des systèmes sources, elle est constituée d'un ensemble de processus appelé *ETL*, « *Extract, Transform and Load* » [MATÉ, 2012]. Ces processus permettent la préparation de données avant leur intégration dans l'ED.

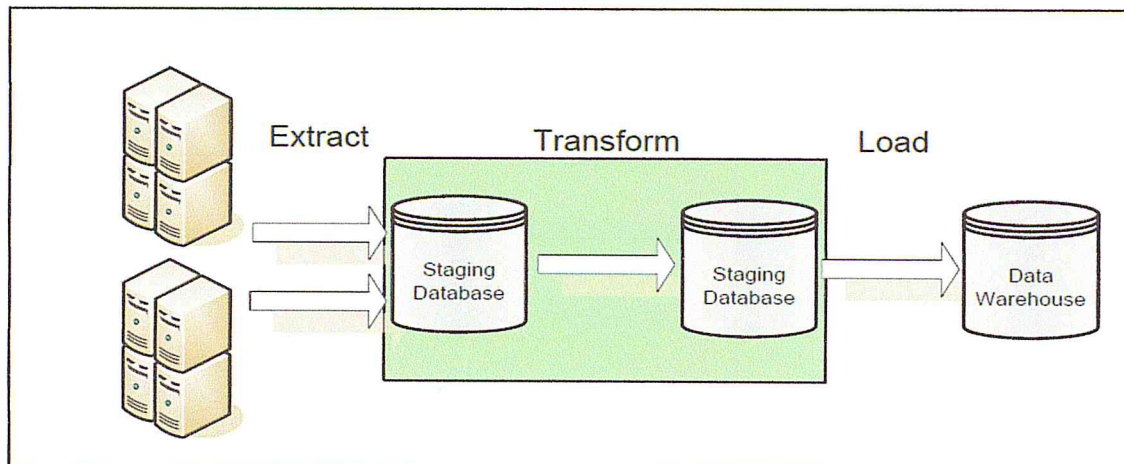


Figure 9: Plateforme du processus ETL [DAVENPORT, 2008]

Les outils ETL rassemblent et nettoient les données provenant de différentes sources afin de les charger dans l'entrepôt de données. Les phases constructives d'un ETL sont :

- **Extraction :**

Cette première phase du processus consiste à extraire les données de sources internes et externes à l'entreprise. La difficulté de cette phase réside dans le choix des données à extraire et des filtres à appliquer.

- **Transformation :**

Afin de garantir la fiabilité des données extraites, cette étape exige beaucoup de réflexion. Elle consiste à nettoyer et à consolider les données pour les transformer en information utilisable, fiable et pertinente.

- **Chargement :**

Afin d'optimiser au mieux le processus de chargement, cette dernière étape nécessite une certaine connaissance des structures du SGBD de l'entreprise. Elle consiste à insérer les données collectées dans une source d'information commune, unifiée et susceptible de masquer la diversité de l'origine des données [DAVENPORT, 2008]. Cette source est l'entrepôt de données.

2.4.3. Zone de présentation des données

Une fois les données nettoyées, elles sont intégrées dans l'entrepôt [BELLATRECHE, 2000]. Cette zone est l'endroit où les données sont organisées et stockées afin d'être interrogées facilement par les utilisateurs à l'aide des outils d'accès.

2.4.4. Outils d'accès aux données

Le besoin de l'entreprise en matière d'informations chiffrées sur ses activités, sur un pas temporel l'oblige à faire recours à des outils d'accès aux données, tels que les outils d'analyse en ligne OLAP et des outils de restitution de données plus sophistiqués. Parmi ces outils nous pouvons citer le Reporting et les tableaux de bord :

- **Reporting :**

Le reporting est un outil d'information de la hiérarchie chargé de restituer les résultats des processus achevés et peut donc être compris comme : « *un outil de contrôle de gestion à posteriori des responsabilités déléguées* » [ZIAN, 2013]. Cet outil permet l'élaboration de comptes rendus pertinents et fournit des bilans analytiques offrant une vision opérationnelle de l'activité et assurant son suivi. Il existe deux types de Reporting :

- Le Reporting Ad Hoc : Les rapports sont créés par l'utilisateur lui-même et non par l'informaticien. Il peut créer un rapport en fonction de ses besoins et avec les informations qu'il souhaite y mettre ou modifier un rapport déjà existant,

- Le Reporting de masse : Ces rapports de suivi d'activité sont créés automatiquement par le service informatique, envoyés par mail ou publiés sur le portail de l'entreprise à disposition des utilisateurs concernés.

- **Tableau de bord :**

Plusieurs consultants et analystes proposent une définition d'un TB. Nous pouvons citer celle de **Fernandez** : « *Un tableau de bord est un instrument de mesure de la performance facilitant le pilotage proactif d'une ou plusieurs activités dans le cadre d'une démarche de progrès. Le tableau de bord contribue à réduire l'incertitude et facilite la prise de risque inhérente à toute décision. Le tableau de bord est un instrument d'aide à la décision* » [FERNANDEZ, 2013].

Donc un TB est un outil d'aide à la décision constitué d'indicateurs permettant l'évaluation fiable des performances de l'entreprise à une certaine période par rapprochement entre les résultats réalisés et les objectifs tracés. Ces outils mettent en évidence les écarts, permettent de : connaître la situation de l'entreprise, anticiper les risques, attirer l'attention sur des phénomènes anormaux, communiquer ces résultats et enfin simuler des décisions.

2.4.5. Métadonnées

Présentes à tous les niveaux, les métadonnées sont des données sur les données (brutes ou transformées), elles sont les informations concernant l'entrepôt de données et son environnement ; par exemple, les métadonnées des applications opérationnelles incluent les schémas, les règles et les fichiers du processus d'extraction, quant aux métadonnées de la zone de préparation comprennent des informations sur les transformations requises pour l'intégration des données [KIMBALL, 2002].

2.5. Approches d'entrepôt de données

La plupart des approches existantes pour le développement d'un ED concernent la façon dont les données devraient être structurées, stockées, et contrôlées [GAM, 2006]. Kimball et Inmon s'entendent sur le fait qu'un entrepôt de données intègre les données des diverses sources de données du système opérationnel. Ces penseurs d'entrepôt de données distinguent trois approches académiques à l'état actuel :

2.5.1. Approche guidée par les sources de données

Cette approche utilise les sources de données pour concevoir les schémas multidimensionnels [INMON, 2002]. Selon Inmon, l'approche met en œuvre l'entrepôt de données comme une seule entité normalisée. Le processus d'un entrepôt de données peut se résumer comme suit : « *On ne fait rien tant que tout n'est pas désigné, l'entrepôt de données doit être exhaustif* » [INMON, 2002] ; c'est-à-dire qu'un entrepôt de données global qui couvre tous les sujets de l'entreprise est construit.

L'objectif de cette approche, c'est de livrer une solution technologiquement saine basée sur des méthodes et technologies éprouvées des bases de données.

2.5.2. Approche guidée par les besoins d'analyse

Kimball décrit le processus d'un entrepôt de données comme suit : « *Que chacun construise ce qu'il veut, on intégrera ce qu'il faudra quand il faudra* » [KIMBALL, 2002] ; c'est-à-dire que des magasins de données sont construits d'une façon incrémentale, ce qui s'adapte bien aux méthodologies agiles.

Cette approche met en œuvre l'entrepôt de données dans des entités de stockage dimensionnelles : magasins de données, qui sont construits à partir des besoins des utilisateurs [LAHMER, 2011]. Un bus de dimension permet de relier les magasins de données afin de constituer l'entrepôt de données global.

2.5.3. Approche hybride

Les méthodes mixtes combinent les deux démarches précédentes et essayent de combler les lacunes de chacune d'elles [INMON, 2002]. Cette approche se décompose en trois phases :

- L'analyse des besoins utilisateurs produit un ou plusieurs schémas idéaux,
- L'étude des sources se caractérise par un ou plusieurs schémas candidats,
- La phase de confrontation permet de comparer les schémas idéaux et candidats pour aboutir à un schéma final.

Cette approche permet de développer un modèle de données d'entreprise de manière itérative, mais parfois elle implique des compromis de découpage (dupliquer les dimensions identiques pour des besoins pratiques).

3. modélisation dimensionnelle

3.1. Définition de la modélisation dimensionnelle

La modélisation dimensionnelle est une méthode de conception logique qui vise à présenter les données sous une forme standardisée intuitive et qui permet des accès hautement performants. [KIMBALL, 2011].

Le modèle dimensionnel désigne la structure de données la plus adéquate pour modéliser la grande masse de données stockées dans un entrepôt de données. Chaque modèle dimensionnel se compose d'une table contenant une clé multiple, *la table des faits*, et d'un ensemble de tables plus petites nommées *tables dimensionnelles*. Chacune de ces dernières possède une clé primaire unique, qui correspond exactement à l'un des composants de la clé multiple de la table des faits. Les concepts de la modélisation dimensionnelle sont les suivants :

- **Table de faits :**

La table de fait ou la table centrale du modèle dimensionnel représente selon Kimball [KIMBALL, 2011] l'observation du marché sur un sujet étudié selon plusieurs axes d'analyse.

Elle sert à stocker des données permettant de mesurer l'activité de l'entreprise, appelées *indicateurs* ou *mesures*, elle permet également de lier les dimensions entre elles.

- **Mesure :**

Les mesures sont des données volumineuses d'un SI qui reflètent les activités de l'entreprise, elles sont généralement des valeurs numériques ou valorisées par des informations textuelles [KIMBALL, 2011].

- **Dimension :**

Les tables dimensionnelles sont les tables qui constituent les différents axes d'analyse accompagnant la table de faits. Elles décrivent selon Kimball le métier de l'entreprise de manière textuelle. « *Le sujet à analyser, c'est-à-dire le fait, est analysé suivant différentes perspectives. Ces perspectives correspondent à une catégorie utilisée pour caractériser les mesures d'activité analysées* » [MARCEL, 1998].

3.2. Modèles et formes de la modélisation dimensionnelle

3.2. Modèles et formes de la modélisation dimensionnelle

Afin de modéliser l'ED, la table de faits et les tables dimensionnelles sont assemblées dans une structure de données, qui correspond aux besoins de l'entreprise et qui facilite la navigation et la bonne exploitation de données de l'ED. Pour cela, il existe plusieurs modèles dont les plus utilisés sont :

- **Modèle en étoile :**

Ce modèle défini par **Kimball** est constitué d'une table au centre : *table de fait* reliée à des tables dimensionnelles, qui ne sont pas reliées entre elles qu'à travers cette table de fait. L'identifiant de la table de faits est la concaténation des clés de chacune des dimensions. Ce modèle est le plus utilisé car :

- Il est très simple et compréhensible par l'utilisateur,
- La navigation dans la base de données est facile (Performances des requêtes).

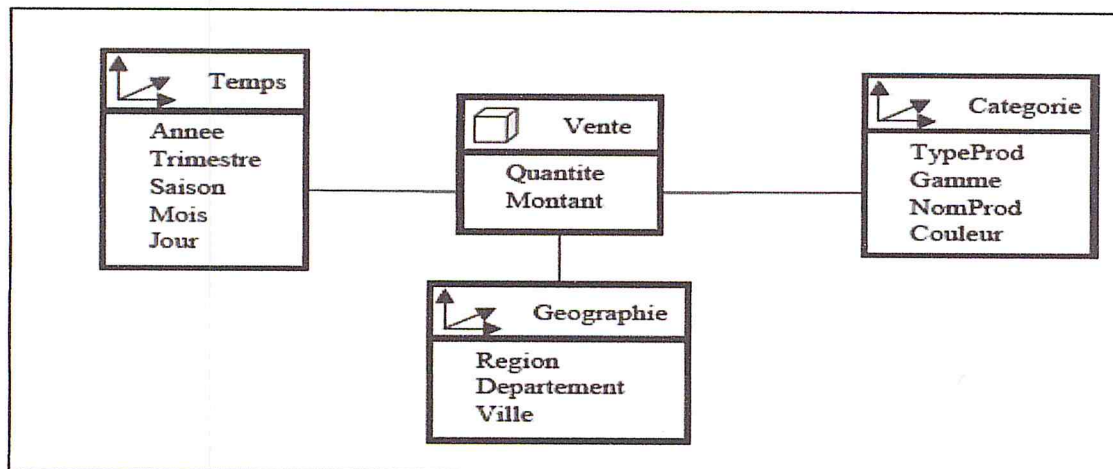


Figure 10 : Exemple d'une modélisation en étoile [TESTE, 2000]

- **Modèle en flocon :**

C'est une variante du modèle en étoile défini par **Inmon**, dont le principe est de normaliser le schéma en étoile correspondant. Cette normalisation consiste à mettre les attributs de niveau hiérarchique dans une nouvelle table dimensionnelle. Ce modèle peut s'avérer utile car :

- Il réduit la taille des dimensions et donc l'espace de stockage,
- Il normalise les dimensions, ce qui élimine les redondances qui pourraient s'y produire et donc performances des mises à jour.

En même temps, **Kimball** affirme que la normalisation des dimensions peut augmenter la complexité de lisibilité et de gestion des données, ainsi que le nombre de jointures de tables nécessaires à l'exécution d'une requête.

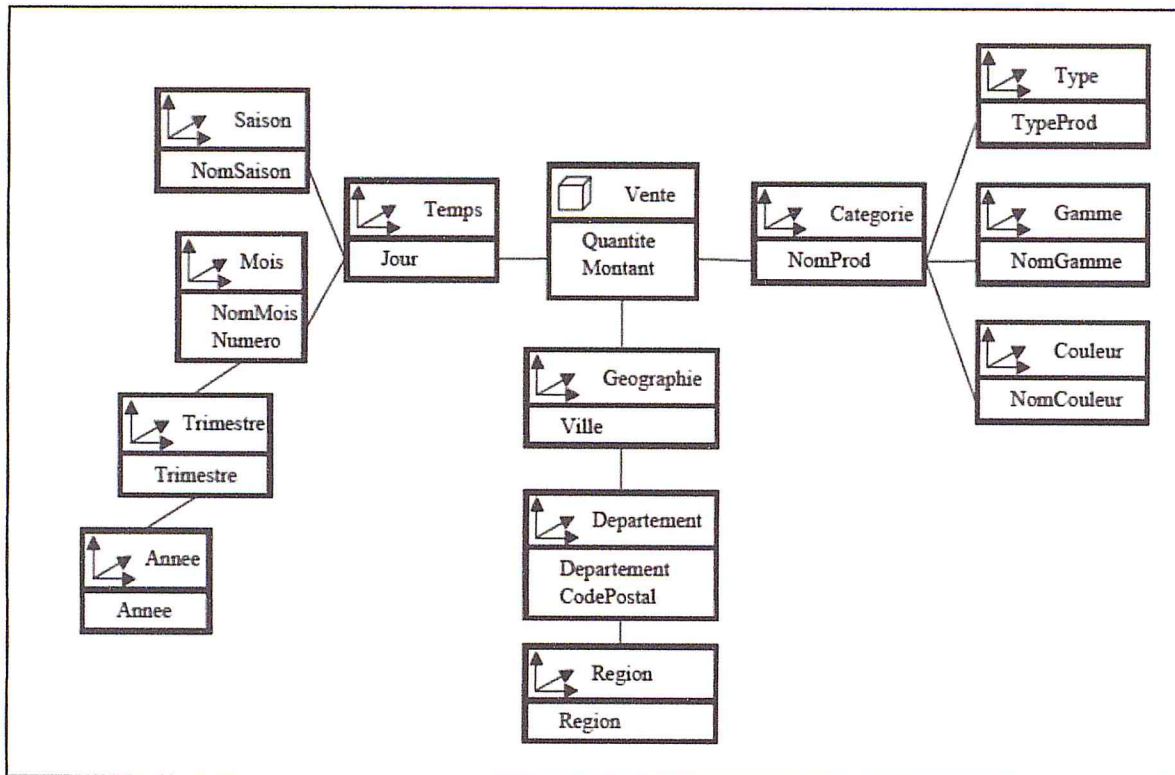


Figure 11 : Exemple d'une modélisation en flocon [TESTE, 2000]

- **Modèle en constellation :**

Le modèle en constellation est composé à partir de plusieurs modèles en étoiles reliés entre eux par le biais de dimensions communes, donc il comporte plusieurs faits.

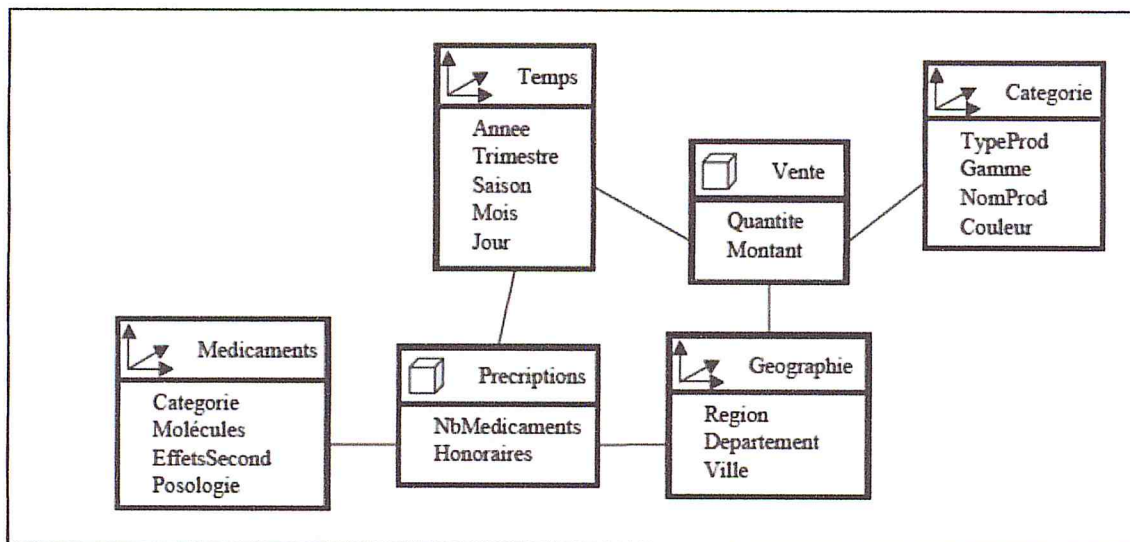


Figure 12 : Exemple d'une modélisation en constellation [TESTE, 2000]

3.3. Architecture des serveurs OLAP

OLAP (Acronyme d'On-Line Analytical Processing) est une activité globale de requêtage et de présentation de données textuelles et numériques contenues dans l'ED ; style d'interrogation et de présentation spécifiquement dimensionnel. La technologie OLAP est non relationnelle et presque toujours basée sur un cube de données multidimensionnel explicite [KIMBALL, 2011]. Les systèmes OLAP peuvent être classés comme suit :

3.3.1. ROLAP : OLAP Relationnel

Cette architecture exploite les bases de données relationnelles pour stocker un gros volume de données dimensionnelles, elle garde en mémoire des tables d'agrégats contenant des données sommaires. Elle n'est pas coûteuse mais le temps de réponse est long à cause de la sollicitation de la base de données à chaque appel aux tables relationnelles

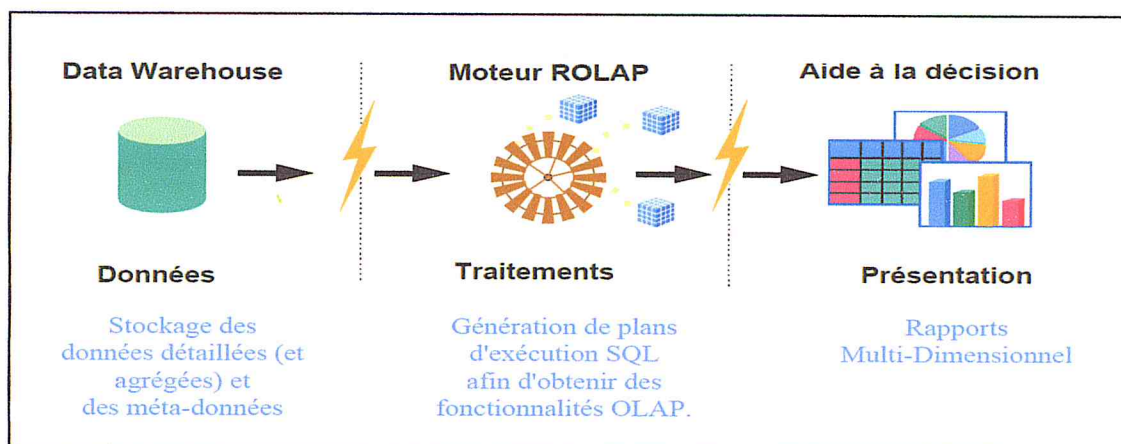


Figure 13 : Architecture ROLAP [NAKACHE, 1998]

3.3.2. MOLAP: OLAP Multidimensionnel

Cette architecture s'appuie sur une base de données multidimensionnelle qui permet de stocker les données fréquemment utilisées nécessitant un temps de réponse minimal directement dans un format permettant des opérations matricielles. Elle permet d'effectuer des calculs très poussés en un temps record, vu que tous les résultats sont pré-calculés mais elle montre ses limites devant une grande masse de données

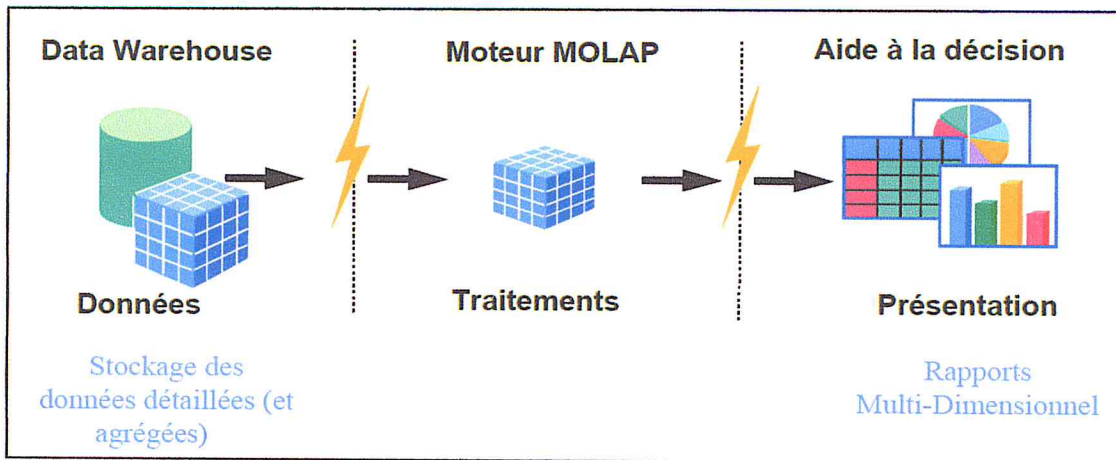


Figure 14: Architecture MOLAP [NAKACHE, 1998]

3.3.3. HOLAP: OLAP hybride

Cette architecture est une hybridation entre les deux architectures précédentes, dont les données du cube non fréquemment utilisées sont stockées dans l'ED relationnel et les agrégations dans une structure multidimensionnelle. Elle sert à minimiser le temps de réponse mais elle reste limitée si les rapports sont trop complexes.

4. Conduite et construction d'un entrepôt de données

Les étapes de conduite et de construction d'un entrepôt de données sont similaires à celles de la conduite de projets informatiques.

Selon Kimball, l'entrepôt de données se constitue peu à peu par les magasins de données de l'entreprise, c'est-à-dire par une approche guidée par les besoins des utilisateurs. Cette approche est illustrée grâce à son cycle de vie dimensionnel comme suit :

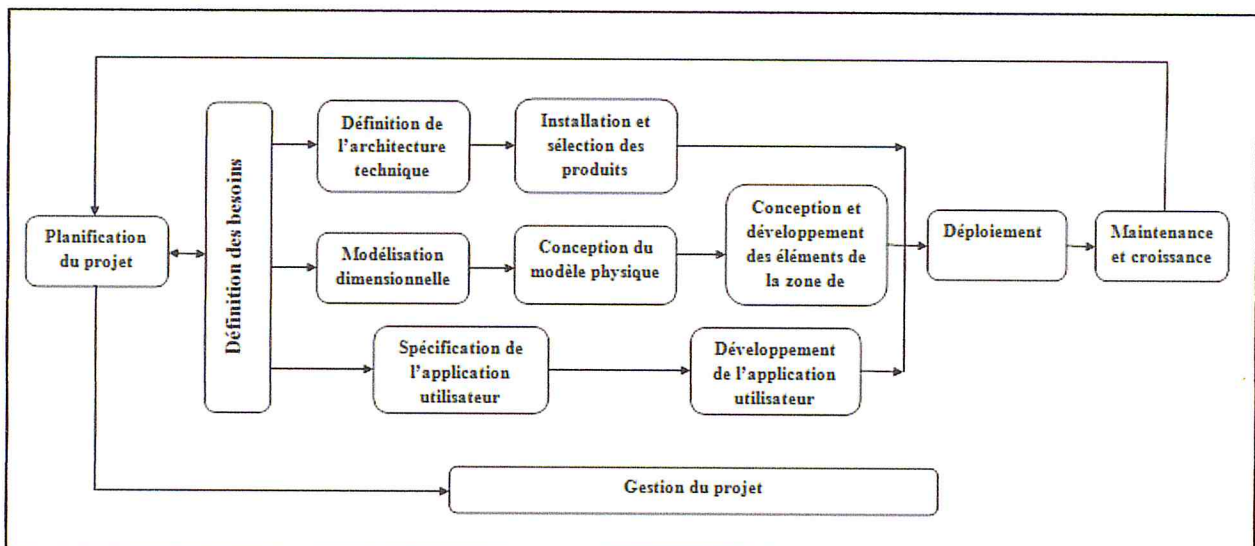


Figure 15 : Schéma du cycle de vie dimensionnel [KIMBALL, 2011]

Conclusion

Dans ce chapitre, nous avons évoqué la notion d'informatique décisionnelle, représentée par les systèmes décisionnels qui permettent aux décideurs de restituer au bon moment l'information sous une forme utilisable, ils servent à contrôler la performance de l'entreprise, coordonner les actions à entreprendre ou récolter des décisions d'ordre tactique et stratégique. À travers une comparaison des systèmes décisionnels aux systèmes opérationnels, nous avons pu déduire que si les derniers font fonctionner l'organisation, les premiers observent son fonctionnement, l'analysent et l'améliorent. Ensuite, nous avons étudié le concept d'entrepôt de données qui est étroitement lié aux systèmes décisionnels. Et après nous avons évoqué les concepts clés de la modélisation dimensionnelle, qui est une étape cruciale de l'implémentation de l'ED et qui répond parfaitement aux attentes des utilisateurs en termes de simplicité et de stabilité. Elle permet de structurer les données sous une forme intuitive et compréhensible permettant ainsi des accès performants. Dans le chapitre suivant, nous allons faire la partie d'analyse de notre projet.

Chapitre 02 : Analyse des besoins

Introduction

Dans la première partie de ce mémoire, nous avons présenté une synthèse bibliographique sur les systèmes décisionnels et les entrepôts de données.

Le point de départ de chaque projet est la connaissance du périmètre d'étude ou de l'organisme d'accueil. Et ce, afin de définir ses besoins et de déterminer la capacité de la solution à répondre aux besoins des utilisateurs finaux et d'apporter des changements aux anciens systèmes. Donc dans le deuxième chapitre on va faire la partie d'analyse de notre travail, ce chapitre est organisé de la manière suivante :

Dans la section 1 nous allons présenter la **Direction du cadastre de Blida**, nous présentons dans un premier temps la **Direction du cadastre national (ANC)**, nous introduisons ensuite la structure générale de l'organisation où nous allons fournir une carte géographique du cadastre, l'organigramme et les missions. Puis, nous présentons la situation informatique, et aussi les différents critiques des systèmes.

Dans la section 2 nous allons présenter la démarche d'identification des besoins, qui consiste en préparation et conduite des entretiens, étude des sources de données de l'organisation. Et nous analysons les besoins fonctionnels du cadastre qui sont classés selon trois sujets d'analyse, les besoins techniques, et les utilisateurs concernés par l'entrepôt.

1.1. Définition du cadastre national(ANC)

La direction de CADASTRE de BLIDA commence de puis la rédaction d'EPA par décret exécutif n°89-234 du 19/12/1989 modifie et complète et cette direction contient deux missions qui sont une mission principale et secondaire, la première fait l'établissement du cadastre général sur l'ensemble du territoire national et la deuxième fait les travaux spéciaux de topographe délimitation bornage pour le compte de service collectivité et organisme public concerne les organisations de la direction qui sont : la direction générale-siège-Alger,8 directions régionales(Alger, Bechar, Constantine, Oran, Ouargla, Biskra, Chleff, Sétif) en plus de 48 directions de wilaya (DCW) qui utilisent les moyens suivants :topographie, photogrammétrie, reprographie informatique.

1.2. Organigramme général de la direction du cadastre(ANC)

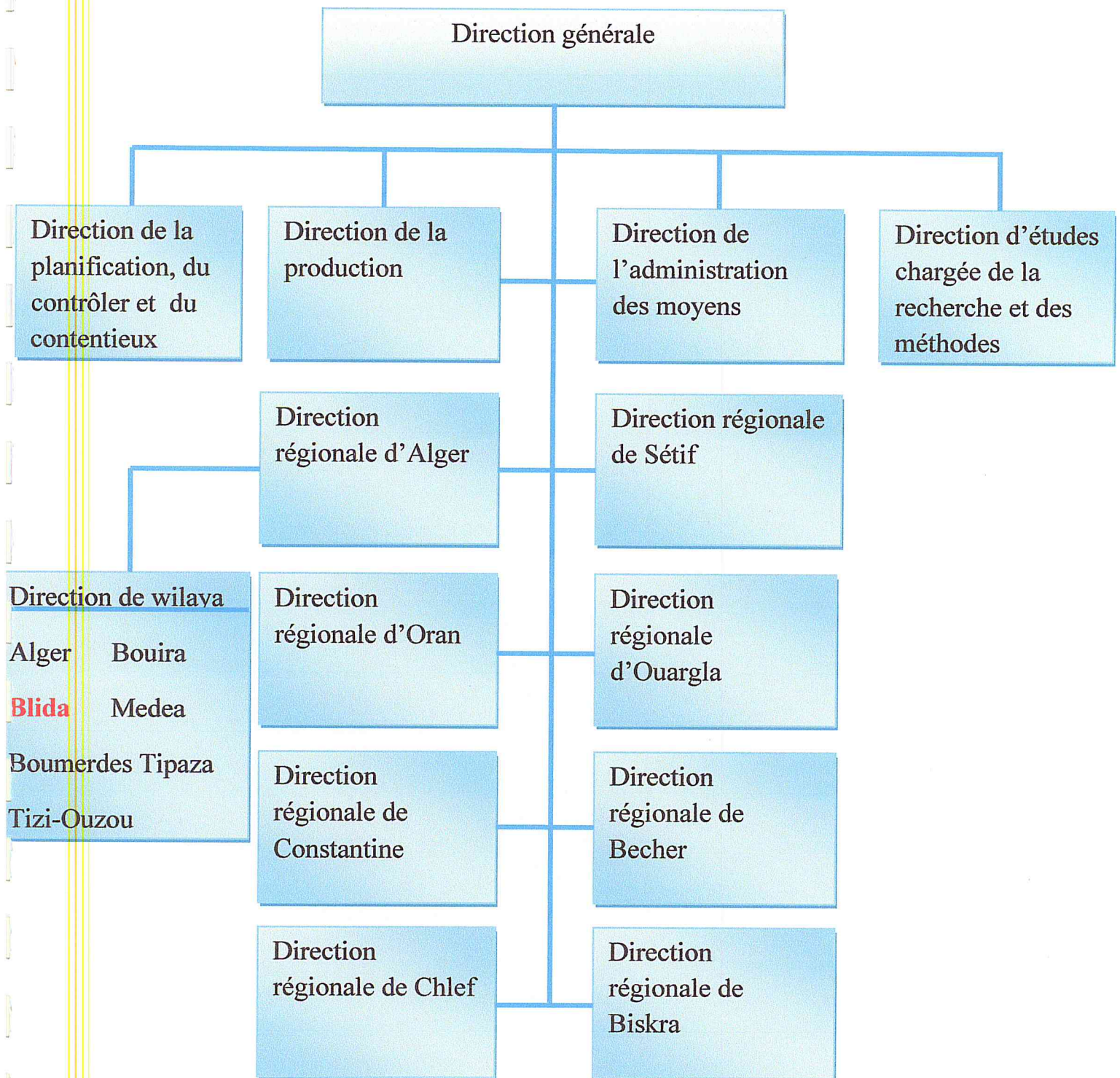


Figure 16: Organigramme général de la direction du cadastre(ANC)

1.3. Définition de cadastre de wilaya de Blida(DCW)

Le cadastre est le service du service public fédéral des finances qui a trois charges fiscales :

- Evaluation fiscale des biens fonciers bâtis et non bâtis, afin de servir de base de calcul des impôts locaux
- Foncière : identification et description des propriétés, recherche de leur propriétaire.
- Technique : établissement des mises à jour permanentes des plans cadastraux.

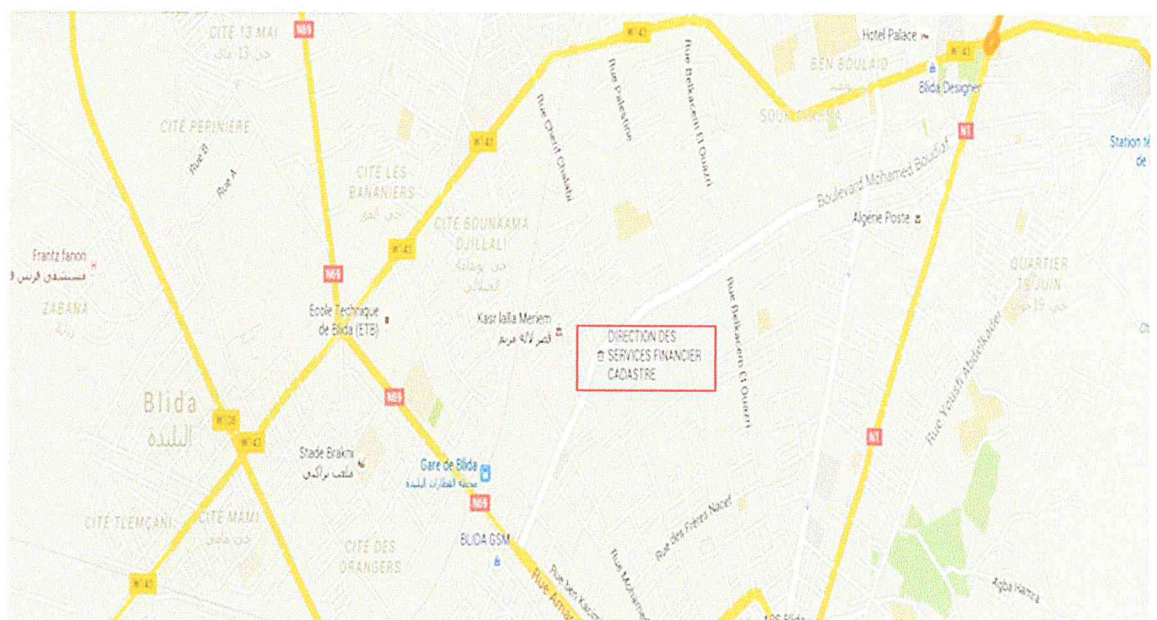


Figure 17: Carte géographique de cadastre de Blida

1.4. Organigramme général de la direction du cadastre Blida :

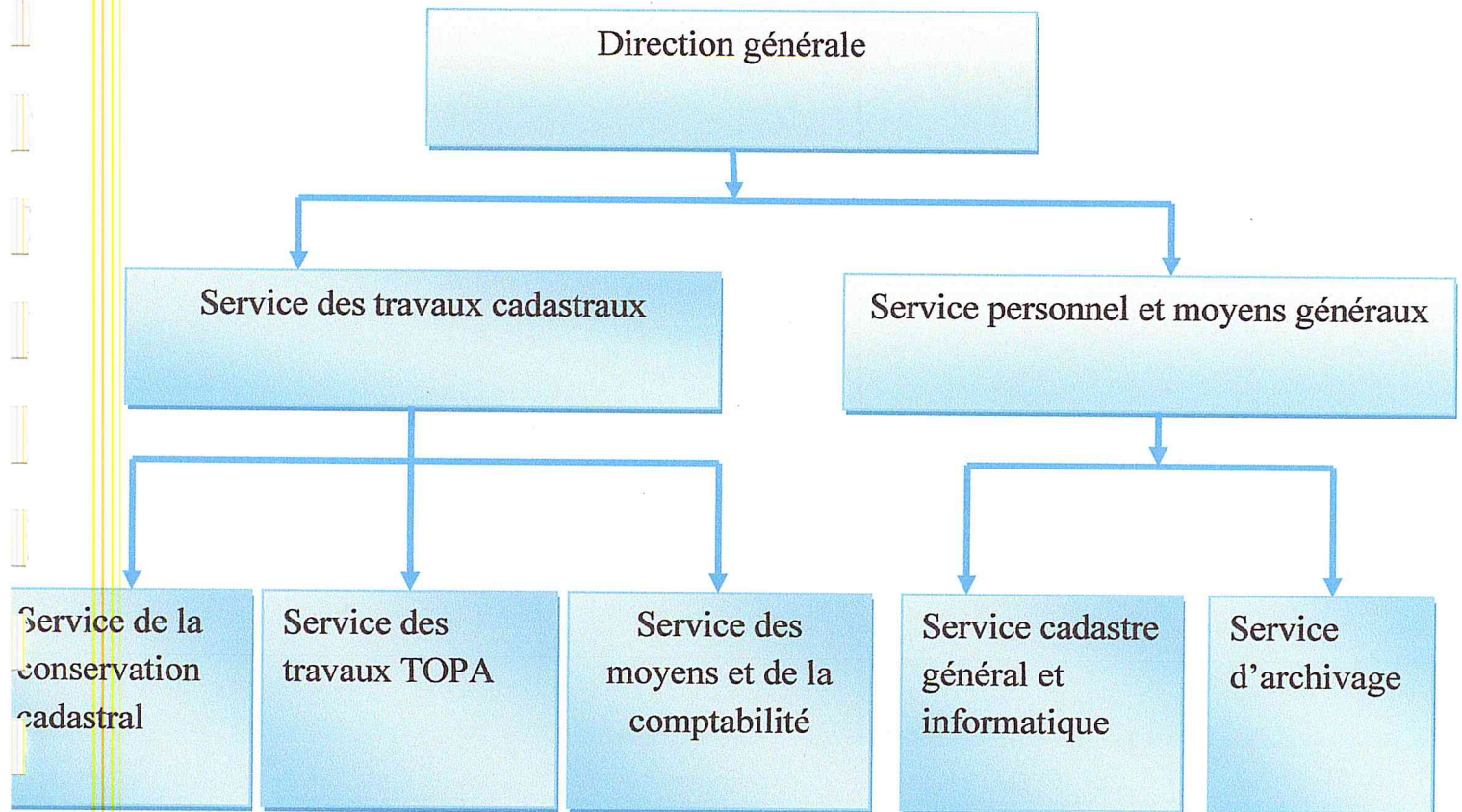


Figure 18: Organigramme général de la direction du cadastre de Blida.

1.5. Mission du cadastre

La direction du cadastre s'intéresse aux grandes principales missions qu'on peut le résumer comme suite:

- représentation graphiquement sur un plan tous les immeubles.
- préciser toutes les informations concernant :
 - La consistant physique ou matériel des immeubles.
 - Les propriétaires apparents de droits réel reconnus sur le terrain.

2.1. Situation informatique

2.1.1. Système de gestion de base de données(GIC) :

Le produit Gestion de l'Information Cadastre (GIC), est composé principalement de trois modules différents, à savoir :

1. Module Production:

Ce module est conçu pour permettre :

- La saisie de la donnée attributaire récoltée sur le terrain, lors de l'enquête cadastrale.
- L'édition de la documentation cadastrale tel que:
 - ✓ La matrice cadastrale concernant

L'impression de la matrice cadastrale se fait par :

- Commune (toute la donnée produite)
- Une ou plusieurs sections
- Un ou plusieurs comptes donnés
 - ✓ La table des comptes

L'impression de la table des comptes se fait par :

- Commune (toute la donnée produite)
- Une ou plusieurs sections (pour un dépôt partiel)
 - ✓ L'état de section

L'impression de l'état de section se fait par :

- commune
- Une ou plusieurs sections (pour un dépôt partiel)

2. Module mise à jour

Ce module permet la mise à jour de la donnée cadastrale, en traitant les cas de mutation, tel que:

Le partage, l'héritage, l'acquisition, donation, changement de limite,.....

3. Module consultation

Ce module permet la consultation de la donnée produite, tel que la recherche des îlots par nature juridique, par superficie, par localisant, affichage de la liste des biens d'un propriétaire donnée,.....

La base de donnée associée à cette application, est conçue pour représenté d'une manière transparente la logique cadastrale (une section est un ensemble d'îlot, un îlot est un ensemble de parcelles, une parcelle est un ensemble de bâtiments, un bâtiment est un ensemble de lots,....) elle comporte :

- La donnée en cour de production.
- L'archive de toute la donnée mis à jour (pour garder l'historique de la donnée)

2.1.2. Système d'information graphique (ArcGIC) :

Le produit Gestion de l'Information Cadastre (GIC), est composé principalement de deux modules différents, à savoir :

1. **Arc catalogue:** utilisé pour gérer la structure de base de données, l'ajout, la modification et suppression des couches des bases (chaque base de données contient plusieurs couches)
2. **Arc Map :** utilisé pour la manipulation de :
 - Production des dessins a partir des images satellites, des restitutions, et google earth.....etc .
 - Mise à jour et les modifications effectués (partage, changement des limites.....).

2.2. Pourquoi utiliser les deux logiciels :

Le logiciel le plus important dans le cadastre c'est l'Arc GIS, qui utilisé pour l'édition des extrait des plans cadastraux. Mais l'Arc gis n'a pas la possibilité de présenter les ilots, les propriétaires et leurs droit et les éditions. Donc le cadastre utilise un notre logiciel (GIC) pour représenter l'ilot, propriétaire, droit de propriétaire et les éditions.

2.3. Postes de travail :

Avant de traiter les différents processus métiers de cadastre on a besoin de connaître les tâches des différents postes actifs dans ces processus. Donc dans cette partie nous avons représenté les différentes tâches de chaque poste.

2.3.1. Poste externes :

- Service de la conservation foncière.
- Domaine et inspections.
- Notaire.
- Géomètre.

2.3.2. Poste internes :

- Directeur de cadastre.
- Service de secrétariat.
- Service des travaux topo.
- Service de la conservation cadastrale.

Poste de travail	Taches
Directeur de cadastre	<ul style="list-style-type: none"> • consulter tous les informations concernant le cadastre
Service de secrétariat	<ul style="list-style-type: none"> • la réception des documents cadastraux. • l'enregistrement des documents cadastraux dans le registre de départ ou bien dans le registre d'arrivé (accusé de réception). • la signature des documents (bordereaux d'envoi) effectué par le directeur su service.
Service travaux topo	<ul style="list-style-type: none"> • faire une sortie sur terrain pour réaliser : • un levé topographique • un croqué de délimitation
Service conservation cadastrale	<ul style="list-style-type: none"> • consultations des documents cadastraux. • réception (recevoir les clients). • Délivrance des documents. • Archivage des documents cadastraux (PR4bis les fiches rose).

Tableau 2: Postes de travail concernés par le cadastre

Remarque :

- Service de secrétariat c'est l'intermédiaire qui gère la correspondance entre l'interne et l'externe
- Service conservation cadastrale : s'occupe l'établissement et la délivrance des documentations cadastrales (état de section, feuille de matrice, plan cadastral) selon les opérations demandée par les services externes (domaine, conservation foncière, notaire, géomètre)

2.4. Description du processus métier

Afin de bien comprendre le processus métier actuel de la direction du cadastre, nous allons le décomposer en deux sous processus (processus de production, processus de mise à jour). Pour chaque sous processus, nous citerons son objectif, les acteurs intervenant pour son accomplissement ainsi que le flux d'information.

2.4.1. Processus de production

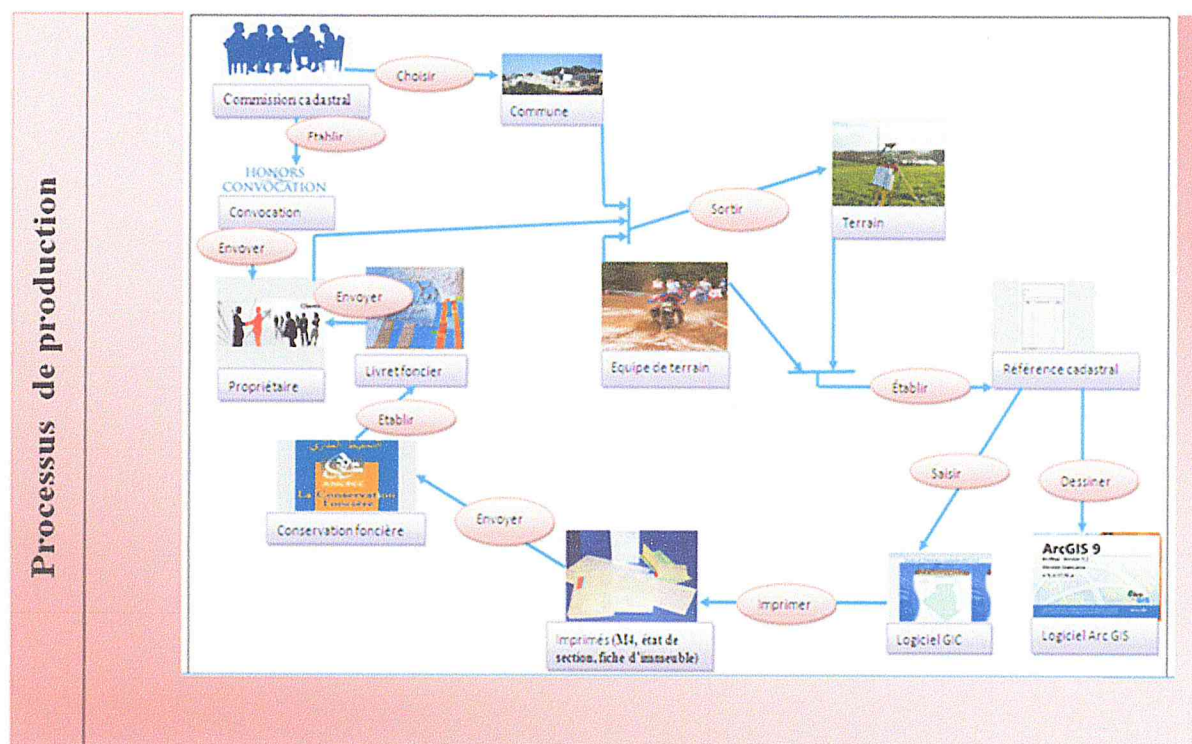


Figure 29: Processus de Production

But du sous processus : faire la délimitation de terrain pour l'immatriculation des propriétés et la délivrance d'un livret foncier titre de propriété

Acteurs impliqués :

- L'équipe de terrain (délimitateur et enquêteur).
- Commission cadastrale.
- Conservation foncière.
- Propriétaire.

Description du sous processus :

Le processus de production est un processus principal concernant le fonctionnement de cadastre, il inclut plusieurs opérations pour construire un plan cadastral au but d'établir un livret foncier de chaque propriétaire. Ces opérations peuvent se résumer comme suite :

Après le choix d'une commune par la commission cadastrale, elles effectuent les travaux à l'équipe de terrain pour la délimitation et l'enquête, au but de déterminer la consistance matérielle qui est la détermination des limites du terrain. Pour une identification adéquate en procède par les formalités suivantes :

- convocation des propriétaires
- Vérification de l'identité des propriétaires
- Reconnaissances des biens de chaque propriétaire
- Rapport des limites

Après la délimitation et enquête, les limites sont figurées au plan cadastral. Ce plan envoyer au niveaux de la direction cadastral pour l'établir la référence cadastral de chaque propriétaire ensuite, le bureau des applications informatiques fait la saisit les travaux dans le GIC et en même temps le dessin dans le ArcGIS. Après, la base de données de cette commune doit être fermé et mettre en circulation pour envoyer les fiches nécessaires à la conservation foncière qui est lui même établit un livret foncier de chaque propriétaire.

Schéma du processus

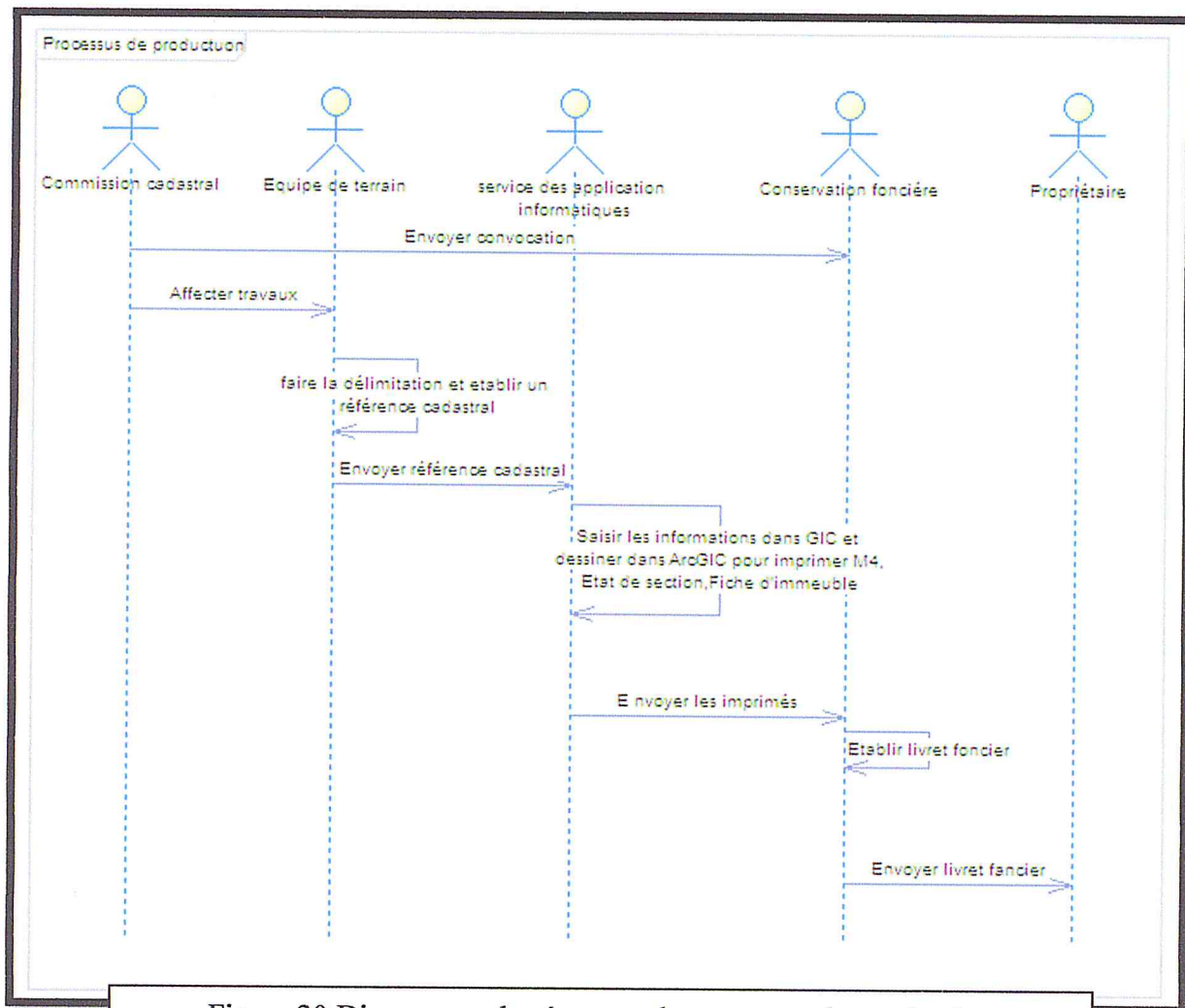


Figure 20:Diagramme de séquence de processus du production

2.4.2 Processus de mise à jour

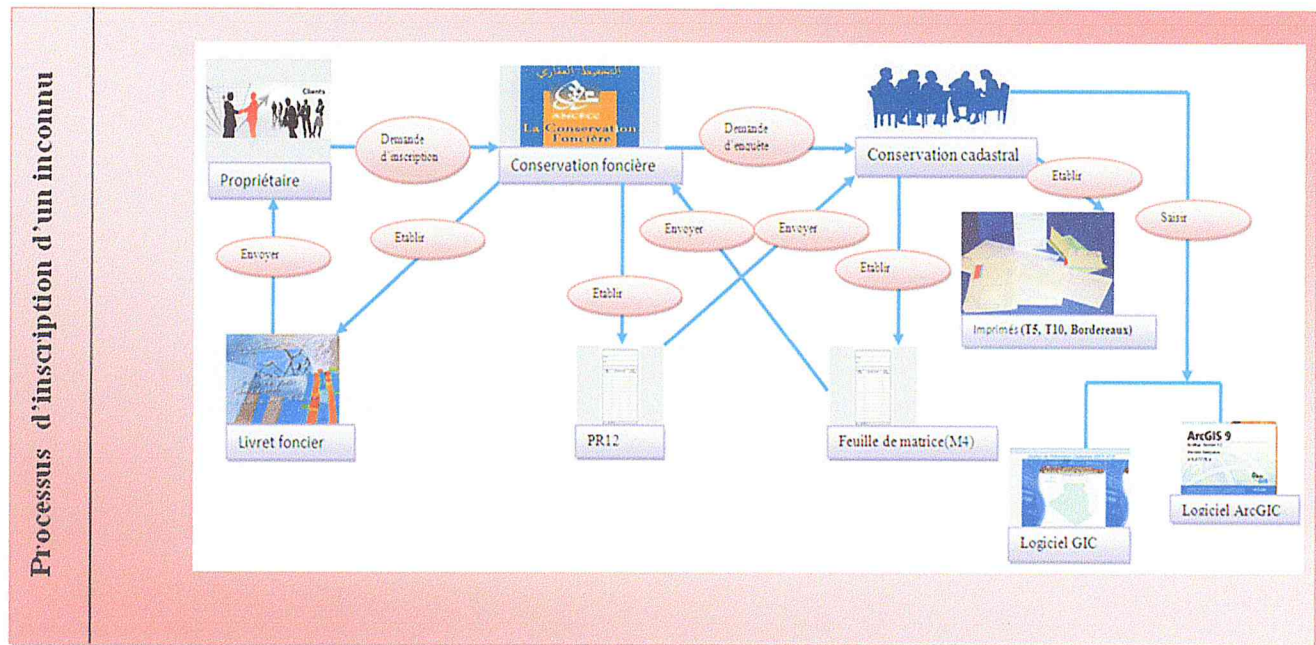
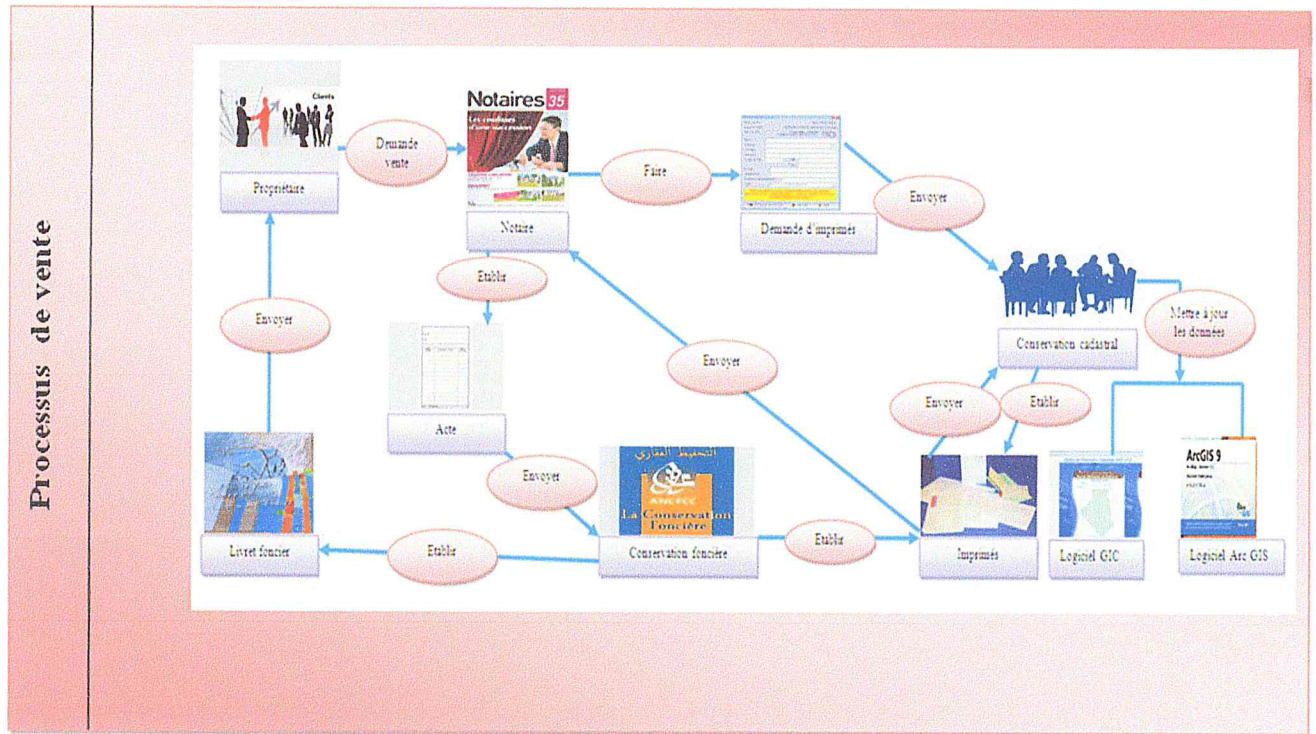


Figure 21:Processus de mise à jour

But du sous processus : faire la mise à jour des données des propriétés après une opération (vent, inscription....)

Acteurs impliqués :

- Conservation foncière.
- Conservation cadastral.
- Notaire.
- Propriétaire.

Description du processus :

Processus de vente : c'est une opération qui inclut deux cas :

Cas1 : opération de vente dans une zone non cadastrée :

On peut se résumer les étapes de cette opération comme suite :

- Propriétaire faire une demande de vente au notaire.
- Notaire envoi une demande de CC6 au service de cadastre.
- Cadastre établit CC6 et CC9 et pour envoyer au notaire.

Cas2 : opération de vente dans une zone cadastrée :

Qui contient les opérations suivantes :

- Propriétaire faire une demande de vente au notaire.
- Notaire envoi une demande de PR4bis au service de cadastre.
- Cadastre établit PR4bis et CC9 et bordereau et envoyer au notaire.
- Notaire établit l'acte pour envoyer à la conservation foncière.
- Conservation foncière publié l'acte et PR4bis et Bordereau pou envoyer au cadastre.
- Conservation cadastral faire la mise a jour dans GIC et ArcGIC et imprime feuille de matrice pour envoyer a la conservation foncière.
- Conservation foncière établit un livret foncier pour envoyer au propriétaire.

Processus d'inscription d'un inconnu : c'est une opération qui inclut deux cas :

Cas1 : Régularisation total :

- Propriétaire demande l'inscription à la conservation foncière.
- Conservation foncière demande l'enquête à la conservation cadastrale.
- Conservation cadastrale faire l'enquête et établit T5 et T10 et Bordereau et envoi T10 et Bordereau à la conservation foncière.
- Conservation foncière envoi Bordereau signé à la conservation cadastrale et établit livret foncier pour envoyer au propriétaire.
- Conservation cadastrale faire la saisit des données de propriétaire après la réception de Bordereau signé.

Cas2 : Régularisation partiel :

- Propriétaire demande l'inscription à la conservation foncière.
- Conservation foncière envoie l'acte et demande d'inscription nffa la conservation cadastrale.
- Conservation cadastrale faire l'enquête et établit T5 et T10 et Bordereau et envoi T10 et Bordereau à la conservation foncière.
- Conservation foncière établit PR12, PR14 publié pour envoyer à la conservation cadastrale et établit livret foncier pour envoyer au propriétaire.
- Conservation cadastrale faire la saisit des données de propriétaire et imprime feuille de matrice pour envoyer à la conservation foncière.

2.5 Critique de l'existant

Le diagnostic est nécessaire avant tout développement d'une application, et toute prise de décision au sein d'une organisation. Ce diagnostic permettra de dépister les différentes anomalies constatées et dégager les insuffisances dans le fonctionnement du système existant. Après l'étude de l'existant du cadastre, nous avons pu identifier les insuffisances qui caractérisent le système actuel selon trois volets :

2.5.1 Le système d'information opérationnel :

Le système actuel du cadastre présente un certain nombre d'insuffisances, qui peuvent se résumer en :

➤ La sécurité :

- Les utilisateurs des logiciels utilisent un même mot de passe pour la production, la mise à jour et la délivrance.
- Base de données Access 2000 qui implique :
 - ✓ Manque de sécurité.
 - ✓ Architecture monoposte donc la duplication de la base dans plusieurs PC.
 - ✓ La base de données défectueuse après plusieurs utilisations.
- Incrémentation de volume de donnée donc un temps énorme de recherche et d'insertion.

➤ Problème de codification :

- Un même propriétaire peut avoir un code dans chaque commune.
- Chaque commune représente une base donc y a pas la possibilité de faire la recherche par wilaya.

2.5.2 Le système décisionnel

La cadastre est une entreprise d'une bonne solution décisionnelle a un avantage certain, mais comme nous l'avons constaté, le système d'information actuel de la direction du cadastre ne répond pas à toutes les demandes quotidiennes des utilisateurs, en termes d'analyse. Ils se basent sur des statistiques et des rapports établis à partir des systèmes opérationnels et des données enregistrées dans les bases de données pour prendre des décisions influentes et assez importantes (c'est vrai ils font des statistique demandés par les directions national et régional manuellement). Cette méthode n'aboutit pas à des résultats fiables.

1 La base de données

Au début, la conception des bases de données utilisées par la direction du cadastre a été faite par un prestataire de service, qui est ensuite transformées en schémas de données et rassemblées dans une base de données.

La direction du cadastre utilise deux bases de données (littéral et graphique) qui sont représentés par les logiciels GIC et l'ArcGIS.

1.1 Base de données littéral

La base de données littérale permet de mettre des données du cadastre à la disposition d'utilisateurs pour une consultation, une saisie ou bien une mise à jour, tout en s'assurant des droits accordés à ces derniers. Cette base est centralisée, c'est-à-dire que les informations sont stockées sur des machines distantes et accessible par un réseau (sûr). Elle contient 26 tables, 14 tables vides et 11 tables utilisées. Le schéma ci-dessous représente le diagramme de classe de cette base.

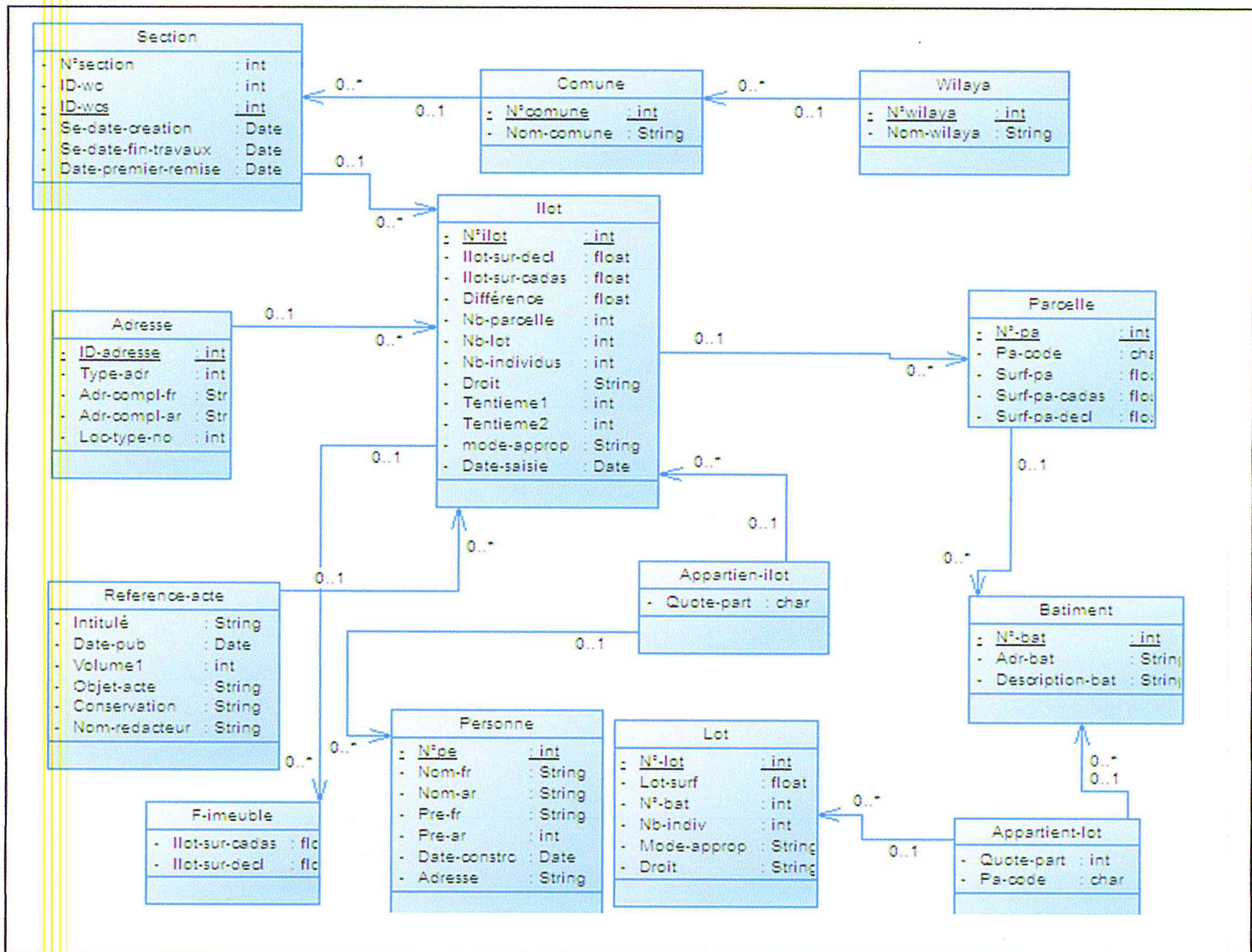


Figure 22:Diagramme de classe de la base littéral (Base GIC)

Nous avons pu recenser quelques anomalies concernant la base de données littérale qui sont :

- **Présence de tables non utilisées ou vides** : 53.84% des tables présentes dans la base de données littérale du cadastre sont vides. Si nous prenons l'exemple suivant :

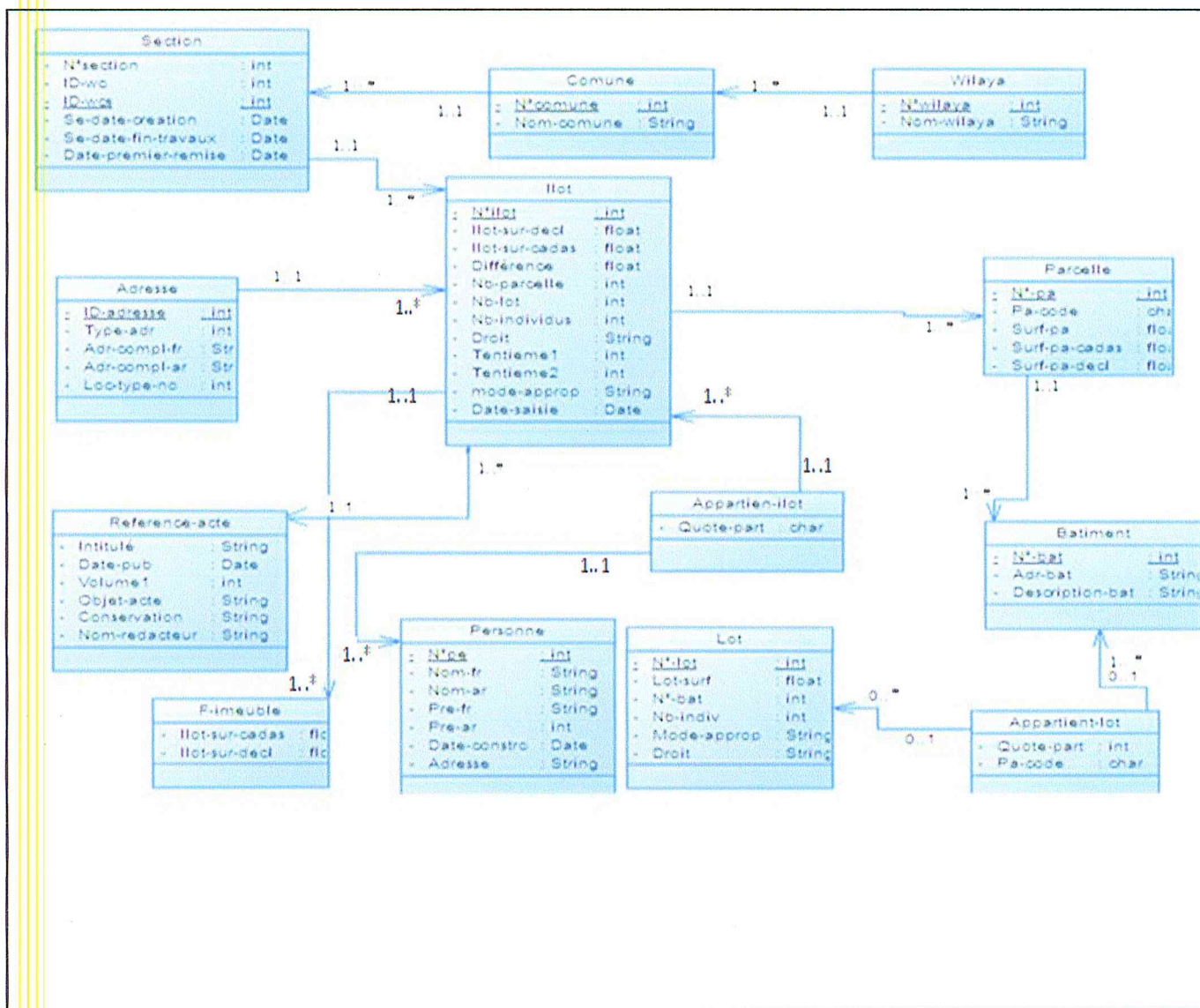


Figure 22:Diagramme de classe de la base littéral (Base GIC)

Nous avons pu recenser quelques anomalies concernant la base de données littérale qui sont :

- **Présence de tables non utilisées ou vides** : 53.84% des tables présentes dans la base de données littérale du cadastre sont vides. Si nous prenons l'exemple suivant :

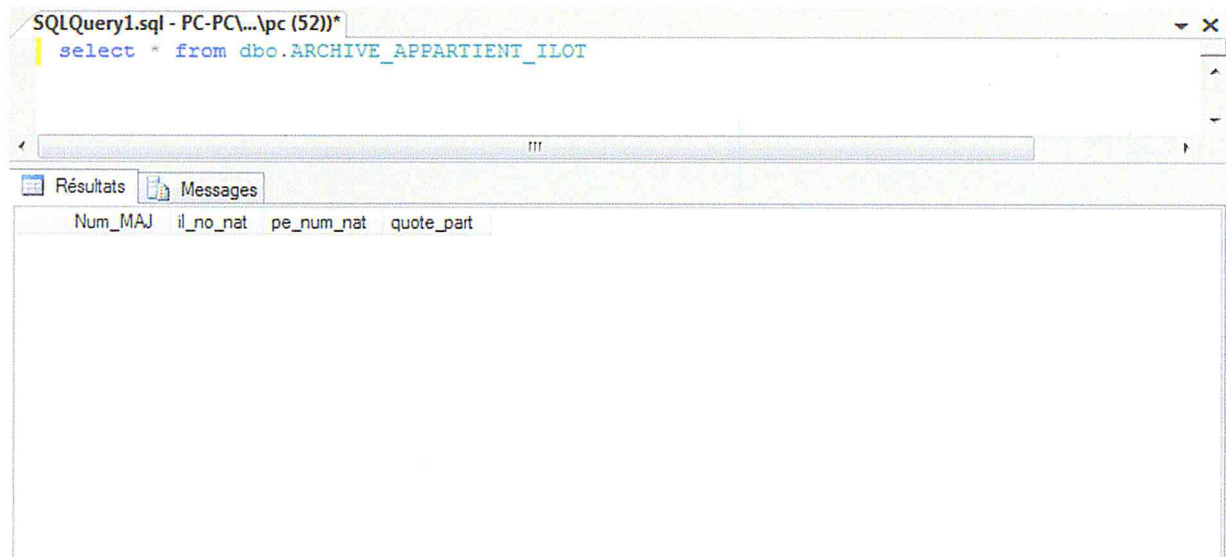


Figure 23 : Table ARCHIVE_APPARTIENT_ILOT

- **Plusieurs champs vides dans les tables :** Dans la table personne qui contient des informations sur les propriétaires, nous constatons 87.5 % de champs vides.

La figure ci-dessous illustre cette anomalie :

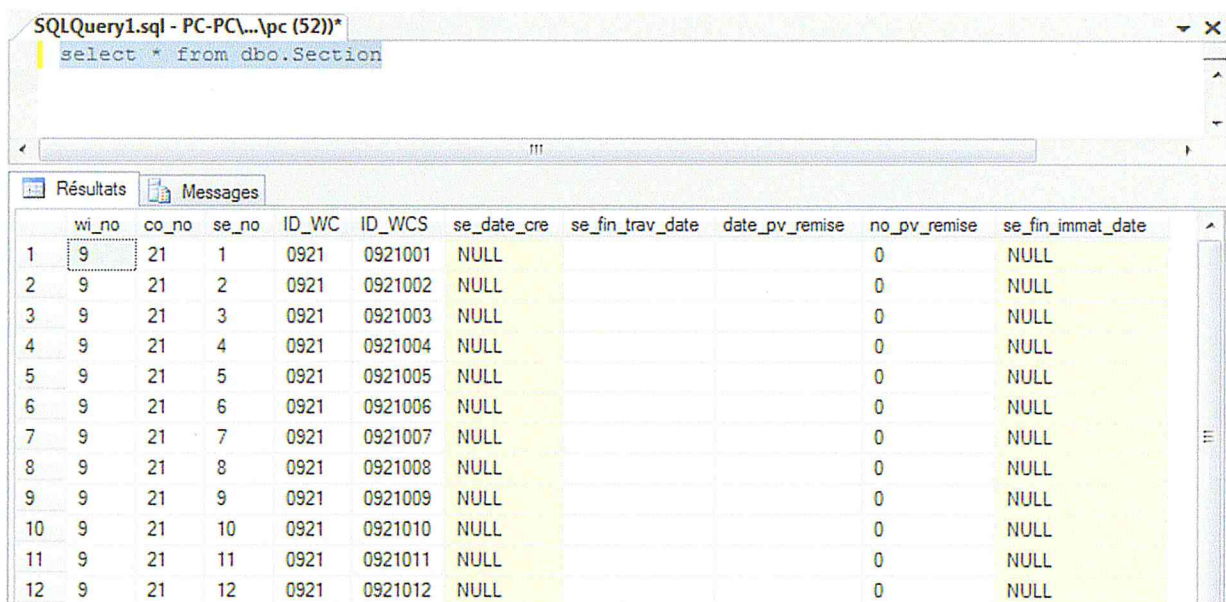


Figure 24: Table section

Les tables utilisées :

Nom-table	Nombre champs total	Nombre champs vide	Pourcentage de champs vides
Wilaya	3	0	0%
Commune	5	0%	0%
Ilot	26	8	30.77%
Parcelle	11	0	0%
Personne	32	28	87.5%
Section	10	5	50%

Tableau 3: les tables utilisées dans la BDD littéral

1.2 Bases de données graphique

La base de données graphique est une base de données qui contient tous les bases de données des communes, elle permettre aux utilisateurs de dessiner des plans à partir des images satellitaires. Elle contient 67 tables, 51 tables vides et 16 tables utilisées. Le schéma ci-dessous représente le diagramme de classe de cette base.

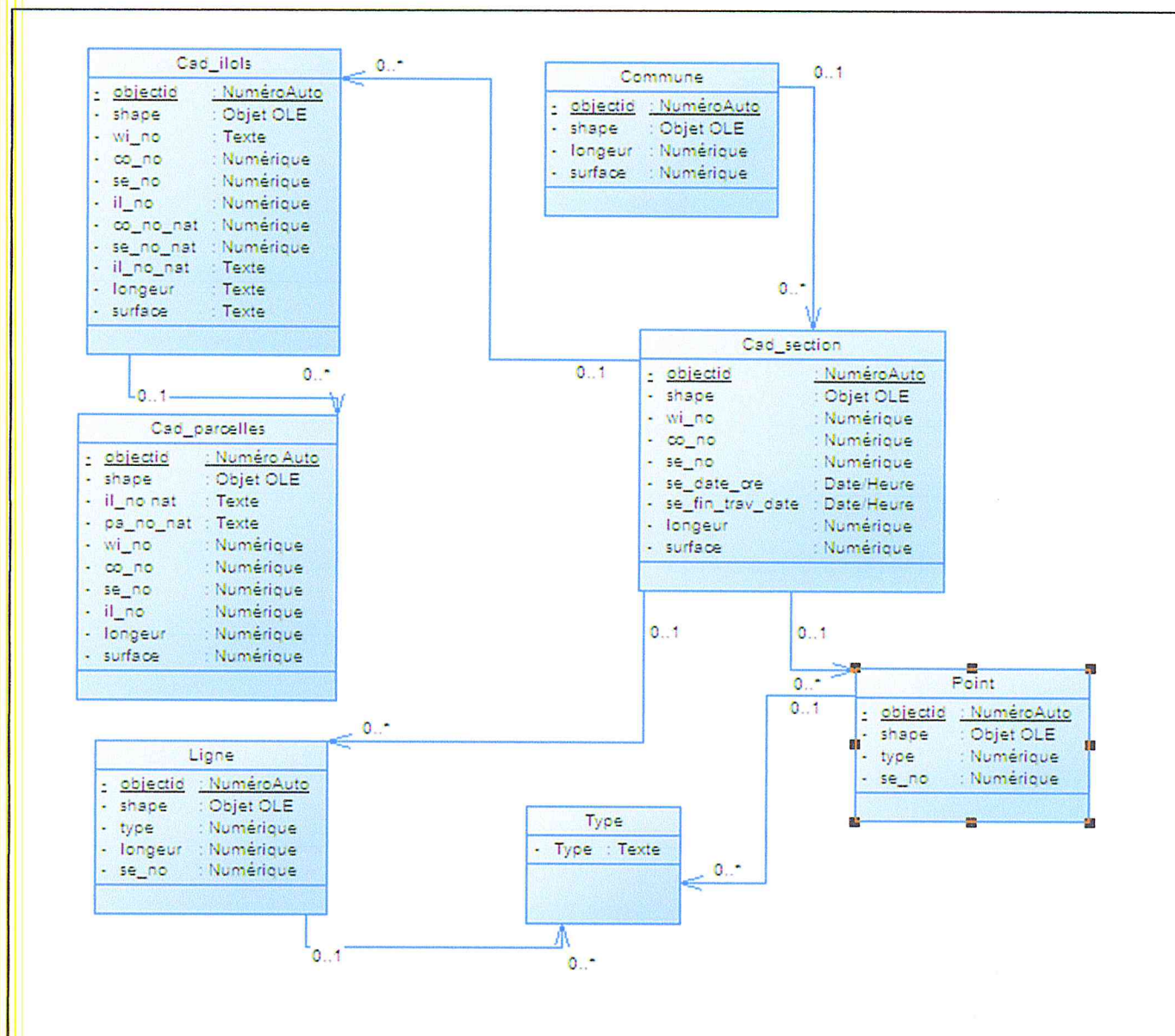


Figure 25: Diagramme de classe de base graphique (Base ArcGIC)

Nous avons pu recenser quelques anomalies concernant la base de données graphique qui sont :

- **Présence de tables non utilisées ou vides** : 76% des tables présentes dans la base de données graphique du cadastre sont vides ou non utilisées. Si nous prenons l'exemple suivant :

SQLQuery2.sql - PC-PC\...\pc (53))* SQLQuery1.sql - PC-PC\...\pc (52))*

```
select * from dbo.Cad_IlotArchive
```

Résultats Messages

OBJECTID	SHAPE	CodeBienCF	wi_no	co_no	se_no	il_no	site_id	bien_no	co_no_nat	se_no_nat	Num_maj	il_no_nat	ilo
----------	-------	------------	-------	-------	-------	-------	---------	---------	-----------	-----------	---------	-----------	-----

Figure 26 : Table Cad_IlotArchive

- **Plusieurs champs vides dans les tables** : Dans la table cad_section contenant des informations sur les sections, nous constatons 45 % de champs vides.

La figure ci-dessous illustre cette anomalie (les champs wi_no, ID_WC, ID_WCS, se_fin_trav_date,... sont vides)

SQLQuery2.sql - PC-PC\...\pc (53))* SQLQuery1.sql - PC-PC\...\pc (52))*

```
select * from dbo.Cad_Section
```

Résultats Messages

	wi_no	ID_WC	ID_WCS	co_no	se_no	se_date_cre	se_fin_trav_date	pv_remise_date	pv_remise_no	s
1	1986E128...	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	↑
2	117F6ABC...	NULL	NULL	NULL	15	NULL	NULL	NULL	NULL	↑
3	14A0C02A...	NULL	NULL	NULL	16	NULL	NULL	NULL	NULL	↑
4	1713D0A5...	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	↑
5	FCA9F152...	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	↑
6	9CC420B0...	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	↑
7	18B6CE77...	9	NULL	21	46	2013-09-04 00:00:00.000	NULL	NULL	NULL	↑
8	1F4FDD4F...	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	↑
9	11A69BC4...	NULL	NULL	NULL	14	NULL	NULL	NULL	NULL	↑
10	1A69BC4A...	NULL	NULL	NULL	14	NULL	NULL	NULL	NULL	↑
11	175931804...	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	↑
12	4185EB51...	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	↑

Figure 27 : Table Cad_Section

Les tables utilisées :

Nom table	Les champs vides	Les champs remplis	Pourcentage de remplissage
Cad_Ilots	51	11	18%
Cad_Parcelles	10	10	50%
Cad_Section	7	8	53%

Tableau 4:les tables utilisées dans BDD graphique

2.6. Présentation de la démarche d'identification des besoins

L'identification des besoins selon une démarche mixte consiste à interroger les personnes du métier et à analyser les sources de données. Elle paraît évidente et simple mais en réalité c'est une étape rigoureuse qui nécessite la réflexion, afin de bien comprendre le travail, les objectifs et les enjeux du métier de la direction du cadastre, ainsi que la manière actuelle et à venir de prendre les décisions. Pour cela, nous allons suivre la démarche ci-dessous :

2.7. Préparation et conduite d'entretiens

La méthode de collecte des besoins et d'étude des données opérationnelles pour laquelle nous avons opté est les entretiens. Elle permet de ne prendre que peu de temps aux personnes interrogées et génère une grande quantité de données. Pour mener à bien ces entretiens, nous avons procédé comme suit :

- **Identification des personnes à interroger :** Ces personnes ayant différentes fonctions dans l'entreprise pouvant nous aider à comprendre le métier et recenser les informations nécessaires.
 - Chef de bureau du cadastre général,
 - Chef de bureau de la conservation de Boufarik,
 - Membres des travaux de terrain,
 - L'Informaticien.

2.8. Documentation

Bien que les entretiens représentent la méthode d'identification des besoins la plus utilisée, les informations fournies par les utilisateurs sont insuffisantes pour la construction d'un entrepôt de données. De ce fait, une documentation utilisée par les différents services du cadastre, nous a été fournie afin de compléter et clarifier les informations recueillies à travers les entretiens.

2.9. Etude des sources de données

Cette phase consiste à détecter les besoins relatifs aux données qui n'ont pas été exprimés par les utilisateurs finaux et de valider les sources de données afin de vérifier leur fiabilité. Dans un premier temps, nous avons fait une analyse générale des sources de données, afin d'avoir une idée globale sur chacune d'entre elles ainsi que leur disponibilité, et passer par la suite à une étude plus approfondie.

La base de données utilisée par le cadastre permet de fournir presque toute les informations nécessaires pour l'alimentation de l'ED. Lors de notre étude détaillée des bases de données sources ; nous avons identifié deux types de base de données : base de données littérale et base de données graphique.

2.10. Analyse des besoins

1. Les besoins fonctionnels

Le tableau ci-dessous représente une synthèse sur les besoins fonctionnels détectés et exprimés lors de la collecte d'informations et classés par sujet d'analyse.

Sujet d'analyse	Besoins analytiques
Section	<ul style="list-style-type: none"> • Nombre de section GIC par commune, wilaya, date de création. • Nombre de section graphique par commune, wilaya, date de création. • Surface cadastrée graphique par section, commune, wilaya.
Ilot	<ul style="list-style-type: none"> • Nombre d'ilot littéral par section, commune, wilaya. • Nombre d'ilot graphique. • Surface ilot cadastré dans GIC par section, commune, wilaya. • Surface ilot cadastré dans ArcGIS par section, commune, wilaya. • Nombre des parcelles, personnes, bâtiments par section, commune, wilaya.

Parcelle	<ul style="list-style-type: none"> • Nombre de parcelles littérales par ilot, section, commune, wilaya. • Nombre de parcelles graphique. • Surface cadastré des parcelles dans GIC par ilot. • Surfaces cadastrées des parcelles dans ArcGIS.
-----------------	---

Tableau 5 : Besoins analytiques classés par sujet d'analyse.

2. Identification des utilisateurs finaux du nouveau système

Après avoir recensé les différents besoins analytiques exprimés par l'entreprise, il est essentiel d'identifier les utilisateurs finaux du nouveau système et qui sont les suivants :

Analyste cadastre de Blida	qui a besoin d'avoir des statistiques sur les travaux réalisées par l'équipe de terrain(les sections, les ilots, et les surfaces cadastrées).
Les directions régional et national(ANC)	qui a besoin d'avoir des statistiques sur les travaux réalisées par la direction de la wilaya de Blida.

Tableau 6:Utilisateurs finaux de l'ED

Conclusion

À travers ce chapitre, nous avons introduit le périmètre de notre étude, ses missions et sa structure. En analysant l'existant informatique de cette direction, nous avons pu comprendre son processus métier, que nous avons décomposé en deux sous processus, ainsi que les faiblesses et les forces du système utilisé (GIC et ArcGIS). Et aussi, nous étudierons les besoins des utilisateurs finaux de notre nouveau système. Dans le chapitre suivant on va entamer la conception de la solution de notre projet.

Chapitre 03 : Conception de l'entrepôt de données

Introduction

Après avoir analysé les besoins des utilisateurs pour lesquels l'ED doit répondre, nous voilà prêts à lancer la conception de notre solution qui doit satisfaire les besoins exprimés avec l'existant informatique et organisationnel de la direction du cadastre.

La conception physique et la conception logique représentent la pierre angulaire de l'entrepôt. Elles permettent de planifier l'extraction des données et les étapes de la transformation des données sources en structures d'ED finalisées [KIMBALL, 2011]. Cette partie est consacrée pour la conception de l'entrepôt de données que nous avons effectuée en trois étapes :

- Conception de la zone d'entreposage,
- Conception de la zone d'alimentation,
- Conception des cubes dimensionnels.

1. Conception de la zone d'entreposage

1.1. Matrice de l'architecture en bus décisionnel

Le meilleur moyen de commencer la conception d'un entrepôt de données est de construire une matrice de magasins de données et de dimensions [KIMBALL, 2011]. L'élaboration de cette matrice consiste en premier lieu à recenser l'ensemble des magasins de données qui constituent une collection de faits numériques, pour ensuite identifier les dimensions impliquées dans ces magasins.

Table de faits Table de dimensions	Suivi des sections	Suivi des ilots	Suivi des parcelles
Section	✓	✓	✓
Commune	✓	✓	✓
Wilaya	✓	✓	✓
Ilot		✓	✓
Parcelle			✓
Dimension graphique	✓	✓	✓

Tableau 7: Matrice de l'architecture en bus dimensionnel

1.2. Modélisation des activités

Après avoir défini les besoins de l'entreprise ainsi que les magasins de données et les dimensions dans une matrice de bus décisionnel, nous passons maintenant à la construction du modèle dimensionnel. Selon **Kimball**, le processus de modélisation se décompose en quatre étapes :

- **Choix du magasin de données** : cette étape consiste, en premier temps, à choisir la source de données opérationnelle, c'est-à-dire construire un magasin de données mono source afin de réduire les risques liés au développement. Pour passer en second temps aux magasins de données multi sources.
- **Définition du grain** : la définition de la granularité de la table de faits, revient à spécifier exactement ce que sera un enregistrement de la table de faits dans la dimension concernée. Et ce, en déclarant l'expression la plus fine qui contiendra le besoin de l'utilisateur de manière plus détaillée.
- **Choix des dimensions** : Après le choix de la granularité, l'identification des dimensions est assez simple. Selon **Kimball** : « *Le temps est la seule dimension qui figure systématiquement dans tout entrepôt de données, et plus souvent la première dimension dans le classement sous-jacent de la base de données* » [KIMBALL, 2011]. Un grand nombre de dimensions supplémentaires peut être rajouté, dans la mesure où le concepteur prend une seule valeur dans le contexte des mesures fournies.
- **Choix des faits** : La granularité de la table de faits permet aussi d'identifier les faits à mesurer. Les tables de faits peuvent intégrer un grand nombre de faits qui contiendront les mesures avec lesquelles l'utilisateur pourra analyser son activité.

1.2.1. Activité de suivi des sections

- **Déclaration du grain** :

Le grain de cette activité se situe dans le suivi de nombre de section réalisées ainsi que la surface cadastrée. Ce suivi s'effectue selon la commune, et la wilaya.

- **Choix des dimensions** :

Les dimensions liées à la table de faits 'FAIT Section' sont :

- Dimension section

Désignation	Détails
Wi_no	Numéro de wilaya
Co_no	Numéro de commune
Se_no	Numéro de section
Id_wc	Identificateur wilaya commune
Id_wcs	Identificateur wilaya commune section

Tableau 8: Dimension section

- Dimension commune

Désignation	Détails
Id_wc	Identificateur wilaya commune
wi	Nom de Wilaya
Co_no	Numéro de commune
Co_nom_fr	Nom commune en français
Co_no_ar	Nom commune en arabe

Tableau 09: Dimension commune

- Dimension wilaya

Désignation	Détails
Wi_no	Numéro de wilaya
Wi_nom_fr	Nom de wilaya en français
Wi_nom_ar	Nom de wilaya en arabe

Tableau 10: Dimension wilaya

- Dimension géographique_section

Désignation	Détails
ID_WCS	Identificateur de section
SHAPE_Area	Surface de section

Tableau 11: Dimension géographique_section

• **Choix des faits :**

Les mesures que nous pouvons déduire à partir des faits précédents sont :

- nb_section : correspond au nombre de sections.
- Surf_cadastree_gic : correspond à la surface cadastrée littérale des sections.
- Surf_section_gra: correspond à la surface graphique des sections.

• Représentation graphique :

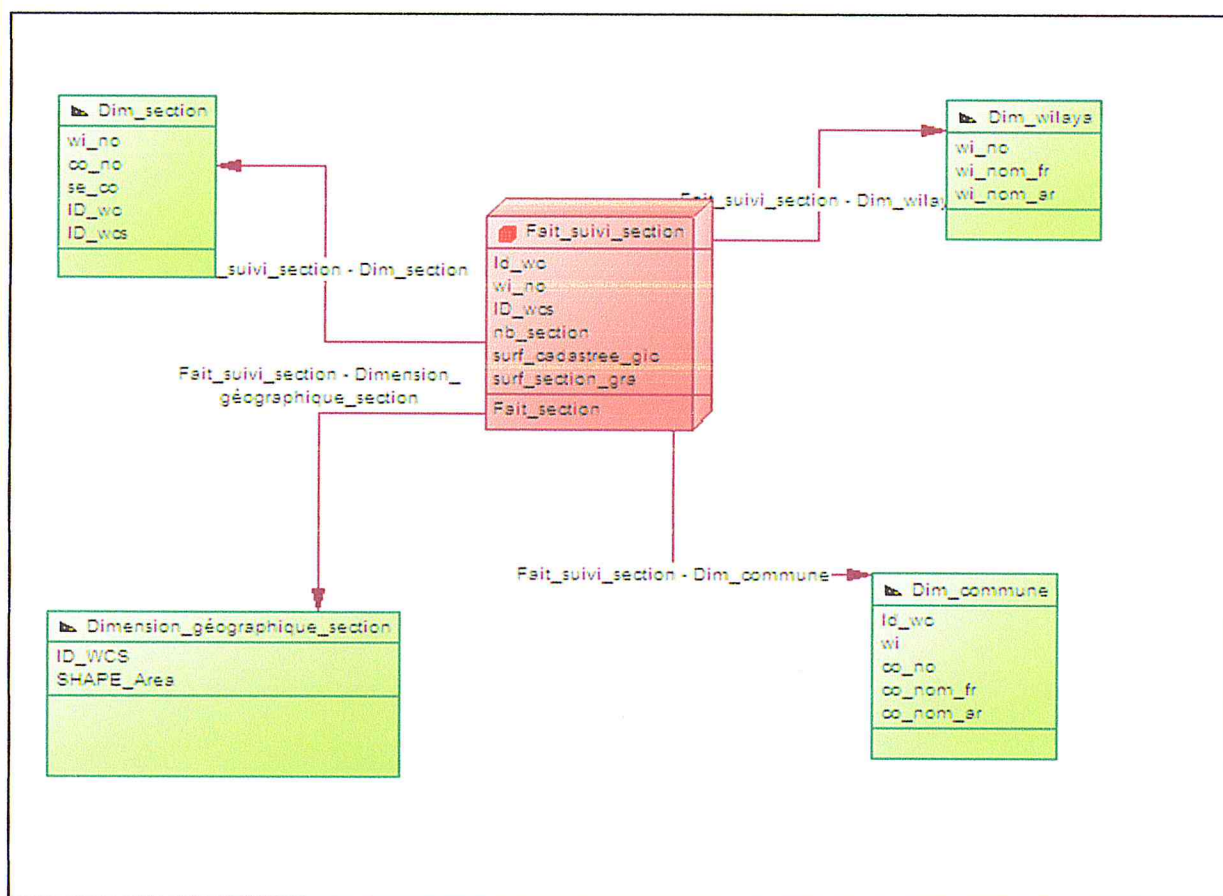


Figure 28 : Schéma en étoile de suivi des sections

1.2.2. Activité de suivi des ilots

• Déclaration du grain :

Le grain de cette activité se situe dans le suivi du nombre d'ilots littéral et graphique, la différence d'ilots entre les deux logiciels, ainsi que la différence surface entre les deux logiciels.

• Choix des dimensions :

Les dimensions liées à la table de faits 'FAIT_Ilots sont :

- Dimension ilot

Désignation	Détails
il_no_nat	Numéro d'ilot
il_surf_cadastree	Surface cadastré d'ilot
nb_parcelle	Nombre de parcelles
nb_indiv	Nombre des individus
nb_batis	Nombre des bâtiments
date_sais	Date de sais

Tableau 12: Dimension ilot

- Dimension section

Désignation	Détails
Wi_no	Numéro de wilaya
Co_no	Numéro de commune
Se_no	Numéro de section
Id_wc	Identificateur wilaya commune
Id_wcs	Identificateur wilaya commune section

Tableau 13: Dimension section_a

- Dimension commune

Désignation	Détails
Id_wc	Identificateur wilaya commune
wi	Nom de Wilaya
Co_no	Numéro de commune
Co_nom_fr	Nom commune en français
Co_no_ar	Nom commune en arabe

Tableau 14: Dimension commune_a

- Dimension wilaya

Désignation	Détails
Wi_no	Numéro de wilaya
Wi_nom_fr	Nom de wilaya en français
Wi_nom_ar	Nom de wilaya en arabe

Tableau 15: Dimension wilaya_a

- Dimension géographique_ilot

Désignation	Détails
il_no_nat	Numéro d'ilot
SHAPE_Area	Surface d'ilot

Tableau 16: Dimension géographique_ilot

• **Choix des faits :**

Les mesures que nous pouvons déduire à partir des faits précédents sont :

- Nbr_ilot_gic : nombre d'ilot dans le la base littérale (GIC).
- Nbr_ilot_gra : nombre d'ilot dans le la base graphique(ArcGIS).
- Suraface_ilot_gic: surface d'ilot dans le la base littérale (GIC).
- Surface_ilot_graphique : surface d'ilot dans le la base graphique (ArcGIS).
- Nb_parcelle : nombre de parcelle.
- Nb_individus : nombre_d'individus.
- Nb_batis : nombre de bâtiments

• **Représentation graphique :**

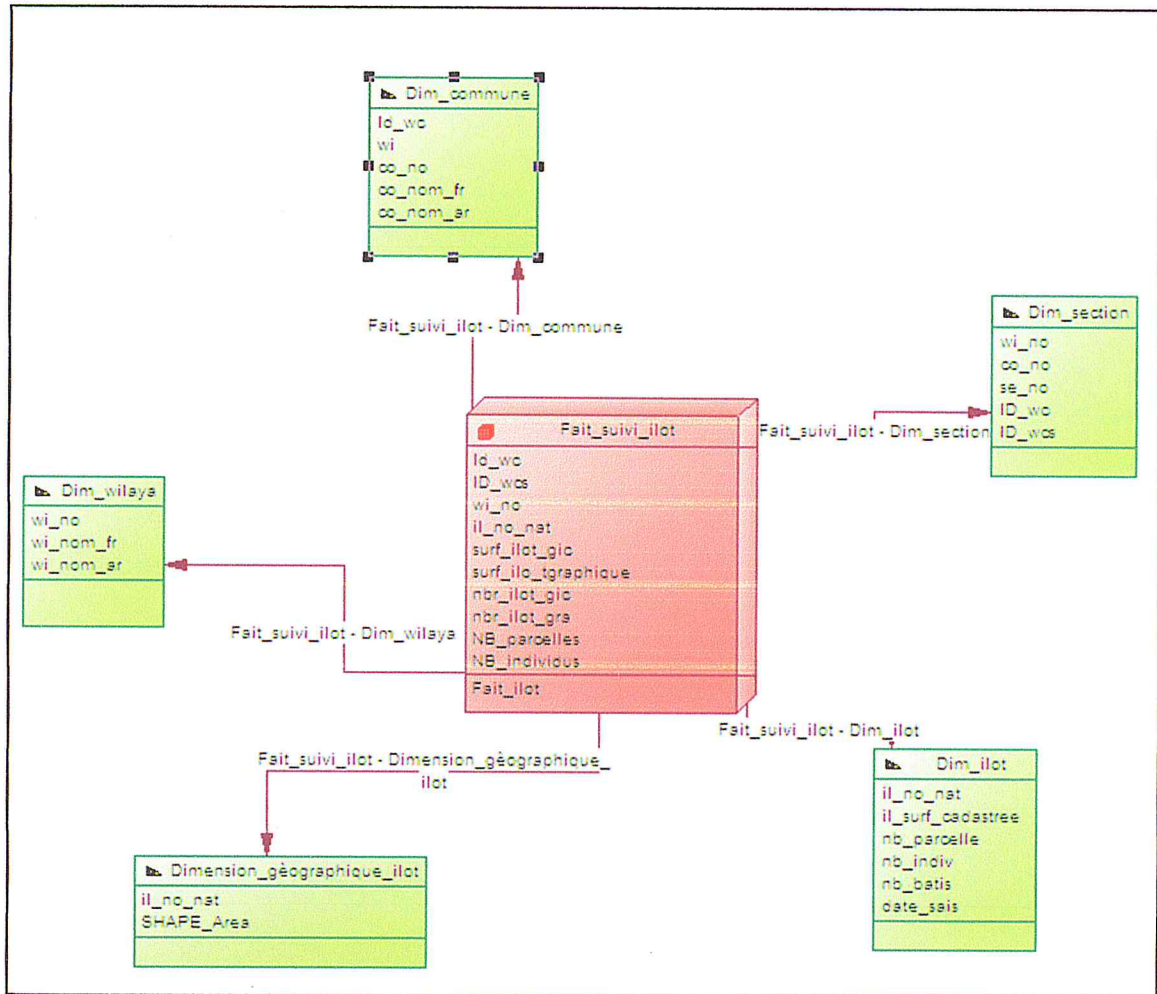


Figure 29 : Schéma en étoile de suivi des ilots

1.2.3. Activité de suivi des parcelles

La facturation représente la dernière activité du processus métier de la capitainerie, qui consiste à établir les factures, communiquées par la suite au client concerné et à la direction finance et comptabilité.

• **Déclaration du grain :**

Le grain de cette activité se situe dans la réalisation des factures clientes, en estimant le chiffre d'affaire moyen, mensuel et annuel, selon le consignataire, le type du navire, le pavillon et la nature des prestations fournies.

• **Choix des dimensions :**

Les dimensions liées à la table de faits 'FAIT_parcelle' sont :

- Dimension parcelle

Désignation	Détails
pa_no_nat	Numéro de parcelle
il_no_nat	Numéro d'ilot
pa_surf_cadastree	Surface parcelle cadastrée

Tableau 17: Dimension parcelle

- Dimension ilot

Désignation	Détails
il_no_nat	Numéro d'ilot
il_surf_cadastree	Surface ilot cadastrée
nb_parcelle	Nombre parcelle
nb_indiv	Nombre individus
nb_batis	Nombre batis
date_sais	Date de saisie

Tableau 18: Dimension ilot_a

- Dimension géographique_parcelle

Désignation	Détails
pa_no_nat	Numéro de parcelle
SHAPE_Area	Surface de parcelle

Tableau 19: Dimension géographique_parcelle

- Dimension_section

Désignation	Détails
Wi_no	Numéro de wilaya
Co_no	Numéro de commune
Se_no	Numéro de section
Id_wc	Identificateur wilaya commune
Id_wcs	Identificateur wilaya commune section

Tableau 20: Dimension section_b

- Dimension_commune

Désignation	Détails
Id_wc	Identificateur wilaya commune
wi	Nom de Wilaya
Co_no	Numéro de commune
Co_nom_fr	Nom commune en français
Co_no_ar	Nom commune en arabe

Tableau 21: Dimension commune_b

- Dimension_wilaya

Désignation	Détails
Wi_no	Numéro de wilaya
Wi_nom_fr	Nom de wilaya en français
Wi_nom_ar	Nom de wilaya en arabe

Tableau 22: Dimension wilaya_b

• Choix des faits :

Les mesures que nous pouvons déduire à partir des faits précédents sont :

- Nbr_parcelle_litt : nombre de parcelle littérale.
- Nbr_parcelle_gra : nombre de parcelle graphique.
- Surf_parcelle_gra : surface cadastrés des parcelles dans la base graphique.
- Surf_parcelle_cadastrée : surface cadastrés des parcelles dans la base littérale

- Représentation graphique :

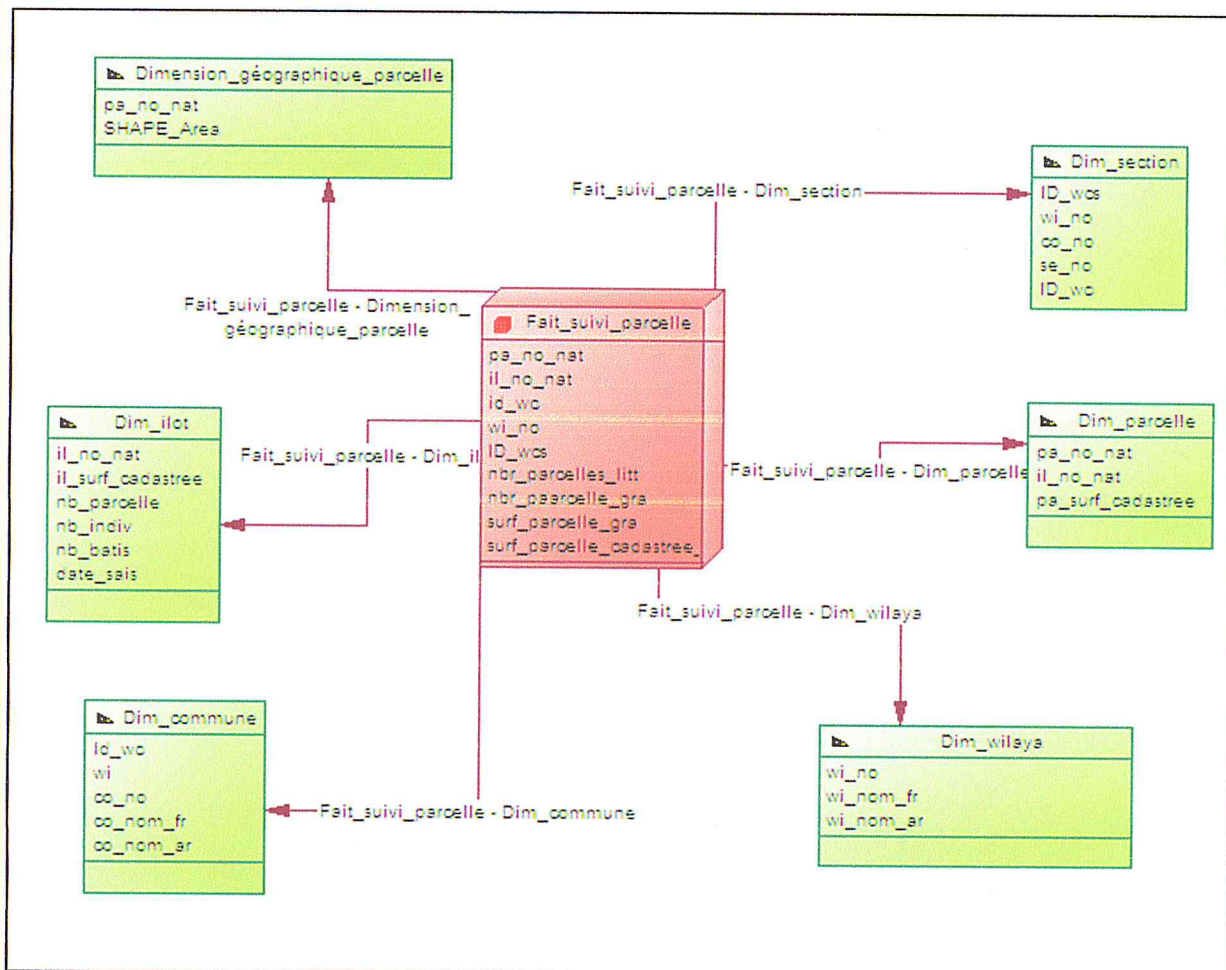


Figure 30: Schéma en étoile de suivi des parcelles

2. Conception de la zone d'alimentation

2.1. Etude des sources de données

Le modèle de données du système source peut aller du fichier séquentiel hiérarchique complètement dénormalisé à une structure en troisième forme normale, en fonction du système source. Pour créer l'architecture de la zone de construction (back room = zone d'entreposage et zone d'alimentation), il est essentiel de comprendre la nature des systèmes sources [KIMBALL, 2011].

Afin de déceler les tables de données porteuses d'information, nous avons examiné les bases de données (base GIC et base graphique) utilisées dans le cadastre. Ces bases de données sont gérées par le Microsoft Access, utilisés par différentes directions.

Cette phase nécessite beaucoup de réflexion et de concentration. Pour la mener à bien, nous avons listé en premier lieu les données nécessaires ainsi que leurs emplacements. Ensuite, nous avons

préparé les transformations nécessaires à appliquer et les liens logiques entre les données sources et cibles. Voici les tables nécessaires pour l'entrepôt :

Nom de la table	Description
SECTION	Il s'agit de la table qui permet de connaître l'identificateur de section, et le numéro de section dans une commune
ILOT	Il s'agit de la table qui permet de connaître le numéro d'ilot dans la section, le nombre de parcelle, nombre d'individus, nombre de bâtiment dans cet ilot. Chaque ilot est identifié par un numéro (il_no_nat).
PARCELLE	Il s'agit de la table qui permet de connaître le numéro de parcelle, l'ilot correspond a cet ilot, et la surface de parcelle. Chaque parcelle est identifiée par un numéro (pa_no_nat).
PERSONNE	Cette table regroupe toutes les personnes qui font inscrits dans le cadastre. Chaque ligne de la table comprend un ensembles de champs, (pe_num_nat) qui est le numéro de personne, (nom_fr) le nom en français ,(adresse_fr) la adresse en français.
COMMUNE	Il s'agit de la table qui permet de connaître le numéro de la commune, le nom de la commune.
DIMENSION Géographique	Il s'agit de la table qui permet de connaître la surface cadastré pour les sections et ilots, parcelles.

Tableau 23: Tables nécessaires de la source de données

2.2. Processus d'alimentation de l'ED

Après avoir conçu le modèle de données, comment alimenter l'ED ? Cette étape n'est pas un simple programme d'extraction, transformation et de chargement. Il s'agit plutôt de l'activité la plus importante du projet en termes de charge de travail et de complexité des tâches, car elle permet de passer de la structure de données source à la structure dimensionnelle de l'ED. L'alimentation d'un ED est un processus qui s'effectue en plusieurs étapes :

2.2.1. Processus d'extraction des données

La zone de préparation des données, transparente pour les utilisateurs, est le chantier de l'entrepôt de données. Elle constitue l'endroit où s'effectue la transformation des données et où une grande partie de la valeur ajoutée de l'ED se crée [KIMBALL, 2011].

Cette zone joue un rôle central car, le plus souvent, elle évolue pour devenir le « système source d'enregistrements » de tous les environnements en aval et des interfaces entre les systèmes.

L'avantage de son utilisation est de ne pas faire les transformations en même temps que les extractions et donc la libération du système source dans les plus brefs délais [KIMBALL, 2011]. Tout au long de la construction de l'ED, plusieurs types d'extraction sont envisageables, qui sont :

Types d'extraction	Description
Chargements incrémentaux	Vu le volume important de l'entrepôt de donnée, le chargement de toutes les données des tables de faits dans une seule fenêtre de chargement est délicat, voire impossible dans certains cas. Pour cela, des chargements sont effectués de manière incrémentale.
Événements transactionnels	Il est important de savoir les différentes modifications effectuées sur les enregistrements (création, mises à jour, ou suppression) depuis le chargement précédent. Il est donc nécessaire de faire recours à d'autres méthodes afin d'identifier les enregistrements effectués. Pour cela un fichier journal de transactions pourra être sollicité.
Actualisation complète	Ce type d'extraction consiste à remplacer intégralement le contenu des tables plutôt que de chercher à identifier les changements survenus. Elle est utilisée comme première extraction sur le système source qui concerne les données les plus récentes disponibles sur ce système. Dans certains cas, le système transactionnel modifie l'historique sans laisser de traces de la mise à jour, il est alors recommandé de faire une actualisation complète périodiquement.

Tableau 24: Types d'extraction de données [KIMBALL, 2011]

Afin de mener à bien le processus d'extraction, notre choix s'est porté sur la technique d'Actualisation complète comme première extraction, afin de charger initialement l'ED. Cette technique consiste à extraire la totalité des données à partir du système source.

Vu le volume de données du système transactionnel (les deux bases de données contiennent 120 tables et certaines tables contiennent plus de 1000 enregistrements).

2.2.2. Transformation des données

Avant de commencer l'étape de transformation, nous avons d'abord vérifié la cohérence des données existantes dans la source : Vérification des clés primaires des tables (unique et non nulles), des clés étrangères dans les tables référencées, ...etc. les données de chaque champ sont saisies sous le même format, même taille et sous les mêmes contraintes d'intégrité.

Une fois les données extraites, elles subissent une série de transformations destinées à les modifier en informations purifiées, consolidées et présentables aux utilisateurs. Nous avons alors réalisé les différents types de transformations des données suivantes :

- **Traitement des champs vides**

Nous avons supprimé plusieurs champs vides pour notre entrepôt, dans les tables sources dont nous avons besoin. Par exemple dans la table personne, nous disposons de 21 colonnes, mais pour notre entrepôt nous n'avons besoin que de 5 colonnes :

1 • SELECT * FROM caefastre_gic.personnes;

pe_num	nom_fr	adresse_fr	code_ad	nom_pere_fr	nom_pere_ar	pre_nere_fr	pre_nere_ar
092800000100018	MINISTERE DE LA DEFENSE NATIONALE	AIN ROMANA					
092800000100019	E.A.C N°01 DAS KHELIFI ALI	AIN ROMANA					
092800000100020	E.A.C N°04 KHELIFA ALI	AIN ROMANA					
092800000100021	E.A.C N°03 DAS KHELIFI ALI	AIN ROMANA					
092800000100022	E.A.C N°07 DAS KHELIFI ALI	AIN ROMANA					
092800000100023	E.A.C N°05 DAS KHELIFI ALI	AIN ROMANA					
092800000100024	E.A.C N°04 DAS HADJELLAH MOHAMED	AIN ROMANA					
092800000100025	E.A.C N°02 DAS HADJELLAH MOHAMED	AIN ROMANA					
092800000100026	E.A.C N°01 DAS HADJELLAH MOHAMED	AIN ROMANA					

Table après la suppression des champs inutiles

1 • SELECT * FROM entrepot_application.td_personnes;

pe_num_nat	nom_fr	adresse_fr
092800000100018	MINISTERE DE LA DEFENSE NATIONALE	AIN ROMANA
092800000100019	E.A.C N°01 DAS KHELIFI ALI	AIN ROMANA
092800000100020	E.A.C N°04 KHELIFA ALI	AIN ROMANA
092800000100021	E.A.C N°03 DAS KHELIFI ALI	AIN ROMANA
092800000100022	E.A.C N°07 DAS KHELIFI ALI	AIN ROMANA
092800000100023	E.A.C N°05 DAS KHELIFI ALI	AIN ROMANA
092800000100024	E.A.C N°04 DAS HADJELLAH MOHAMED	AIN ROMANA
092800000100025	E.A.C N°02 DAS HADJELLAH MOHAMED	AIN ROMANA
092800000100026	E.A.C N°01 DAS HADJELLAH MOHAMED	AIN ROMANA

Figure 31: Suppression des champs inutiles dans la table personne

- **Contrôle d'intégrité référentielle**

Ce contrôle nous a permis de vérifier que les données contenant dans chaque table correspondent aux données équivalentes situées dans toutes les autres tables référencées. Et dans notre cas, ce contrôle est parfaitement respecté.

2.2.3. Chargement des données

Au moment des chargements des données, il est indispensable d'identifier les différences entre les magasins de données cibles en termes de structure et syntaxe, savoir les utiliser ou bien les éviter, ainsi que les tâches à exécuter avant et après le chargement. L'utilisation des techniques ou des index afin d'assurer l'optimisation des performances est également recommandée. Pour cette étape de chargement nous avons procédé comme suit :

- **Chargement des dimensions**

Cette étape consiste en premier lieu à charger les tables dimensionnelles statiques, qui représentent toutes les tables référentielles de la base de données source. Puis, des clés de substitution seront affectées, afin que l'administrateur de l'entrepôt puisse réagir à d'éventuels changements et à la détection d'anomalies dans les données brutes.

- **Tables de faits et automatisation**

Après avoir chargé toutes les dimensions, nous passons au chargement des différentes tables des faits. Cette étape commence par l'extraction de l'historique des tables de faits et passage à leurs traitements, en recherchant des clés de substitution, éliminant les valeurs nulles, optimisant le contenu, et restructurant les données, ... etc.

Ensuite, le processus de chargement incrémental est lancé, et qui consiste à ne charger que les données nouvelles et celles qui ont été modifiées, puis les tables d'agrégats. Il s'agit simplement d'exécuter une requête d'agrégation et d'en stocker les résultats dans une table.

2.2.4. Diagramme d'activité du processus ETL

La méthode d'extraction des données à partir des sources ainsi que le chargement des dimensions et des faits que nous avons adapté pour notre solution est illustrée par le diagramme d'activité représenté ci-dessous :

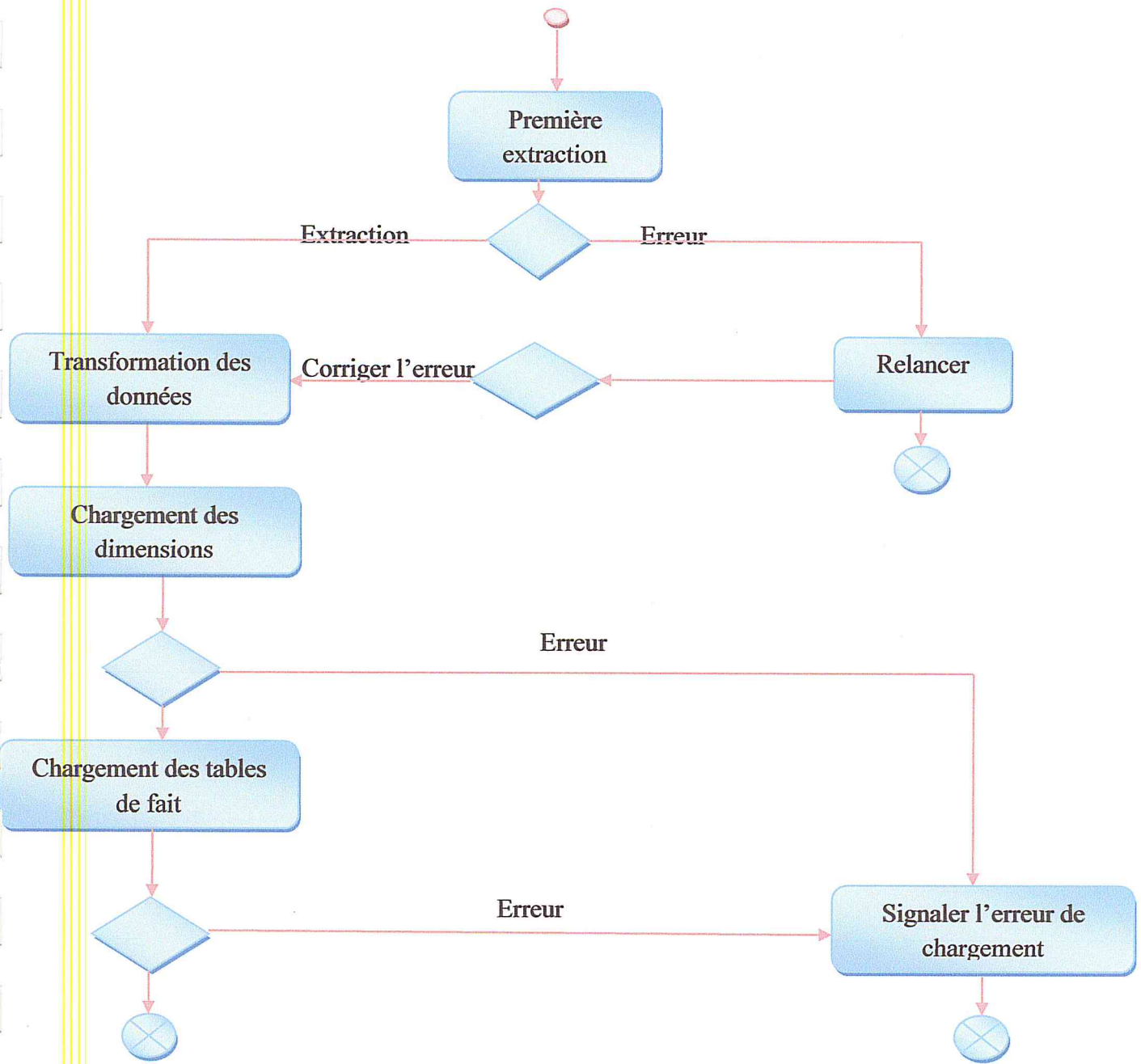


Figure 32 : Diagramme d'activité du processus ETL

3. Conception des cubes dimensionnels

3.1. Niveaux et hiérarchies des dimensions

La définition des niveaux et hiérarchies d'une dimension est une étape très importante dans la conception des cubes dimensionnels. En effet, c'est grâce à ces hiérarchies et dimensions que les utilisateurs peuvent naviguer facilement dans les données et exploiter de manière efficace les informations contenues dans l'ED, avec le niveau de détail souhaité.

Une hiérarchie se compose de plusieurs niveaux et représente le classement des attributs selon une série de relations 1-n, suivant plusieurs niveaux où le niveau « ALL » est le niveau le plus haut d'une hiérarchie, il représente le niveau le plus agrégé. Par conséquent, nous allons présenter dans le tableau ci-dessous, les différents niveaux et leurs attributs, ensuite nous passons aux hiérarchies.

Dimension	Attribut	Niveau	Hiérarchie
DIM_WILAYA	wi_no	N1	N1 → ALL
	wi_nomfr		
DIM_COMMUNE	ID_WC	N1	N1 → N2 → ALL
	Co_nom_fr	N2	
DIM_SECTION	ID_WCS	N1	N1 → N2 → ALL
	se_no	N2	
DIM_ILOT	il_no_nat	N1	N1 → N2 → ALL
	ID_WCS	N2	
DIM_PARCELLE	pa_no_nat	N1	N1 → N2 → ALL
	il_no_nat	N2	
DIM_SECTION_géographique	ID_WCSG	N1	N1 → N2 → ALL
	Co_no	N2	
DIM_ILOT_géographique	il_no_gra	N1	N1 → N2 → ALL
	ID_WCS	N2	

DIM_PARCELLE_ géographique	Pa_no_gra	N1	N1 → N2 → ALL
	il_no	N2	

Tableau 25: Liste des niveaux et des hiérarchies des dimensions de l'entrepôt

3.2. Liste des cubes dimensionnels

Après avoir identifié les niveaux et hiérarchies de chaque dimension, nous dressons la liste de tous les cubes à réaliser, avec tous les détails concernant chacun d'eux, les mesurables ainsi que les dimensions participantes dans ces cubes. Le tableau ci-dessous représente la liste de tous les cubes dimensionnels :

Nom du cube	Description cube	Mesurables	Dimensions participantes
CUBE_SUIVI_SECTION	Suivi de la réalisation des sections	Nombre de section GIC Nombre section graphique Surface section graphique	DIM_WILAYA DIM_COMMUNE DIM_SECTION DIM_géographique_section
CUBE_SUIVI_ILOT	Suivi de la réalisation des ilots	Nombre ilot (GIC, graphique) Surface ilot (GIC, graphique) Nombre parcelle Nombre individus Nombre bâtiments	DIM_WILAYA DIM_COMMUNE DIM_SECTION DIM_ILOT DIM_géographique_ilot
CUBE_SUIVI_PARCELLE	Suivi de la réalisation des parcelles	nombre parcelles littéral nombre parcelles graphique surface parcelle Cadastré GIC surface parcelle graphique	DIM_WILAYA DIM_COMMUNE DIM_SECTION DIM_ILOT DIM_PARCELLE DIM_géographique_parcelle

Tableau 26: Liste des cubes

3.3. Représentation des cubes dimensionnels

Après avoir listé les différents cubes dimensionnels, nous passons à une représentation graphique de ces derniers.

3.3.1. Représentation du cube CUBE_SUIVI_SECTION

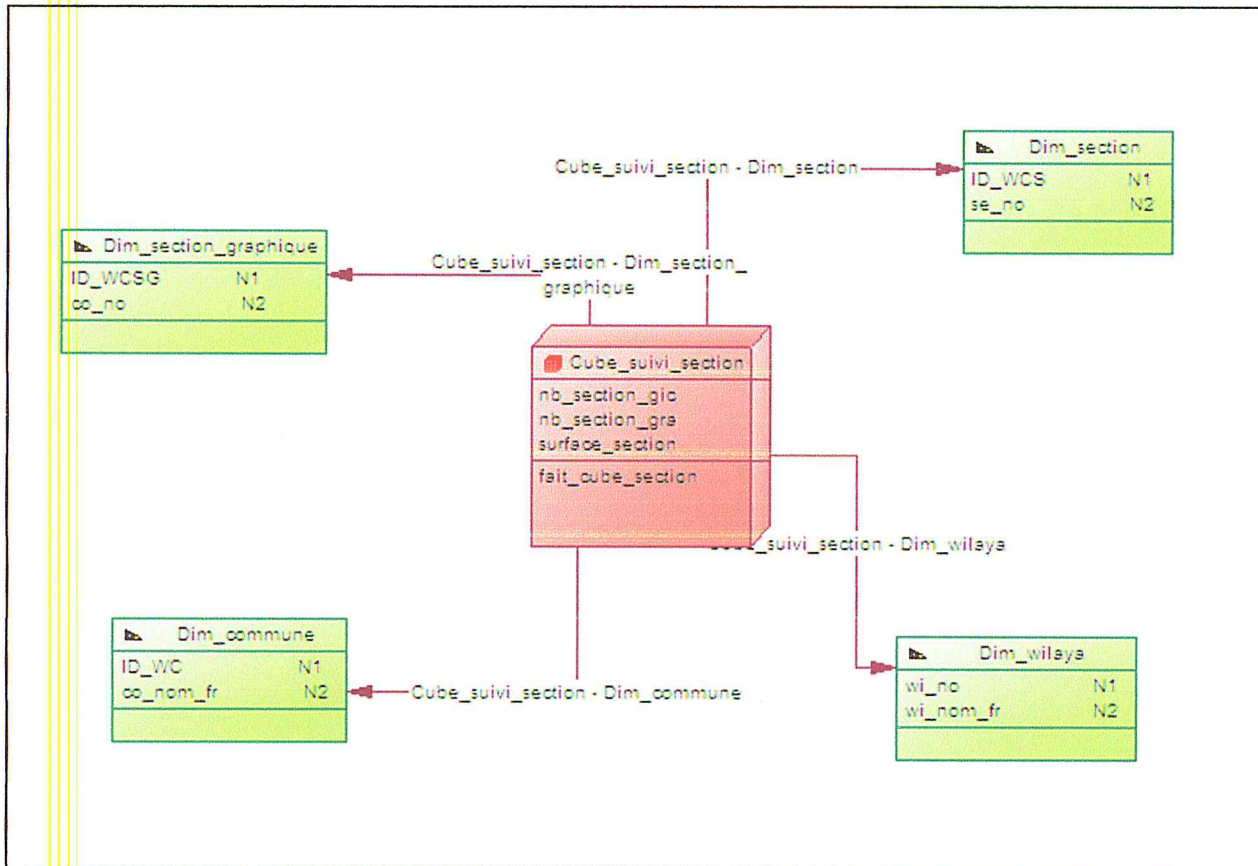


Figure 33 : Cube dimensionnel "CUBE_SUIVI_SECTION"

3.3.2. Représentation du cube CUBE_SUIVI_ILOT

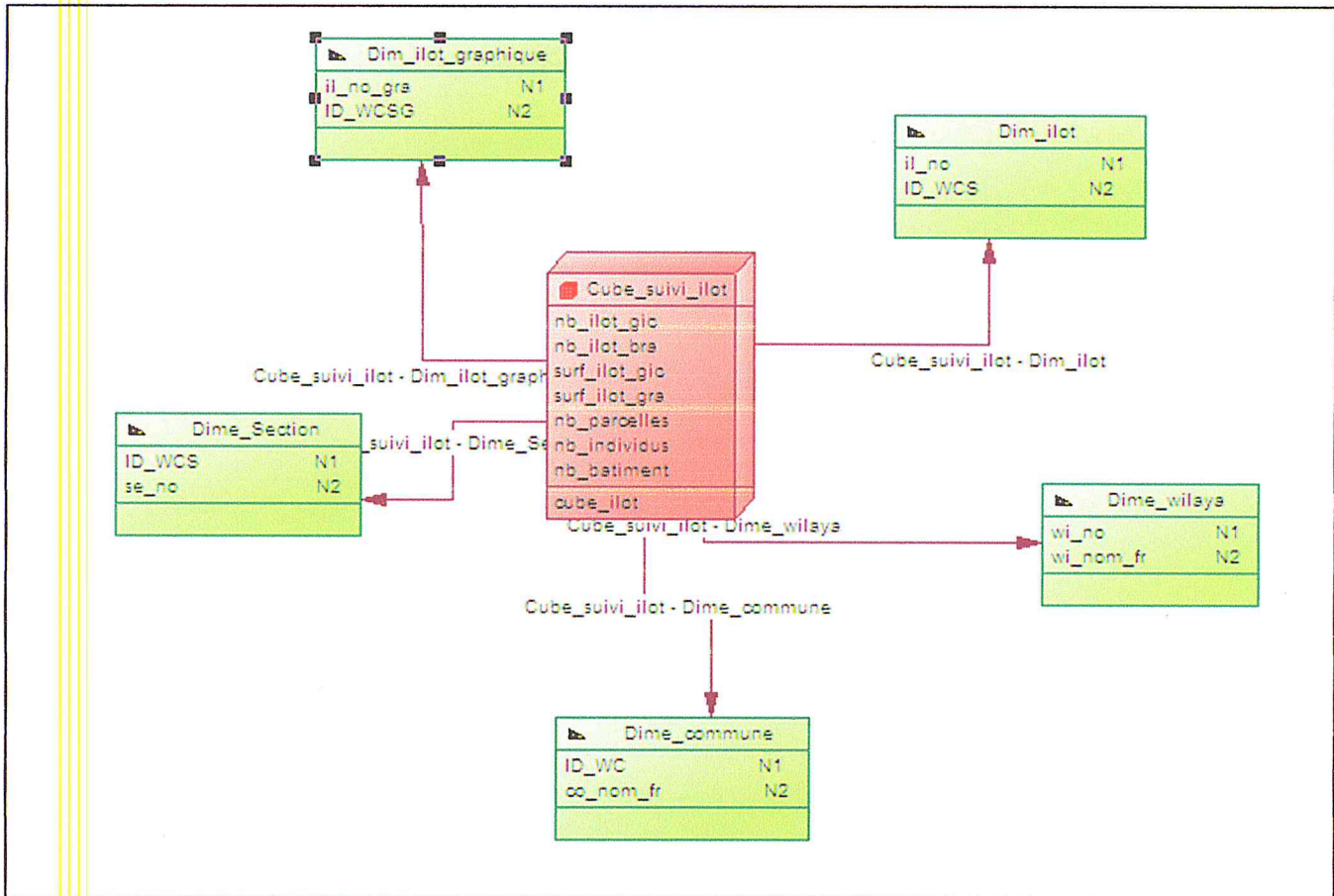


Figure 34: Cube dimensionnel "CUBE_SUIVI_ILOT"

3.3.3. Représentation du cube CUBE_SUIVI_PARCELLE

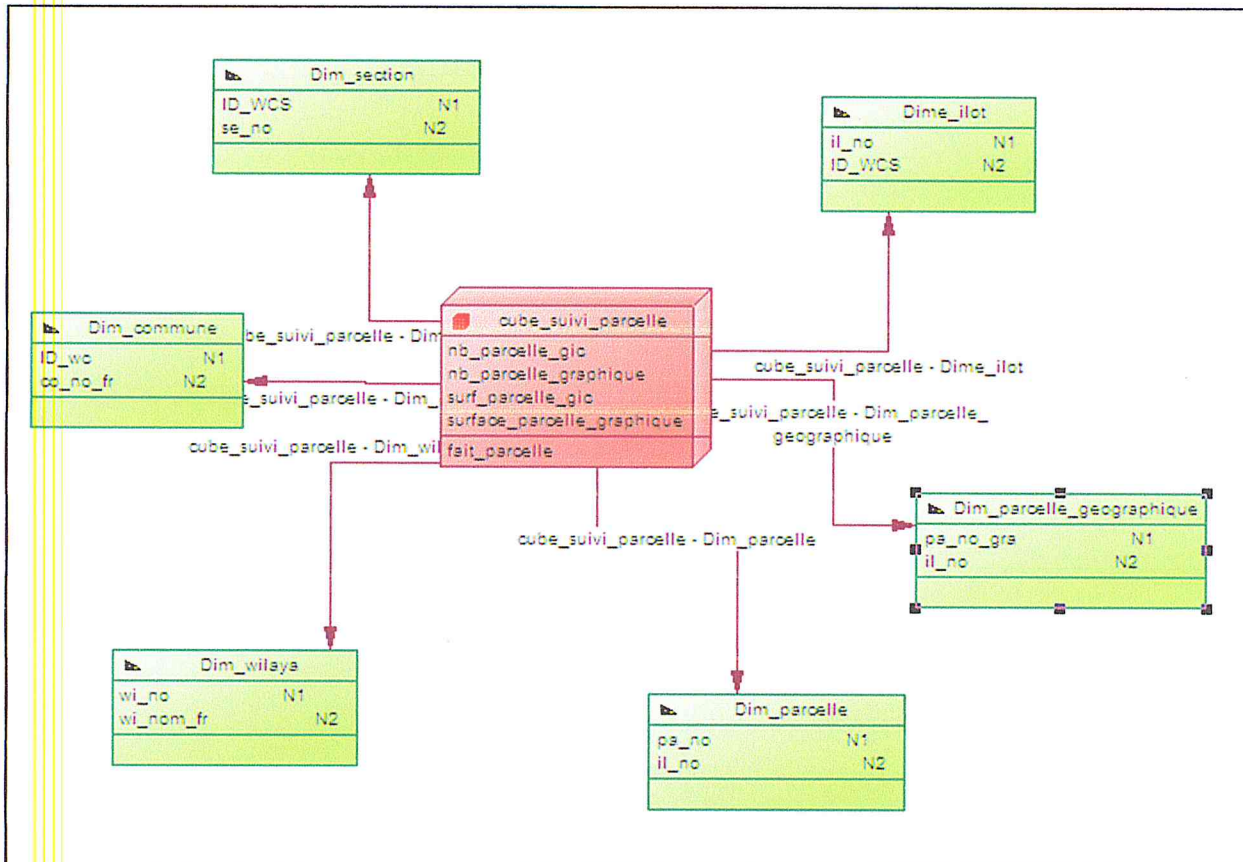


Figure 35 : Cube dimensionnel "CUBE_SUIVI_PARCELLE"

Conclusion

A travers ce chapitre nous allons faire la conception de la zone d'entreposage qui est une étape essentielle du projet. Elle se fait grâce à la modélisation dimensionnelle, qui offre aux utilisateurs des modèles intuitifs, la conception de la zone d'alimentation, et la conception des cubes dimensionnels qui est une étape très importante dans tout projet de réalisation d'un entrepôt. Dans le chapitre suivant on va faire la réalisation de notre projet.

Chapitre 4 : Mise en œuvre de notre système décisionnel

Introduction

Une fois la conception est terminée, nous passons à la phase de réalisation de la solution. Elle s'agit de l'étape de développement de l'ouvrage proprement dite, où nous allons décrire l'architecture technique de l'entrepôt dans la section 1. Nous présentons ensuite dans la section 2 les différents outils utilisés, ainsi que les étapes de réalisation du nouveau système dans la section 3. Dans la section 4, nous allons nous intéresser à la réalisation de notre solution.

1. Architecture technique de la solution

L'architecture est la traduction technique des besoins métier exprimés par les utilisateurs en informations de l'entreprise. Elle détermine les fonctionnalités nécessaires à la satisfaction de ces besoins [KIMBALL, 2011]. Le processus de développement de l'architecture technique doit produire à la fin deux résultats :

- **Le plan d'architecture technique** qui décrit la vision de la structure du futur ED, il permet par ailleurs de construire les composants de manière modulaire et de préparer l'intégration de nouveaux produits et services à l'infrastructure existante.
- **Le plan d'infrastructure** qui décrit les éléments matériels de base de l'ED.

La mise en œuvre d'un entrepôt de données requiert l'élaboration de plan d'infrastructure physique, qui identifie le matériel nécessaire, les systèmes d'exploitation et les logiciels. Elle dépend essentiellement de la plateforme logicielle et l'architecture réseau de l'entreprise. La figure suivante représente l'architecture technique de notre solution :

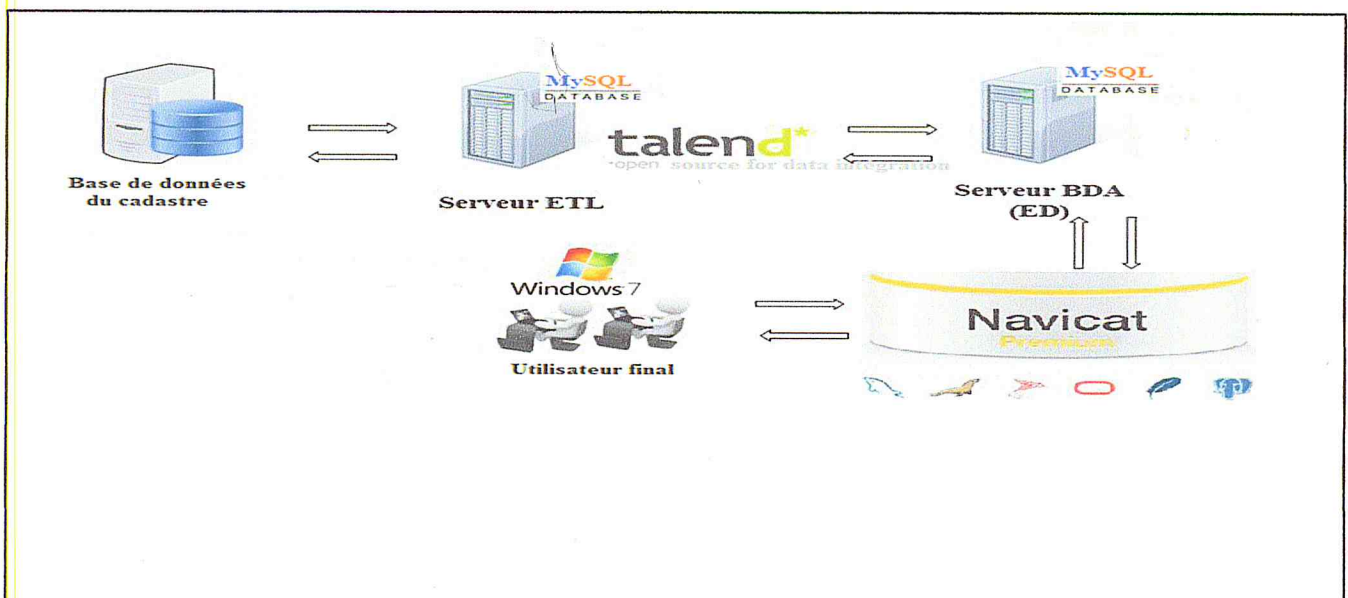


Figure 36: Architecture technique du système proposé

2. Sélection des outils

Au cours du processus d'évaluation technique, les besoins métier et les besoins techniques sont deux éléments essentiels, qui sont aptes à guider la sélection des produits. Les supports fournis par les distributeurs (documentation, formations et supports techniques) ainsi que les supports externes (forums de support technique en ligne) sont également à prendre en considération. Et ce, afin d'éviter tout problème lié à la compatibilité entre la plateforme existante et la nouvelle solution implémentée.

L'étude de marché accompagnée par l'élaboration de la matrice d'évaluation des outils, nous a permis d'identifier les produits les plus intéressants, qui sont classifiés en trois domaines :



Figure 37 : Outils utilisés pour la réalisation de notre système

2.1. Plate-forme du SGBD

Nous avons implémenté notre entrepôt de données sous *MySQL Workbench* (logiciel de gestion et d'administration de bases de données MySQL).

Ce choix *MySQL workbench* comme SGBD, il permet, entre autres, de créer, modifier ou supprimer des tables, des comptes utilisateurs, et d'effectuer toutes les opérations inhérentes à la gestion d'une base de données. Pour ce faire, il doit être connecté à un serveur MySQL.

2.2. Outils de préparation des données

Pour la zone d'alimentation, nous avons choisi **Talend Open Studio (TOS)**, qui permet de créer graphiquement des processus de manipulation et de transformation de données. Il est basé sur l'IDE « Eclipse » et intègre un ensemble de composants implémentés en JAVA. Talend Open Studio offre une très grande connectivité aux progiciels, bases de données, entrepôts de données, applications OLAP pour analyse, ...etc. [TALEND, 2010]

Après une étude comparative (ANNEXE 3), le choix a été porté sur Talend Open Studio dans sa version « 5.3.1 ». Ce choix se justifie par les avantages offerts par cette solution :

- Solution logicielle open source, développée par la société Talend en France, qui offre les fonctionnalités nécessaires pour répondre aux besoins d'extraction et chargement des données. Son usage permet de se doter d'un outil puissant sans engager des frais de licence,
- Solution facile à utiliser avec une interface très agréable et très maniable (beaucoup d'actions se font en glisser-déposer), incluant une documentation complète et riche,
- TOS est adaptable et polyvalent, admettant une communauté très réactive et participative pour son développement,
- Il possède une très grande gamme de composants (246 composants), permettant de se connecter à quasiment toutes les sources de données, applications métier et type de fichiers existants,
- Talend est un code générateur java toujours visible par l'utilisateur, c'est-à-dire que l'installation de TOS n'est pas obligatoire sur les serveurs de production. Il est possible d'enregistrer les transformations sous forme de scripts, qui sont alors exécutables à partir de n'importe quelle machine disposant d'une JVM (pour Java).

2.3. Outils d'accès aux données

Navicat est une suite logicielle graphique de gestion et de développement de bases de données produits par PremiumSoft CyberTech Ltd. pour MySQL, MariaDB, Oracle, SQLite, PostgreSQL. Son interface utilisateur graphique est similaire à celle d'Explorer et il supporte des connexions multiples vers des bases de données locales et distantes. Il est conçu pour répondre aux besoins d'un public diversifié allant de programmeurs et administrateurs de bases de données à des entreprises de tous types qui servent des clients et partagent des informations avec des partenaires.

Navicat est un outil multiplate-forme qui fonctionne sous Microsoft Windows. Notre choix se justifie par les avantages offerts par cette suite logicielle, qui peuvent se résumer en :

- permet de se connecter simultanément à différents types de bases de données dont MySQL.
- permet aux utilisateurs d'effectuer des migrations de données entre des bases de données croisées.
- générateur de rapport.

3. Réalisation de la solution

La réalisation de la solution consiste à transformer le modèle logique en implémentation physique, en se basant sur les caractéristiques de la BDD, des données, des flux, la stratégie de sécurité...etc. Dans ce qui suit la construction de chacune des zones de l'ED.

3.1. Construction de la zone d'entreposage

Les données issues de la base de données opérationnelle du cadastre, déployée sous MySQL, sont intégrées dans une base de données relationnelle conforme aux schémas en étoiles conçus auparavant, qui représentent les magasins de données de notre entrepôt.

Pour mener à bien cette étape, nous avons créé d'abord les dimensions avec des clés de substitution, ensuite les tables de faits ainsi que les clés étrangères et enfin les jobs de transformation et de chargement.

3.2. Construction de la zone d'alimentation (ETL)

La préparation de données, en vue de leur intégration dans l'ED, est une étape ardue et constitue l'activité la plus importante en termes de charge de travail. Avant de commencer le processus d'alimentation, le chargement des données dans une zone de staging est recommandé, afin de libérer le plus tôt possible le système source. Cette phase d'alimentation des magasins de données se déroule en plusieurs étapes, qui sont : planification, chargement des dimensions, et chargement des tables de faits.

Comme nous l'avons mentionné précédemment, nous avons opté pour **Talend Open Studio** pour la réalisation des différents programmes ETL. Les trois figures suivantes illustrent, respectivement, le mapping du chargement des dimensions parcelle (par exemple), ainsi que le mapping du chargement les tables de faits.

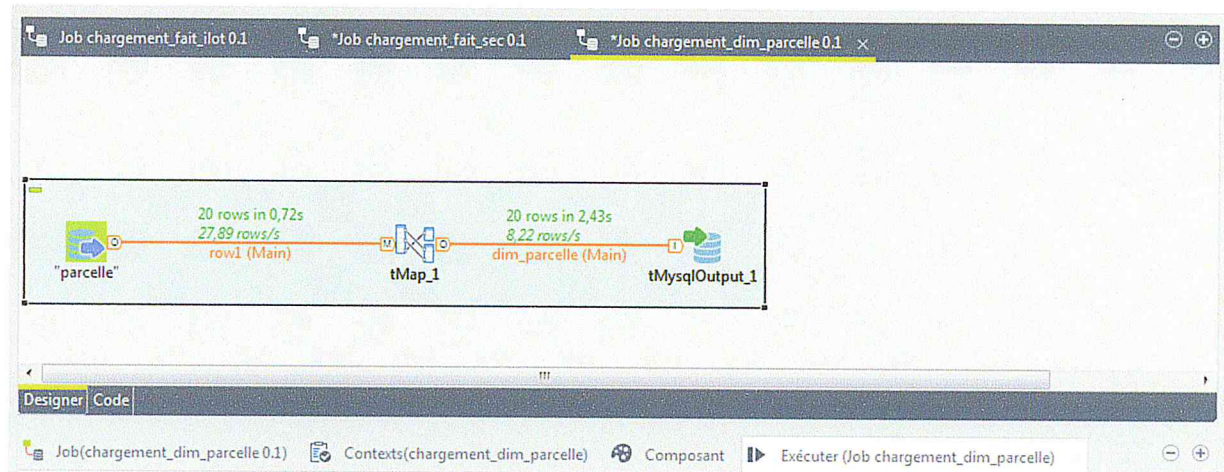


Figure 38 : Mapping du chargement de la dimension parcelle

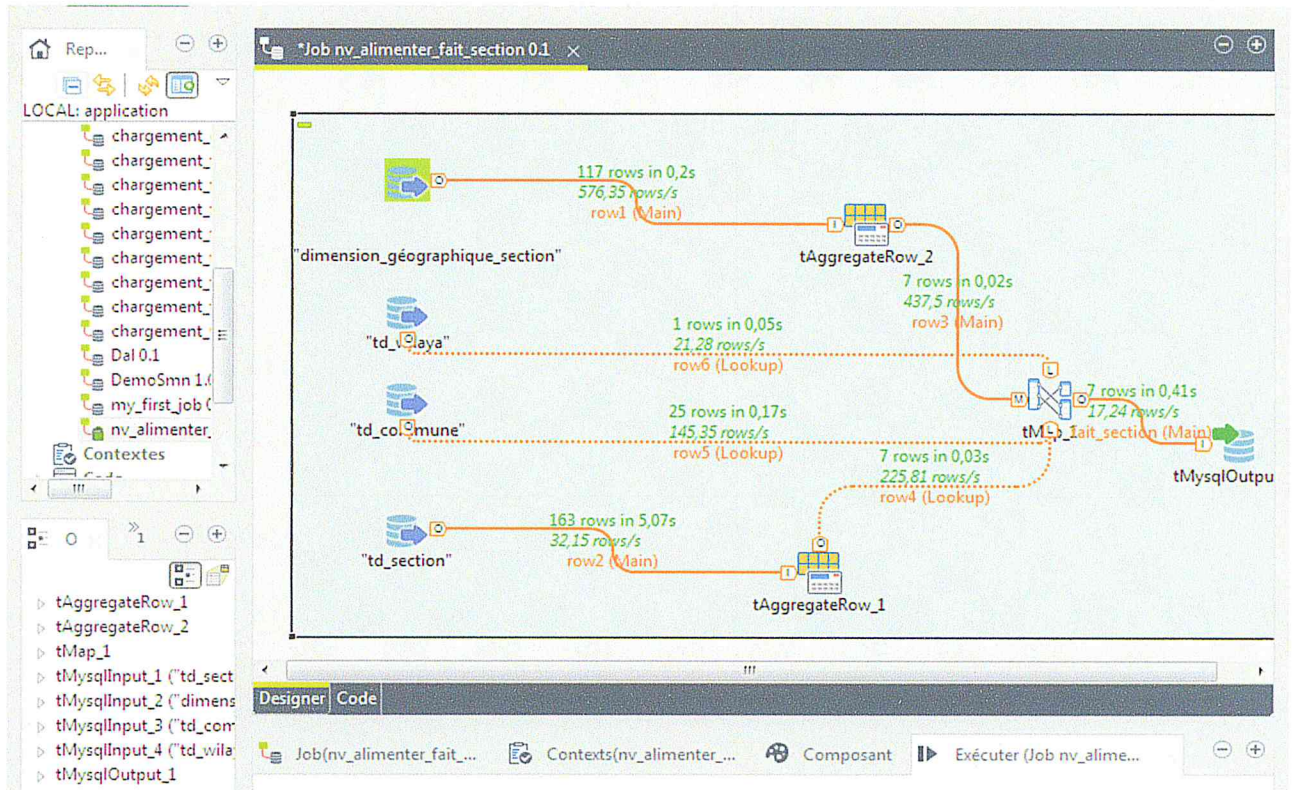


Figure 39 : Mapping du chargement de la table de faits suivi_section

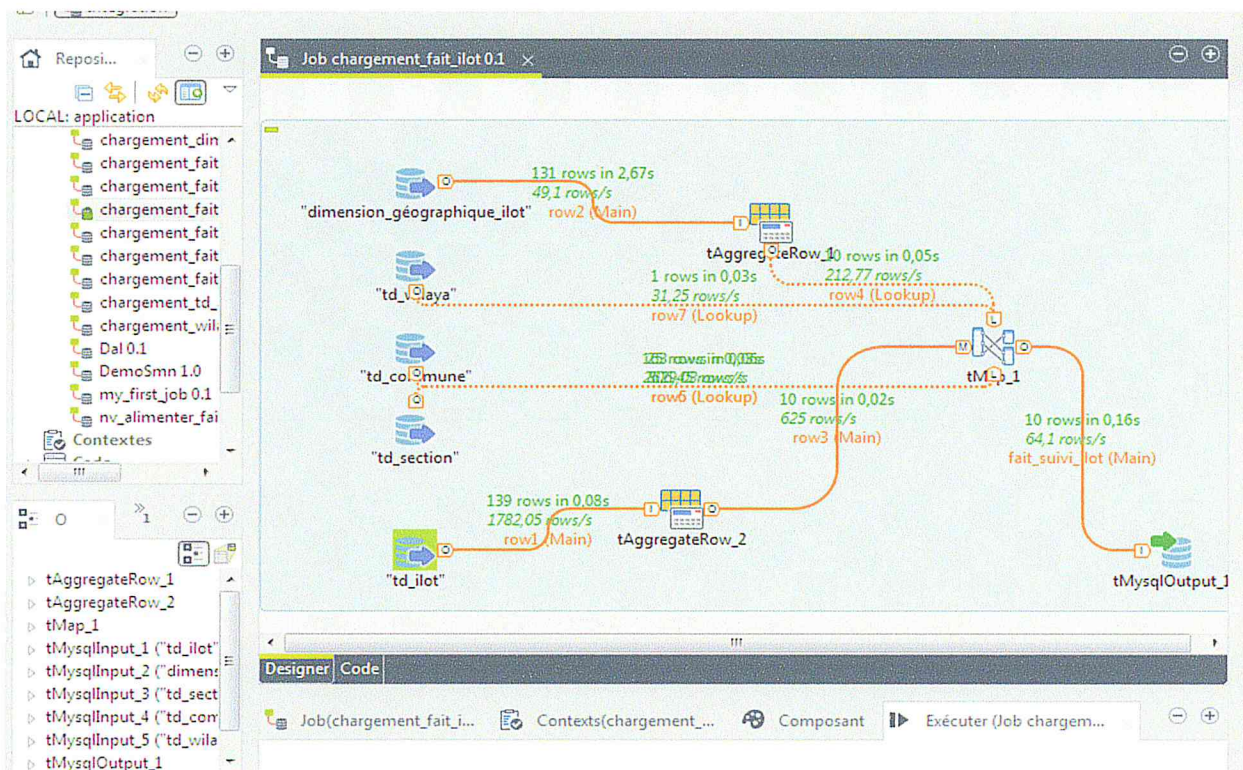


Figure 40: Mapping du chargement de la table de faits suivi_ilot

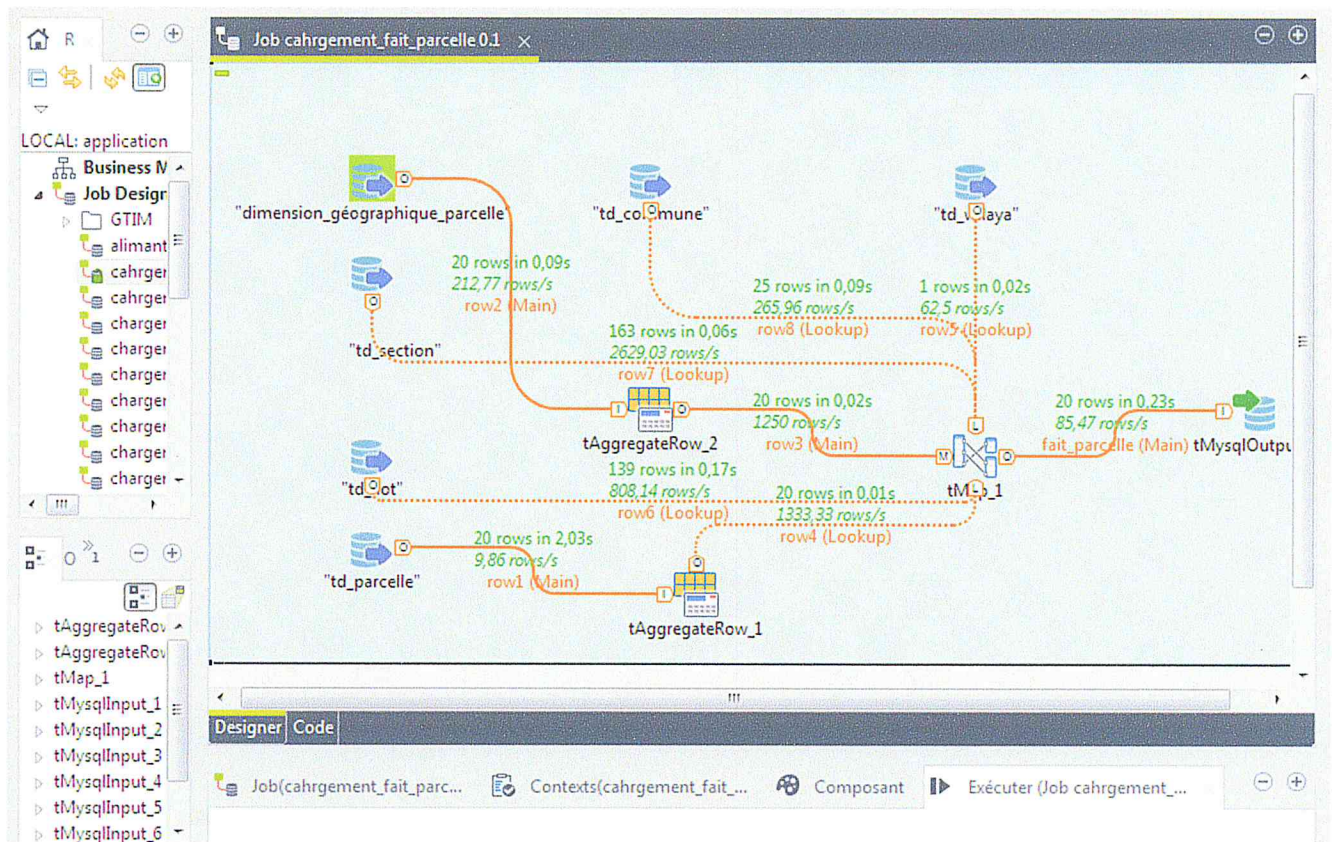


Figure 41 : Mapping du chargement de la **table de faits** suivi_parcelle

3.3. Construction des cubes dimensionnels

Une fois notre entrepôt est chargé, nous passons à l'alimentation de nos cubes dimensionnels. Pour la construction de ces derniers, nous utilisons **pentaho workbench**, et nous avons suivi les étapes suivantes :

- Connexion de Pentaho à la base de données référentielle (nommée : **entrepot_application**),
- Création de l'ensemble des dimensions et leurs hiérarchies, et ce en utilisant le module de Pentaho (**Schema Workbench**), Création de l'ensemble des cubes ainsi que leurs mesures,
- Construction de la correspondance avec les dimensions des magasins de données,

La figure suivante illustre l'ensemble des cubes et dimensions créés, en utilisant le module de Pentaho Schéma WorkBench :

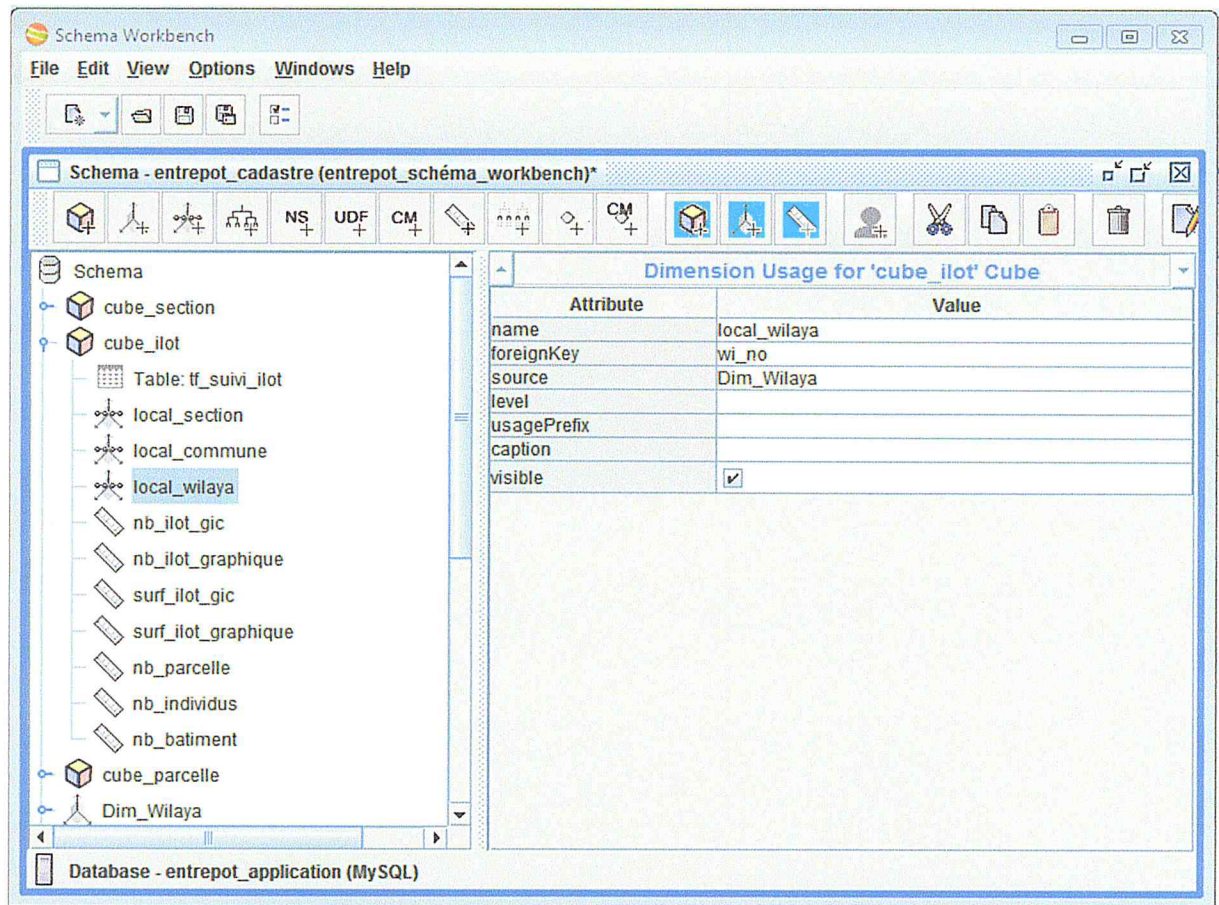


Figure 42: Ensemble de Cubes et Dimensions créés sous Pentaho Schéma WorkBench

3.4. Construction des outils de restitution

Pour la représentation des analyses nous avons utilisé le reporting qui désigne une famille d'outils de Business intelligence destinés à assurer la réalisation, la publication et la diffusion de rapports d'activité selon un format prédéterminé. Ils sont essentiellement destinés à faciliter la communication de résultats chiffrés ou d'un suivi d'avancement.

Pour cela on utilise le logiciel graphique de gestion et de développement Navicate Premium. Voici quelques captures d'écran des résultats réalisées :

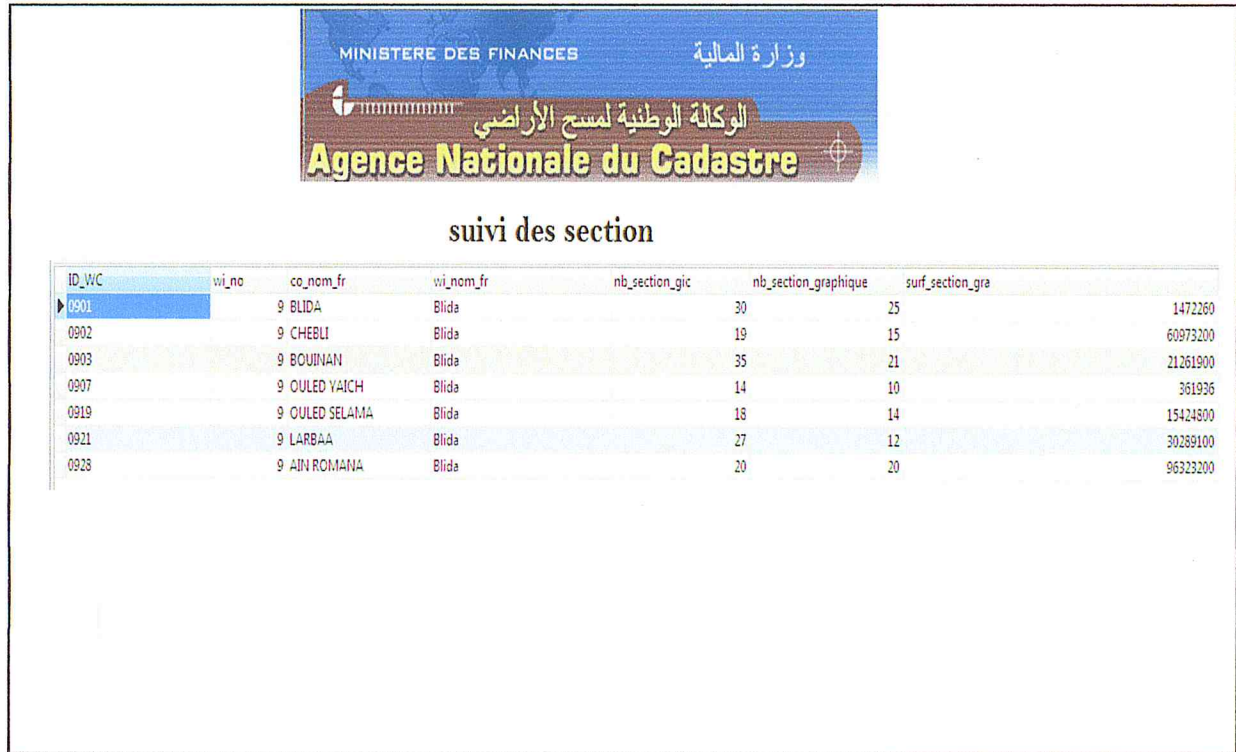


Figure43: statistiques sur les sections cadastrées par commune



Figure44 : statistiques sur les ilots cadastrés

éliminer

Statistiques sur le nombre des sections cadastrées

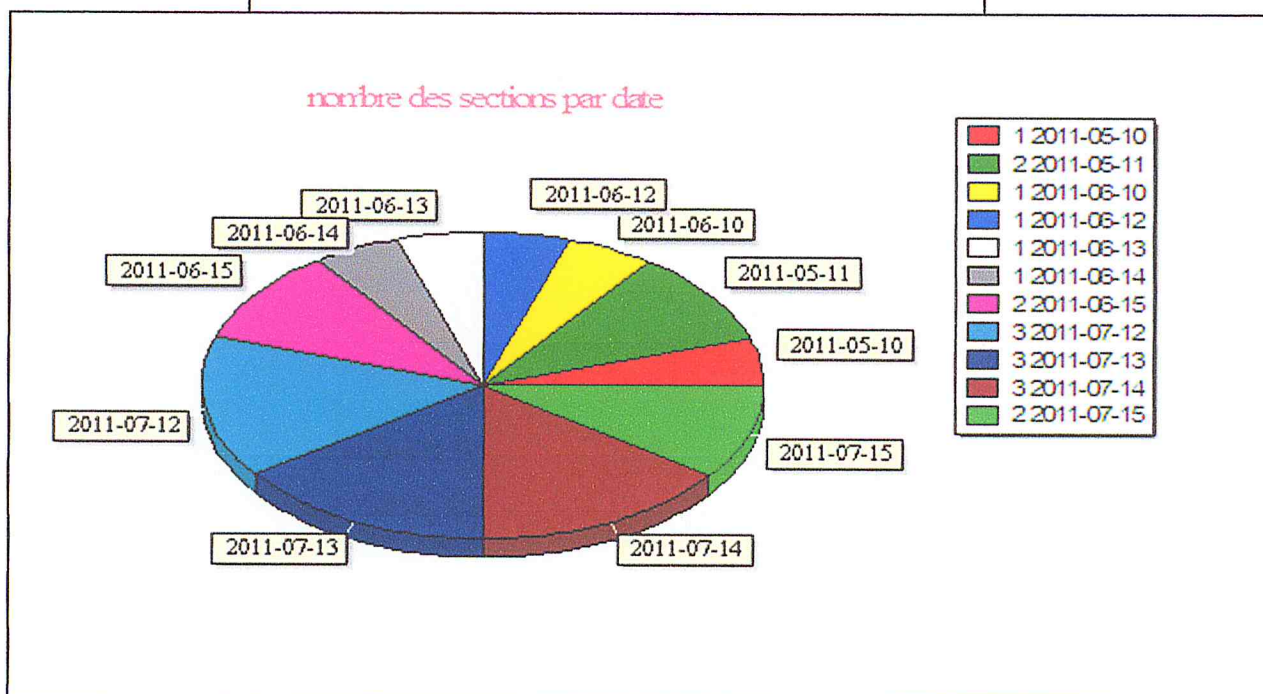


Figure45: statistiques sur le nombre des sections cadastrées de la commune Ain Romana

Statistiques sur les ilots cadastrés

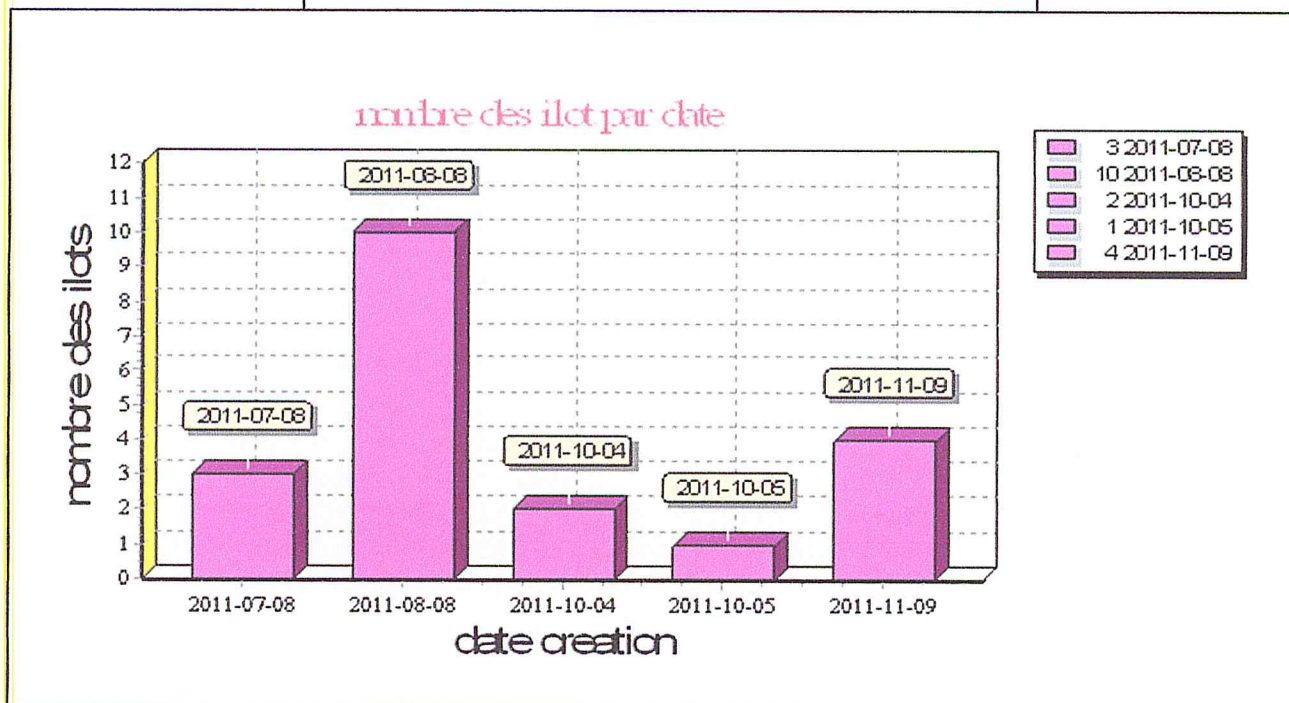


Figure 46 : statistiques sur le nombre des ilots cadastrés dans la section 1 de la commune Ain Romana

4. Conduit du projet

Il existe plusieurs méthodes pour mettre en place un entrepôt de données, mais trois seulement sont les plus répandues. Il s'agit de l'approche guidée par les sources de données prônée par **Inmon**, l'approche guidée par les besoins d'analyse de **Kimball**, ainsi que l'approche hybride qui dérive des deux.

Afin de bien mener notre projet, nous avons choisi de suivre la démarche qui s'inspire de la méthode de **Kimball** décrite dans le **chapitre 01** : Conduite et construction d'un entrepôt de données, qui est l'approche la plus utilisée pour la conception, le développement et le déploiement des entrepôts de données.

Cette méthode se base sur la description du cycle de vie dimensionnel (Illustré dans la **Figure 15**) qui matérialise l'enchaînement des grandes étapes de l'implémentation d'un entrepôt de données. Pour l'étape d'identification des besoins, nous avons choisi une démarche mixte alliant les besoins des utilisateurs et les sources de données.

Conclusion

Dans ce chapitre, nous avons décrit l'architecture technique de notre solution, puis nous avons abordé l'implémentation. Nous avons présenté également les outils utilisés, qui nous ont permis la collecte, l'intégration, l'organisation des données ainsi que la restitution des informations résultantes de l'analyse sous différentes formes de rapports. La partie de restitution a été mise en place en intégrant des solutions Open Source qui nécessitent une étude comparative des différents outils existants et une bonne détermination des facteurs de choix.

Conclusion Générale et perspectives

Jusqu'à présent, la direction du cadastre ne dispose d'aucun système décisionnel permettant d'améliorer son activité. Et dans le but de répondre à cette problématique et palier aux difficultés récurrentes dans le processus de prise de décision, le cadastre a initié le projet d'entrepôt de données de la gestion de l'information cadastral textuelle et géographique, qui permet de transformer les données de production en informations statistiques.

Afin de bien mener notre projet, nous avons en premier temps effectué une étude bibliographique, où nous nous sommes concentrées sur les points de vue des pionniers du domaine des entrepôts, **Ralph Kimball** et **William H. Inmon**. Cette étude se limite au concept de l'informatique décisionnelle, aux entrepôts de données, à la modélisation dimensionnelle et aux étapes de construction et de réalisation d'un tel projet.

La mise en place d'un entrepôt de données nécessite une parfaite compréhension du métier de l'entreprise et une analyse détaillée des besoins des utilisateurs, qui déterminent de manière significative le déroulement de toutes les étapes suivantes. C'est dans la deuxième partie que nous présentons l'environnement professionnel dans lequel nous avons mené notre stage, son métier ainsi que les besoins exprimés par les utilisateurs finaux de notre système.

En ce qui concerne la conception de la zone d'entreposage, nous avons préconisé la modélisation dimensionnelle, compte tenu de ses avantages. Le but principal de la modélisation de la zone de stockage consiste à produire un schéma graphique dimensionnel, qui offre une vision claire tout en satisfaisant les besoins analytiques des utilisateurs.

Au cours de notre travail, nous avons accordé une attention particulière à la conception de la zone de préparation des données, qui a été l'étape la plus longue et la plus complexe du projet. Cette étape assure la livraison de données conformes, cohérentes et correctes, en collaboration avec l'administrateur des bases de données du cadastre.

La démarche adoptée pour l'implémentation de notre système est la démarche de **Kimball**. Nous avons commencé par présenter d'abord la disposition physique du système, choisir ensuite les outils nécessaires pour assurer un environnement adéquat et favorable à l'exécution.

Bien que les objectifs fixés soient atteints, il est utile de rappeler qu'un projet d'ED n'est jamais complètement terminé, le travail est appelé à être amélioré et enrichi. Un ensemble de perspectives est envisageable, afin de continuer et d'améliorer ce travail :

- Réaliser les tableaux de bord permettant de suivre et d'anticiper le fonctionnement et l'activité de la direction du cadastre, afin de mesurer la performance et de faciliter le pilotage de cette dernière,
- Apporter les modifications nécessaires au niveau des bases de données (Gic et Arcgic), et ce dans le but d'éviter les doublons et assurer une meilleure qualité de données,

REFERENCES BIBLIOGRAPHIQUES

[ATOL, 2008]

Atol Conseils et Développements. Les ETL Open Source : Une réelle alternative aux solutions propriétaires. Publié le : 25 juin 2008.

[BELLATRECHE, 2000]

BELLATRECHE, Ladjel. Utilisation des vues matérialisées, des index et de la fragmentation dans la conception logique et physique d'un entrepôt de données. Thèse de doctorat, Université de Clermont-Ferrand 2, Clermont-Ferrand, France, 2000.

[DAVENPORT, 2008]

DAVENPORT, Robert J. ETL vs ELT a subjective view. Insource Commercial aspects of BI whitepaper, 2008.

[FERNANDEZ, 2013]

FERNANDEZ, Alain. Les nouveaux tableaux de bord des managers : le projet Business Intelligence clés en main. Editions Eyrolles, 2013.

[GAM, 2006]

GAM, Inès, SALINESI, Camille, et al. Analyse des Exigences pour la Conception d'Entrepôts de Données. Proceedings of Informatique des Organisations et Systèmes d'Information et de Décision, 2006, p. 1023-1038.

[INMON, 2002]

INMON, William H. Building the data warehouse. John Wiley & sons, 2005.

[JUKIC, 2006]

JUKIC, Nenad. Modeling strategies and alternatives for data warehousing projects. Communications of the ACM, 2006, vol. 49, no 4, p. 83-88.

[KHOURI, 2009]

KHOURI, Selma. Modélisation conceptuelle à base ontologique d'un entrepôt de données : Sciences et Technologies de l'Information et de la Communication. Magistère, Ecole Doctorale, Ecole Nationale Supérieure d'Informatique (ESI), Alger, Algérie, 2009.

[KIMBALL, 2011]

KIMBALL, Ralph, et ROSS, Margy, et al. Le data warehouse: Guide de conduite de projet. Eyrolles, 2011.

[KIMBALL, 2002]

KIMBALL, Ralph et ROSS, Margy. The data warehouse toolkit: the complete guide to dimensional modeling. John Wiley & Sons, 2002.

[LAHMER, 2011]

LAHMER, Fatima. Une approche Hybride d'intégration de sources de données hétérogènes dans les data warehouses. Thèse de doctorat, Université Mentouri de Constantine, Algérie, 2011.

[MARCEL, 1998]

MARCEL, Patrick. Manipulations de données multidimensionnelles et langages de règles. Thèse de doctorat, Institut National des Sciences Appliquées de Lyon, Lyon, France, 1998.

[NAKACHE, 1998]

NAKACHE, Didier, CAULIER DONNEGER, Anne, RIVELOIS DUGRESSON, Pascale, et al. Delebecq, Jean-Louis. Data warehouse et data mining. CNAM de Lille, 1998.

[POLETTTO, 2012]

POLETTTO, Maxime. L'informatique décisionnelle. Thèse professionnelle. Ecole Supérieure d'informatique (CESI exia), 2012.

[PONNIAH, 2004]

PONNIAH, Paulraj. Data warehousing fundamentals: a comprehensive guide for IT professionals. John Wiley & Sons, 2004.

[TALEND, 2010]

TALEND. Guide utilisateur : Talend Open Studio 4.X, 2010, p.286.

[TESTE, 2000]

TESTE, Olivier. Modélisation et manipulation d'entrepôts de données complexes et historiées. 2000. Thèse de doctorat. Université Paul Sabatier-Toulouse III.

[ZIAN, 2013]

ZIAN, Houda. Contribution à l'étude des tableaux de bord dans l'aide à la décision des PME en quête de performances. 2013. Thèse de doctorat. Bordeaux 4.

REFERENCES WEB GRAPHIES

[METAIS, 2014]

METAIS, Elisabeth. Systèmes informatiques : systèmes d'aide à la décision. Encyclopædia Universalis [en ligne]. Consulté le 28/09/2014. Disponible sur : <http://www.universalis.fr/encyclopedie/systemes-informatiques-systemes-d-aide-a-la-decision/>.

[TASLIMANKA SYLLA, 2007]

TASLIMANKA SYLLA, Mohamed. Initiation au décisionnel (Business Intelligence, Data Warehouse, OLAP) [en ligne]. Publié le 20/11/2007. Consulté le 05/10/2014. Disponible sur : <http://taslimanka.developpez.com/tutoriels/bi/#LIIC>

https://help.talend.com/display/TalendOpenStudioComponentsReferenceGuide521EN/18.1+AggregateRow?thc_login=Ddone

<https://help.talend.com/display/talendopenstudiocomponentsreferenceguidebFR>

Annexe 1: Liste des questions pour les interviews

Question générale :

- Quelle sont les principaux problèmes existant dans votre société ?
- Quelle sont vos objectifs attendre ?
- Quelle sont les processus métier existant dans le cadastre ?
- Quelles sont vos sources de données ?
- Quelle sont les analyses que vous souhaiteriez faire ?
- Qu'attendez-vous de l'entrepôt de données ?

Annexe 2 : Compte rendu d'entretien

Contexte et objectifs :

L'objectif de l'entretien avec le responsable du service informatique, de la direction du cadastre de Blida, est de comprendre le fonctionnement du logiciel Gic et Arcgic (ce sont des logiciels pour notre travail), qui est utilisé par les différents services de la direction du cadastre. Afin d'assimiler l'enchaînement des différents processus métier de l'entreprise.

Besoins en informations

- Informations sur les sections, les ilots, les parcelles..., etc.
- Informations sur le Gic.
- Informations sur l'ArcGic.
- Information sur les processus métiers existes.

Annexe 3 : Etude comparative entre les outils ETL

Les outils ETL sont généralement utilisés dans les systèmes décisionnels afin de permettre l'alimentation des entrepôts de données. Parmi les outils ETL open source nous pouvons citer Talend Open Studio, Pentaho Data Integration, Clover ETL, Kettle, ...etc. Nous avons réalisé une comparaison entre Talend Open Studio et Pentaho Data Integration (Inspiré de [ATOL, 2008]), qui sont les plus répondu dans le marché :

	Fonctionnalités	Pentaho	Talend
Accès aux données	Appel de procédures stockées	OUI	Seulement pour certains SGBD
	Exécution des requêtes	OUI	OUI
	Outil de création de requêtes	NON	OUI
Traitement des données	Fonctions de transformations des dates et des nombres	OUI	OUI OUI
	Génération de documentation technique et fonctionnelle	NON	OUI
	Gestion des erreurs d'intégration	OUI, pour certaines étapes	OUI
Gestion de la sécurité	Sécurité sur l'accès aux métadonnées	OUI	OUI
	Sécurité sur la console d'administration	OUI	OUI